

CENTRO DE INVESTIGACION Y DE ESTUDIOS  
AVANZADOS DEL INSTITUTO POLITECNICO NACIONAL

DEPARTAMENTO DE INGENIERIA ELECTRICA  
SECCION DE ELECTRONICA DEL ESTADO SOLIDO

TITULO DE LA TESIS:

**APROXIMACION BASADA EN LOGICA DIFUSA DEL  
COMPORTAMIENTO DE UN SISTEMA DE TRATAMIENTO  
DE CRUDO**

QUE PRESENTA:

M.C. FERNANDO HERNANDEZ ROSALES \*

PARA OBTENER EL GRADO DE  
DOCTOR EN CIENCIAS  
EN LA ESPECIALIDAD EN INGENIERIA ELECTRICA

ASESOR:

JOSE ANTONIO MORENO CADENAS

México, D. F., diciembre de 2004.

\*Becario de CONACyT



## ***Agradecimientos:***

*Quiero manifestar mi agradecimiento al Consejo Nacional de Ciencia y Tecnología, por el apoyo económico brindado para la realización de esta tesis; al Centro de Investigación y de Estudios Avanzados, por proveer excelentes condiciones para el estudio y el trabajo.*

*Quiero hacer extensivo mi más grande aprecio al Dr. José A. Moreno Cadenas por su amistad y su continuo estímulo brindados durante el desarrollo de esta tesis, la cual no hubiera sido posible sin su acertada dirección y ayuda.*

*Mi agradecimiento al comité de tesis:*

- Dr. José Angel Ortega Herrera
- Dr. Francisco José Ruiz Sánchez
- Dr. Alfredo Reyes Barranca
- Dr. Felipe Gómez Castañeda

*Cuyos comentarios ayudaron a concluir exitosamente esta tesis.*

*Un agradecimiento a los integrantes del grupo de investigación del laboratorio de VLSI, por su ayuda y amistad brindada durante mi estadía en este laboratorio.*

*Esta tesis está dedicada a mi esposa Norma por su amor, paciencia y ayuda; a mis padres, a quienes la vida les cristaliza sus plegarias y rezos; a mis hermanos, por brindarme la fuerza de la esperanza; y a todos mis sobrinos, por quienes vale la pena trabajar por hacer de este, un mundo mejor.*

Agradezco a la Dra. Norma Gabriela Rojas Avelizapa el haber proporcionado los datos utilizados para nuestras demostraciones. Estos datos corresponden al proyecto D.00023 “Atenuación natural de sitios contaminados (biopilas)”, del Programa de Biotecnología del Petróleo en el Instituto Mexicano del Petróleo.



## RESUMEN

En este trabajo se establece una metodología basada en el paradigma de los sistemas difusos, para abordar el modelado de procesos complejos integrando conocimiento experto y datos de mediciones entrada-salida.

Los sistemas difusos, que gracias a sus propiedades matemáticas atractivas, posibilitan la integración de varias fuentes de información lógicas, permiten comportarse como aproximadores universales de funciones; aunado a esto, el empleo de términos lingüísticos en la estructura de su base de reglas los hace relativamente fáciles de abordar, analizar y adecuar.

Los conjuntos difusos que conforman las reglas, representan términos lingüísticos cualitativos que aproximan la forma en que los humanos cuantifican la información.

La parte principal del trabajo se relaciona con la obtención de un marco teórico de modelado basado en técnicas de clustering difuso para identificar los sub-modelos subyacentes en los datos entrada-salida. Cada sub-modelo se modela mediante un modelo local lineal. Estos modelos locales se conjuntan mediante una técnica de interpolación suave. La metodología obtenida es flexible con respecto al conocimiento *a priori* y con los datos entrada-salida del proceso. La obtención del modelo es computacionalmente de bajo costo; con una alta calidad del modelo obtenido. En general, los sistemas a modelar presentan un limitado conocimiento sobre su funcionamiento y disponibilidad de una cantidad razonable de datos del proceso. Se discuten en detalle las propiedades de identificación de la estructura y los parámetros del modelo a partir de los datos del proceso, mediante la metodología propuesta.

Las propiedades del marco teórico de modelado se muestran con la utilización de la metodología propuesta en el modelado de un sistema de biopilas para la remediación de un suelo contaminado, el cual por si mismo es de importancia trascendental dado que en nuestro país se encuentran muchas zonas con este problema. Este proceso se selecciona por el grado de dificultad que presenta para ser modelado mediante técnicas convencionales.

La metodología de modelado se desarrolla en dos pasos:

Primeramente, se realiza una exploración de los datos por medio de un algoritmo difuso de clustering. De sus resultados se obtiene la información que se requiere para la creación del modelo difuso.

Posteriormente, se realiza la identificación de los parámetros del modelo (forma y posición de los conjuntos difusos), usando técnicas de mínimos cuadrados.

En este trabajo también se muestra cómo la metodología se enriquece con el conocimiento del experto, pues además de establecer los valores para la inicialización del proceso de identificación, en general, se comprueba que, una de las propiedades más importantes de los sistemas difusos es su capacidad para formalizar el conocimiento empírico en todas las etapas del proceso de modelado.

# TEMARIO

RESUMEN.....	5
INTRODUCCIÓN.....	11
CAPITULO 1.....	19
CONJUNTOS DIFUSOS, LÓGICA DIFUSA Y SISTEMAS DIFUSOS.....	19
1.1 CONJUNTOS DIFUSOS .....	19
1.2 PROPIEDADES BÁSICAS Y TIPOS DE CONJUNTOS DIFUSOS.....	21
1.3 PARTICIÓN DIFUSA .....	21
1.4 OPERACIONES CON CONJUNTOS DIFUSOS .....	23
1.5 SISTEMAS DIFUSOS.....	26
1.6 DISEÑO DE SISTEMAS DIFUSOS.....	28
1.7 CONCLUSIONES .....	30
CAPITULO 2.....	31
METODOLOGÍA PROPUESTA .....	31
2.1 INTRODUCCIÓN .....	31
2.2 MODELADO E IDENTIFICACIÓN DE SISTEMAS COMPLEJOS.....	31
2.3 DIFERENTES PARADIGMAS DE MODELADO.....	32
2.4 METODOLOGÍA GENERAL PARA LA CONSTRUCCIÓN DE MODELOS.....	33
2.5 METODOLOGÍA PROPUESTA PARA LA CONSTRUCCIÓN DE MODELOS DIFUSOS .....	36
2.6 MODELADO DIFUSO.....	38
2.7 IDENTIFICACIÓN DIFUSA.....	39
2.8 RESUMEN DE LOS MODELOS DIFUSOS.....	40
2.8.1 Modelo Lingüístico .....	40
2.8.2 Modelo Impulso.....	41
2.8.3 Modelo Takagi-Sugeno.....	41
2.8.3.1. Inferencia en el Modelo TS.....	43
2.9 CONCLUSIONES .....	45
CAPITULO 3.....	47
PARTICIÓN DEL ESPACIO-PRODUCTO ENTRADA-SALIDA .....	47
EN REGIONES DIFUSAS .....	47
3.1 INTRODUCCIÓN .....	47
3.2 TÉCNICAS DE AGRUPAMIENTO ( <i>CLUSTERING</i> ).....	48
3.3 UTILIZANDO MÉTODOS DE CLUSTERING.....	48
3.4 MEDIDAS DE SIMILITUD.....	50
3.4.1 Distancia Euclídea y Distancia Euclídea ponderada .....	50
3.4.2 Distancia de Mahalanobis .....	50
3.5 CRITERIOS DE CLUSTERING .....	51
3.6 MÉTODOS DE CLUSTERING .....	51
3.7 CLUSTERING CERTERO Y DIFUSO.....	52
3.7.1 Clusters Certeros.....	52
3.7.2 Clusters Difusos.....	54
3.7.3. Clusters Posibilísticos .....	55
3.7.4. Obtención de Clusters con Fuzzy c-means .....	56
3.7.5 La funcional difusa c-means .....	56

3.7.6. El algoritmo fuzzy c-means .....	56
<b>3.8 NORMAS PRODUCTO-INTERNO.</b> .....	57
<b>3.9. CLUSTERING CON MATRICES DIFUSAS DE COVARIANZA</b> .....	60
3.9.1 El Algoritmo Gustafson-Kessel .....	60
<b>3.10 CONCLUSIONES</b> .....	62
<b>CAPITULO 4</b> .....	64
<b>CONSTRUCCIÓN DE SISTEMAS DIFUSOS</b> .....	64
<b>4.1 INTRODUCCIÓN</b> .....	64
<b>4.2 MODELADO DEL SISTEMA.</b> .....	65
<b>4.3 CONSTRUCCIÓN DEL MODELO DIFUSO BASADO EN EL CONOCIMIENTO.</b> .....	66
<b>4.4 CONSTRUCCIÓN DEL MODELO DIFUSO BASADO EN DATOS.</b> .....	67
4.4.1 Construcción de modelos difusos basados en plantillas.....	67
4.4.2 Construcción de modelos difusos a través de Clustering Difuso. ....	68
4.4.2.1. Construcción del Modelo para Sistemas Dinámicos.....	70
<b>4.5 CONCLUSIÓN.</b> .....	80
<b>CAPITULO 5</b> .....	81
<b>DESCRIPCIÓN DEL BIOPROCESO</b> .....	81
<b>5.1 INTRODUCCIÓN</b> .....	81
<b>5.2 COMPONENTES ESENCIALES PARA LA BIOREMEDIACIÓN</b> .....	82
<b>5.3 CARACTERIZACIÓN DEL SUELO Y ESTUDIOS DE TRATABILIDAD</b> .....	83
<b>5.4 DESCRIPCIÓN DEL SISTEMA CONSTRUIDO</b> .....	83
<b>5.5 FACTORES AMBIENTALES A CONSIDERAR EN UN PROCESO DE BIOPILAS.</b> .....	84
5.5.1 Nutrientes.....	84
5.5.2 Requisitos ambientales.....	84
<b>5.6 USO DE MATERIALES ORGÁNICOS EN UN SISTEMA DE BIOPILAS</b> .....	86
<b>5.7 DISEÑO DE UN SISTEMA DE BIOPILAS</b> .....	87
5.7.1 Mezclado y aireación de las pilas.....	90
5.7.2 Muestreo .....	90
5.7.3 Adición de nutrientes y residuo agrícola.....	90
<b>5.8 CONCLUSIONES</b> .....	91
<b>CAPITULO 6</b> .....	93
<b>MODELADO DIFUSO DEL BIOPROCESO</b> .....	93
<b>6.1 INTRODUCCIÓN</b> .....	93
<b>6.2 DESCRIPCIÓN DEL PROCESO DE MODELADO PROPUESTO PARA LA BIOPILA ESTUDIADA.</b> .....	94
<b>6.3 ANÁLISIS DEL PH.</b> .....	98
<b>6.4 HUMEDAD.</b> .....	100
<b>6.5 OXIGENACIÓN</b> .....	103
<b>6.6 TEMPERATURA DE LA BIOPILA</b> .....	103
<b>6.7 H TOTALES</b> .....	107
<b>6.8 COORDINADOR DIFUSO</b> .....	110
<b>6.9 CONCLUSIONES</b> .....	112
<b>CAPITULO 7</b> .....	113
<b>CONCLUSIONES GENERALES</b> .....	113
<b>REFERENCIAS</b> .....	115
<b>APÉNDICE A.</b> .....	119



<b>IMPLEMENTACIÓN EN MATLAB DE LOS ALGORITMOS UTILIZADOS.....</b>	<b>119</b>
Implementación del algoritmo difuso de clustering fuzzy c-means.....	120
Implementación del algoritmo difuso de clustering Gustafson-Kessel.....	122
Archivo SCRIP con Ejemplo del Algoritmo Fuzzy c-means.....	126
Archivo SCRIP para el Modelado de TEMPERATURA Vs TPHs.....	128



# INTRODUCCIÓN

Durante los últimos veinte años, la Lógica Difusa ha sido aplicada exitosamente a muchos problemas de modelado y control. Una de las razones de este éxito es que la lógica difusa provee una representación del conocimiento acerca de la operación y funcionamiento de sistemas complejos, semejante a la forma en que el humano realiza razonamientos con información ambigua e incompleta y los describe lingüísticamente en forma comprensible y amigable al ser humano. Así, uno de los objetivos de esta tesis es utilizar estas habilidades de la lógica difusa para extraer y explicar el comportamiento complejo de los sistemas, en términos lingüísticos. En esta introducción se describe un panorama genérico del tema tratado en este trabajo, así como el problema de investigación que presenta y se explica cómo se aborda su solución por medio del desarrollo de una metodología desarrollada.

## **Fundamentos generales**

Partimos de la premisa de que la característica más importante del pensamiento humano es que es predominantemente inexacto. Sin embargo, esta inexactitud, esencial para el manejo de sistemas reales, provee la habilidad para resumir los datos y enfocarse en aquella información que es relevante para la toma de decisiones. Esta inexactitud es algo muy opuesto a lo que una computadora actual puede hacer; de esto, se puede plantear la necesidad de contar con técnicas de Inteligencia Artificial (IA) que imite el razonamiento humano.

De las técnicas de IA que han emergido en los últimos 50 años, dos son las que más sobresalen: Las *Redes Neuronales Artificiales* (RNA).

Las redes neuronales artificiales realizan dos funciones principalmente: Aprender y Recordar. Recordar es el proceso de aceptar un estímulo de entrada y producir una respuesta de salida de acuerdo con la estructura de los pesos de las conexiones internas de la red (estos pesos en la red representan el “conocimiento distribuido”). El aprendizaje es el proceso de adaptar estos pesos de las conexiones internas de la red para producir el vector de salida deseado, en respuesta a un vector de estímulos presente en la entrada. Actualmente, las RNA son muy simples comparadas con el cerebro, pues no se autoregulan y no piensan.

La *Lógica Difusa* es otra de las técnicas importantes de IA. Mientras que las RNA emulan el aspecto físico del cerebro humano, la lógica difusa o lógica multi-valuada imita el modo de pensamiento humano. Dado que la fusividad es una propiedad del lenguaje, la lógica difusa se puede usar para razonar sobre conceptos inherentemente vagos e inciertos, ya que provee un esquema de representación y de cálculo para dichos conceptos.

Fundamentalmente, la lógica difusa es una generalización de la lógica clásica (o Booleana), en la cual un concepto puede poseer un grado de veracidad continuo en cualquier valor entre 0.0 y 1.0 a diferencia de la lógica Booleana, la cual aplica únicamente a conceptos que son completamente ciertos (teniendo un grado de verdad de 1.0) o completamente falsos (teniendo un grado de verdad de 0.0).

Lofti A. Zadeh concibió la técnica de la lógica difusa como una manera de solucionar problemas presentes en las ciencias suaves, particularmente aquellas que involucran interacción entre personas y/o entre personas y máquinas. La lógica difusa es particularmente útil en aquellas situaciones en las cuales la lógica convencional no es efectiva, tales como sistemas y dispositivos que no pueden describirse con precisión mediante modelos matemáticos, aquellos que presenten condiciones con incertidumbres significativas o condiciones contradictorias, y dispositivos o sistemas controlados lingüísticamente.

Hoy, ambas técnicas de la IA ya están en uso, en artículos comerciales y además, algunos científicos están intentando crear robots con capacidades de tipo-humanos.

La interpretación lingüística es una poderosa herramienta para analizar datos numéricos y se puede usar para obtener información útil sobre un sistema desconocido a modelar. Sin embargo, es conveniente remarcar que la lógica difusa ciertamente no es la solución para los problemas de modelado en general, pero es difícil negar su considerable potencial para aplicaciones prácticas – esto, básicamente, debido a que los métodos de diseño comunes están más cercanos al modo de pensar y de percibir de los humanos.

En términos generales, usar el marco teórico de la representación difusa, permite incluir diferentes fuentes de información (vaga e incierta) durante la obtención de un modelo, por ejemplo, la información adquirida por la experiencia a través del tiempo (conocimiento empírico), a la información que se obtiene a partir del tratamiento de datos representativos de un sistema mediante técnicas de identificación, lográndose un modelo más fidedigno de la realidad representada.

Considerando las propiedades del modelado difuso, el propósito de este trabajo, es el desarrollar una metodología que sea capaz de detectar las relaciones relevantes entre los datos correspondientes a una situación concreta; que sea universalmente aplicable (independiente del área de aplicación específica), y que permita resolver cada una de las características presentes en la información, utilizándose para esto, la Teoría de los Conjuntos Difusos y de la Lógica Difusa. Esta metodología debe permitir la incorporación de la información proveniente de varias fuentes, tales como el conocimiento de un experto, la información obtenida a partir de datos representativos, etc.. Además, la forma en que se debe presentar el modelo obtenido debe ser lo más claro posible a fin de que el experto lo

valide y, de ser necesario, lo complemente añadiendo información para situaciones no consideradas en dicho modelo.

Básicamente, existen dos categorías de modelos difusos las cuales difieren fundamentalmente en su habilidad para representar diferentes tipos de información.

La primera categoría incluye a los Modelos Lingüísticos (ML), los cuales están basados en colecciones de reglas del tipo *si-entonces*, que emplean predicados vagos y hacen uso del razonamiento difuso. En estos modelos las cantidades difusas están asociadas con etiquetas lingüísticas, y el modelo difuso es esencialmente una expresión cualitativa del sistema. Modelos de este tipo forman las bases para el modelado cualitativo que describe el comportamiento de un sistema usando un lenguaje natural.

La segunda categoría de modelos difusos está basada en el método de razonamiento de Takagi-Sugeno-Kang (TSK). Estos modelos están formados por reglas lógicas que tienen una parte antecedente difusa y un consecuente funcional. Esencialmente, son una combinación de modelos difusos y no difusos. Los modelos difusos basados en el método de razonamiento TSK integran la habilidad de los modelos lingüísticos para representar el conocimiento cualitativo, con un potencial efectivo para representar también información cuantitativa. Además, este tipo de modelo difuso permite una aplicación relativamente fácil de técnicas de aprendizaje para realizar identificación a partir de datos.

Los principios básicos de los modelos difusos fueron creados por Zadeh en [1], donde estableció tres características distintivas de esta nueva metodología, la cual provee una forma aproximada y efectiva para describir el comportamiento de sistemas, sean estos complejos o mal definidos, para admitir el uso de un análisis matemático preciso.

Estos principios fueron:

1. El uso de variables lingüísticas, en lugar de, o, en adición, a las variables numéricas.
2. La caracterización de relaciones simples entre variables por *declaraciones difusas condicionales*.
3. La caracterización de relaciones complejas por medio de *algoritmos difusos*.

Los primeros ejemplos de modelado difuso, inspirados por el trabajo contiguo sobre sistemas expertos, implementaban las ideas de Zadeh tratando de extraer el modelo difuso directamente del conocimiento del experto. Este método original de construcción de modelos difusos se inspiró en lo que llamaremos el *método directo*. En esta metodología, primeramente se describe el sistema de manera lingüística usando términos del lenguaje natural y posteriormente se traslada a la estructura formal de un modelo difuso con la ayuda de la capacidad representativa de la teoría del razonamiento aproximado. La descripción lingüística se construye subjetivamente, basándose en el conocimiento *a priori* del sistema.

Así, la fuente para derivar las reglas lingüísticas es el conocimiento empírico del sistema, directo del experto; es este conocimiento el que se expresa en forma de reglas

lógicas. Este método se puede considerar como una versión cualitativa de la construcción tradicional del modelo usado en ciencias de sistemas.

La primera dirección significativa de la lógica difusa en el modelado de sistemas complejos, en la duplicación de las acciones del operador; fue una manifestación de la gran capacidad de esta metodología para tratar con la complejidad del mundo real. Sin embargo, esta metodología directa para el modelado difuso, basada únicamente en el uso de la descripción del funcionamiento del sistema dada por un experto, presenta algunas limitaciones inherentes. Si el conocimiento experto del sistema es erróneo podemos obtener un mal modelo; buscando mayor objetividad en la construcción de modelos difusos, se propone tratar de desarrollar técnicas más formales que puedan usar los datos disponibles para incrementar el conocimiento humano, o aún generar conocimiento nuevo.

La segunda dirección en el desarrollo de modelos difusos, inspirados por la teoría de sistemas clásicos y desarrollos recientes en redes neuronales, está basada en el uso de los datos de entrada-salida de un proceso. En el lenguaje de teoría de sistemas, se puede considerar como un proceso de identificación.

La identificación de un modelo difuso consiste principalmente de dos fases: La primera, la identificación de la estructura del modelo difuso (identificación de la estructura) y la segunda, la estimación de los valores de los parámetros del modelo difuso (identificación de los parámetros). Hablando a *grosso modo*, la identificación de la estructura incluye la determinación de las variables de entrada y salida, las relaciones entre las variables (la estructura de las reglas), el número de reglas en la base de reglas, y la partición de las variables de entrada y salida por medio de conjuntos difusos. En general, la identificación de la estructura es un proceso difícil y extremadamente mal definido, y no fácilmente adecuado a técnicas automatizadas; es parcialmente simplificado si algún conocimiento experto sobre el sistema se usa en conjunción con los datos observados.

El problema de obtener la estructura de modelos difusos a partir de datos se presenta como uno de los temas importantes en la teoría de conjuntos difusos en los años posteriores a los conceptos originales de Zadeh sobre el modelado difuso. Las primeras ideas fundamentales para establecer las propiedades de las reglas difusas y su determinación a partir de los datos de entrada-salida provienen de Zadeh. El fundamento de esto se puede considerar el método de modelado denominado *modelado basado en una plantilla*. En esta metodología, la cual combina el uso del conocimiento del experto y de los datos, el experto provee valores lingüísticos en una plantilla, los cuales particionan el espacio de entrada-salida; los sub-conjuntos difusos son dados a priori. Estos valores en la plantilla se usan para definir las reglas potenciales para el modelo difuso del sistema.

Si la información sobre los valores lingüísticos de la plantilla no está disponible y solo se tiene datos de entrada-salida, entonces la estructura del sistema (relaciones entre las variables, estimación aproximada de las funciones de membresía de los conjuntos difusos antecedentes y consecuentes, y el número de reglas), se puede obtener usando algoritmos difusos de clustering para particionar el espacio entrada-salida. El método de clustering difuso Gustafson-Kessel provee una metodología sistemática para identificar las reglas más importantes a partir de los datos de entrada-salida del sistema.

En el modelado difuso de sistemas, el problema de identificación de parámetros está estrechamente relacionado a la estimación de las funciones de membresía de los conjuntos difusos o, alternativamente, a la relación difusa asociada con el modelo difuso.

## **Objetivo de la tesis.**

Nuestro objetivo es obtener una metodología que sea *amigable* para la integración del conocimiento *a priori*, durante la tarea de la búsqueda del modelado de un sistema. Consideramos que una metodología *fácil de usar* puede ser bien aceptada por aquellas personas encargadas de plasmar el conocimiento de los ingenieros directamente al proceso de modelado, considerando además, una metodología que tenga la característica de contar con algoritmos de aprendizaje para resolver el problema de la incertidumbre contenida en el proceso.

Nuestra propuesta es usar las ventajas que ofrecen los métodos de clustering pues son métodos no supervisados cuya principal característica es la búsqueda de aquella estructura contenida en los datos y cuyos resultados se utilizan para crear los conjuntos difusos y las reglas que especificarán el modelo difuso representativo del proceso generador de los datos.

De lo anterior, el objetivo planteado en nuestro trabajo es:

*Desarrollar una metodología basada en la teoría de la Lógica Difusa y en los Algoritmos Difusos de Clustering para generar el modelo de un sistema basado en información cualitativa y cuantitativa de su comportamiento.*

## **El caso de estudio.**

El caso de estudio utilizado para generar la metodología de modelado es el problema descrito brevemente a continuación.

En la bioindustria, existe una fuerte demanda por sistemas, que optimicen y controlen automáticamente los procesos biológicos, debido a que la complejidad operacional de estos, resulta en una intensa labor para su operación. Se han estudiado varios métodos de modelado, optimización y control para procesos microbianos, basados principalmente en modelos matemáticos determinísticos, sin embargo, la metodología determinística rara vez ha resultado exitosa para describir la operación de los procesos microbianos dada la dificultad para desarrollar un modelo matemático capaz de describir las complejas reacciones intracelulares de los microorganismos involucrados. Las condiciones operativas del bioproceso a menudo se optimizan a través del conocimiento adquirido durante las operaciones repetidas bajo el control de operadores habilidosos en efecto, en los procesos biológicos, la información cualitativa, tal como el conocimiento del operador del estado fisiológico de los microorganismos juega un papel significativo en la operación. Mientras que la operación basada en el conocimiento empírico se puede considerar útil desde un

punto de vista práctico, es difícil adaptarlo a un control por computadora debido a que las habilidades involucradas a menudo no se pueden cuantificar o ser incorporadas a un sistema de control tradicional. Para resolver este inconveniente en el campo de la biotecnología, se ha introducido una metodología basada en el conocimiento y en el modelado difuso.

La bioremediación es un proceso controlado o espontáneo en el cual las actividades microbiológicas se usan para destruir contaminantes. La bioremediación por biopilas involucra mecanismos para estimular y mantener la actividad de procesos biológicos intrínsecos, mediante los cuales microbios indígenas, vía la adición de nutrientes inorgánicos, convierten contaminantes en productos inertes. En el caso específico de suelos contaminados con hidrocarburos, en condiciones óptimas, los microorganismos indígenas utilizan a los hidrocarburos como su fuente de alimento y los convierten a dióxido de carbono y agua. Las biopilas son sistemas que usan el proceso de bioremediación para limpiar de manera económica suelos contaminados con hidrocarburos. Estos sistemas están proveídos de un medio para aireación que provee oxígeno a los microbios, un sistema de inyección y/o irrigación para proveer de nutrientes y humedecer las pilas y un sistema de colección de lixiviados para controlar el exceso de humedad.

En México existen muchos sitios contaminados con hidrocarburos y/o productos derivados de la industria del petróleo, por lo que en los últimos años se ha incrementado el interés sobre la remediación de estos sitios contaminados a través de la búsqueda de alternativas factibles y de bajo costo para implementarlas a gran escala. En el proceso que se desea modelar en este trabajo, se construyen las biopilas adicionando desechos agrícolas sobre los hidrocarburos sustraídos de un suelo contaminado con lodos de perforación; para realizar esto, se excavó y homogeneizó el suelo y se formaron cuatro biopilas de 1 tonelada. Se añadió paja a una razón de 97/3, se ajustó a un contenido de humedad del 30% y a una razón de C/N/P de 100/3/0.5. Las biopilas fueron aireadas mecánicamente y humedecidas con agua. Las concentraciones iniciales de los TPH (Total Petroleum Hydrocarbons) en el suelo fueron de 99,000 mg TPHs/kg. de suelo. Después de seis meses de tratamiento, las concentraciones de TPHs disminuyeron a 5,528 mg TPHs/kg. de suelo para la biopila B, esto es, se logró reducir la concentración de los TPHs a un 5% de su valor inicial.

Este proceso involucra una serie de incertidumbres debido a que es difícil cuantificar los mecanismos biológicos básicos y predecir las reacciones internas potenciales; por lo que, es difícil formular un modelo matemático bien-definido para el proceso; entre las fuentes de incertidumbre involucradas en este proceso, se cuentan entre otras, datos experimentales imprecisos e imprecisión inherente en los juicios de los expertos, en cambio, usualmente, los expertos asumen la responsabilidad de determinar los parámetros tales como la cantidad de nutrientes a añadir.

En este trabajo se desarrolla un modelo basado en lógica difusa que considera las incertidumbres mencionadas y permite representar mediante reglas el complejo proceso de bioremediación. La base de reglas del sistema consiste de una colección de reglas *si-entonces* obtenidas mediante el método de clustering difuso Gustafson-Kessel.



## Síntesis de la tesis

A continuación se hace una breve descripción del contenido de cada capítulo: primero se da una breve introducción a la teoría de los conjuntos difusos, posteriormente se describe la metodología para obtener un algoritmo difuso a partir de los datos representativos del sistema mediante el uso de algoritmos difusos de clustering; luego, se describen las características de complejidad de los sistemas biológicos, resaltando su dificultad para ser caracterizados por modelos determinísticos y, finalmente, se muestran los resultados obtenidos con nuestra propuesta

En el Capítulo 1, se mencionan brevemente algunos conceptos importantes de la teoría de los sistemas difusos necesarios para abordar con estos conceptos la metodología propuesta los capítulos siguientes. En el Capítulo 2 se hace énfasis respecto de los problemas encontrados para obtener el modelo de sistemas complejos; una vez conocidos estos problemas se hace un planteamiento respecto de la forma en que se pueden abordar dichos problemas. Este capítulo contiene el planteamiento de la metodología propuesta para resolver el problema de modelado de sistemas complejos; concluye con la descripción del sistema difuso Takagi-Sugeno. El Capítulo 3, contiene la descripción de los métodos difusos de clustering; esta descripción es completa pues estos métodos son fundamentales para la búsqueda del modelo difuso, por lo que se dan todas las características de los diferentes métodos de clustering que se pueden aplicar dependiendo del problema a resolver. La obtención del modelo difuso, a partir de los resultados del proceso de clustering, es el material contenido en el Capítulo 4; aquí, se mencionan las técnicas que se pueden usar para la obtención de la estructura y de los parámetros del modelo difuso. Se describe básicamente la operación de proyección para la obtención de las funciones de membresía y los métodos de mínimos cuadrados para la obtención de sus parámetros. El Capítulo 5, incluye una descripción detallada de las características del proceso, objeto de nuestra ejemplificación, y de lo adecuado de nuestra metodología para resolver la tarea de obtención del modelo a partir de datos. En el Capítulo 6, se muestran los resultados del proceso de modelado utilizando clustering difuso. En este capítulo, se muestra también que debido a sus características de funcionamiento, algunas de las partes del proceso se modelan mediante el método directo, en tanto que otras partes se modelan usando una plantilla, por lo que, para el modelo final, se recurre a los tres tipos de modelado. Dada la complejidad del proceso, se recurre a un sistema coordinador encargado de administrar los diferentes submodelos del sistema. El Capítulo 7 incluye las conclusiones generales.



# CAPITULO 1

## Conjuntos Difusos, Lógica Difusa y Sistemas Difusos.

En este capítulo se presenta una breve introducción a los conceptos de conjuntos difusos, lógica difusa y de sistemas difusos. El material presentado, se considera que es fundamental para abordar los capítulos posteriores. Para una descripción exhaustiva de los mismos se recomienda recurrir a la literatura existente en el tema, por ejemplo [2], [3].

### 1.1 Conjuntos Difusos

L.A. Zadeh en 1965 [1], introdujo el concepto de conjunto difuso y su teoría respectiva, la cual se considera como una extensión de la teoría de conjuntos clásica. En la teoría clásica de los conjuntos, denominados certeros, un elemento  $x$  cualesquiera dentro de un Universo  $X$ , se considera como miembro o no-miembro de un sub-conjunto  $A$  dentro del citado universo  $X$ . La pertenencia, determinada por una función característica o de discriminación  $\mu_A(x)$  de  $x$  en  $A$  está dada por:

$$\mu_A(x) = \begin{cases} 1, & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} . \quad (1.1)$$

Así, la función que transforma o proyecta elementos del conjunto universal  $X$  hacia el sub-conjunto que contiene sólo los elementos 0 y 1, se puede representar como:

$$\mu_A : X \rightarrow \{0,1\}.$$

Sin embargo, en la vida real se presenta un conjunto de situaciones donde una pertenencia bi-valuada (1.1) no es suficiente para realizar una descripción real de muchos fenómenos, debido a que fuerza a que exista una transición abrupta entre la pertenencia y la no-pertenencia de un elemento en un conjunto. Un ejemplo típico podría ser el problema donde, dada la edad de una persona, se requiera determinar si tal persona es o no joven (por ejemplo, para calcular algún riesgo en su salud). En otras palabras, se debe definir un conjunto que permita se caracterice con mayor realidad si una persona es o no joven. Para tal fin, podría ser obvio que las personas menores de 20 años son incondicionalmente

jóvenes, mientras que la gente mayor a 40 años es completamente gente adulta. Sin embargo, al evaluar a las personas con edades entre los 20 y los 40 años, estas tienen una evaluación diferente, así, pareciera ser más adecuada una transición gradual; esta transición gradual se puede implementar permitiendo un grado de pertenencia que varíe en el intervalo  $[0, 1]$ , como se muestra en la figura 1.1.

La función dada por la ecuación 1.1 puede ser generalizada de tal manera que los valores asignados a los elementos del conjunto universal puedan ser cualquier número dentro de un intervalo específico, en nuestro caso entre 0 y 1, indicando el grado de membresía de estos elementos en el sub-conjunto tratado. A la función que transforma elementos  $x$  del conjunto universal  $X$ , dentro del intervalo  $[0,1]$  para un sub-conjunto  $A$ , se le denomina función de membresía, FM, y al sub-conjunto  $A$  se le denomina conjunto difuso. Esto se representa como:

$$\mu_A : X \rightarrow [0,1].$$

Los conjuntos difusos se representan por conjuntos de pares ordenados, como:

$$\mu_A(x) = \{\mu_1/x_1, \mu_2/x_2, \dots, \mu_n/x_n\}, \quad (1.2)$$

donde cada valor de  $x$  se relaciona con su valor de membresía en  $A$ . Esta representación es muy flexible pues permite que la forma de la FM sea arbitraria. Una representación más práctica es del tipo funcional, tal como:

$$\mu_A(x) = f(x). \quad (1.3)$$

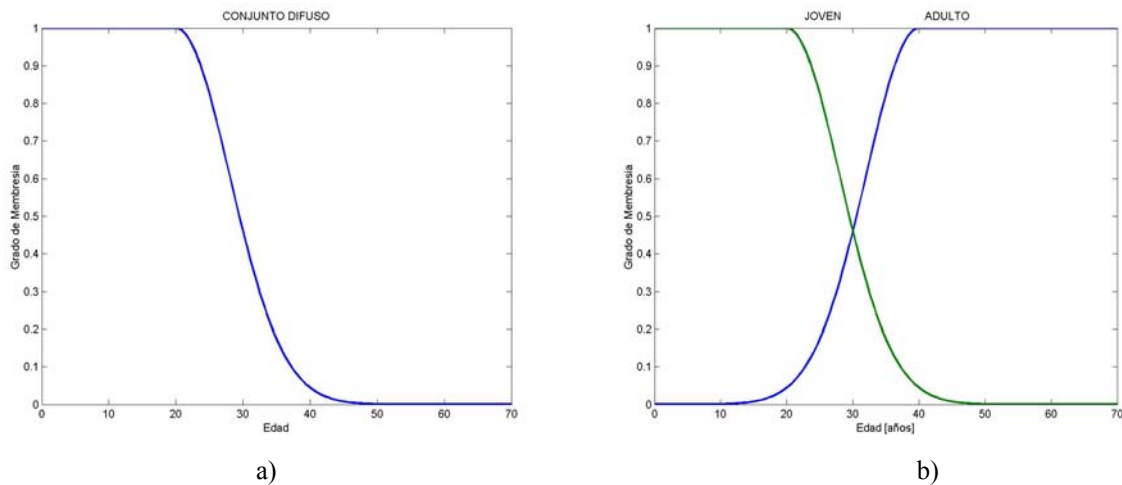


Figura 1.1 a) Conjunto Difuso caracterizando el concepto GENTE JOVEN. b) Conjuntos difusos Definidos para caracterizar los conceptos GENTE JOVEN y GENTE ADULTA

## 1.2 Propiedades básicas y tipos de Conjuntos Difusos

A continuación, se presentan algunas propiedades básicas de los conjuntos difusos.

a) La *altura* de un conjunto difuso  $A$ ,  $\text{alt}(A)$ , se define como

$$\text{alt}(A) = \sup_{x \in X} \mu_A(x). \quad (1.4)$$

Al conjunto difuso con una altura igual a 1, se le llama *normal*.

b) El *núcleo* de un conjunto difuso, es un sub-conjunto certero de  $X$ :

$$\text{nuc}(A) = \{x \in X \mid \mu_A(x) > 0\}. \quad (1.5)$$

c) *Soporte*. El soporte de un conjunto difuso  $A$  dentro de un universo de discurso  $X$  es el conjunto certero que contiene a todos los elementos de  $X$  que tienen grado de membresía diferente de cero en  $A$ . Si el soporte de un conjunto difuso es finito, se le llama *soporte compacto*.

Un conjunto difuso convexo se caracteriza por

$$\forall x_1, x_2, x_3 \in X, x_1 \leq x_2 \leq x_3 \rightarrow \mu_A(x_2) \geq \min(\mu_A(x_1), \mu_A(x_3)). \quad (1.6)$$

Los conjuntos difusos más usados son funciones de membresía trapezoidales o triangulares, esto por su facilidad de representación matemática en programas de cómputo.

Las funciones de membresía (FM) trapezoidales se definen por cuatro parámetros  $a \leq b \leq c \leq d$ , donde  $a = \min(\text{supp}(A))$ ,  $b = \min(\text{nuc}(A))$ ,  $c = \max(\text{nuc}(A))$ ,  $d = \max(\text{supp}(A))$  como se muestra en la figura 1.2. Las FM triangulares se consideran un caso particular de la función descrita con  $b = c$ .

Un segundo tipo de FM's son las FM's tipo "gaussianas", las cuales son más "suaves" y se determinan por dos parámetros. Estas FM's son diferenciables y tienen una representación compacta.

$$\mu_A(x) = e^{-\frac{(x-c)^2}{2\sigma^2}}. \quad (1.7)$$

## 1.3 Partición Difusa

Si se quiere generalizar la caracterización de un universo  $X$  de edades de personas en gente joven, adulta y anciana, la partición resultante, figura 1.3, puede denominarse una *partición difusa*, la cual consiste de conjuntos difusos que se identifican mediante la asignación de *etiquetas lingüísticas* (términos). Como se observa en la figura 1.3, se denomina variable

lingüística, a aquella variable cuyo universo de discurso está caracterizado mediante conjuntos difusos.

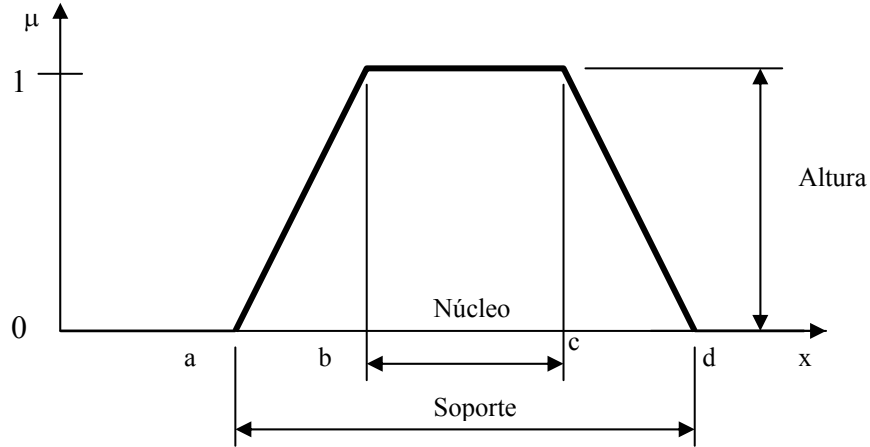


Figura 1.2. Representación de la altura, soporte y núcleo de un conjunto difuso.

El traslape entre los conjuntos difusos es la propiedad que permite caracterizar la incertidumbre entre dos conceptos. En efecto, la membresía parcial (difusa) de un elemento, en este caso una edad, en dos conjuntos al mismo tiempo, permite describir conceptos en los cuales no se define con certeza la frontera entre tener y no tener una propiedad (en nuestro ejemplo, no es fácil decidir si una persona de 60 años es adulta o anciana). Algo muy importante es que la utilización del marco teórico de los conjuntos difusos y las etiquetas lingüísticas permite el traslado del ámbito de los números (las edades) a un ámbito abstracto (los conjuntos difusos), lo cual es una forma más natural en que las personas tratan la información.

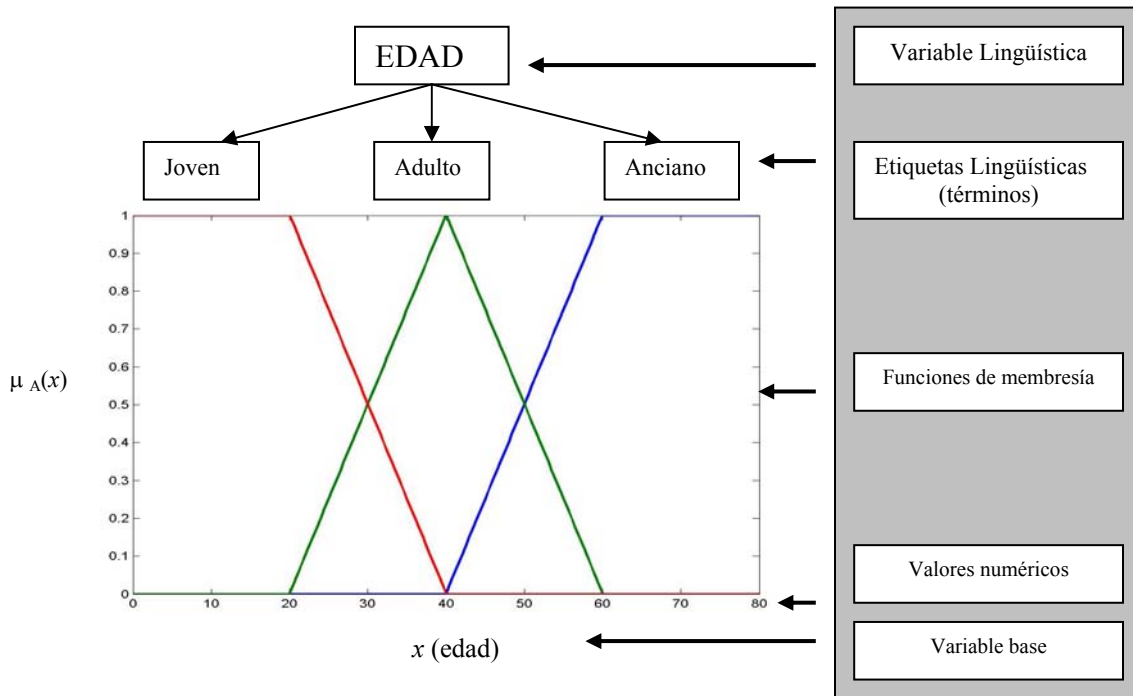


Figura 1.3. Caracterización del Universo de discurso edad de una persona

En los conjuntos bien definidos cada  $x$  tiene un valor de membresía diferente de cero para al menos un conjunto difuso, esto es:

$$\forall x \in X, \exists i, \mu_{A_i}(x) > 0, \quad (1.8)$$

o, escrito de manera alternativa:

$$\forall x \in X: \sum_{s=1}^S \mu_{A_s}(x) > 0, \quad (1.9)$$

donde  $S$  es el número de sub-conjuntos difusos que forman la partición. Una partición que satisface la ecuación (1.9) tiene la propiedad llamada de *cobertura*.

Un tipo de partición particularmente interesante es aquella en la que:

$$\sum_{s=1}^S \mu_{A_s}(x) = 1, \quad \forall x \in X, \quad (1.10)$$

la cual, se usa a menudo para referirse a una *partición difusa*. En el caso de una partición difusa, en la cual se considera el traslape de los conjuntos que la forman, para cada elemento  $x$  la suma total de membresías es igual a 1, en tanto que puede pertenecer a dos sub conjuntos difusos como máximo. Por ejemplo, la partición en la figura 1.3 es una partición difusa.

## 1.4 Operaciones con Conjuntos Difusos

Consideramos las operaciones básicas con conjuntos conocidas de la teoría de conjuntos clásicos tales como unión, intersección y complemento. Las extensiones de estas operaciones a la teoría de conjuntos difusos no son definidas unívocamente (como en la teoría de conjuntos clásicos) debido al hecho de que una función de membresía puede tomar cualquier valor en el intervalo entre  $[0, 1]$ .

Las operaciones de intersección y unión, en general, se representan por normas triangulares (T-normas) y conormas triangulares (T-conormas o S-normas), respectivamente.

Las t-normas más comunes son el operador mínimo y el operador producto, ilustrados en la figura 1.4.

$$A \cap B = \min(\mu_A(x), \mu_B(x)), \quad (1.11)$$

$$A \cap B = \mu_A(x) \mu_B(x). \quad (1.12)$$

Las s-normas más comunes son respectivamente el operador máximo y el operador suma probabilística. (Figura 1.5).

$$A \cup B = \max(\mu_A(x), \mu_B(x)), \quad (1.13)$$

$$A \cup B = \mu_A(x) + \mu_B(x) - \mu_A(x) \mu_B(x). \quad (1.14)$$

El complemento de un conjunto difuso  $A$ , se define típicamente como

$$\bar{A} = 1 - \mu_A(x). \quad (1.15)$$

Esta operación se muestra gráficamente en la figura 1.6

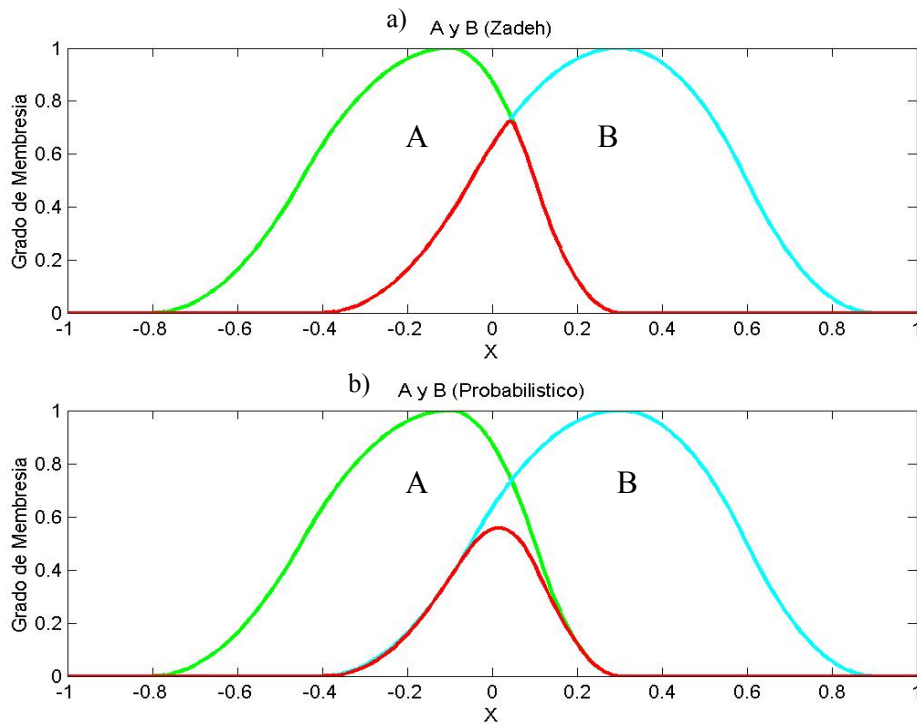


Figura 1.4 a) Operador Mínimo aplicado a los conjuntos difusos A y B; b) Operador producto aplicado a los conjuntos difusos A y B



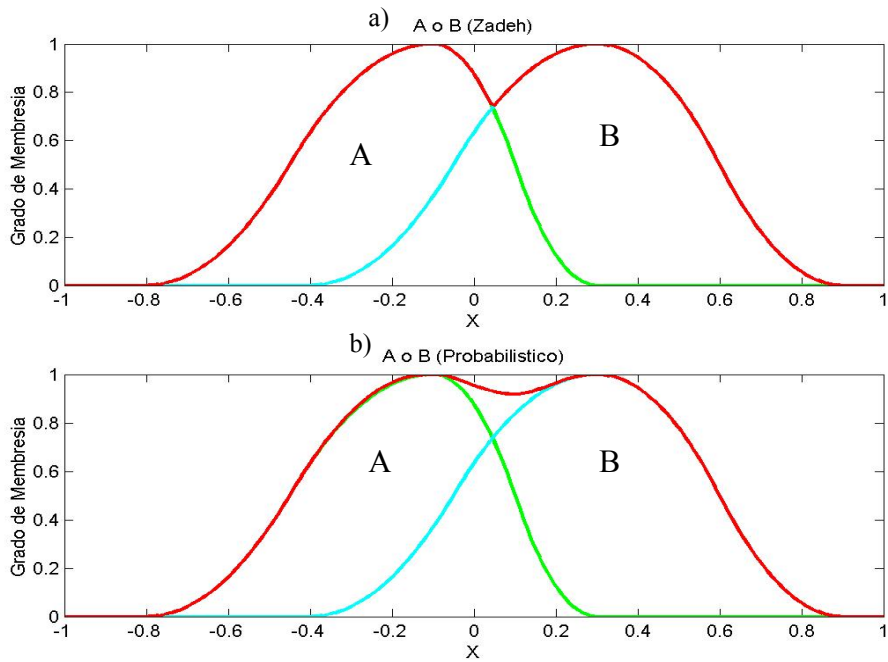


Figura 1.5 a) Operador Máximo aplicado a los conjuntos difusos A y B; b) Operador suma probabilística aplicado a los conjuntos difusos A y B.

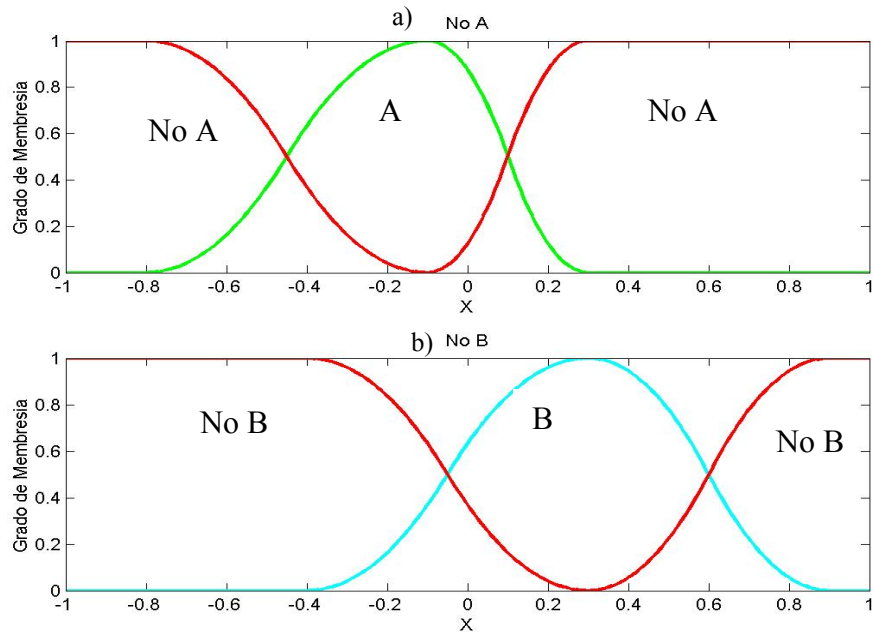


Figura 1.6 Complemento de los conjuntos difusos A y B.

De la misma manera que la teoría de los conjuntos clásicos sirve como base para la lógica clásica, la teoría de los conjuntos difusos sirve como base para la lógica difusa. Las operaciones teóricas con conjuntos difusos constituyen la base para las operaciones en la lógica difusa. Esto significa que las operaciones definidas para conjuntos (unión, intersección y complemento), tienen una operación lógica correspondiente – “o”, “y” y “no”- respectivamente y, de manera similar, están basadas en una t-norma, s-norma y en el complemento.

## 1.5 Sistemas Difusos

Un sistema difuso, es aquel sistema estático o dinámico el cual hace uso de los conjuntos difusos y su correspondiente marco teórico matemático. Los conjuntos difusos pueden estar involucrados de varias maneras en el modelo del sistema.

- Los sistemas difusos más comunes son aquellos definidos por medio de reglas *si – entonces*: denominados *sistemas difusos basados en reglas*. Este trabajo se enfoca en este tipo de sistemas difusos únicamente.
- Los sistemas difusos pueden servir para diferentes propósitos, tales como modelado, análisis de datos, predicción o control de procesos. En este trabajo al sistema difuso basado en reglas se le llama simplemente *modelo difuso* independientemente de su eventual propósito.

La teoría de los Conjuntos Difusos y de la Lógica Difusa provee los medios para la construcción de sistemas difusos [1]. Los sistemas difusos a obtener consisten de una serie de reglas que especifican una relación lingüística entre las etiquetas lingüísticas de las variables de entrada y de salida del sistema. Una regla difusa, como se muestra en la ecuación (1.16), es una declaración donde las premisas y los consecuentes consisten de *proposiciones difusas* que son declaraciones como “*x es grande*”, conectando una variable real con una etiqueta lingüística definida para tal variable. Por ejemplo, una regla esta definida como,

$$\text{Si } U_1 \text{ es } A_{1r} \text{ Y } U_2 \text{ es } A_{2r} \dots \text{ Y } U_N \text{ es } A_{Nr} \text{ ENTONCES } V_1 \text{ es } B_{1r}. \quad (1.16)$$

$A_{ir}$  y  $B_{jr}$  denotan las etiquetas lingüísticas de la *iésima* variable de entrada lingüística  $x_i$  y la *jésima* variable de salida lingüística  $y_j$  asociadas con la *résima* regla.

Mediante la ecuación (1.16) se dan las relaciones en términos lingüísticos. A fin de asociarlas con los valores numéricos, se debe realizar un proceso de inferencia, el cual se describe en la siguiente página. Este proceso de inferencia implementa la transformación o proyección entre las variables lingüísticas  $U_i$  y  $V_j$ , y las correspondientes variables  $x_i$  e  $y_j$ .

Una representación de un sistema que utiliza una representación basada en sub-conjuntos difusos se describe en la figura 1.7, donde la base de reglas contiene un conjunto de reglas lógicas SI-ENTONCES definidas sobre las variables del sistema; la base de datos almacena el conjunto de FMs de las etiquetas lingüísticas usadas en el conjunto de reglas. Estas dos

bases se puede considerar que conforman la capa lingüística (o base de conocimiento) del sistema.

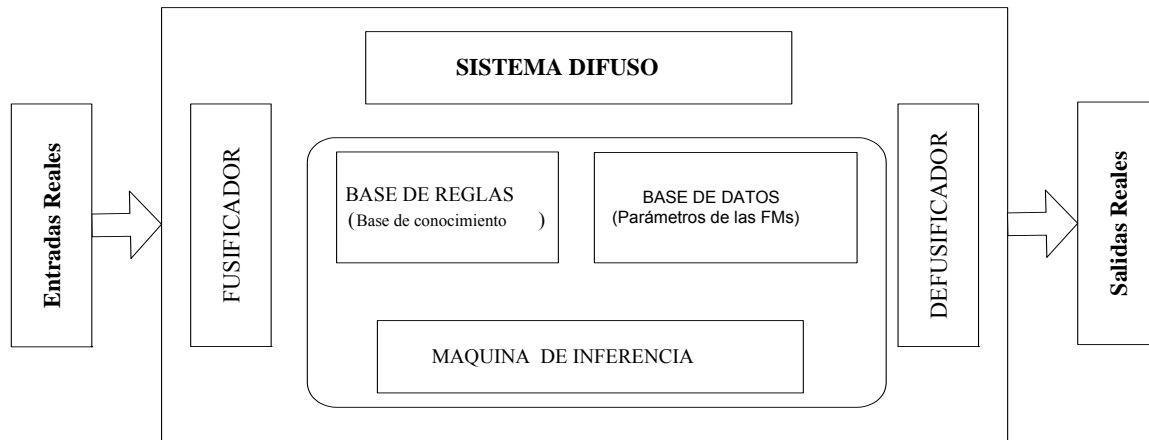


Figura 1.7. Esquema a bloques de un Sistema Difuso genérico

El proceso de inferencia consiste de un fusificador (convierte una variable certera en una variable difusa para permitir la aplicación de las reglas lógicas), una máquina de inferencia difusa (el algoritmo que calcula en qué grado se activa cada regla y las combina en un sistema difuso de salida) y un defusificador (convierte el conjunto de variables difusas en un valor certero para permitir que la salida de un sistema difuso se pueda aplicar a un sistema no difuso). El bloque de razonamiento de un sistema difuso está formado por la máquina de inferencia difusa, la base de reglas y la base de datos.

El algoritmo de inferencia consiste de seis pasos y para su análisis nos auxiliaremos de la figura 1.8. En esta figura se muestra la forma en que se desarrolla este proceso de inferencia.

1. El mecanismo de inferencia opera sobre conjuntos difusos para producir otros conjuntos difusos. En general, las variables de entrada al sistema son valores certeros y por lo tanto, estos se deben convertir a conjuntos difusos. Este es el primer paso en el mecanismo de inferencia, y se le denomina *fusificación*. En esta acción, la incertidumbre, la imprecisión o inexactitud de la variable de entrada se puede modelar usando conjuntos difusos para representarla.
2. *Evaluar las proposiciones*, es decir, evaluar las proposiciones en las premisas de las reglas en términos numéricos. Se determina el valor de membresía de  $x$  respecto de los conjuntos difusos definidos. Acción de *fusificación*.
3. *Conjunción* de las premisas. El operador  $\wedge$  que concatena las proposiciones en las premisas corresponde a una  $t$ -norma. Aquí se calcula el grado de activación (también, *grado de satisfacción*, *intensidad de encendido*) de una regla.

\* En este escrito usaremos el término “fusificador” para la palabra en Inglés “fuzzifier” y la palabra “defusificador” para “defuzzifier”



*identificación difusa* se usa para denominar a la tarea de adquirir o entonar sistemas difusos a partir de datos.

Considerando el proceso clásico para modelado de sistemas difusos, se puede ver un cierto paralelismo entre éste y el modelado con lógica difusa, pues el diseño basado en el conocimiento es análogo al modelado mediante principios básicos, mientras que el proceso de identificación difusa se puede considerar como perteneciente a la misma clase que los métodos estadísticos usados en la identificación de sistemas. En el modelado clásico, a menudo se efectúa una combinación de metodologías en las cuales se usa la física para especificar una ecuación diferencial que se cree representa el comportamiento del sistema y, posteriormente, se realizan experimentos para determinar algunos parámetros o funciones del sistema. De manera similar, en el diseño de sistemas difusos, cada vez más, se usan combinaciones de metodologías a fin de aprovechar sus mejores cualidades para la obtención del modelo.

Según Yager y Filev [4], el diseño de sistemas difusos se puede considerar como un algoritmo genérico, el cual consiste de seis pasos:

- i) Seleccionar las variables de entrada y salida;
- ii) Determinar el universo de discurso para cada variable;
- iii) Seleccionar el mecanismo de razonamiento adecuado para la formalización del modelo difuso;
- iv) Determinar las etiquetas lingüísticas que particionarán las variables;
- v) Crear el conjunto de reglas lingüísticas que representen las relaciones entre las variables del sistema;
- vi) Evaluar lo adecuado del sistema resultante para representar el problema a modelar.

A menudo, el resultado de este algoritmo consiste únicamente en un sistema difuso preliminar, pues durante la fase de evaluación del sistema el índice de desempeño puede no ser lo que se esperaba.

Grosso modo, los pasos i-v se pueden considerar como la determinación de la estructura del sistema difuso, por lo que posteriormente se requiere de una fase de identificación de parámetros, durante la cual, se obtienen los valores óptimos de los parámetros de las FMs. Desde luego, se parte de la premisa de que los datos de entrada-salida del sistema son representativos del comportamiento óptimo del mismo.

Sin embargo, para sistemas complejos, no siempre queda claro cuáles variables se deben usar como entradas al sistema. En este caso, las fuentes de información típicas para hacer la elección son el conocimiento a priori, las consideraciones sobre el comportamiento del proceso y el proceso de modelado.

Para la selección del mecanismo de razonamiento adecuado (que incluye la especificación del tipo de sistema, los operadores de inferencia, el método de defusificación y el tipo de FMs) los factores decisivos son nuevamente el propósito del modelado y el tipo de conocimiento disponible.

El diseño de sistemas difusos aún es algo muy dependiente de la aplicación por lo que no se puede definir un algoritmo de diseño exacto. Sin embargo, la guía genérica mencionada provee un marco teórico confiable para el diseño de sistemas difusos.

## **1.7 Conclusiones**

En este capítulo se presentaron algunos conceptos considerados relevantes para la construcción de sistemas difusos. El material presentado constituye sólo una pequeña parte de la teoría de los conjuntos difusos y de la lógica difusa, pero es suficiente para abordar los temas correspondientes al modelado de sistemas difusos, resaltando lo correspondiente a sistemas representados por reglas del tipo SI-ENTONCES. Se describe un procedimiento de seis pasos para definir el proceso de inferencia, mediante el cual se evalúan las reglas del sistema y se obtiene un resultado de ello.

El empleo de conjuntos difusos para la caracterización de variables lingüísticas permite modelar información incierta.

La arquitectura de los sistemas difusos, los hace muy útiles para problemas con interacción hombre-máquina pues posibilitan que la experiencia y el conocimiento humano sean expresados en términos lingüísticos, los cuales son difíciles de implementar en otras metodologías.

## CAPITULO 2

### Metodología Propuesta

#### 2.1 Introducción

En este capítulo se aborda el problema de como obtener el modelo de sistemas complejos, con comportamientos no lineales o parcialmente desconocidos, mediante el uso de técnicas basadas en la teoría de los conjuntos difusos y de la lógica difusa. Se demuestra que la metodología propuesta, denominada modelado difuso, permite obtener soluciones de modelado para aquellos sistemas que presentan dificultades en el uso de los métodos convencionales, principalmente debido a la incertidumbre por comportamientos no lineales y a la carencia de un conocimiento preciso sobre el comportamiento. Se describe asimismo un método para desarrollar modelos difusos a partir de los datos experimentales. En general, para la construcción, la validación y el análisis de modelos, la información disponible es de tipo heterogéneo (datos numéricos, conocimiento cualitativo y principios básicos del sistema); no obstante, nuestro marco teórico permite usar estos tipos de información en forma integrada de manera efectiva. El modelo obtenido puede ser parte de un algoritmo de control en tiempo real, o puede servir para analizar el proceso bajo estudio, esto con el fin de obtener una mejor comprensión del mismo y mejorar las tareas de monitoreo o diagnóstico.

#### 2.2 Modelado e identificación de sistemas complejos.

El desarrollo de modelos matemáticos es una tarea crucial para abordar el estudio de sistemas reales. Estos modelos se pueden usar por ejemplo, para realizar simulaciones, analizar el comportamiento, coadyuvar a una mejor comprensión de los mecanismos subyacentes, durante el diseño de nuevos procesos, y para el control de sistemas. El desarrollo de un modelo matemático que represente adecuadamente la realidad es una tarea importante. Si el modelo no es suficientemente preciso, los subsecuentes pasos de análisis, predicción, síntesis del controlador, etc., no pueden ser exitosos. Sin embargo, existe un compromiso obvio entre la precisión necesaria del modelo y su complejidad. Los modelos deben proveer información al nivel de precisión (abstracción) más relevante, y cuando sea apropiado, eliminar detalles innecesarios. Si el modelo es muy simple, no puede representar adecuadamente las características del sistema a estudiar y no sirve a su propósito. Sin

embargo, cuando el modelo tiene que implementarse prácticamente no debe ser muy complejo.

Muchos sistemas no son adecuados a las metodologías convencionales de modelado debido entre otras causas a la falta de conocimiento formal y preciso sobre el sistema, a un comportamiento fuertemente no lineal, a un alto grado de incertidumbre en su comportamiento, a características variantes en el tiempo, etc. Ejemplos de tales sistemas se pueden encontrar en la industria de procesos, en sistemas de manufactura flexible, en ingeniería aeroespacial, en ingeniería (bio)química, y también en dominios como el de la ecología, en los sistemas sociales y en el ámbito financiero.

### **2.3 Diferentes paradigmas de modelado.**

Tradicionalmente, la tarea de modelado es vista como la conjunción de, por una parte, la comprensión completa de la naturaleza y el comportamiento del sistema, y por otra, de un tratamiento matemático adecuado, lo cual de origen a un modelo útil. Esta metodología usualmente es denominada modelado tipo “*caja blanca*” (principios básicos, físicos, mecánicos). Sin embargo, en la práctica, cuando se consideran sistemas complejos y pobremente comprendidos, un factor limitador muy severo es el requerimiento de una buena comprensión de los fundamentos físicos del problema. Por ejemplo, las dificultades encontradas en el modelado tipo caja blanca convencional pueden ser debido a: una pobre comprensión del fenómeno subyacente, valores imprecisos de varios parámetros del problema, o a la complejidad del modelo resultante. Una comprensión total del mecanismo subyacente es virtualmente imposible para la mayoría de los sistemas reales, sin embargo, la tarea de conjuntar un grado de conocimiento aceptable necesario para el modelado físico puede ser muy difícil, consumir mucho tiempo y ser costosa. Aún si se determina la estructura del modelo, queda el problema de obtener valores precisos para los parámetros, es necesaria una etapa de identificación en donde, la tarea de identificación del sistema consiste en estimar los parámetros a partir de los datos medidos del sistema.

La metodología denominada como tipo “*caja negra*” supone que el proceso bajo estudio se puede aproximar usando alguna estructura suficientemente genérica, la cual se puede usar como un aproximador genérico de funciones. Consecuentemente, el proceso de modelado se reduce a postular una estructura adecuada del aproximador a fin de capturar correctamente la dinámica y las no linealidades del sistema. En el modelado tipo caja negra, la estructura del modelo difícilmente se relaciona a la estructura del sistema real. El problema de identificación consiste en estimar los parámetros del modelo. Usualmente, se pueden desarrollar modelos tipo caja negra con relativa facilidad si se dispone de datos representativos del proceso y sin requerir un conocimiento específico del problema. Un inconveniente severo de esta metodología es que la estructura y los parámetros de estos modelos comúnmente no tienen significado físico alguno. Tales modelos no pueden usarse para analizar el comportamiento del sistema de otra forma que no sea por simulación numérica, tampoco pueden escalarse hacia arriba o hacia abajo cuando se operen a una escala diferente los procesos, y por lo tanto, son menos útiles para la práctica industrial.



Existe un rango de técnicas de modelado que combinan las ventajas de las metodologías tipo caja blanca y caja negra tal, que las partes del sistema conocidas se modelan usando el conocimiento físico y, las partes desconocidas o menos ciertas, se aproximan de una manera tipo caja negra, usando datos del proceso. Estos métodos a menudo se denotan como *modelos híbridos*, *semi-mecanísticos* o tipo *caja gris*.

Un inconveniente común en la mayoría de las metodologías clásicas de modelado es que no pueden usar de manera efectiva la información extra disponible tal como el conocimiento y la experiencia de los ingenieros y los operadores, adquiridos a través del tiempo durante la operación diaria del proceso, pues dicha información a menudo es de naturaleza imprecisa y cualitativa. El hecho de que los humanos seamos capaces de manejar tareas complejas bajo considerable incertidumbre, ha estimulado la búsqueda de paradigmas alternativos para realizar la tarea de modelado. Se han desarrollado una serie de métodos también conocidos como metodologías inteligentes, las cuales emplean técnicas inspiradas en los sistemas biológicos y en la inteligencia humana, para desarrollar modelos y controladores para sistemas dinámicos. Estas técnicas exploran esquemas de representación alternos usando, por ejemplo, lenguaje natural, reglas, redes semánticas o modelos cualitativos, y métodos formales para incorporar información extra relevante

## **2.4 Metodología general para la construcción de modelos.**

Estableciendo que se dispone de un conjunto de datos empíricos representativos del sistema y de algún conocimiento a priori del mismo, el problema de *modelado* consiste en encontrar un *conjunto de modelos*, que contenga aquel que describe adecuadamente los aspectos deseados del sistema [5].

Conceptualmente, el problema de modelado del sistema y el problema de identificación del sistema están íntimamente ligados. Si se considera que ambos transforman o proyectan el conocimiento *a priori* y los datos empíricos del sistema, a un modelo. Esto es, denominamos como el *proceso de identificación*, a las actividades de búsqueda de la estructura del modelo y búsqueda de un vector que contenga los parámetros del sistema.

Así, se puede especificar el problema de identificación de la siguiente manera:

Dados, un conjunto de modelos, un conjunto de datos empíricos y un criterio objetivo que evalúe el error entre la predicción de un modelo dado y los datos empíricos, encontrar un modelo dentro del conjunto de modelos que minimice tal criterio.

Típicamente, la construcción del modelo es un proceso altamente iterativo; esto se puede observar en el esquema de la figura 2.1, el cual describe las diversas tareas que se deben realizar y la forma jerárquica en que se resuelven.

Algunas generalidades respecto del proceso de construcción de un modelo son:

1. Formalización y análisis del conocimiento *a priori*, diseño de experimentos (cuando aplique) y colección y análisis de datos. En el proceso de modelado, este paso es muy importante, puesto que el éxito de las tareas de modelado e identificación dependen fuertemente de la información contenida en los datos y, de la cantidad y confiabilidad del conocimiento *a priori*.
2. La acción de modelado ha sido definida como la elección de un modelo óptimo a partir de un conjunto de modelos genéricos, y usualmente se basa en el conocimiento *a priori* y/o en algún análisis preliminar de los datos. Por un lado, con un buen conocimiento *a priori*, el conjunto de modelos típicamente contendrá modelos mecanísticos (formales), por otro lado, un conjunto de modelos tipo caja negra es útil puesto que son intrínsecamente genéricos, esto es, existe una razonable oportunidad de que exista un modelo dentro del conjunto de modelos tipo caja negra, que pueda describir cualquier sistema dentro de una gran clase de sistemas. Típicamente, el conjunto de modelos tipo caja negra es muy grande, aunque se puede reducir considerablemente por la aplicación de algún conocimiento *a priori*, esto es importante, puesto que mientras que es necesario que el conjunto de modelos sea grande, al mismo tiempo, éste debe ser tan restringido como sea posible para reducir el costo de la búsqueda a través de dicho conjunto. Se puede establecer que uno puede poner una gran cantidad de recursos en la parte de modelado para reducir la complejidad en la identificación y vice versa. En la práctica, para la elección del conjunto de modelos, siempre influyen las restricciones de tiempo, de tipo operacional y económicas, entre otras.
3. Típicamente, el conjunto de modelos está parametrizado por elementos de espacios función, espacios discretos y espacios euclidianos. La identificación de la estructura del modelo trata con la búsqueda y optimización en espacios función y espacios discretos. Típicamente este es un proceso iterativo, y la estructura del modelo en el paso previo de la iteración se usa usualmente como base para la selección de una estructura del modelo más promisoría, ya sea refinada o simplificada. A menudo, el problema de la identificación de la estructura es un problema combinatorio, por lo que, el conocimiento heurístico evita muchas iteraciones en este lazo.
4. El problema de *identificación de parámetros* se define como la identificación de parámetros en un espacio euclidiano y es un problema que ha sido muy estudiado en la literatura.
5. Después de que aparentemente se ha encontrado el mejor modelo en el conjunto de modelos, éste se debe *validar*. Una prueba de validación puede incluir pruebas de predicción con un conjunto de datos de validación separados. Si el modelo no pasa la prueba de validación, se debe tratar con un conjunto de modelos más general. Otra alternativa puede ser revisar y mejorar el conocimiento *a priori*, o bien, se deben coleccionar datos más significativos del fenómeno.

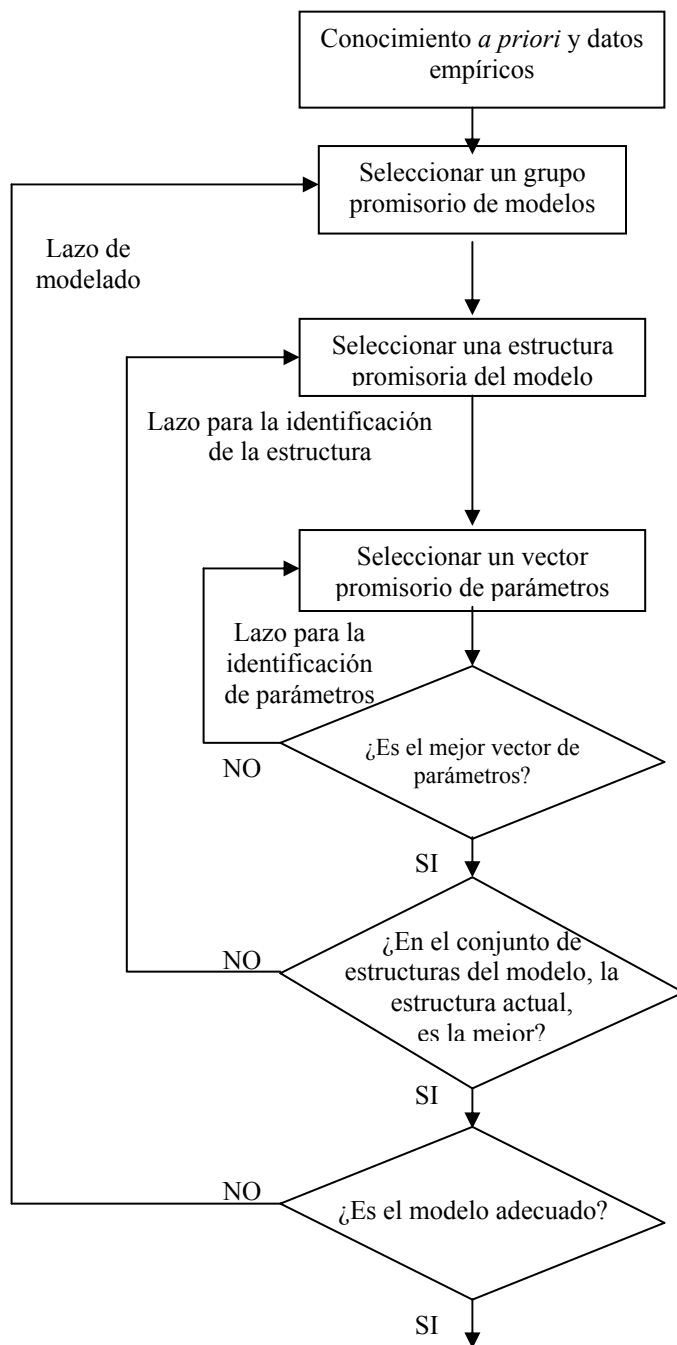


Figura 2.1. Esquema que muestra el proceso iterativo de modelado e identificación

Obsérvese que cada uno de los pasos en el procedimiento se puede realizar ya sea manualmente o bien automáticamente por una computadora. Debido a que los lazos internos realizan tareas bien definidas, basadas en información cuantitativa y en datos del sistema, estos son más adecuados para ser implementados en un programa de cómputo. Por

el contrario, y debido a que está basado en conocimiento cualitativo, el lazo externo es más difícil de formalizar e implementar en un programa de cómputo.

Algunas de las desventajas de esta metodología son su falta de capacidad para incorporar el conocimiento *a priori* de una manera fácil y transparente por que el proceso de modelado puede ser muy demandante de recursos o no factible si el sistema no está muy bien entendido, etc.

## 2.5 Metodología propuesta para la construcción de modelos difusos

En este caso, se aprovechan las ventajas de la teoría de los conjuntos difusos y de la lógica difusa para representar información y se resuelve el problema de la construcción del modelo de un sistema, empleando algoritmos difusos de clustering para realizar el proceso de identificación.

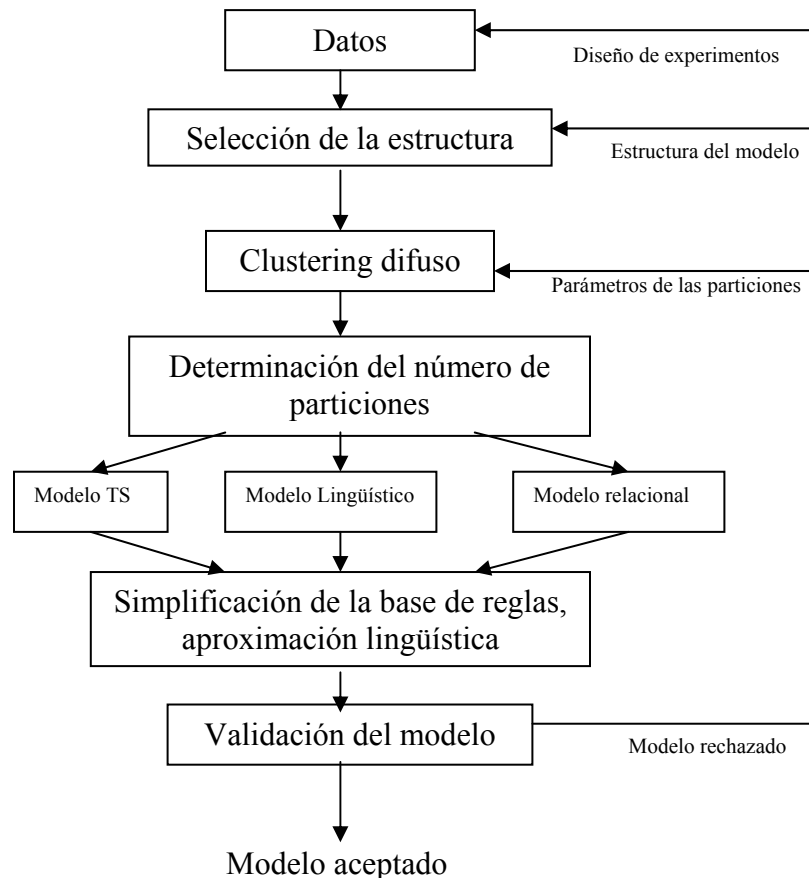


Figura 2.2 Esquema en que se muestra la metodología propuesta para la construcción de modelos.

En este trabajo se considera que los datos disponibles para representar el sistema, son representativos del sistema, esto es, que han sido adecuados y por lo tanto, no se hará ningún tratamiento a los mismos por nuestra parte.

El esquema general de búsqueda se adecúa para poder considerar los diferentes tipos de modelos difusos existentes y también para poder incorporar las diferentes familias de operadores difusos. La figura 2.2 muestra el esquema que se sigue para la construcción del modelo difuso empleando algoritmos difusos de clustering en las tareas de identificación. Se ha mostrado que los modelos difusos se pueden aplicar eficientemente a sistemas no lineales complejos, los cuales no son tratables por medio de métodos lineales convencionales.

Se describen a continuación las partes de esta metodología y se establecen algunos aspectos prácticos concernientes al diseño de experimentos y a la validación del modelo.

1. **Diseño de experimentos de identificación y colección de un conjunto de mediciones representativas del proceso.** Este es un paso inicial importante para cualquier método de modelado puesto que determina el contenido de información en el conjunto de datos. Generalmente se usan los datos obtenidos durante la rutina del proceso de operación del sistema. Si el conjunto de datos de identificación no cubre el intervalo completo de operación, el usuario puede añadir reglas adicionales. No es necesario ningún pre-procesamiento adicional a los datos, pues dada la robustez de los métodos de clustering, estos pueden tratar con datos con ruido. Naturalmente, con niveles de ruido crecientes, se incrementará la varianza de los parámetros estimados, el número de reglas, la posición de las funciones de membresía, etc.
2. **Selección de la estructura del modelo.** Como entradas y salidas del modelo, se deben seleccionar las variables relevantes del sistema y, cuando se estén identificando sistemas dinámicos, se debe elegir la estructura y el orden del modelo dinámico. Después de que se defina una estructura adecuada, se construirá una matriz  $Z$  con los datos de identificación.
3. **Clustering de los datos.** Antes de realizar la partición, se debe definir un número inicial de clusters  $c$ , y los parámetros de clustering  $m$  y  $\epsilon$ . Nótese que para conjuntos grandes de datos, el procedimiento de clustering puede ser computacionalmente demandante.
4. **Determinación del número de clusters.** El número adecuado de clusters se puede determinar usando medidas de validez de los clusters, la técnica de conjunción de clusters compatibles o una combinación de ambas. Obviamente, este paso involucra varias repeticiones del paso 3 con diferentes valores para  $c$  y diferentes matrices de partición iniciales del universo de datos entrada-salida.
5. **Generación de las reglas.** Las funciones de membresía se obtienen proyectando los clusters encontrados, sobre las variables antecedentes. Los parámetros consecuentes se extraen de la estructura propia de la matriz de covarianza del cluster o bien se estiman por mínimos cuadrados.
6. **Validación del modelo.** La validación del modelo difuso se puede realizar de varias maneras, tales como la validación estándar a través de simulaciones numéricas comparándola con los datos del proceso, el análisis de los modelos consecuentes

lineales y el análisis de la cobertura del espacio de entrada por las reglas. Para una base de reglas incompleta, se pueden proveer reglas adicionales con base en el conocimiento *a priori* o por la linealización local de modelos de principios básicos. Los antecedentes de estas reglas se pueden crear de las combinaciones no usadas de las funciones de membresía del modelo inicial.

## 2.6 Modelado difuso.

Los sistemas se pueden representar por modelos matemáticos de diferentes formas, tales como ecuaciones algebraicas, ecuaciones diferenciales, máquinas de estados finitos, etc. El marco teórico de modelado considerado en este trabajo se fundamenta en modelos difusos basados en reglas, las cuales describen la relación entre variables por medio de reglas *si-entonces*, tales como:

*Si el calor es alto entonces la temperatura se incrementará rápido.*

Estas reglas establecen relaciones lógicas entre las variables del sistema relacionando valores cualitativos (el calor es alto) de una variable, a valores cualitativos (la temperatura se incrementará rápido) de otra variable. Típicamente, los valores cualitativos tienen una interpretación lingüística clara, tal como en el ejemplo anterior, y se les nombra como *términos lingüísticos* (etiquetas, valores). El concepto de modelado y análisis de sistemas por medio de variables lingüísticas fue introducido por Zadeh [1], y se ha desarrollado considerablemente en años recientes. El significado de los términos lingüísticos con respecto a las variables de entrada y salida, las cuales pueden ser numéricas, (en nuestro ejemplo, calor, temperatura) se define por los conjuntos difusos elegidos. En este sentido, los conjuntos difusos, o más precisamente, sus funciones de membresía proveen una interfaz entre las variables numéricas de entrada y salida y los valores cualitativos lingüísticos en las reglas.

La estructura lógica de las reglas facilita la comprensión y el análisis del modelo de una manera semi-cualitativa, cercana a la forma en la que los humanos razonamos respecto del mundo real. En determinado contexto, la caracterización de los valores numéricos por términos lingüísticos puede ser más apropiada que el valor numérico preciso. El traslape deliberado de las funciones de membresía asegura la generalización para situaciones no completamente capturadas por las reglas. En términos matemáticos, el proceso de inferencia en modelos difusos puede considerarse como una interpolación entre los resultados de las reglas individuales.

La metodología de conjuntos difusos presenta varias ventajas sobre otras técnicas de modelado inteligentes, tales como las redes neuronales [6] o las redes de función de base radial [7]:

- Los modelos difusos integran el procesamiento lógico de la información con atractivas propiedades matemáticas, al ser considerados como aproximadores generales de funciones. Los modelos difusos se pueden ver como sistemas basados en reglas adecuados para

formalizar el conocimiento de expertos y, al mismo tiempo, como estructuras matemáticas flexibles, las cuales pueden representar transformaciones no lineales complejas [8], [9]. Debido a que el modelado difuso integra procesamiento numérico y simbólico en un marco teórico de referencia común, no se restringe a áreas que requieran experiencia y conocimientos humanos. Los modelos difusos también pueden hacer uso de algoritmos de aprendizaje manejados por datos y se pueden combinar con técnicas convencionales de regresión [10], [9].

- La estructura basada en reglas de los sistemas difusos, es útil durante el análisis de los modelos difusos obtenidos a partir de datos numéricos, ya que las reglas pueden revelar una descripción cualitativa útil del sistema que generó los datos. Tal descripción se puede confrontar y posiblemente combinar con el conocimiento de los expertos, lo cual ayuda a comprender el sistema y a validar el modelo al mismo tiempo.
  
- El uso de términos cualitativos lingüísticos en las reglas puede considerarse como un tipo de cuantización de la información. Dependiendo del número de valores cualitativos considerados (la granularidad), se pueden desarrollar modelos de diferentes niveles de abstracción y precisión para un sistema dado. Cada uno de los modelos puede servir para propósitos diferentes (predicción, análisis, diseño de un controlador, monitoreo, etc.).

## 2.7 Identificación difusa.

El término *identificación difusa* se refiere usualmente al conjunto de técnicas y algoritmos usados para construir modelos difusos a partir de los datos. Dos principales metodologías se pueden distinguir para la integración del conocimiento y de datos en un modelo difuso:

1. El conocimiento experto expresado en forma verbal se traduce en una colección de reglas *si-entonces*; de esta forma, se crea una cierta estructura del modelo. Los parámetros en esta estructura (funciones de membresía, pesos de las reglas, etc.) se pueden determinar usando los datos de entrada-salida. Los algoritmos particulares de entonación explotan el hecho de que en el ámbito computacional, un modelo difuso puede verse como una estructura por capas (red), similar a las redes neuronales artificiales, a las cuales se les pueden aplicar algoritmos estándar de aprendizaje. Esta metodología es llamada usualmente *modelado neuro-difuso*.
  
2. El modelo difuso se construye usando únicamente datos numéricos, es decir, inicialmente ningún conocimiento previo del sistema bajo estudio se usa para formular las reglas. Se espera que las reglas y las funciones de membresía extraídas puedan proveer una interpretación *a posteriori* del comportamiento del sistema. Un experto puede confrontar esta información con su propio conocimiento, puede modificar las reglas, o suplirlas con unas nuevas, y puede diseñar experimentos adicionales a fin de obtener datos con mayor información. Esta metodología se puede denominar *extracción de reglas*.

Obviamente, dependiendo de la aplicación particular, se pueden combinar las dos técnicas. Este trabajo se enfoca principalmente en el desarrollo de métodos y algoritmos para la segunda metodología, es decir, para la obtención automática de modelos difusos a partir de los datos. Se cree que esta técnica es más útil en la práctica, ya que puede favorecer el proceso de adquisición del conocimiento, pues se sabe que éste es un cuello-de-botella para las aplicaciones prácticas de sistemas basados en el conocimiento [11]. A su vez, el experto puede asumir un papel más activo de análisis y validación del modelo, lo cual puede conducir a revelar nuevas piezas de información, representando otro tipo de adquisición del conocimiento.

Los sistemas difusos se consideran como una clase de metodología de modelado local en donde se plantea la solución de un problema de modelado complejo dividiéndolo en una serie de sub-problemas más simples. La teoría de los conjuntos difusos ofrece una herramienta excelente para representar la incertidumbre asociada con la tarea de particionar, al proveer transiciones suaves entre los sub-modelos locales individuales, y poder integrar varios tipos de conocimiento dentro de un marco teórico común. Desde esta perspectiva, la identificación difusa se puede considerar como una búsqueda de la división de un sistema no lineal, en regiones lineales, la cual proporciona el balance deseado entre la complejidad y la precisión del modelo, explotando el hecho de que la complejidad de los sistemas no es usualmente uniforme en todo su intervalo de operación. Concerniente a esta división, y ya que no puede esperarse a que suficiente conocimiento *a priori* esté disponible, se desarrollan métodos para la generación automática de la partición, básicamente a partir de datos del sistema.

## 2.8 Resumen de los Modelos Difusos

En todos los sistemas difusos, la proposición antecedente siempre es una proposición difusa del tipo “ $x$  es  $A$ ” donde  $x$  es una variable lingüística y  $A$  es un conjunto difuso. Dependiendo de la estructura particular de la proposición consecuente, se pueden distinguir principalmente, dos tipos de modelos:

- *Modelo difuso lingüístico (ó Mamdani)* [1],[12]: tanto el antecedente como el consecuente son proposiciones difusas. Un caso especial de este tipo de modelo es el modelo difuso *impulso*, en el cual los consecuentes son conjuntos difusos del tipo impulso (una constante real).
- *Modelo difuso Takagi-Sugeno (TS)* [10]: el consecuente es una función certera de las variables antecedentes y no una proposición difusa.

### 2.8.1 Modelo Lingüístico

El modelo difuso lingüístico se ha introducido como una forma de capturar el conocimiento cualitativo mediante reglas del tipo *si-entonces*:



$$R_i : \text{Si } x \text{ es } A_i \text{ entonces } y \text{ es } B_i, \quad i = 1, 2, \dots, K. \quad (2.1)$$

Aquí,  $x$  es la *variable lingüística* de entrada (antecedente) y los  $A_i$  son los *términos lingüísticos* de los antecedentes (etiquetas). De manera similar,  $y$  es la *variable lingüística* de salida (consecuente) y los  $B_i$  son los *términos lingüísticos* de los consecuentes. Los valores de los términos lingüísticos  $A_i$  ( $B_i$ ) son conjuntos difusos.

El proceso de inferencia en sistemas basados en reglas, se define como el proceso de derivar un consecuente (conjunto difuso de salida) dado un conjunto de reglas y los datos de entrada. En el modelo lingüístico el mecanismo de inferencia está basado en la *regla composicional de inferencia* (Zadeh, 1973[1]), también conocida como la *regla del modus ponens generalizada*.

Para una completa descripción de este modelo ver Klir (Klir and Yuan, 1995[3]).

### 2.8.2 Modelo Impulso

Se obtiene un caso especial del modelo difuso lingüístico cuando los conjuntos difusos en los consecuentes  $B_i$  son conjuntos difusos del tipo impulso. Estos conjuntos se pueden representar como números reales  $b_i$ , lo cual genera las siguientes reglas:

$$R_i : \text{Si } x \text{ es } A_i \text{ entonces } y = b_i, \quad i = 1, 2, \dots, K. \quad (2.2)$$

A este tipo de modelo se le llama *modelo difuso tipo impulso*.

Una ventaja del modelo tipo impulso sobre el modelo lingüístico es que los parámetros consecuentes  $b_i$  se pueden estimar a partir de los datos, usando técnicas por mínimos cuadrados. El modelo difuso tipo impulso pertenece a una clase general de aproximadores genéricos de funciones, llamados *expansión de funciones base*, que toman la forma:

$$y = \sum_{i=1}^K \Phi_i(\mathbf{x}) b_i \quad (2.3)$$

En el modelo difuso tipo impulso, las funciones base  $\Phi_i(\mathbf{x})$  se determinan por los grados de satisfacción (normalizado) de los antecedentes de las reglas, siendo las constantes  $b_i$  los consecuentes. La mayoría de las estructuras usadas en la identificación de sistemas no-lineales, tales como las redes neuronales artificiales, las redes de función de base radial, o splines, pertenecen a esta clase de sistemas.

### 2.8.3 Modelo Takagi-Sugeno

El modelo difuso Takagi-Sugeno (TS) usa funciones certeras en los consecuentes. Así, el modelo TS se puede considerar como una combinación de modelos lingüísticos con modelos matemáticos de regresión, en el sentido de que los antecedentes describen regiones difusas en el espacio de entrada, en las cuales son válidas las funciones consecuentes. Las reglas TS tienen la siguiente forma:

$$R_i: \underbrace{\text{Si } \mathbf{x} \text{ es } A_i}_{\text{antecedente}} \text{ entonces } \underbrace{y_i = \mathbf{a}_i^T \mathbf{x} + b_i}_{\text{consecuente}}, \quad i = 1, 2, \dots, K \quad (2.4)$$

Donde  $\mathbf{x} \in X \subset \mathbb{R}^p$  es un vector certero de entrada,  $A_i$  es un conjunto difuso (multidimensional):  $\mu_{A_i}(\mathbf{x}): X \rightarrow [0, 1]$ ,  $y_i$  es la salida (escalar) de la *i*ésima regla,  $\mathbf{a}_i \in \mathbb{R}^p$  es un vector de parámetros y  $b_i$  es un desplazamiento (offset) escalar.

Este modelo puede representar sistemas estáticos y dinámicos múltiples-entradas, múltiples-salidas (MIMO). Dadas las salidas de los consecuentes individuales  $y_i$ , la salida global  $y$  del modelo TS se calcula usando la fórmula de la media ponderada (difusa).

$$y = \frac{\sum_{i=1}^K \beta_i(\mathbf{x}) y_i}{\sum_{i=1}^K \beta_i(\mathbf{x})}. \quad (2.5)$$

Aquí,  $\beta_i(\mathbf{x})$  denota el grado de satisfacción del antecedente de la *i*ésima regla calculada simplemente como el grado de membresía de  $\mathbf{x}$  en el conjunto difuso  $A_i$ :  $\mu_{A_i}(\mathbf{x})$ . Puesto que puede ser difícil implementar e interpretar conjuntos difusos multidimensionales, la proposición antecedente se expresa usualmente como una combinación de proposiciones simples con conjuntos difusos unidimensionales definidos para las componentes individuales de  $\mathbf{x}$ . El grado de satisfacción del antecedente se calcula de la combinación de los grados de membresía de las proposiciones individuales mediante operaciones de conjunción, disyunción y complemento, la más común es la proposición que utiliza la forma conjuntiva de los antecedentes, esto es:

$$\text{Si } x_1 \text{ es } A_{i,1} \text{ y } x_p \text{ es } A_{i,p} \text{ entonces } y_i = \mathbf{a}_i^T \mathbf{x} + b_i, \quad (2.6)$$

Donde el grado de satisfacción  $\beta_i(\mathbf{x})$  se calcula como:

$$\beta_i(\mathbf{x}) = \mu_{A_{i,1}}(x_1) \wedge \mu_{A_{i,2}}(x_2) \wedge \dots \wedge \mu_{A_{i,p}}(x_p). \quad (2.7)$$

En donde el operador mínimo ( $\wedge$ ) se puede reemplazar por cualquier otra t-norma, tal como el operador producto, el operador de intersección bold, etc. Denotando el grado de satisfacción normalizado por  $\gamma_i(\mathbf{x}) = \beta_i(\mathbf{x}) / \sum_{j=1}^K \beta_j(\mathbf{x})$ , la salida global del modelo TS se puede escribir como un modelo global pseudo-lineal:

$$y = \left( \sum_{i=1}^K \gamma_i(\mathbf{x}) \mathbf{a}_i^T \right) \mathbf{x} + \sum_{i=1}^K \gamma_i(\mathbf{x}) b_i = \tilde{\mathbf{a}}^T(\mathbf{x}) \mathbf{x} + \tilde{b}(\mathbf{x}). \quad (2.8)$$

donde,  $\tilde{\mathbf{a}}(\mathbf{x})$  y  $\tilde{b}(\mathbf{x})$  son parámetros dependientes de la entrada, calculados como combinaciones lineales convexas de los parámetros constantes  $\mathbf{a}_i$  y  $b_i$ , esto es:

$$\tilde{\mathbf{a}}(\mathbf{x}) = \sum_{i=1}^K \gamma_i(\mathbf{x}) \mathbf{a}_i, \quad \tilde{\mathbf{b}}(\mathbf{x}) = \sum_{i=1}^K \gamma_i(\mathbf{x}) b_i. \quad (2.9)$$

Una posible representación del resultado de modelar una función usando este tipo de modelo difuso se puede observar en la figura 2.3.

Un caso especial de la función consecuente ocurre cuando el desplazamiento  $b_i = 0$ , para  $i = 1, \dots, K$ , y al modelo se le llama, el modelo TS *homogéneo*, esto es:

$$\text{Si } \mathbf{x} \text{ es } A_i \text{ entonces } y_i = \mathbf{a}_i^T \mathbf{x}. \quad (2.10)$$

Este modelo sólo puede aproximar una cierta clase de sistemas, puesto que en general, todos los modelos consecuentes contienen al origen.

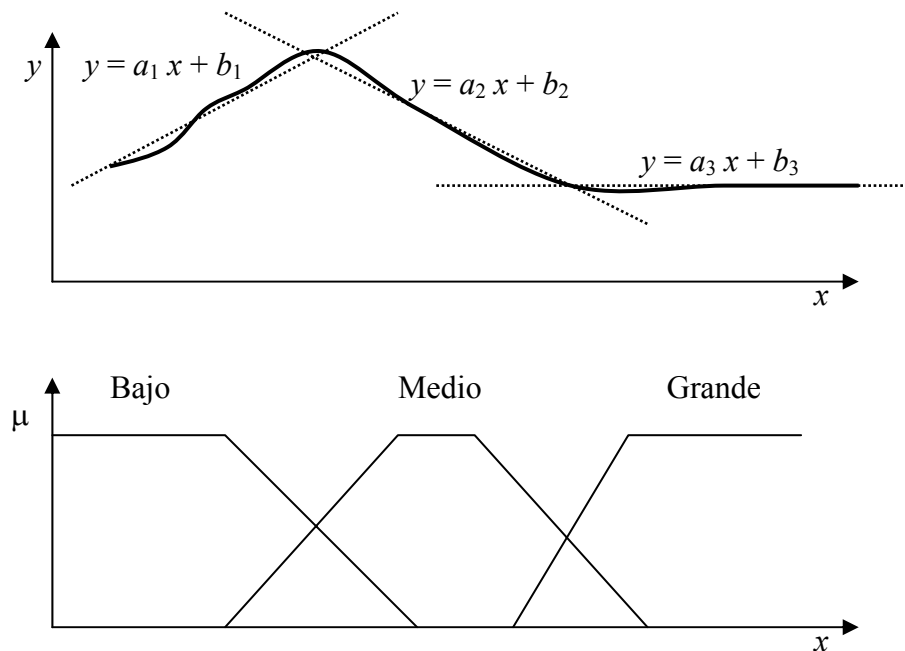


Figura 2.3. Resultado de aproximación de una función usando tres reglas difusas T-S afines

### 2.8.3.1. Inferencia en el Modelo TS

Antes de que se pueda inferir el valor de salida, se debe calcular el *grado de satisfacción* (*fulfillment*) de los antecedentes, denotado por  $\beta_i(\mathbf{x})$ . Para reglas con conjuntos difusos multidimensionales en los antecedentes, el grado de satisfacción se obtiene utilizando conectivas lógicas, esto es, el grado de satisfacción del antecedente se calcula como una

combinación de los grados de membresía de las proposiciones individuales, usando operadores de lógica difusa.

En el modelo Takagi-Sugeno-Kang, el proceso de inferencia se reduce a una simple expresión algebraica, similar a la fórmula de defusificación de la media-difusa [3]:

$$y = \frac{\sum_{i=1}^K \beta_i(\mathbf{x}) y_i}{\sum_{i=1}^K \beta_i(\mathbf{x})}. \quad (2.11)$$

Si se denota el grado de satisfacción normalizado por:

$$\lambda_i(\mathbf{x}) = \frac{\beta_i(\mathbf{x})}{\sum_{j=1}^K \beta_j(\mathbf{x})}, \quad (2.12)$$

entonces, el modelo TS afín con una estructura consecuente común se puede expresar como un modelo pseudo lineal con parámetros dependientes de la entrada, esto es:

$$y = \left( \sum_{i=1}^K \lambda_i(\mathbf{x}) \mathbf{a}_i^T \right) \mathbf{x} + \sum_{i=1}^K \lambda_i(\mathbf{x}) b_i = \mathbf{a}^T(\mathbf{x}) \mathbf{x} + b(\mathbf{x}). \quad (2.13)$$

Los parámetros  $\mathbf{a}(\mathbf{x}) + b(\mathbf{x})$  son combinaciones lineales convexas de los parámetros consecuentes  $\mathbf{a}_i$  y  $b_i$ , esto es:

$$\mathbf{a}(\mathbf{x}) = \sum_{i=1}^K \lambda_i(\mathbf{x}) \mathbf{a}_i, \quad b(\mathbf{x}) = \sum_{i=1}^K \lambda_i(\mathbf{x}) b_i. \quad (2.14)$$

Esta propiedad facilita el análisis de los modelos TS en el marco teórico de los sistemas politopos. Un modelo T-S se puede considerar como una transformación del espacio de los antecedentes (entrada) a una región convexa (polítopo) en el espacio de los parámetros consecuentes de un sistema cuasi-lineal, como se muestra esquemáticamente en la figura 2.4.

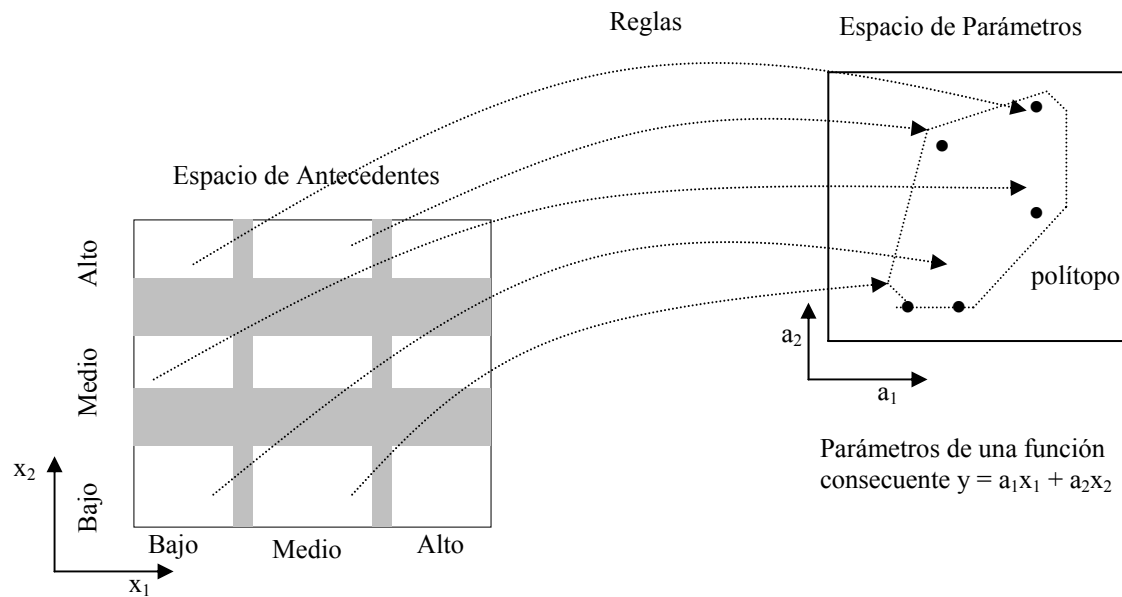


Figura 2.4. El modelo TS se puede considerar como una transformación del espacio de los antecedentes al espacio de los parámetros consecuentes.

Esta propiedad facilita el análisis de los modelos TS en un marco teórico similar al de los sistemas lineales.

## 2.9 Conclusiones

En este capítulo se hace una breve introducción al problema de la obtención del modelo de un sistema, mencionando las distintas metodologías disponibles y la forma en que se pueden construir. Se describe un procedimiento general considerado como la guía que debe seguir cualquier metodología para la obtención de un modelo.

Siguiendo el esquema de la metodología genérica, se provee una metodología para la construcción de Modelos Difusos, en la cual la estructura se identifica mediante el empleo de algoritmos difusos de clustering y los parámetros mediante la utilización del método por mínimos cuadrados. Aquí, se considera fundamental la utilización de los datos en el proceso de identificación.

Al describir los diferentes tipos de sistemas difusos a los que podemos recurrir para resolver un problema, se describe con más detalle el modelo difuso tipo Takagi-Sugeno, pues con el fin de resolver la complejidad del sistema, objeto de aplicación de este trabajo, se hace necesario emplear estos distintos modelos disponibles; esto nos permitirá cubrir el Universo de Variables de entrada-salida al bioproceso.

## CAPITULO 3

### Partición del Espacio-producto Entrada-salida en Regiones Difusas

#### 3.1 Introducción

En este capítulo se trata el tema de los Algoritmos Difusos de Clustering, los cuales son una forma efectiva para la identificación de sistemas complejos no lineales, utilizándose una aproximación de aprendizaje no supervisado, esto es, se determinan los clusters sin información *a priori*, a partir de datos no etiquetados. Dos de las razones para justificar el uso de estos métodos, las cuales pueden ser muy variadas, pueden ser:

1. La recolección de datos etiquetados es algo que puede ser muy costoso; generalmente es más fácil obtener datos que prototipos, ya que en este último caso, debe mediar la intervención de un experto.
2. El reconocimiento de la estructura de los datos se utiliza para determinar las particiones difusas que representarán el sistema a modelar.

Presumiblemente, los datos llevan información sobre el proceso que los genera. Por estructura entendemos la manera en la cual se puede organizar esta información tal que se puedan identificar las relaciones entre las variables del proceso. Las representaciones de la estructura identificada dependen de los datos, del método de búsqueda y del modelo usado. En términos de información, los datos la contienen, la búsqueda la reconoce y la estructura la representa.

Las técnicas que vamos a estudiar se centran en descubrir la estructura de los datos. Ya es sabido que los datos tienden a presentarse *agrupados* en regiones del espacio de representación y que idealmente, cada agrupación se asocia a una clase informativa. La aproximación no supervisada tiene como objetivo encontrar y caracterizar estos agrupamientos con base en los datos sin considerar ninguna información *a priori*.

### 3.2 Técnicas de Agrupamiento (*clustering*).

Una técnica de clustering se puede definir como una técnica diseñada para realizar una partición, asignando datos a grupos de tal forma que cada grupo sea más o menos homogéneo y distinto de los demás [13].

Intuitivamente, un conjunto de datos se consideran un *grupo* si forman un conjunto homogéneo y diferenciado de los demás en el espacio de representación. El criterio de homogeneidad más simple está basado en la *distancia*: se espera que la distancia entre los datos de un mismo agrupamiento sea significativamente menor que la distancia entre datos de agrupamientos diferentes. Como se puede observar, el concepto de grupo está íntimamente relacionado con el de *clase*, de hecho, idealmente, cada clase (informativa) debería tener asociado un grupo.

Definitivamente, una técnica de clustering debe encontrar y caracterizar las agrupaciones subyacentes en los datos. Funcionalmente, un método de clustering se puede esquematizar como se muestra en la figura 3.1, en la que a partir de un conjunto de  $M$  datos no etiquetados  $x_i, i=1, 2, \dots, M$ , se encuentran  $k$  agrupamientos  $S_j, j = 1, 2, \dots, k$ .

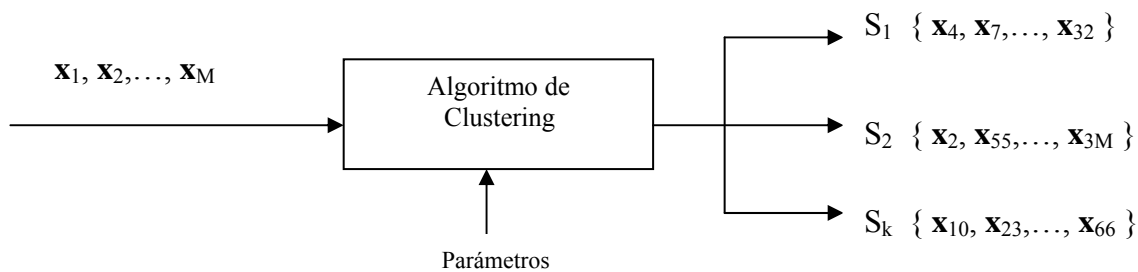


Figura 3.1. Esquema funcional de un algoritmo de agrupamiento

### 3.3 Utilizando Métodos de clustering

El uso de métodos de clustering, entre otras razones, se puede argumentar cuando se carece del conocimiento suficiente para determinar el número de particiones en que se pueden dividir los datos de las variables de interés y que, por medio de estos, sea posible determinar automáticamente las relaciones difusas

**Nota:** Se suele emplear indistintamente los términos *grupo*, *agrupamiento*, *agrupación* o el término inglés *cluster*, y también se suele hablar de *clustering* en lugar de técnicas de agrupamiento.



El resultado de un algoritmo de clustering (los agrupamientos detectados) depende de diversos factores:

1. El algoritmo concreto empleado para encontrar los agrupamientos.

Existen diferentes aproximaciones a este problema y numerosos algoritmos de clustering. Esta variedad hace que la elección de una estrategia de agrupamiento e incluso entre diferentes implementaciones de un mismo algoritmo, proporcione resultados diferentes.

2. El valor de los parámetros del algoritmo.

En algunos algoritmos, la elección de los valores adecuados de los parámetros resulta difícil. Si a esto le sumamos que algunos algoritmos pueden tener muchos parámetros, es fácil deducir que los resultados pueden ser muy diferentes. En algunos casos el valor de éstos depende de las unidades en las que se cuantifican las variables, lo que hace aún más difícil el ajuste.

3. Los datos utilizados y, en algunas ocasiones, hasta el orden en que se procesan.

Muchos algoritmos utilizan sólo una parte de los datos disponibles para el aprendizaje (caracterización de los agrupamientos). Usualmente se emplea alguna técnica de muestreo para esta selección. En consecuencia, frecuentemente diferentes permutaciones del mismo conjunto de datos dan lugar a diferentes resultados. Además, muchos algoritmos se ven influenciados por los primeros datos que se procesan, lo que hace que para una misma permutación, el orden en que se presentan estos, determina, a veces de manera significativa, el resultado final.

4. La medida de similitud adoptada.

El criterio general es que los datos asociados a un agrupamiento son relativamente *similares* entre sí, mientras que son relativamente *disimiles* a los de otros agrupamientos. A la hora de concretar este criterio se requiere especificar la medida de similitud (normalmente alguna medida de *distancia*) y un valor de umbral asociada a ésta. Es evidente que diferentes criterios de distancia proporcionarán diferentes resultados al igual que diferentes umbrales para un mismo criterio de distancia.

La aplicación práctica de los algoritmos de agrupamiento conlleva una dificultad inherente respecto del establecimiento de la medida de similitud apropiada, así como para establecer un buen criterio de partición adecuado al sistema a resolver.

### 3.4 Medidas de similitud.

Como se ha mencionado, los datos asociados en un agrupamiento son relativamente *similares* entre sí, mientras que son relativamente *disímiles* a los de otros agrupamientos.

Si  $\xi$  es una medida de disimilitud definida entre dos datos, resulta evidente que las siguientes dos propiedades se cumplen:

$$1. \quad \xi(x_i, x_i) = 0 \quad (3.1)$$

$$2. \quad \xi(x_i, x_j) \neq 0 \quad \forall j \neq i \quad (3.2)$$

Una medida de este tipo se expresa de manera numérica para indicar el grado de asociación entre dos datos en un grupo, o entre un dato y un grupo, o entre dos grupos de datos. Cuando se mide la similitud entre un dato y un grupo, este último se resume habitualmente en su valor medio por lo que el cálculo de similitud se simplifica notablemente.

En la literatura, se pueden encontrar distintas expresiones para  $\xi$ , aunque las más empleadas son las medidas de distancia denominadas como: distancia Euclídea, distancia de Mahalanobis, etc.

#### 3.4.1 Distancia Euclídea y Distancia Euclídea ponderada

La distancia Euclídeana entre dos datos  $x_i$  y  $x_j$  se define como:

$$\delta_E^2(x_i, x_j) = \|x_i - x_j\|^2 = (x_i - x_j)^T (x_i - x_j). \quad (3.3)$$

Este cálculo se puede expresar como:

$$\delta_E^2(x_i, x_j) = \sum_{k=1}^d (x_{ik} - x_{jk})^2, \quad (3.4)$$

donde  $x_{ik}$  es la  $k$ -ésima componente del dato  $x_i$ .

#### 3.4.2 Distancia de Mahalanobis

La distancia de Mahalanobis entre dos datos  $x_i$  y  $x_j$  se define como:

$$\delta_M^2(x_i, x_j) = (x_i - x_j)^T \Sigma^{-1} (x_i - x_j) \quad (3.5)$$

donde  $\Sigma^{-1}$  es la inversa de la matriz de covarianza, este término considera la distinta dispersión de las variables en el espacio.

Esta distancia se emplea en situaciones en las que las matrices de covarianza son iguales y si además el espacio de representación no es Euclídeano.

### 3.5 Criterios de Clustering

Con el uso de una medida de similitud sería suficiente para construir un algoritmo de agrupamiento. Sin embargo, en ocasiones es posible encontrar varias maneras, igualmente satisfactorias, de asignar los patrones a diferentes agrupamientos.

Evidentemente se pueden encontrar diferentes asignaciones de un conjunto de  $m$  patrones a  $k$  agrupamientos. Si cada una de estas asignaciones se denomina una **partición**, resta encontrar la mejor partición entre todas las posibles. El objetivo así, consiste en definir una función criterio que pueda aplicarse a una partición para medir su *calidad*. De esta manera, podría seleccionarse la partición que proporcionará el valor óptimo de esta función criterio. La función criterio más simple y más comúnmente empleada es la *suma de los errores al cuadrado* o simplemente criterio SSE (del inglés, Sum of Squared Errors) es:

$$J_e = \sum_{i=1}^k \sum_{x \in S_i} \|x - c_i\|^2, \quad (3.6)$$

donde  $c_i$  es el centro del agrupamiento  $S_i$ .

Esta función tiene una interpretación muy simple: para un agrupamiento cualquiera,  $S_i$ , su centro,  $c_i$ , es el mejor representante de todos los  $x \in S_i$ , en el sentido que minimiza la suma de las diferencias (al cuadrado) entre los patrones asignados al agrupamiento  $S_i$  y su centro. En consecuencia,  $J_e$  mide el cuadrado del error que se obtiene cuando se representan los  $m$  datos que se agrupan mediante los  $k$  centros de los agrupamientos. El valor de  $J_e$  depende de la manera en que se organicen los  $m$  datos en los  $k$  agrupamientos, en consecuencia, una partición óptima es la que minimiza  $J_e$ . Estas particiones se dice también que son las *particiones de mínima varianza*.

En la literatura se pueden encontrar otras funciones criterio, sin embargo, en nuestro caso nos centraremos en el criterio SSE, ya que la mayoría de los algoritmos se basan, en última instancia, en este criterio.

### 3.6 Métodos de Clustering

Muchos algoritmos de clustering han sido introducidos en la literatura. Puesto que formalmente, las particiones se pueden ver como sub conjuntos del conjunto de datos, una posible clasificación de los métodos de clustering se puede hacer de acuerdo a si los sub conjuntos son difusos o certeros [14].

Los métodos de clustering certeros están basados en la teoría de conjuntos clásicos, y requieren la definición de que un objeto pertenezca o no a una partición. Una partición

certera significa la partición de los datos en un número específico de sub conjuntos mutuamente exclusivos.

En contraparte, los métodos de clustering difusos, permiten que los objetos pertenezcan simultáneamente a varias particiones, con varios grados de membresía. En muchas situaciones, una partición difusa es más natural que una partición certera, puesto que los objetos que se encuentran en las fronteras entre varias clases no son forzados a pertenecer completamente a una de las clases, sino que se les asignan grados de membresía entre 0 y 1, indicando membresías parciales.

### 3.7 Clustering Certero y Difuso.

Las particiones difusas y posibilísticas se pueden considerar como una generalización de la partición certera, la cual se formula en términos de sub conjuntos clásicos. A continuación, se presentan las propiedades más relevantes de las técnicas de clustering certero y difuso.

#### 3.7.1 Clusters Certeros

El objetivo de un proceso de clustering es particionar un conjunto de  $\mathbf{Z}$  datos en  $c$  grupos, dentro de un espacio. Por el momento, se considera que  $c$  se conoce, por ejemplo, basándose en un conocimiento a priori. Usando conjuntos clásicos, una partición certera de  $\mathbf{Z}$  se puede definir como una familia de sub conjuntos  $\{A_i | 1 \leq i \leq c\} \subset P(\mathbf{Z})$ , con las siguientes propiedades [14]:

$$\bigcup_{i=1}^c A_i = \mathbf{Z}, \quad (3.7a)$$

$$A_i \cap A_j = \emptyset, \quad 1 \leq i \neq j \leq c, \quad (3.7b)$$

$$\emptyset \subset A_i \subset \mathbf{Z} \quad 1 \leq i \leq c, \quad (3.7c)$$

La ecuación (3.7a) significa que los sub conjuntos  $A_i$  contienen colectivamente todos los datos de  $\mathbf{Z}$ . Los sub conjuntos deben ser disjuntos, tal como lo establece (3.7b) y ninguno de ellos está vacío ni tampoco contiene a todos los datos de  $\mathbf{Z}$  (3.7c). En términos de *funciones de membresía*, las ecuaciones (3.7) se pueden expresar como:

$$\bigvee_{i=1}^c \mu_{A_i} = \mathbf{1}, \quad (3.8a)$$

$$\mu_{A_i} \wedge \mu_{A_j} = \mathbf{0}, \quad 1 \leq i \neq j \leq c, \quad (3.8b)$$

$$\mathbf{0} \subset \mu_{A_i} \subset \mathbf{1} \quad 1 \leq i \leq c, \quad (3.8c)$$

Aquí,  $\mathbf{0}$  denota la función cero y  $\mathbf{1}$  denota la función uno, y  $\mu_{A_i}$  es la función de membresía de  $A_i$ . Con el fin de simplificar la notación, en lo subsecuente se usará  $\mu_i$  en lugar  $\mu_{A_i}$ . Además, si se denota a  $\mu_i(\mathbf{z}_k)$  por  $\mu_{ik}$ , las particiones se pueden representar convenientemente en notación matricial. Una matriz  $\mathbf{U} = [\mu_{ik}]$  de tamaño  $c \times N$ , representa una partición certera si y sólo si sus elementos satisfacen las siguientes condiciones:

$$\mu_{ik} \in \{0,1\} \quad 1 \leq i \leq c, 1 \leq k \leq N, \quad (3.9a)$$

$$\sum_{i=1}^c \mu_{ik} = 1, \quad 1 \leq k \leq N, \quad (3.9b)$$

$$0 < \sum_{k=1}^N \mu_{ik} < N, \quad 1 \leq i \leq c, \quad (3.9c)$$

las cuales se obtienen directamente de la ecuación (3.8). El  $i$ -ésimo renglón de la matriz de partición certera  $\mathbf{U}$  contiene los valores de la función característica del  $i$ -ésimo sub conjunto  $A_i$  de  $\mathbf{Z}$ . La siguiente definición establece el concepto de espacio de partición certera [14]:

**Definición .1 (Espacio de partición certera).** Sea  $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$  un conjunto finito y  $2 \leq c < N$ , un entero. El espacio de partición certera para  $\mathbf{Z}$  es el conjunto:

$$M_{hc} = \left\{ \mathbf{U} \in \mathbb{R}^{c \times N} \mid \mu_{ik} \in \{0,1\}, \forall i,k; \sum_{i=1}^c \mu_{ik} = 1, \forall k; 0 < \sum_{k=1}^N \mu_{ik} < N, \forall i \right\} \quad (3.10)$$

Una aplicación de este concepto, se puede ilustrar mediante el siguiente ejemplo: supóngase que se tiene un conjunto de datos bi-dimensionales con una distribución como se muestra en la figura 3.2

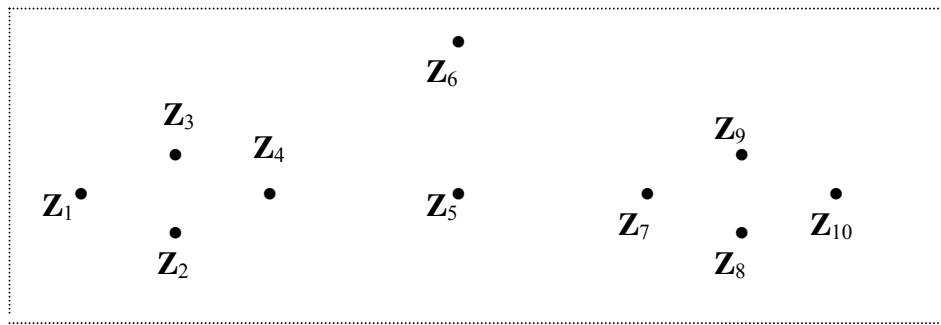


Figura 3.2. Se muestran los datos bi-dimensionales a agrupar en los ejemplos.

Una inspección visual de los datos puede sugerir dos clusters bien separados (los puntos  $\mathbf{z}_1$  a  $\mathbf{z}_4$  y de  $\mathbf{z}_7$  a  $\mathbf{z}_{10}$ ), un punto entre las dos clusters ( $\mathbf{z}_5$ ), y un externo, “outlier”,  $\mathbf{z}_6$ . Una posible partición certera  $\mathbf{U} \in M_{hc}$  de los datos en dos sub conjuntos está dada por:

$$\mathbf{U} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix}$$

En donde en el primer renglón de  $\mathbf{U}$  define de manera puntual la función característica para el primer sub conjunto de  $\mathbf{Z}$ ,  $A_1$ , y el segundo renglón define la función característica del segundo sub conjunto de  $\mathbf{Z}$ ,  $A_2$ .

Cada muestra se debe asignar exclusivamente a un sub conjunto de la partición. En este caso, tanto el punto en la frontera  $\mathbf{z}_5$  como el punto externo  $\mathbf{z}_6$  han sido asignados a  $A_1$ . De esto, se observa que una partición certera no puede dar una imagen real de los datos, pues los puntos en la frontera de los sub-conjuntos pueden representar patrones con una mezcla de propiedades de los datos en  $A_1$  y  $A_2$ , y por lo tanto no se pueden asignar completamente a cualquiera de estas clases, o bien, constituyen clases separadas; este inconveniente se puede solucionar usando particiones difusas y posibilísticas.

### 3.7.2 Clusters Difusos

La generalización de la partición certera al caso difuso se obtiene directamente, permitiendo pertenencias parciales, esto es, que  $\mu_{ik}$  obtenga valores reales en el intervalo  $[0, 1]$ , [15]. Las condiciones para una matriz de partición difusa, análoga a (3.7) están dadas por:

$$\mu_{ik} \in [0,1], \quad 1 \leq i \leq c, 1 \leq k \leq N, \quad (3.11a)$$

$$\sum_{i=1}^c \mu_{ik} = 1, \quad 1 \leq k \leq N, \quad (3.11b)$$

$$0 < \sum_{k=1}^N \mu_{ik} < N, \quad 1 \leq i \leq c. \quad (3.11c)$$

**Definición .2 (Espacio de partición difusa)** Sea  $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$  un conjunto finito y  $2 \leq c < N$  un entero. El espacio de partición difusa para  $\mathbf{Z}$  es el conjunto:

$$M_{fc} = \left\{ \mathbf{U} \in \mathbb{R}^{c \times N} \mid \mu_{ik} \in [0,1], \forall i,k; \sum_{i=1}^c \mu_{ik} = 1, \forall k; 0 < \sum_{k=1}^N \mu_{ik} < N, \forall i \right\}. \quad (3.12)$$

El  $i$ -ésimo renglón de la matriz de partición difusa  $\mathbf{U}$  contiene los valores de la  $i$ -ésima función de membresía del sub conjunto difuso  $A_i$  de  $\mathbf{Z}$ . La ecuación (3.11b) restringe la suma de cada columna a 1, y así la membresía total de cada  $\mathbf{z}_k$  en  $\mathbf{Z}$  es igual a uno.

Si ahora consideramos que la matriz de partición correspondiente a la figura 3.2 es de la forma:

$$\mathbf{U} = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 0.8 & 0.5 & 0.5 & 0.2 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.5 & 0.5 & 0.8 & 1.0 & 1.0 & 1.0 \end{bmatrix}$$

El punto en la frontera  $\mathbf{z}_5$  ahora tiene un grado de membresía de 0.5 en ambas clases, lo cual refleja correctamente su posición a la mitad entre los dos clusters. Sin embargo, se observa que el dato  $\mathbf{z}_6$  (externo) tiene el mismo par de grados de membresía, aún cuando está más allá de los dos clusters, y así se puede considerar como menos típico tanto de  $A_1$  como de  $A_2$  comparado con  $\mathbf{z}_5$ ; esto se debe a que la condición (3.11b) requiere que la suma de las membresías de cada punto sea igual a uno. En este caso podría considerarse que tres particiones son más apropiadas que dos por que eliminaría la ambigüedad en estos puntos, sin embargo, en general es difícil detectar los datos “externos” y asignarlos a particiones extras. El uso de una partición posibilística resuelve este inconveniente de las particiones difusas.

### 3.7.3. Clusters Posibilísticos

Una forma más genérica de partición difusa, la *partición posibilística*, se puede obtener relajando la restricción (3.11b). Sin embargo, esta restricción no se puede eliminar por completo, a fin de asegurar que cada punto sea asignado a al menos uno de los subconjuntos difusos, con una membresía mayor que cero. La ecuación (3.11b) se puede reemplazar por una restricción menos severa tal que  $\forall k, \exists i, \mu_{ik} > 0$ . Las condiciones para una matriz de partición difusa posibilística, análoga a (3.11) son:

$$\mu_{ik} \in [0,1] \quad 1 \leq i \leq c, 1 \leq k \leq N, \quad (3.13a)$$

$$\exists i, \mu_{ik} > 0, \quad \forall k, \quad (3.13b)$$

$$0 < \sum_{k=1}^N \mu_{ik} < N, \quad 1 \leq i \leq c, \quad (3.13c)$$

**Definición .3 (Espacio de partición posibilística)** Sea  $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$  un conjunto finito de muestras y  $c, 2 \leq c < N$ , un entero. El espacio de partición posibilística para  $\mathbf{Z}$  es el conjunto:

$$M_{pc} = \left\{ \mathbf{U} \in \mathbb{R}^{c \times N} \mid \mu_{ik} \in [0,1], \forall i,k; \exists i, \mu_{ik} > 0, \forall k; 0 < \sum_{k=1}^N \mu_{ik} < N, \forall i \right\}. \quad (3.14)$$

Usando este tipo de partición, la matriz correspondiente a la figura 3.2, puede quedar como:

$$\mathbf{U} = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 0.8 & 0.5 & 0.2 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.5 & 0.2 & 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix}.$$

Puesto que no se restringe la suma de los elementos en cada columna de  $\mathbf{U} \in M_{pc}$ , el dato externo  $\mathbf{z}_6$  tiene una membresía de 0.2 en ambos clusters, la cual es menor que la membresía del punto en la frontera  $\mathbf{z}_5$ , reflejando con ello el menor grado de “tipicidad” de este punto en ambos conjuntos.

### 3.7.4. Obtención de Clusters con Fuzzy c-means

La mayoría de los algoritmos difusos analíticos de clustering, están basados en la optimización de la función objetivo básica c-means, o alguna modificación de ésta. A continuación se presenta la funcional c-means difusa [16].

### 3.7.5 La funcional difusa c-means

Una gran familia de algoritmos difusos de partición están basados en la minimización de la funcional *fuzzy c-means*, formulada como [10]:

$$J(\mathbf{Z}; \mathbf{U}, \mathbf{V}) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m \|\mathbf{z}_k - \mathbf{v}_i\|_{\mathbf{A}}^2. \quad (3.15a)$$

Donde:

$$\mathbf{U} = [\mu_{ik}] \in M_{fc}, \quad (3.15b)$$

es una matriz de partición difusa de  $\mathbf{Z}$ ,

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c], \quad \mathbf{v}_i \in \mathbb{R}^n \quad (3.15c)$$

es un vector de *prototipos de los clusters* (centros), el cual se debe determinar,

$$D_{ik\mathbf{A}}^2 = \|\mathbf{z}_k - \mathbf{v}_i\|_{\mathbf{A}}^2 = (\mathbf{z}_k - \mathbf{v}_i)^T \mathbf{A} (\mathbf{z}_k - \mathbf{v}_i), \quad (3.15d)$$

es una norma distancia producto interno cuadrática, y

$$m \in [1, \infty).$$

es un exponente de ponderación, el cual determina la fusividad de los clusters resultantes.

La medida de disimilitud en (3.15a) es la distancia cuadrada entre cada punto  $\mathbf{z}_k$  de los datos y el  $\mathbf{v}_i$  prototipo del cluster, esta distancia es ponderada por el exponente del grado de membresía  $(\mu_{ik})^m$  en tal punto. El valor de la función costo (3.15a) se puede considerar como una medida de la varianza total de  $\mathbf{z}_k$  a  $\mathbf{v}_i$ .

### 3.7.6. El algoritmo fuzzy c-means

La minimización del funcional c-means (3.15a), representa un problema de optimización no lineal, que se puede solucionar usando una variedad de métodos disponibles, variando desde una minimización coordinada agrupada [17], pasando por recocido simulado [18], hasta algoritmos genéticos [19]. Sin embargo, el método más popular es una iteración



simple de Picard a través de las condiciones de primer orden para puntos estacionarios de (3.15a), conocido como el algoritmo fuzzy c-means (FCM) [17].

Los puntos estacionarios de la función objetivo (3.14a) se pueden encontrar agregando la restricción (3.11b) a  $J$ , por medio de los multiplicadores de Lagrange:

$$\bar{\mathbf{J}}(\mathbf{Z}; \mathbf{U}, \mathbf{V}, \boldsymbol{\lambda}) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ik\mathbf{A}}^2 + \sum_{k=1}^N \lambda_k \left[ \sum_{i=1}^c \mu_{ik} - 1 \right], \quad (3.16)$$

e igualando con cero los gradientes de  $\bar{\mathbf{J}}$  con respecto a  $\mathbf{U}$ ,  $\mathbf{V}$  y  $\boldsymbol{\lambda}$ . Si  $D_{ik\mathbf{A}}^2 > 0, \forall i, k$  y  $m > 1$ , entonces  $(\mathbf{U}, \mathbf{V}) \in M_{fc} \times \mathbb{R}^{n \times c}$  puede minimizar a (3.15a) solamente si:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c (D_{ik\mathbf{A}} / D_{jk\mathbf{A}})^{2/(m-1)}}, \quad 1 \leq i \leq c, \quad 1 \leq k \leq N, \quad (3.17a)$$

y

$$\mathbf{v}_i = \frac{\sum_{k=1}^N (\mu_{ik})^m \mathbf{z}_k}{\sum_{k=1}^N (\mu_{ik})^m}, \quad 1 \leq i \leq c. \quad (3.17b)$$

Esta solución también satisface las restricciones restantes (3.11a) y (3.11c). Nótese que la ecuación (3.17b) genera un  $\mathbf{v}_i$  como la media ponderada de los elementos de los datos que pertenecen a un cluster, donde los pesos son los grados de membresía. Esta es la razón por la que el algoritmo es nombrado “c-means”. El algoritmo fuzzy c-means itera a través de las ecuaciones (3.17a) y (3.17b).

### 3.8 Normas Producto-interno.

La forma de cada partición se determina por la elección de la matriz  $\mathbf{A}$  en la medición de la distancia dada por (3.15d). Una elección común es proponer  $\mathbf{A} = \mathbf{I}$ , la matriz identidad, la cual induce la norma Euclidiana estándar:

$$D_{ik}^2 = (\mathbf{z}_k - \mathbf{v}_i)^T (\mathbf{z}_k - \mathbf{v}_i). \quad (3.18)$$

Clustering con el Algoritmo fuzzy c-means (FCM).

Si se considera un conjunto de datos en  $\mathbb{R}^2$ , los cuales contienen dos grupos de formas diferentes, bien separados, figura 3.3a, donde las muestras en ambas particiones se obtienen de una distribución normal y la desviación estándar para el grupo superior es de 0.2 para ambos ejes, mientras que para el grupo inferior es de 0.2 para el eje horizontal y de 0.05 para el eje vertical, entonces, de las curvas de nivel para las membresías, como se observan

en la figura 3.3b, se puede ver que el algoritmo FCM impone estrictamente una forma circular en ambas particiones, aún cuando el grupo inferior es más bien elongado.

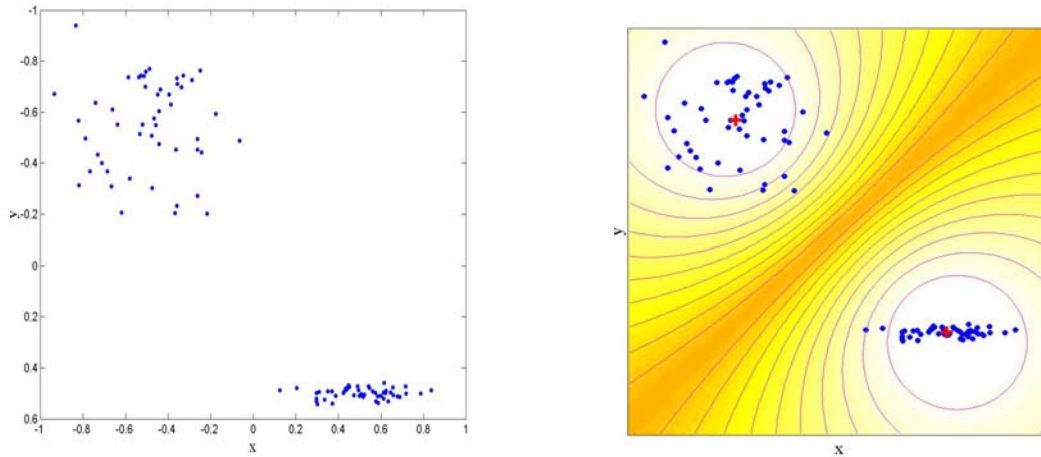


Figura 3.3 a) Distribución de datos bidimensionales. b) Se muestran los centros obtenidos con FCM así como las curvas de nivel respecto de los gradientes de los grados de membresía resultantes.

La matriz norma inductora fue establecida como  $\mathbf{A} = \mathbf{I}$  para ambos grupos, el exponente de peso fue  $m = 2$ , y el criterio de terminación  $\epsilon = 0.001$ . El algoritmo se inició con una matriz de partición aleatoria y convergió después de cuatro iteraciones

La matriz  $\mathbf{A}$  se puede elegir también como una matriz diagonal de tamaño  $n \times n$ , la cual considera diferentes varianzas en las direcciones de los ejes coordenadas de  $\mathbf{Z}$ :

$$\mathbf{A}_D = \begin{bmatrix} (1/\sigma_1)^2 & 0 & \cdots & 0 \\ 0 & (1/\sigma_2)^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (1/\sigma_n)^2 \end{bmatrix} \quad (3.19)$$

Esta matriz induce una norma diagonal en  $\mathbb{R}^n$ .

Finalmente, la matriz  $\mathbf{A}$  se puede definir como la inversa de la matriz de covarianza de la muestra  $\mathbf{Z}$  de tamaño  $n \times n$ :  $\mathbf{A} = \mathbf{R}^{-1}$ , con

$$\mathbf{R} = \frac{1}{N} \sum_{k=1}^N (\mathbf{z}_k - \bar{\mathbf{z}})(\mathbf{z}_k - \bar{\mathbf{z}})^T. \quad (3.20)$$

Aquí  $\bar{z}$  denota la media de la muestra de los datos. En este caso,  $\mathbf{A}$  induce la norma de Mahalanobis en  $\mathbb{R}^n$  [10].

La norma métrica influye en el criterio de partición cambiando la medida de disimilitud. La norma Euclidiana induce clusters hiper-esféricos, esto es, clusters cuyas superficies de membresías constantes son hiper-esferas. En cambio, tanto la norma diagonal como la norma de Mahalanobis generan grupos hiper-elipsoidales. La diferencia es que con la norma diagonal, los ejes de las hiper elipsoides son paralelos a los ejes coordenados mientras que con la norma de Mahalanobis la orientación de los hiper-elipsoides es arbitraria, como se muestra en la figura 3.4.

Una limitación común de los algoritmos de partición que están basados en una norma distancia fija, es que tal norma induce una estructura topológica fija en  $\mathbb{R}^n$  y fuerza a la función objetivo a preferir grupos con dicha forma aún cuando no estén presentes en los datos. Posteriormente se mostrará que la matriz norma inductora  $\mathbf{A}$  se puede adaptar usando estimaciones de la covarianza de los datos, y que se puede usar para estimar la dependencia de los datos en cada cluster.

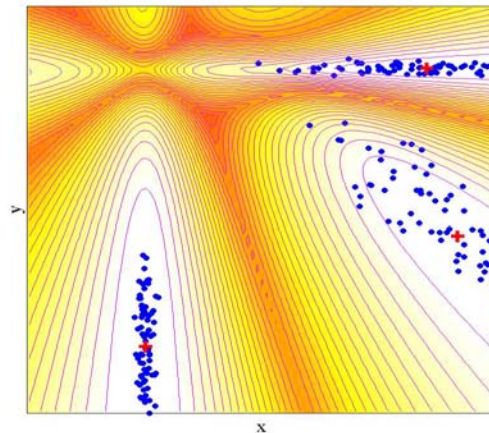


Figura 3.4 Imagen que muestra la forma en que cada Clusters se adapta a la forma de los datos

Nótese que no es útil usar cualquier matriz  $\mathbf{A}$ , puesto que los grupos pueden diferir tanto en forma como en orientación. En general, se requieren diferentes matrices  $\mathbf{A}_i$  para cada partición diferente, pero no existe una guía respecto a cómo elegirlas *a priori*.

A continuación, se presentan varias extensiones del algoritmo básico c-means. Una característica común de estos algoritmos es que pueden detectar grupos los cuales subyacen en el espacio de datos. A *grosso modo*, estos métodos se pueden caracterizar en dos grupos:

- Algoritmos que usan una medida de distancia adaptable, tales como el algoritmo Gustafson-Kessel [20], o el algoritmo difuso de estimación de máxima verosimilitud [21].

- Algoritmos basados en prototipos hiper planares o funcionales, o prototipos definidos por funciones. En estos, se incluyen los modelos difusos c-varieties [10], los modelos difusos c-eliptotipos [22], y los modelos difusos de regresión [23].

### 3.9. Clustering con Matrices Difusas de Covarianza

Se puede derivar una familia de algoritmos a partir del esquema básico de FCM adaptando la norma producto–interno (3.15d). Dos de ellas son el algoritmo Gustafson-Kessel y el algoritmo basado en estimación difusa de máxima verosimilitud. A continuación se describe el algoritmo Gustafson-Kessel que será el algoritmo que emplearemos en nuestro trabajo.

#### 3.9.1 El Algoritmo Gustafson-Kessel

Gustafson y Kessel, extendieron el algoritmo estándar fuzzy c-means empleando una norma distancia adaptable, con el fin de detectar particiones con diferentes formas geométricas en un conjunto de datos. Cada partición tiene su propia matriz norma–inductora  $\mathbf{A}_i$ , lo cual produce la siguiente norma producto–interno:

$$D_{ik\mathbf{A}_i}^2 = (\mathbf{z}_k - \mathbf{v}_i)^T \mathbf{A}_i (\mathbf{z}_k - \mathbf{v}_i). \quad (3.21)$$

Las  $\mathbf{A}_i$  matrices se usan como variables de optimización en el funcional c-means, permitiendo así que cada partición adapte la norma distancia a la estructura topológica de los datos. Si  $\mathbf{A}$  denota un  $c$ -tuple de las matrices norma-inductora:  $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_c)$ . El funcional objetivo del algoritmo Gustafson-Kessel está definida por:

$$J(\mathbf{Z}; \mathbf{U}, \mathbf{V}, \mathbf{A}) = \sum_{i=1}^c \sum_{k=1}^N (\mu_{ik})^m D_{ik\mathbf{A}_i}^2, \quad (3.22)$$

donde  $\mathbf{U} \in M_{fc}$ ,  $\mathbf{V} \in \mathbb{R}^{n \times c}$  y  $m > 1$ .

Las soluciones,

$$(\mathbf{U}, \mathbf{V}, \mathbf{A}) = \arg \min_{M_{fc} \times \mathbb{R}^{n \times c} \times PD^n} J(\mathbf{Z}; \mathbf{U}, \mathbf{V}, \mathbf{A}) \quad (3.23)$$

son puntos estacionarios de  $J$ , donde  $\text{PD}^n \in \mathbb{R}^{n \times n}$  denota un espacio de matrices de tamaño  $n \times n$  definidas positivas. Para una  $\mathbf{A}$  fija, se pueden aplicar directamente las condiciones (3.9). Sin embargo, la función objetivo (3.21) no se puede minimizar directamente con respecto a  $\mathbf{A}_i$ , puesto que ésta es lineal en  $\mathbf{A}_i$ .  $J$  se puede hacer tan pequeña como se desee haciendo a  $\mathbf{A}_i$  menos definida positiva. Para obtener una solución factible, la matriz  $\mathbf{A}_i$  se debe restringir de alguna manera; la forma usual de hacer esto es restringir el determinante de  $\mathbf{A}_i$ . Permitir que la matriz  $\mathbf{A}_i$  varíe con su determinante fijo corresponde a optimizar la forma del grupo mientras su volumen permanece constante, esto es:

$$|\mathbf{A}_i| = \rho_i, \quad \rho_i > 0, \quad \forall_i. \quad (3.24)$$

Usando el método de multiplicadores de Lagrange, se obtiene la siguiente expresión para  $\mathbf{A}_i$ :

$$\mathbf{A}_i = [\rho_i \det(\mathbf{F}_i)]^{1/n} \mathbf{F}_i^{-1}, \quad (3.25)$$

donde  $\mathbf{F}_i$  es la *matriz difusa de covarianza* del  $i$ -ésimo grupo definida por:

$$\mathbf{F}_i = \frac{\sum_{k=1}^N (\mu_{ik})^m (\mathbf{z}_k - \mathbf{v}_i)(\mathbf{z}_k - \mathbf{v}_i)^T}{\sum_{k=1}^N (\mu_{ik})^m}. \quad (3.26)$$

Nótese que la sustitución de las dos ecuaciones previas en la ecuación de la norma producto interno, genera una norma distancia de Mahalanobis cuadrática generalizada entre  $\mathbf{z}_k$  y la media  $\mathbf{v}_i$  del grupo, donde la covarianza es ponderada por los grados de membresía en la matriz  $\mathbf{U}$ . El algoritmo Gustafson-Kessel se usa para particionar los datos usados en la figura 3.3, lo cual se muestra en la figura 3.5 donde se puede apreciar la forma en que los clusters resultantes se adecúan a la forma en que están distribuidos los datos. Al final del capítulo se presenta el algoritmo de Gustafson-Kessel.

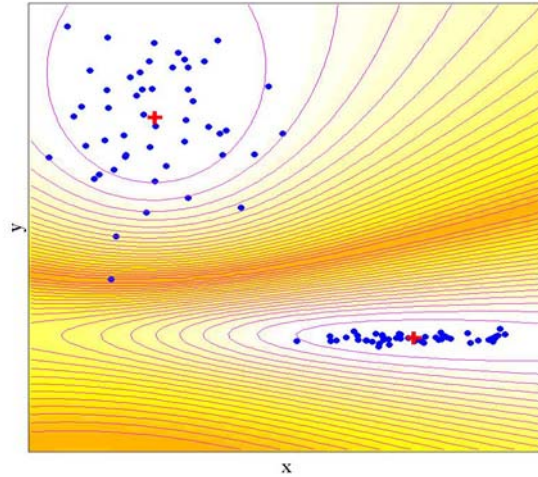


Figura 3.5 La figura muestra el resultado de usar el algoritmo Gustafson-Kessel para particionar el mismo tipo de datos usados en la figura 3.3. Se muestran los centros de los clusters y sus curvas de nivel.

### 3.10 Conclusiones

En este capítulo se describió la técnica utilizada para buscar aquella estructura subyacente en los datos de entrada-salida y considerados como representativos del comportamiento de un sistema. Esta técnica denomina clustering difuso, presenta la propiedad de ser no supervisada y en el caso del algoritmo Gustafson-kessel, de adecuarse a la forma geométrica que siguen los datos. Esta técnica es la parte medular de la metodología propuesta para modelar un sistema, pues permite describir lingüísticamente, mediante los conjuntos difusos obtenidos, el comportamiento del proceso.

Se describen también, los tres tipos de técnicas de clustering a los que podemos recurrir para abordar un mayor número de casos de estudio. Aunque nuevamente, estas técnicas requieren de la especificación de los valores de algunos parámetros a su inicialización, pues su desempeño se ve fuertemente influenciado, precisamente, por estos parámetros iniciales.

Estos métodos de clustering o tienen problemas de dimensionalidad pues actualmente los equipos de cómputo tienen mayores recursos, lo cual les permiten realizar todos los cálculos necesarios a un bajo costo.

El origen de estos métodos de análisis fue en el área de reconocimiento de patrones, así que actualmente se continúan desarrollando variaciones de los algoritmos descritos en este capítulo para resolver problemas de reconocimiento de formas, entre otros.

### Algoritmo Gustafson-Kessel (G-K)

Dado el conjunto de datos  $\mathbf{Z}$ , elija el número de clusters  $1 < c < N$ , el exponente de ponderado  $m < 1$  y la tolerancia para terminación  $\epsilon > 0$ . Inicialice la matriz difusa de partición tal que  $\mathbf{U}^{(0)} \in \mathbf{M}_{fc}$ .

**Repetir para**  $l = 1, 2, \dots$

**Paso 1: Calcular los prototipos de los grupos (medias):**

$$\mathbf{v}_i^{(l)} = \frac{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m \mathbf{z}_k}{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m}, \quad 1 \leq i \leq c.$$

**Paso 2: Calcular las matrices de covarianza de los grupos:**

$$\mathbf{F}_i = \frac{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m (\mathbf{z}_k - \mathbf{v}_i^{(l)}) (\mathbf{z}_k - \mathbf{v}_i^{(l)})^T}{\sum_{k=1}^N (\mu_{ik}^{(l-1)})^m} \quad 1 \leq i \leq c.$$

**Paso 3: Calcular las distancias:**

$$D_{ikA_i}^2 = (\mathbf{z}_k - \mathbf{v}_i)^T [(\rho_i \det(\mathbf{F}_i))^{1/n} \mathbf{F}_i^{-1}] (\mathbf{z}_k - \mathbf{v}_i). \quad 1 \leq i \leq c, 1 \leq k \leq N.$$

**Paso 4: Actualizar la matriz de partición:**

Sí  $D_{ikA_i} > 0$  para  $1 \leq i \leq c$ ,  $1 \leq k \leq N$ ,

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^c (D_{ikA_i} / D_{jkA_j})^{2/(m-1)}},$$

de otra manera

$$\mu_{ik}^{(l)} = 0 \text{ sí } D_{ikA_i} < 0, \text{ y } \mu_{ik}^{(l)} \in [0, 1] \text{ con } \sum_{i=1}^c \mu_{ik}^{(l)} = 1.$$

hasta que

$$\|\mathbf{U}^{(l)} - \mathbf{U}^{(l-1)}\| < \epsilon.$$

## CAPITULO 4

### Construcción de Sistemas Difusos

#### 4.1 Introducción

En este capítulo se presentan los métodos y algoritmos para la construcción de sistemas difusos. Ya se ha mencionado que, para la construcción de estos, las dos fuentes de información son el *conocimiento a priori* y los *datos* (las mediciones). Usualmente, el conocimiento a priori es originado por el “experto” (el operador, el diseñador del proceso, etc.) y su naturaleza es más bien del tipo aproximado (conocimiento cualitativo o heurístico), por lo que en este sentido, a estos modelos difusos se les considera como sistemas expertos difusos [24].

En el caso de la construcción de modelos difusos a partir de datos, este proceso involucra métodos tanto de tipo lógicos como de razonamiento aproximado, pero también incluye algunas ideas originadas en el área de las Redes Neuronales, en el análisis de datos y en la identificación de sistemas convencionales. Se ha mencionado también que a la tarea de adquisición o sintonizado de modelos difusos por medio de datos, se le denomina de forma genérica como *identificación difusa*.

Si se integran en un mismo marco teórico el conocimiento cualitativo y los datos para crear un modelo difuso, entonces se pueden establecer dos principales metodologías para obtenerlo:

1. El conocimiento experto, expresado en forma verbal, se traduce en una colección de reglas *si-entonces*. De esto, se obtiene una estructura del modelo. El entonado de los parámetros de esta estructura (las funciones de membresía, los parámetros consecuentes o consecuentes tipo impulso), se puede realizar usando los datos de entrada-salida proporcionados por mediciones realizadas sobre el sistema. Algunos algoritmos particulares de entonado aprovechan el hecho de que en el ámbito computacional, un sistema difuso se puede visualizar como una estructura por capas (tipo red) semejante a una red neuronal artificial, a la cual se le pueden aplicar algoritmos de aprendizaje estándares. A esta metodología se le denomina usualmente *modelado neuro-difuso*.



2. Si ningún conocimiento a priori respecto del sistema se usa inicialmente para formular las reglas, y el modelo difuso se construye por completo a partir de los datos, a esta metodología se le puede denominar *extracción de reglas*. En esta, se espera que las reglas y las funciones de membresía extraídas puedan proporcionar una interpretación *a posteriori* sobre el comportamiento del sistema, que sea amigable y legible, de tal forma que el experto pueda confrontar los resultados con su propio conocimiento, además de poder modificar las reglas o añadir algunas si lo considera necesario. En esta metodología, los algoritmos difusos de clustering, son una de las técnicas que se usan para resolver el problema de la extracción de reglas [25].

Considerando la naturaleza de la aplicación particular, estas técnicas se pueden combinar para buscar la mejor manera de obtener el modelo buscado. A continuación, se describen los principales pasos para la construcción de modelos difusos basados en el conocimiento, así como las técnicas principales para extraer o sintonizar los modelos difusos mediante datos.

## 4.2 Modelado del Sistema.

Diseñar un modelo difuso requiere resolver dos problemas: identificar la estructura del modelo e identificar los parámetros del modelo. La estructura determina la flexibilidad del modelo para la aproximación de la transformación de las entradas a las salidas (lo cual se desconoce). Los parámetros se entonan para adecuarse a los datos disponibles. Un modelo con una estructura compleja es capaz de aproximar funciones complicadas, pero, al mismo tiempo, sus propiedades de generalización disminuyen. Una buena capacidad de generalización significa que un modelo que se ajusta bien a un conjunto de datos funcionará igualmente bien con otro conjunto de datos del mismo proceso. En los modelos difusos, la selección de la estructura involucra seleccionar entre:

- *Variables de entrada y salida.* Cuando el sistema es complejo, no siempre queda claro cuáles variables se deben usar como entradas al modelo. En general, para realizar esta elección, las fuentes de información son el conocimiento *a priori*, algunas restricciones sobre el comportamiento del proceso y el propósito del modelado. Si el sistema es dinámico, se deben hacer también estimaciones del orden del sistema.
- *Estructura de las reglas.* Esta elección involucra el tipo de modelo (lingüístico, impulso o Takagi-Sugeno) y la forma de los antecedentes y los consecuentes. La elección de las reglas se ve fuertemente influenciada por el propósito del modelo así como por el tipo de conocimiento disponible.
- *Número y tipo de Funciones de Membresía para cada variable.* La elección de estas funciones determina el nivel de detalle (granularidad) del modelo. En este caso, también la elección de éstas se ve fuertemente influenciada por el propósito del modelo así como por el tipo de conocimiento disponible. Para añadir o remover funciones de membresía del modelo, se pueden usar métodos automatizados controlados por datos.

Aquí también se debe considerar que un aumento en el número de funciones, aumenta el costo computacional.

- *El tipo de Mecanismo de Inferencia.* Esta elección consiste en seleccionar entre los tipos de modelos, Mamdani, Takagi-Sugeno, y sus variaciones.

Un comentario respecto de estas selecciones; para poder realizar el aprendizaje a partir de datos de un modelo difuso, es preferible utilizar los operadores diferenciables, el operador producto y el operador suma, en lugar de los operadores estándar min y max, pues éstos no son diferenciables; esto permite construir leyes convergentes sencillas.

Una vez que se establece la estructura del modelo, el desempeño del modelo se puede optimizar ajustando sus parámetros. Los parámetros sintonizables de los modelos lingüísticos son los parámetros de las funciones de membresía antecedentes y consecuentes (esto consiste en determinar su forma y posición), y las reglas (estas determinan la transformación entre las regiones difusas antecedentes y consecuentes). En el caso de los modelos difusos Takagi-Sugeno, éstos tienen parámetros en las funciones de membresía antecedentes y en las funciones de los consecuentes.

### **4.3 Construcción del Modelo Difuso basado en el conocimiento.**

Este tipo de método es también referido como *método directo*, pues utiliza el marco teórico de los Sistemas Difusos para formalizar el conocimiento del experto; éste fue el primer método que se utilizó para la creación de un modelo difuso.

Si se dispone de un experto que pueda dictar las reglas de comportamiento del sistema, se deben seguir los siguientes pasos para diseñar un modelo difuso (lingüístico):

1. Elegir las variables de entrada y salida, la estructura de las reglas y el método de inferencia y de defusificación.
2. Especificar el conocimiento disponible en forma de reglas difusas *si-entonces*
3. Decidir sobre el número de términos lingüísticos para cada variable y especificar sus correspondientes funciones de membresía.
4. Validar el modelo. En caso de que el modelo no satisfaga los objetivos de modelado, entonces se procede a iterar sobre los pasos previos.

El éxito de esta metodología depende en gran medida del tipo de problema a resolver y de la cantidad y calidad del conocimiento disponible. Para ciertos problemas, se pueden obtener modelos útiles con relativa facilidad, pero para otros puede ser un procedimiento tardado e ineficiente (especialmente si se entonan manualmente los parámetros del modelo). De esto, se deduce que es útil combinar el diseño basado en el conocimiento con algún método controlado por datos para el entonado de los parámetros del modelo.

## 4.4 Construcción del Modelo Difuso basado en datos.

Cuando la obtención del modelo difuso se realiza conjugando el conocimiento heurístico, expresado en forma de reglas, con la información contenida en los datos, se puede recurrir a varias técnicas para la estimación de los parámetros del modelo. En nuestro caso, recurrimos a las técnicas denominadas: *Modelado basado en plantilla* y obviamente a nuestra propuesta denominada *Construcción de modelos difusos a partir de algoritmos de clustering*.

En general, para la construcción de un sistema difuso se considera que se dispone de un conjunto de  $N$  pares de datos de entrada-salida  $\{(\mathbf{x}_i, y_i) \mid i = 1, 2, \dots, N\}$ , donde  $\mathbf{x}_i \in \mathbb{R}^p$  son vectores de entrada y  $y_i$  son los escalares de salida.  $\mathbf{X} \in \mathbb{R}^{N \times p}$  denota una matriz que contiene los posibles  $\mathbf{x}_k^T$  vectores de entrada en sus renglones y  $\mathbf{y} \in \mathbb{R}^N$  denota el vector que contiene las salidas  $y_k$  correspondientes, esto es:

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T, \quad \mathbf{y} = [y_1, \dots, y_N]^T. \quad (4.1)$$

### 4.4.1 Construcción de modelos difusos basados en plantillas.

Esta técnica se emplea para crear sistemas difusos tipo Takagi-Sugeno. En este caso, primeramente los dominios de las variables antecedentes son particionados mediante un número específico de funciones de membresía uniformes y equidistantes. Posteriormente, se establece la base de reglas de tal manera que cubra aquellas combinaciones de los términos antecedentes que representen el comportamiento del sistema. Finalmente, los parámetros consecuentes se estiman por el método de mínimos cuadrados.

Esta técnica es la que utilizaremos en el capítulo 6, para obtener los modelos correspondientes a las variables fósforo y H Totales en el proceso de la biopila. Esto, por adecuarse a sus características de comportamiento.

La estimación de los parámetros consecuentes se realiza a partir de los datos de entrada-salida, mediante la técnica de mínimos cuadrados. Un inconveniente de la técnica de modelado basada en una plantilla es que el número de reglas en el modelo crece muy rápidamente.

Si no existe conocimiento respecto de qué variables provocan la no-linealidad del sistema, entonces todas las variables antecedentes se particionan uniformemente; sin embargo, en general, la complejidad del comportamiento global del sistema no es uniforme. En efecto, algunas regiones pueden ser muy bien aproximadas por un sólo modelo, mientras que en otras regiones se requerirá una partición más fina, por lo que, a fin de tener una representación eficiente con el menor número de reglas posibles, las funciones de membresía se deben colocar de tal manera que capturen el comportamiento no uniforme del sistema. Esto se resuelve haciendo que las mediciones del sistema se usen para crear estas

funciones de membresía, utilizando pseudo particiones como se mostrará en nuestra metodología propuesta.

#### 4.4.2 Construcción de modelos difusos a través de Clustering Difuso.

En el caso de que la no linealidad impida el uso distribuciones que particionen el espacio uniformemente, recurrimos a particionar el espacio de los datos representativos del sistema mediante sub conjuntos difusos, los cuales, a su vez, representan sub modelos locales lineales. Si el sistema a modelar es un sistema dinámico, se debe seleccionar una estructura de regresión antes de proceder a realizar el proceso de clustering, esto con el fin de representar adecuadamente las dinámicas del sistema. Cuando exista poco conocimiento *a priori* respecto del funcionamiento del sistema, éste se puede representar mediante una estructura NARX (el cual se describe en la sección 4.4.2.1) usando los pares de datos entrada-salida. Los parámetros consecuentes de las reglas se obtienen usando métodos de mínimos-cuadrados.

Tal como se estableció en el capítulo 3, el concepto de grado de membresía se usa para representar el grado al cual un objeto dado, representado como un vector de características, es similar o compatible con un objeto prototipo. Este grado de similitud se calcula usando una medida de distancia adecuada. En la figura 4.1, se muestra esquemáticamente la forma en que las reglas difusas *si-entonces* se pueden extraer al proyectar los clusters obtenidos sobre cada eje; en este caso, se trata de un conjunto de datos en  $\mathbb{R}^2$  particionado en tres grupos cuyos prototipos son  $v_1$ ,  $v_2$  y  $v_3$ . Se usa la distancia Euclidiana.

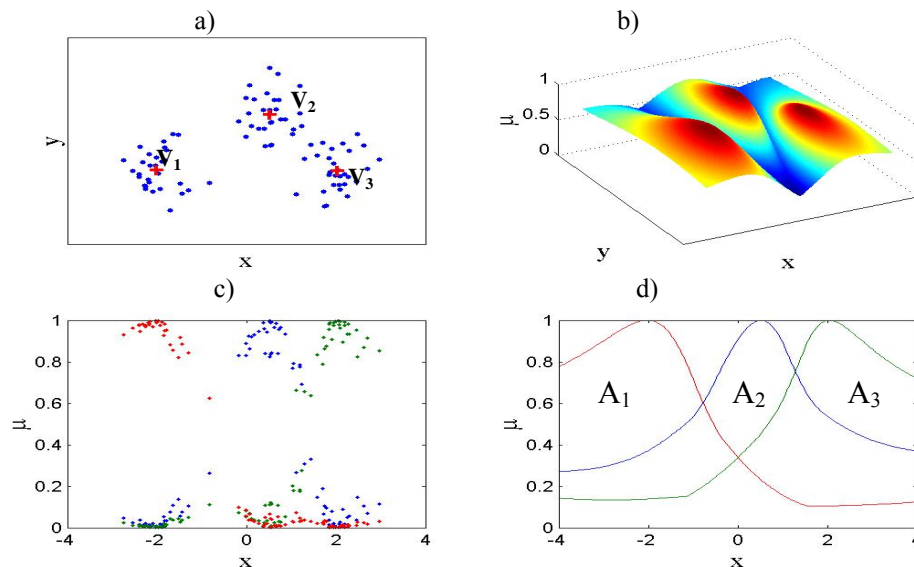


Figura 4.1 a) Datos bi-dimensionales a particionar. b) Resultado de evaluar el continuo de datos en x e y en los centros obtenidos del proceso de clustering. c) Los datos proyectados sobre el eje x. d) Las Funciones de membresía obtenidas a partir de las proyecciones.

Para crear el modelo difuso Takagi-Sugeno, las funciones de membresía antecedentes y los parámetros consecuentes se pueden extraer de cada cluster obtenido como resultado del proceso de clustering, estas reglas quedarían como:

- Si**  $x$  es  $A_1$  **entonces**  $y = a_1 x + b_1$ ,
- Si**  $x$  es  $A_2$  **entonces**  $y = a_2 x + b_2$ ,
- Si**  $x$  es  $A_3$  **entonces**  $y = a_3 x + b_3$ ,

En el modelo Takagi-Sugeno, cada cluster obtenido se representa por una regla. Las funciones de membresía para los conjuntos difusos  $A_1$ ,  $A_2$ , y  $A_3$  se generan por la proyección puntual (figura 4.1 c) de la matriz de partición difusa sobre la variable antecedente. Estos Conjuntos Difusos definidos de manera puntual, se aproximan por una función paramétrica adecuada (figura 4.1 d). Los parámetros consecuentes para cada regla se obtienen mediante la estimación de mínimos cuadrados (sección 4.4.1).

El resultado de aproximar una función a partir del Modelo difuso obtenido mediante clustering se muestra en la figura 4.2. En este caso los datos a aproximar son generados por la ecuación  $y = \sin(7 \cdot x)$ , para valores de  $x$  en el intervalo  $[0 \ 1]$ .

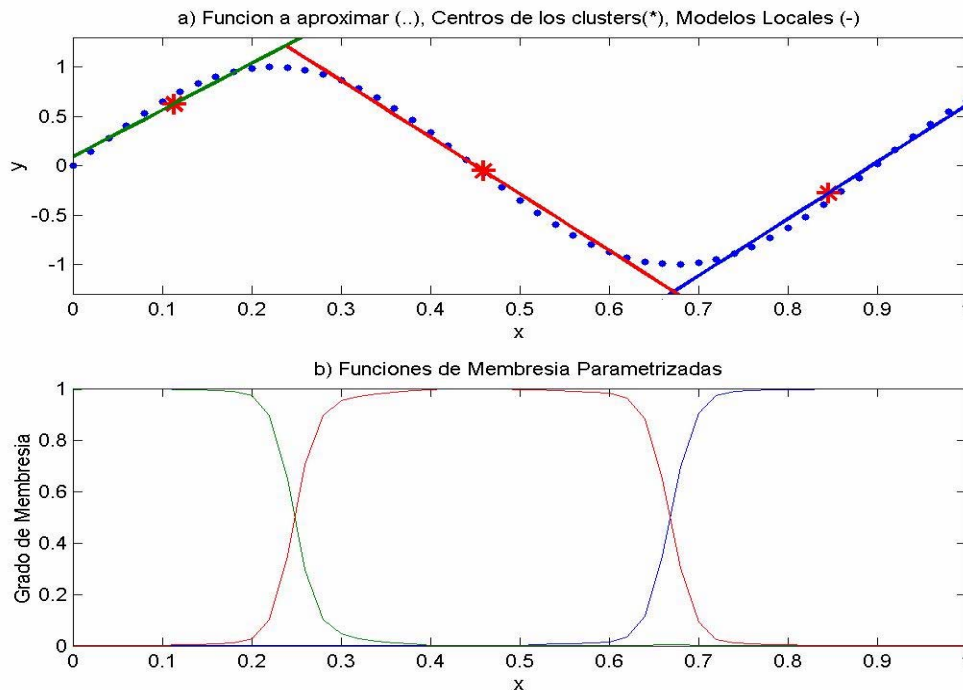


Figura 4.2 En la Figura a) se muestran tanto los datos de la función original (...), los centros obtenidos, en este caso usando el algoritmo Gustafson Kessel (\*) y los modelos locales lineales de cada regla. En la figura b) se muestran las funciones de membresía parametrizadas obtenidas por la proyección de cada cluster sobre el eje  $x$ .

En la figura 4.2a se muestran los modelos locales lineales obtenidos mediante el proceso de clustering difuso y en la figura 4.2b se muestran las funciones de membresía proyectadas y

parametrizadas por una función de membresía. El modelo difuso se expresa, en términos de reglas del tipo Takagi-Sugeno como:

$$\begin{aligned} \text{Si } x \text{ es } C_1 \text{ entonces } y &= 5.77x - 5.16, \\ \text{Si } x \text{ es } C_2 \text{ entonces } y &= 4.73x + 0.08, \\ \text{Si } x \text{ es } C_3 \text{ entonces } y &= -5.73x + 2.58, \end{aligned}$$

Donde  $C_1$ ,  $C_2$ , y  $C_3$ , son los conjuntos difusos correspondientes a los modelos locales.

En estas reglas, cada uno de los consecuentes es una función lineal. Estas funciones lineales son las que se grafican en la figura 4.2a en donde se puede apreciar al alto grado de aproximación obtenido por cada modelo local.

#### 4.4.2.1. Construcción del Modelo para Sistemas Dinámicos

Los principios de identificación utilizados para sistemas estáticos se pueden extender a sistemas dinámicos de una manera relativamente directa. En este caso, el espacio producto se forma por los regresores (datos de entrada-salida retardados) y por el regresando (la variable de salida a predecir).

La regresión no lineal es el modelado de la dependencia estática de una variable de respuesta, llamada el regresando,  $y \in Y \subset \mathbb{R}$  respecto del vector de regresión  $\mathbf{x} = [x_1, \dots, x_p]^T$  sobre algún dominio  $X \subset \mathbb{R}^p$ . A los elementos del vector de regresión se les llama los *regresores* y al dominio  $X$  el *espacio de regresión*. Se supone que el sistema generador de los datos se puede describir por:

$$y \approx f(\mathbf{x}) \tag{4.2}$$

La función determinística  $f(\cdot)$  captura la dependencia de  $y$  sobre  $\mathbf{x}$ , y el símbolo  $\approx$  significa que  $y$  es aproximada por una función de  $\mathbf{x}$ , es decir, no es una función exacta de  $\mathbf{x}$ .

Dado que se ha demostrado que los sistemas difusos son aproximadores genéricos de funciones [8], [9], estos se pueden aplicar a problemas de regresión no lineales genéricos. El objetivo de la regresión es usar los datos para construir una función  $F(\mathbf{x})$  que sirva como una aproximación razonable de  $f(\mathbf{x})$  no únicamente para los datos dados, sino para el dominio completo de  $X$ . La definición de “aproximación razonable” depende del propósito para el cual se construye el modelo. Si el objetivo del modelado es para obtener predicciones de  $y$ , entonces la precisión es el criterio más relevante. Una forma de evaluar el error de precisión entre funciones es mediante la integral del error:

$$I = \int_X \|f(\mathbf{x}) - F(\mathbf{x})\| d\mathbf{x}, \tag{4.3}$$

sobre el dominio total  $\mathbf{X}$ . Dado que el valor de  $f$  se conoce sólo en los puntos disponibles de los datos, este tipo de error no se puede calcular exactamente, por lo que en general se calcula un error de predicción promedio sobre los datos disponibles, esto es:

$$J = \frac{1}{N} \sum_{i=1}^N \|f(\mathbf{x}_i) - F(\mathbf{x}_i)\| \quad (4.4)$$

dónde  $N$  es el número de muestras de datos disponibles.

Obtener un mínimo en (4.3) implica que se tiene el mejor modelo posible con la estructura seleccionada. Sin embargo, con el criterio  $J$  esto no es el caso, pues este criterio sólo garantiza que el modelo ajusta los datos disponibles con el menor error. Esta desventaja implica la necesidad de un paso de validación, a fin de evaluar que tan adecuado es el modelo actual respecto de la región de interés total  $\mathbf{X}$ .

Durante la identificación de un sistema no lineal existen varias posibilidades para la elección de los regresores. Dado que en nuestro caso, se utilizarán algoritmos de clustering para hacer la identificación, se mencionará brevemente a continuación la estructura que se puede usar para hacer la identificación.

Varios de los métodos de identificación no lineales usan el modelo NARX (Nonlinear AutoRegressive with eXogenous input) para representar el sistema bajo estudio, por ejemplo, en las redes neuronales [26], en las funciones de base radial [7], y en los modelos difusos [4], [9].

El modelo NARX, establece una relación entre los datos de entrada-salida pasados y la salida a predecir, esto es:

$$\hat{y}(k+1) = F(y(k), \dots, y(k-n_y+1), u(k), \dots, u(k-n_u+1)), \quad (4.5)$$

donde  $k$  denota las muestras discretas en el tiempo,  $n_u$  y  $n_y$  son enteros relacionados al orden del sistema y  $F$  denota un modelo difuso. En el modelo NARX, el vector de regresión es una colección de un número finito de valores de entrada y salida pasados,  $\mathbf{x}(k) = [y(k), \dots, y(k-n_y+1), u(k), \dots, u(k-n_u+1)]^T$ . El regresando es la salida a predecir  $\hat{y}(k+1)$ . A partir de un conjunto de datos de entrada y salida de un sistema dinámico desconocido,  $S = \{(u(j), y(j)) \mid j = 1, 2, \dots, N\}$ , la función  $F(\cdot)$  en (4.5), se puede aproximar usando regresión no lineal estática. Los retardos de la entrada a la salida se pueden incorporar directamente en el vector de regresión  $\mathbf{x}(k) = [y(k), \dots, y(k-n_y+1), u(k-n_d+1), \dots, u(k-n_d-n_u+2)]^T$ , donde  $n_d$  es el retardo en el muestreo.

En nuestro caso, la determinación de la estructura de un modelo a partir de datos de entrada-salida se centra en la elección del tipo de modelo y a los parámetros relacionados con la estructura  $n_y$ ,  $n_u$  y a los retardos  $n_d$ . En general, se usa el conocimiento *a priori* para establecer una primera aproximación del rango de estos parámetros y se busca una estructura que minimice algún criterio dado.

Para determinar la estructura del sistema, usualmente es mucho más conveniente iniciar con un modelo lineal, puesto que las técnicas de identificación lineales son mucho más sencillas y numéricamente más robustas que los métodos no lineales. Como punto de partida para el modelado no lineal, se puede tomar la estructura del mejor modelo lineal. En nuestro caso, se pueden usar medidas de validez de los clusters para guiar la selección del orden del modelo y el número de clusters dentro de la estructura dada.

En el modelado difuso los sistemas MIMO se descomponen en un conjunto de sistemas MISO (Múltiple Entrada y una Única Salida). La razón de esto es que se tiene una mayor flexibilidad, pues cada salida por si misma puede estar asociada con diferentes tipos de no-linealidad, es decir, una salida puede contener una no-linealidad compleja en una región mientras que otra salida puede ser lineal en la misma región. Además, mediante este tipo de división se reducen tanto el número de funciones de membresía como el número de reglas.

- Identificación del modelo difuso

Si  $\mathbf{Z}$  denota el conjunto de datos a particionar y es construido concatenando la matriz de datos del regresor  $\mathbf{X}$  y el vector del regresando  $\mathbf{y}$ , esto es:

$$\mathbf{Z} = [\mathbf{X}, \mathbf{y}], \quad (4.6)$$

Entonces, este conjunto de datos es un sub conjunto del espacio producto cartesiano  $X \times Y$ , definido por la relación funcional no-lineal:

$$\mathbf{Z} \subset X \times Y \quad \text{tal que } \mathbf{y} \approx f(\mathbf{X}) \quad (4.7)$$

Los algoritmos difusos de clustering se aplican para particionar el conjunto de datos  $\mathbf{Z}$  construido, en sub conjuntos difusos, esto permite detectar sub estructuras lineales subyacentes, las cuales están presentes en los datos.

Las membresías de los datos respecto de cada cluster, se da en la matriz de partición difusa  $\mathbf{U}$ . Cada cluster, se caracteriza por su centro y su matriz de covarianza, la cual representa la varianza de los datos en el cluster. Un algoritmo de clustering difuso  $\mathfrak{G}$  se puede considerar como una transformación:  $(Z \times \mathbb{N}) \rightarrow (M_{fc} \times \mathbb{R}^{n \times c} \times PD^n)$ , esto es:

$$(\mathbf{U}, \mathbf{V}, \mathbf{F}) = \mathfrak{G}(\mathbf{Z}, c; \mathbf{U}^o, m, \epsilon) \quad (4.8)$$

dónde  $c$  es el número de clusters,  $\mathbf{U}^o$  es la matriz de partición inicial y  $m, \epsilon$ , son los parámetros del algoritmo de clustering.

La información sobre la forma y orientación de cada  $i$  cluster está contenida en la matriz de covarianza del  $i$ -ésimo cluster  $\mathbf{F}_i$ . El  $j$ -ésimo valor propio y el  $j$ -ésimo vector propio unitario de  $\mathbf{F}_i$  se denotan por  $\lambda_{ij}$  y  $\phi_{ij}$ , respectivamente. Los valores propios de la matriz de covarianza del  $\mathbf{F}_i$  cluster se arreglan de tal manera que:



$$\lambda_{i1} \geq \lambda_{i2} \geq \dots \geq \lambda_{in}, \quad (4.9)$$

y los vectores propios se etiquetan de la misma manera. Los vectores propios  $\phi_{i1}$  a  $\phi_{in-1}$  contienen el  $i$ -ésimo sub espacio lineal del cluster, siendo el  $n$ -ésimo vector propio  $\phi_{in}$  la componente normal a este sub espacio lineal. Puesto que  $\lambda_{in}$  es el valor propio más pequeño, a  $\phi_{in}$  se le llama el vector propio más pequeño.

Usualmente, la identificación de un modelo Takagi-Sugeno se realiza en dos pasos. En un primer paso, se determinan los conjuntos  $A_{ij}$  correspondientes a los antecedentes de las reglas. Esto se realiza mediante la técnica de clustering difuso y se produce una partición del espacio producto de los datos de entrada-salida. En el segundo paso, se determinan los parámetros consecuentes.

Para determinar las reglas, se propone aplicar la metodología de fuzzy clustering en el espacio producto cartesiano de  $X \times y$  a fin de particionar los datos en regiones características, es decir, regiones en donde el comportamiento del sistema pueda ser aproximado por modelos locales lineales (una regla por cada modelo local). De  $X$  e  $y$  se forma la matriz de datos  $Z$ , la cual será particionada.

$$Z = \begin{bmatrix} X^T \\ \mathbf{y}^T \end{bmatrix}. \quad (4.10)$$

Dados los datos  $Z$  y un número de clusters  $M$ , el algoritmo de clustering difuso particiona  $Z$  en  $M$  clusters difusos.

Una partición difusa se puede representar por la matriz  $U$  de  $N \times M$  elementos, cuyo elemento  $\mu_{ik} \in [0,1]$  es el valor de la membresía de la combinación entrada-salida en la  $k$ -ésima columna de  $Z$  en el  $i$ -ésimo cluster. Para calcular la matriz de partición difusa  $U$ , se aplica el algoritmo de clustering Gustafson-Kessel, presentado en el capítulo anterior. Es de observar, que cada cluster tiene su propia norma distancia, inducida por la matriz de covarianza difusa del cluster, esto permite que el cluster se adapte a la estructura local de los datos. Esta es como se mencionó, la característica que hace al algoritmo GK adecuado para la identificación de sub espacios en los datos tal, que puedan ser modelados efectivamente por las reglas en el modelo TS.

- Generación de las funciones de membresía antecedentes

Las funciones de membresía antecedentes se pueden obtener proyectando la partición difusa sobre las variables antecedentes o calculando los grados de membresía directamente en el espacio producto de las variables antecedentes, esto es, por la proyección de los vectores propios, como se muestra a continuación.

- Generación de las funciones de membresía antecedentes por proyección.

El principio de este método consiste en proyectar los conjuntos difusos multidimensionales, definidos de manera puntual en los renglones de la matriz de partición  $U$ , sobre las variables antecedentes individuales de las reglas. Estas variables pueden ser las variables de regresión originales, en cuyo caso la proyección es una proyección ortogonal de los datos. Se pueden obtener nuevas variables antecedentes transformadas por medio de la proyección de los vectores propios, usando el  $p$  vector propio más grande de las matrices de covarianza de las particiones. La proyección del vector propio es útil cuando las particiones son *opacas* a los ejes del espacio de regresión y no se pueden representar con suficiente precisión por proyecciones ortogonales a los ejes.

- Proyección ortogonal a los ejes.

Este método proyecta la matriz de partición difusa sobre los ejes de las variables antecedentes  $x_j$ ,  $1 \leq j \leq p$ . Las reglas TS se expresan en forma conjuntiva

$$R_i: \text{Si } x_1 \text{ es } A_{i1} \text{ y } \dots \text{ y } x_p \text{ es } A_{ip} \text{ entonces } y_i = \mathbf{a}_i^T \mathbf{x} + b_i, \quad i = 1, 2, \dots, k.$$

A fin de obtener las funciones de membresía para los  $A_{ij}$  conjuntos difusos antecedentes, los conjuntos difusos multidimensionales definidos de manera puntual en el  $i$ -ésimo renglón de la matriz de partición, se proyectan sobre los  $x_j$  regresores usando el operador de proyección:

$$\mu_{A_{ij}}(x_{jk}) = \text{proy}_j(\mu_{ij}). \quad (4.11)$$

Cuando se calcula el grado de satisfacción de la  $i$ ésima regla  $\beta_i(\mathbf{x})$ , la partición original en el espacio producto antecedente se reconstruye aplicando el operador de intersección en el espacio producto cartesiano de las variables antecedentes, esto es:

$$\beta_i(\mathbf{x}) = \mu_{A_{i1}}(x_1) \wedge \mu_{A_{i2}}(x_2) \wedge \dots \wedge \mu_{A_{ip}}(x_p). \quad (4.12)$$

Se pueden usar otras  $t$ -normas para definir la conjunción, tales como el operador producto, en lugar del operador mínimo. En general, esta reconstrucción no es exacta, y puede dar origen a un *error por descomposición* en el caso en el que las particiones sean oblicuas a los ejes. Este error se puede compensar parcialmente, estimando los parámetros consecuentes mediante una operación denominada por mínimos cuadrados global. Este método es útil, a pesar de que alguna información se pierde a través de la proyección ortogonal a los ejes, puesto que provee la posibilidad de interpretar el modelo usando directamente los regresores.

- Proyección de los vectores propios.

El error por descomposición se puede reducir por medio de la proyección de los vectores propios, usando la información sobre la orientación de los clusters. Si  $\mathbf{F}^x$  denota la partición de la matriz de covarianza del cluster, correspondiente al regresor  $\mathbf{x}$ , esto es:

$$\mathbf{F}^x = [f_{ij}], \quad 1 \leq j \leq p, \quad (4.13)$$

donde  $p$  es la dimensión del regresor  $\mathbf{x}$ , entonces, los vectores propios unitarios de  $\mathbf{F}^x$  constituyen una base ortonormal orientada en la dirección de los ejes elipsoidales del cluster. La matriz de regresión  $\mathbf{X}$  se proyecta sobre esta nueva base, mediante:

$$\mathbf{X}'_i = \mathbf{X}\mathbf{H}_i, \quad (4.14)$$

donde  $\mathbf{H}_i \in \mathbb{R}^{p \times p}$  es una matriz de proyección que contiene los vectores propios unitarios de  $\mathbf{F}^x$  en sus columnas. En general, cada cluster tendrá su propia matriz de proyección; consecuentemente, cada regla tendrá diferentes variables antecedentes, calculadas como combinaciones lineales de los regresores, así:

$$R_i: \text{Si } x'_{i1} \text{ es } A_{i1} \text{ y } \dots \text{ y } x'_{ip} \text{ es } A_{ip} \text{ entonces } y_i = \mathbf{a}_i^T \mathbf{x}'_i + b_i, \quad i = 1, 2, \dots, k.$$

Esta transformación es similar al análisis de componentes principales, y se puede considerar como un tipo de pre-procesamiento de datos para cada regla.

- Parametrización de las funciones de membresía.

Proyectando el  $i$ -ésimo renglón  $\mu_i$  de la matriz de partición difusa  $\mathbf{U}$  sobre la variable antecedente  $x_j$ , se obtiene una definición puntual del conjunto difuso  $A_{ij}$ . Con el fin de obtener un modelo de predicción o un modelo adecuado para propósitos de análisis, las funciones de membresía antecedentes se deben expresar de tal forma que permitan calcular los grados de membresía también para datos de entrada que no estén contenidos en el conjunto de datos  $\mathbf{Z}$ . Esto se logra aproximando la función de membresía, definida puntualmente, por alguna función paramétrica adecuada, como se describe en la figura 4.3.

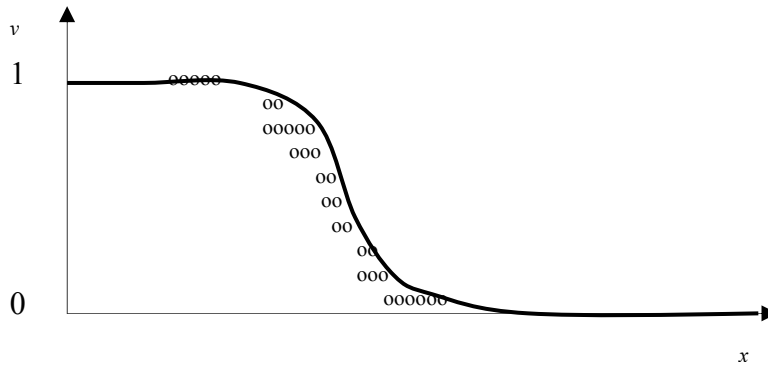


Figura 4.3. Aproximación de los datos proyectados por una función de membresía paramétrica.

Las funciones de membresía exponenciales a trazos han probado ser adecuadas para la representación precisa de la forma de los clusters. Este tipo de función se ajusta a la envolvente de los datos proyectados, optimizando numéricamente sus parámetros. Una ventaja de este método sobre las funciones de membresía multidimensionales, es que las funciones de membresía proyectadas siempre pueden ser aproximadas tal que se obtengan conjuntos difusos convexos. Además, se pueden usar funciones de membresía asimétricas para reflejar la partición del problema de regresión no lineal considerado.

- Estimación de los Parámetros Consecuentes.

Existen varios métodos para obtener los parámetros consecuentes. Los parámetros consecuentes se pueden calcular directamente, con base en la interpretación geométrica del modelo TS, a partir de los puntos prototipos de los clusters y de los vectores propios más pequeños de las matrices de covarianza de los clusters. Este método supone que existe un error tanto en los regresores como en los regresandos, el cual corresponde a la solución por mínimos cuadrados totales de la linealización local alrededor de los centros de los clusters.

También se puede estimar un conjunto de parámetros óptimos con respecto a la salida del modelo a partir del conjunto de datos de identificación por el método de mínimos cuadrados ordinarios. Este método se puede formular como una minimización del error de predicción total usando la fórmula de defusificación TS, o como una minimización de los errores de predicción de los modelos locales individuales, resuelto como un conjunto de  $c$  problemas de mínimos cuadrados ponderados independientes. Si el problema debe servir como un predictor numérico, es preferible el método de mínimos cuadrados global puesto que éste da el menor error de predicción; sin embargo, se pueden obtener modelos locales más precisos usando los métodos de mínimos cuadrados ordinarios ponderados o totales.

- Cálculo de los Parámetros Consecuentes por el método de Mínimos Cuadrados Totales.

Los parámetros  $\mathbf{a}_i$  y  $b_i$  del modelo TS, se pueden derivar a partir de la estructura geométrica de los clusters. Se considera que la colección de  $c$  clusters aproxima la superficie de regresión. Estos clusters se pueden considerar a su vez, aproximadamente como sub espacios lineales  $p$ -dimensionales del espacio de regresión. El vector propio  $\phi_{in}$ , correspondiente al valor propio más pequeño  $\lambda_{in}$ , determina el vector normal al hiperplano formado por los vectores propios restantes del cluster.

El vector propio más pequeño del  $i$ -ésimo cluster se denotará como  $\phi_i$ , omitiendo el sub índice  $n$ . Si el vector de datos está dado por  $\mathbf{z}^T = [\mathbf{x}^T, y]$  y los prototipos de los clusters son  $\mathbf{v}_i$ , la forma normal implícita del hiperplano del consecuente está dada por:

$$\phi_i^T (\mathbf{z} - \mathbf{v}_i) = 0. \quad (4.15)$$

Esta expresión establece que el producto interno del vector  $\phi_i$  con cualquier vector perteneciente al hiperplano es igual a cero. Si se subdivide el prototipo  $\mathbf{v}_i$  en un vector  $\mathbf{v}_i^x$  correspondiente al regresor  $\mathbf{x}$  y a un escalar  $v_i^y$ , correspondiente al regresando  $y$ , esto es,  $\mathbf{v}_i^T = [(\mathbf{v}_i^x)^T, v_i^y]$ , y, de igual manera, se divide el vector propio más pequeño como  $\phi_i^T = [(\phi_i^x)^T, \phi_i^y]$ , entonces, la ecuación (4.15) se puede reescribir como:

$$[(\phi_i^x)^T, \phi_i^y] ([\mathbf{x}^T, y]^T - [(\mathbf{v}_i^x)^T, v_i^y]^T) = 0. \quad (4.16)$$

Desarrollando el producto interno, se obtiene la siguiente igualdad:

$$(\phi_i^x)^T (\mathbf{x} - \mathbf{v}_i^x) + \phi_i^y (y - v_i^y) = 0. \quad (4.17)$$

de la cual, por medio de una manipulación algebraica simple, se obtiene una ecuación explícita para el hiperplano, dada por:

$$y = \underbrace{\frac{-1}{\phi_i^y} (\phi_i^x)^T}_{\mathbf{a}_i^T} \mathbf{x} + \underbrace{\frac{1}{\phi_i^y} \phi_i^T \mathbf{v}_i}_{b_i}. \quad (4.18)$$

Comparando esta expresión con el consecuente de la regla TS, se obtienen directamente las ecuaciones para  $\mathbf{a}_i$  y  $b_i$ :

$$\mathbf{a}_i = \frac{-1}{\phi_i^y} \phi_i^x = \frac{-1}{\phi_{ip+1}} [\phi_{i1}, \phi_{i2}, \dots, \phi_{ip}]^T, \quad (4.19a)$$

$$b_i = \frac{1}{\phi_i^y} \phi_i^T \mathbf{v}_i. \quad (4.19b)$$

\*\* El algoritmo de clustering con distancias adaptativas (GK) se puede considerar como un método que divide un problema de regresión no lineal global en un conjunto de problemas TLS locales lineales. Esta descomposición está basada en la minimización de la varianza de los datos con respecto a los prototipos de los clusters; la varianza se mide mediante la norma producto interno adaptada iterativamente a la forma del cluster.

- Estimación de los parámetros consecuentes por el Método de Mínimos Cuadrados Ordinarios.

Existen dos métodos para estimar los parámetros consecuentes  $\mathbf{a}_i$  y  $b_i$  mediante la técnica de mínimos cuadrados ordinarios.

En el primero, se resuelven  $c$  problemas de mínimos cuadrados ponderados independientes, uno para cada cluster. Esta metodología es independiente del método de agregación/defusificación usado, puesto que estima por separado los parámetros para cada modelo (como el método TLS).

En el segundo, se resuelve únicamente un problema de mínimos cuadrados global junto con la fórmula de defusificación media ponderada. Esta metodología produce un estimado del error mínimo de predicción.

Los datos de identificación y los grados de membresía de la partición difusa se muestran en forma de matrices, esto es:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \mathbf{W}_i = \begin{bmatrix} \mu_{i1} & 0 & \cdots & 0 \\ 0 & \mu_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_{iN} \end{bmatrix}. \quad (4.20)$$

Los parámetros consecuentes de la regla, pertenecientes al  $i$ -ésimo cluster,  $\mathbf{a}_i$  y  $b_i$  se concatenan en un sólo vector de parámetros  $\boldsymbol{\theta}_i$ , donde

$$\boldsymbol{\theta}_i = [\mathbf{a}_i, b_i]^T. \quad (4.21)$$

Creamos una matriz de regresión extendida  $\mathbf{X}_e$  añadiendo un vector columna unitario a  $\mathbf{X}$ ;

$$\mathbf{X}_e = [\mathbf{X}, \mathbf{1}] \quad (4.22)$$

Suponiendo que cada cluster representa un modelo local lineal del sistema, el vector de parámetros consecuentes  $\boldsymbol{\theta}_i$ ,  $i=1,2,\dots,c$ , se puede estimar independientemente, usando el método de mínimos cuadrados ponderado. Los grados de membresía  $\mu_{ik}$  de la partición difusa sirven como los pesos que expresan la relevancia del par de datos  $(\mathbf{x}_k, y_k)$  al modelo local. Si las columnas de  $\mathbf{X}_e$  son linealmente independientes y  $\mu_{ik} > 0$  para  $1 \leq k \leq N$ .

Entonces

$$\boldsymbol{\theta}_i = [\mathbf{X}_e^T \mathbf{W}_i \mathbf{X}_e]^{-1} \mathbf{X}_e^T \mathbf{W}_i \mathbf{y}, \quad (4.23)$$

es la solución por mínimos cuadrados de  $y = \mathbf{X}_e \theta + \epsilon$  donde el  $k$ -ésimo par de datos  $(\mathbf{x}_k, y_k)$  es ponderado por  $\mu_{ik}$ . Los parámetros  $\mathbf{a}_i$  y  $b_i$  están dados por:

$$\mathbf{a}_i = [\theta_1, \theta_2, \dots, \theta_p], \quad b_i = \theta_{p+1}. \quad (4.24)$$

Si las columnas de  $\mathbf{X}_e$  son linealmente dependientes, se debe usar la técnica basada en factorización ortogonal de  $\mathbf{X}_e$ .

La aproximación por mínimos cuadrados ponderados para el primer caso, resulta en un estimado óptimo de los parámetros de los modelos locales, pero no provee un modelo TS óptimo en términos de minimizar el error de predicción. A fin de obtener un predictor global óptimo, en el segundo caso, se debe tomar en cuenta la agregación de las reglas.

Cuando se usa la fórmula media-difusa:

$$y = \frac{\sum_{i=1}^K \beta_i(\mathbf{x}) [\mathbf{a}_i^T \mathbf{x} + b_i]}{\sum_{i=1}^K \beta_i(\mathbf{x})} \quad (4.25)$$

los parámetros consecuentes estimados se pueden obtener solucionando el problema de mínimos cuadrados lineales. El grado de cumplimiento (satisfacción)  $\beta_{ik}$  de la  $i$ -ésima regla se puede obtener de la matriz de partición difusa  $\mathbf{U}$ , pero la situación es un poco más complicada que en el caso de la aproximación local por mínimos cuadrados ponderados.

Debe recordarse que cada renglón de  $\mathbf{U}$  contiene una definición puntual de la función de membresía para los datos en el espacio-producto  $X \times Y$ . A fin de obtener una definición puntual de la función de membresía  $A_i$  en el espacio antecedente  $X$ , el  $i$ -ésimo renglón de  $\mathbf{U}$  se debe proyectar sobre el espacio antecedente:

$$\beta_{ik} = \text{proy}_{\mathbb{N}^p}(\mu_{ik}), \quad k = 1, \dots, N, \quad (4.26)$$

donde  $\text{proy}$  denota un operador de proyección.

El método por mínimos cuadrados global produce un mínimo error de predicción y por lo tanto es adecuado para derivar modelos de predicción. Sin embargo, polariza la estimación de los parámetros del modelo local. Por lo tanto, para propósitos de interpretación y análisis local del modelo TS, es preferible la aproximación por mínimos cuadrados ponderados.

## 4.5 Conclusión.

En este capítulo se integraron los métodos y algoritmos que nos permitirán completar la tarea de Construcción de Modelos Difusos. Se mencionan las dos vías para su construcción, aquella basada completamente en el conocimiento experto (método directo) y la que utiliza los datos de entrada-salida del sistema. El método directo, en general, se emplea en problemas de baja complejidad mientras que el segundo método se puede emplear en problemas de mediana y alta complejidad. Un inconveniente del método directo es el sintonizado de sus parámetros, la cual puede ser una tarea exhaustiva por la cantidad de parámetros a ajustar (tipo y posición de las funciones de membresía, grado de granularidad de las variables, los antecedentes o consecuentes mismos, etc.), además de que el este método se restringe a modelos tipo Mamdani o lingüísticos. En cambio, el método Takagi-Sugeno, es más adecuado de ser identificado a partir de datos, pues en este caso se aprovechan las técnicas desarrolladas en otras áreas de estudio. Por ejemplo, en la identificación de los parámetros consecuentes, se utilizan los métodos de mínimos cuadrados.

Hasta aquí, hemos completado la solución de las diversas tareas contenidas en nuestro planteamiento propuesto en el capítulo 2 para la construcción de Modelos Difusos. En este capítulo entonces, se conjuntas las diversas técnicas desarrolladas en los capítulos previos.

La parte central de nuestra propuesta es el empleo de la técnica de Clustering Difuso para realizar la partición de los datos en búsqueda de esa estructura que nos permita alcanzar nuestro propósito. Aquí hay que resaltar la importancia de un buen resultado en la partición de los datos pues, la exactitud del modelo final resultante, depende en gran medida de los resultados de este proceso de clustering. Esto enfatiza más la conveniencia de contar, no únicamente con datos que sean válidos, sino con el experto que proporcione aquella información “fina” que nos permita establecer los parámetros adecuados a nuestro modelo.



## CAPITULO 5

### Descripción del Bioproceso

#### 5.1 Introducción

En la actualidad, entre los sistemas más difíciles de modelar se encuentran aquellos en los que intervienen procesos biológicos. En efecto, el comportamiento de procesos en donde existe intercambio de materia orgánica presenta características altamente no lineales imposibles de modelar con modelos convencionales. Así, la utilización de la metodología presentada en este trabajo es de gran utilidad para representar el comportamiento específico de un proceso referente a la conservación del medio ambiente por bioremediación de la contaminación producida por derrame de hidrocarburos en suelo.

El proceso de bioremediación en este caso, se refiere a un bioproceso en el cual el suelo contaminado se trata con microorganismos que degradan o neutralizan el efecto nocivo de los productos petroleros primarios o secundarios. Los bioprocesos son ampliamente aplicados en la industria para fermentación de biomásas de alta-densidad, para la producción de metabólicos primarios y secundarios, así como de proteínas y otros productos biotecnológicos. En este caso, se emplea el método denominado como biopila para eliminar los hidrocarburos contenidos en el suelo [28]. Más específicamente, las biopilas se refieren a sistemas en los cuales el suelo a tratar se excava y se somete a un tratamiento del tipo “ex situ”.

Bajo condiciones óptimas de preparación del suelo, el ideal es no-compacto y granulación media (sandy loam), los microorganismos indígenas usan a los hidrocarburos como una fuente de alimento y lo convierten a dióxido de carbono y agua. Estos sistemas cuentan con un sistema de aireación para proveer de oxígeno a los microbios, un sistema de irrigación e inyección de nutrientes para proveer de nutrientes y humedad. Una vez construida la pila, y un sistema de recolección de lixiviados para controlar el exceso de humedad en la pila. El

tiempo de remediación varía dependiendo de la concentración de hidrocarburos, del suelo, de los nutrientes, de la temperatura y de las condiciones microbianas.

Así, en adelante, en este trabajo se denotará como biopila al tratamiento de biorecuperación, de tipo “ex situ”, en condiciones no saturadas. A las biopilas también se les conoce como bioceldas, biomontículos o pilas de composteo [29].

## **5.2 Componentes esenciales para la bioremediación**

Los componentes esenciales para la bioremediación son: el oxígeno, el agua, los nutrientes presentes en el suelo, el pH del suelo y la temperatura.

*Oxígeno*- En las pilas de suelo, se debe mantener un nivel de oxígeno, para realizar una degradación aeróbica de contaminantes por hidrocarburos. El oxígeno se puede suministrar inyectando aire por medio de una red de tuberías instalada en las pilas. Las pilas se deben monitorear frecuentemente para mantener un contenido de oxígeno superior al 15%.

*Agua* – El agua es esencial para la actividad microbiana. Mientras que un bajo contenido de humedad inhibe el crecimiento y la movilidad microbiana, una humedad excesiva bloqueará los poros, restringiendo por lo tanto el flujo de aire necesario. En la pila se mantiene una humedad del suelo de entre el 70% y el 95% de la capacidad del suelo en campo.

*Nutrientes* – Los nutrientes requeridos para el crecimiento microbiano son el nitrógeno, fósforo, potasio, sulfuro, magnesio, calcio, manganeso, hierro, zinc y cobre. Un suelo contaminado con hidrocarburos es muy probable que tenga deficiencias de nitrógeno y fósforo. Durante la construcción de la biopila, estos nutrientes se aplican en forma líquida o sólida y, si es necesario, se pueden aplicar a través de un sistema de irrigación.

*pH* – La mayoría de las bacterias hidrocarbonoclastas crecen mejor en ambientes de pH en un intervalo de neutral a ligeramente alcalino, así, los niveles de pH se mantienen cercanos a 7, pero definitivamente deberán estar dentro del intervalo de 5 a 9. Donde están presentes las concentraciones de compuestos aromáticos y donde el suelo tiene baja alcalinidad puede ser necesario un tratamiento de acidez.

*Temperatura* – La temperatura también afecta la razón de biodegradación. Cuando la temperatura disminuye, la actividad microbiana es baja. En caso contrario, temperaturas muy altas pueden esterilizar el suelo, creando un ambiente adverso para cierta actividad microbiana. Dentro de las biopilas se genera calor cuando los microbios digieren los hidrocarburos, a menudo elevando la temperatura de la biopila de 10° F a 30 °F. La temperatura óptima de las biopilas varía entre 50° F a 100° F, para una instalación de concreto permanente.

### **5.3 Caracterización del suelo y estudios de tratabilidad**

Siempre es necesario realizar una caracterización del suelo antes de cualquier tratamiento, para maximizar la eficiencia de la bioremediación. Esta caracterización permite contar con los valores base del contenido de hidrocarburos de petróleo totales (TPH), la población indígena de microbios (incluyendo los degradadores de hidrocarburos), niveles de nutrientes, pH, la porosidad y el contenido de humedad. Posteriormente, se llevan a cabo análisis de laboratorio, para determinar las condiciones óptimas necesarias para una biodegradación expedita. Durante la bioremediación y a intervalos designados, se pueden monitorear los parámetros identificados en los análisis de laboratorio junto con los niveles de hidrocarburos, esto facilita el seguimiento de la degradación de contaminantes y guía los ajustes del sistema para una bioremediación óptima.

### **5.4 Descripción del sistema construido**

El caso que se trata en este trabajo es del tipo alimentación-lote ya que para mantener la actividad microbiana y lograr la mayor remoción de contaminantes, a las biopilas se les agregan fuentes de nutrientes durante el proceso. En general, esta agregación se controla por un experto, quien usa una especie de receta para hacerlo. Tales recetas están basadas, a menudo, en el conocimiento y la experiencia previa del proceso.

El proceso de biorremediación se implementa mediante la construcción de pilas de dimensiones variables, formadas por suelo contaminado y materia orgánica (composteo), en condiciones favorables para el desarrollo de los procesos de biodegradación de los contaminantes.

Las biopilas han probado ser efectivas en reducir la concentración de casi todos los componentes y derivados del petróleo. Los productos del petróleo más ligeros tienden a ser removidos por evaporación durante los procesos de aireación y en menor proporción son degradados por la respiración microbiana. Los productos intermedios contienen menores porcentajes de compuestos ligeros, siendo así que la biodegradación de estos productos del petróleo es más importante que la evaporación.

Las biopilas de composta pueden airearse en forma activa, volteándolas, o bien en forma pasiva, mediante la colocación de tubos perforados.

Con base en reportes disponibles, entre los principales factores que influyen en la aplicación de las biopilas para suelos contaminados por petróleo destacan:

- Los hidrocarburos deben ser no halogenados y deben encontrarse en el suelo en concentraciones menores a 50,000 ppm.
- Dada la necesidad de excavación y posterior depósito del suelo contaminado, se requiere una superficie de trabajo relativamente grande cuyas dimensiones dependen del volumen de suelo a tratar.

- Se requiere contar con una densidad de poblaciones microbianas (>1,000 UFC/g suelo), condiciones de humedad (40%-85% de capacidad de campo), temperatura (10°C-45°C), textura (baja proporción de arcillas), pH del suelo adecuado (6 a 8) y baja concentración de metales pesados (<2,500 ppm).
- La concentración de nutrientes en el suelo deberá estar en una proporción promedio de los porcentajes de carbono, nitrógeno y fósforo, C: N: P del orden de 100: 10: 1
- El tiempo de tratamiento puede ser largo (meses a años), pero con un costo bajo.

Es de destacar que varios de los factores anteriores se deben evaluar mediante pruebas de tratabilidad, antes de aprobar la aplicación de ésta tecnología. En la Tabla 5.1, se listan las principales ventajas y desventajas de los sistemas de biopilas.

## 5.5 Factores ambientales a considerar en un proceso de biopilas.

Las variables de operación en un sistema de biopilas son: humedad, temperatura, porosidad, aireación, concentración de hidrocarburos, relación C/N, relación C/P, material orgánico del suelo y su pH, entre otros.

### 5.5.1 Nutrientes

Aunque los microorganismos se encuentran presentes en los suelos contaminados, ellos no pueden necesariamente estar en el número requerido para la bioremediación del sitio, así, su crecimiento y actividad deberán ser estimulados. La bioestimulación usualmente involucra la adición de nutrientes y oxígeno para ayudar a los microorganismos nativos. Estos nutrientes son los bloques básicos de construcción celular y permiten a los microorganismos producir las enzimas necesarias para destruir a los contaminantes.

### 5.5.2 Requisitos ambientales

Las condiciones ambientales óptimas para la degradación de los contaminantes se reportan en la tabla 5.2.

El crecimiento y la actividad microbianos son realmente afectados por el pH, la temperatura y la humedad. Aunque los microorganismos han sido también aislados en condiciones extremas, la mayor parte de ellos crecen óptimamente en un rango estrecho, así que es importante considerar las condiciones óptimas.

Ventajas	Desventajas
Relativamente fácil de diseñar e implementar	Reducción en la concentración >95% y

	concentraciones constituyentes <0.1 ppm son difíciles de alcanzar
Periodos cortos de tratamiento: usualmente de 6 meses a 2 años	Puede no ser efectiva para concentraciones >a 50,000 ppm
Costo competitivo \$30-\$90 por ton. de suelo contaminado	Presencia importante de metales pesados >2,500 ppm puede inhibir el crecimiento microbiano
Efectivo para la biodegradación de contaminantes orgánicos con baja biodegradabilidad	Los compuestos volátiles tienden a evaporarse más que a biodegradarse
Requiere menor área para su tratamiento	Requieren un área de tratamiento grande aunque menor a la labranza
Puede ser diseñada en un sistema cerrado, controlando la emisión de vapores	Los vapores generados pueden requerir un postratamiento previo a su descarga
Puede ser diseñado para ser efectivo combinando las condiciones del sitio y los productos del petróleo	Requieren una capa impermeable por si existen lixiviados

Tabla 5.1. Ventajas y desventajas de un sistema de biopilas

Por ejemplo, si el suelo es muy ácido es posible aumentar el pH adicionando cal. La temperatura afecta las velocidades de las reacciones químicas, por arriba de cierta temperatura, sin embargo, las células mueren. El agua disponible es esencial para todos los organismos vivos y se requiere irrigación para alcanzar el nivel óptimo de humedad.

La cantidad de oxígeno disponible determina si el sistema es aerobio o anaerobio. Los hidrocarburos son degradables bajo condiciones aerobias. Para aumentar el nivel de oxígeno en el suelo es posible arar o difundir aire, o bien, en algunos casos se puede introducir peróxido de hidrógeno o peróxido de magnesio al ambiente.

La textura del suelo afecta la permeabilidad, el contenido de humedad y la densidad del mismo. Los suelos con partículas finas son menos permeables que los suelos con partículas grandes. Los suelos con menor permeabilidad son más difíciles de airear pero tienden a retener mejor la humedad que los suelos con permeabilidad mayor. Sin embargo, permeabilidades menores se encuentran usualmente asociadas con suelos compactos, haciendo difícil la distribución de humedad, aire y nutrientes.

El suelo normalmente contiene un número diverso de microorganismos que incluyen bacterias, hongos, algas protozoarios y actinomicetos. En suelos con buen drenado, los cuales son adecuados para las biopilas, estos organismos son aerobios. De éstos, las bacterias son los más abundantes y bioquímicamente más activos. Las bacterias requieren una fuente de carbono para el crecimiento celular y una fuente de energía para sostener las funciones metabólicas requeridas para el crecimiento.

<b>Parámetros</b>	<b>Condición requerida para la actividad microbiana</b>	<b>Valor óptimo para la degradación de petróleo</b>
-------------------	---	---

Humedad del suelo	25-28% de la capacidad de retención de agua	30-90%
pH del suelo	5.5-8.8	6.5-8.0
Contenido de oxígeno	Aeróbico, espacio mínimo de aire 10%	10-40%
Contenido de nutrientes	N y P para crecimiento	C:N:P= 100:10:1
Temperatura (°C)	15-45	20-30
Concentración de contaminantes	No muy tóxico	Hidrocarburo 5-10% del peso seco del suelo
Metales pesados	Contenido total 2000 ppm	700 ppm
Tipo de suelo	Bajo contenido de arcilla y limo	

Tabla 5.2. Condiciones ambientales que afectan la degradación de contaminantes

La estructura del suelo controla la entrada de aire, agua y nutrientes. Para mejorar la estructura del suelo, se pueden utilizar materiales como gypsum o materia orgánica. La baja permeabilidad del suelo puede impedir el movimiento del agua, nutrientes y oxígeno, de tal forma que suelos con baja permeabilidad pueden no ser apropiados para técnicas de limpieza *in situ*.

## 5.6 Uso de materiales orgánicos en un sistema de biopilas

Como parte de los esfuerzos de bioremediación, se han desarrollado tecnologías que utilizan materiales orgánicos. Esta tecnología conocida con el nombre de composteo ha probado ser efectiva para la remediación de suelos contaminados con hidrocarburos aromáticos policíclicos (U.S. EPA, 1996) e hidrocarburos del petróleo, TPHs, especialmente diesel. A pesar de esto, la relación de suelo contaminado-residuos orgánicos debe ser determinada, debido a que una relación inapropiada puede retardar o inhibir la actividad microbiana.

La adición de materiales orgánicos no sólo cumple con la función de mejorar las características del suelo (estructura), sino que además constituye una parte de los nutrientes requeridos para la biodegradación; en general, son materiales de baja densidad que mejoran las propiedades físicas, químicas y biológicas del suelo. Estos materiales además, incrementan la porosidad y permiten una mejor difusión de oxígeno, lo cual favorece la actividad microbiana, estabilidad y estructura mediante la infiltración de agua y el intercambio gaseoso [30].

Los materiales orgánicos de origen biológico, donde el contenido de materia orgánica representa del 95% al 99%, están formados principalmente por los elementos básicos: carbono, hidrógeno, oxígeno y, en menor proporción, nitrógeno, fósforo y azufre.

Los materiales usados, provienen básicamente de actividades como la agricultura, beneficio de granos, ingenios, ganadería, mataderos, residuos forestales y domésticos.

En la tabla 5.3 se mencionan las principales mejoras que implican la adición de materiales orgánicos a una biopila de suelo contaminado.

<b>Propiedades físicas</b>	<ul style="list-style-type: none"> <li>• Formación de agregados y estabilidad estructural.</li> <li>• Se unen a las arcillas formando complejos de cambio, estabilizando la estructura del suelo y su cohesión.</li> <li>• Favorecen la penetración de agua y su retención.</li> <li>• Disminuyen la erosión.</li> <li>• Favorecen el intercambio gaseoso, la aireación y disponibilidad de oxígeno.</li> </ul>
<b>Propiedades químicas</b>	<ul style="list-style-type: none"> <li>• Incrementan la capacidad amortiguadora del suelo.</li> <li>• Mantienen y aumentan el contenido de materia orgánica y nutrientes.</li> <li>• Aumentan la capacidad de cambio y la reserva de nutrientes.</li> <li>• Promueven la formación de compuestos metálicos solubles en forma de materia orgánica-metal.</li> </ul>
<b>Propiedades biológicas</b>	<ul style="list-style-type: none"> <li>• Favorecen los procesos de mineralización.</li> <li>• Facilitan la reactividad y los mecanismos de absorción de sustancias peligrosas.</li> </ul>

Tabla 5.3. Mejoras de la adición de materiales orgánicos a un suelo contaminado.

## 5.7 Diseño de un sistema de biopilas

Estos sistemas consisten de un mecanismo de aireación para proveer oxígeno a los microorganismos, un mecanismo de inyección/irrigación para proveer de nutrientes y humedad después de la construcción de la biopila, y un medio para la colección de lixiviados a fin de controlar el exceso de humedad en la pila. Generalmente, se utiliza una capa impermeable, un soporte y una cubierta para proteger a las pilas de suelo de corrientes de aire y prevenir la dispersión del contaminante. Los tiempos de remediación dependen de la concentración de hidrocarburos, suelo, nutrientes, temperatura y poblaciones microbianas.

### **Establecimiento de un sistema de composteo por biopilas al nivel de campo.**

Para este trabajo se consideró un suelo arcilloso ligeramente alcalino (pH 7.9) proveniente de un sitio cercano a un pozo de producción abandonado. La mezcla de hidrocarburos encontrados en el suelo consistió de hidrocarburos alifáticos sustituidos, poliaromáticos sustituidos y asfáltenos. La relación C:N:P, basada en el contenido original del suelo como carbono, nitrógeno y fósforo correspondió a 9661:195:1. El nivel de contaminación por TPHs correspondió a una concentración de 99,929 +/- 23,046 mg/Kg. suelo.

El suelo contaminado con lodos de perforación fue excavado en los tres puntos con mayor contaminación localizados en el área de Paredón 31 en Huimanguillo, Tabasco. Aproximadamente 4 Toneladas de suelo fueron transportadas a un área cercana en donde se construyeron cuatro contenedores de cemento de 2 m ancho x 3 m largo y 20 cm de alto, los contenedores se encuentran separados 1 m aproximadamente (Fig. 5.1). En la parte frontal de cada contenedor se adaptó un sistema de tuberías de tipo PVC para recolectar los lixiviados generados durante el periodo de tratamiento. En la parte interior de los contenedores se colocó arena con un ángulo de inclinación de 5 grados. Este material permitió soportar el suelo contaminado y además darle inclinación al material y con esto dirigir los lixiviados al sistema recolector. Encima de la arena se colocó una geomembrana impermeable de PVC para soportar el suelo y evitar la migración del material hacia la arena además de evitar la pérdida de humedad. Para evitar la inundación de los sistemas y las condiciones ambientales (lluvias, exceso de sol, vientos) y permitir un mejor control de los sistemas de tratamiento, se construyó un enrejado con techo de lámina sobre la superficie asfaltada (Fig. 5.2).

Una vez que el material contaminado se colocó sobre el plástico impermeable en el área asfaltada, fue mezclado y dividido en 4 montículos de suelo de 1 tonelada cada uno; los cuatro montículos fueron mezclados utilizando palas. Para evitar problemas mecánicos de mezclado, se eliminaron piedras y otros materiales (material de rechazo).

Para su tratamiento por composteo, tres montículos de suelo se prepararon de la siguiente forma (Fig. 5.1): a 1 Ton de suelo se le adicionó paja de alfalfa en una relación suelo/paja de 98/2. Además, se adicionó urea y fosfato monobásico de potasio como fuente nitrógeno y fósforo, los cuales fueron adicionados, en forma sólida, hasta alcanzar una relación C/N/P de 100/3/0.5. Estos materiales fueron mezclados mecánicamente utilizando palas para, posteriormente, ser colocados sobre la geomembrana en los contenedores de cemento. La humedad de los sistemas se ajustó a una humedad entre 25%-30%.

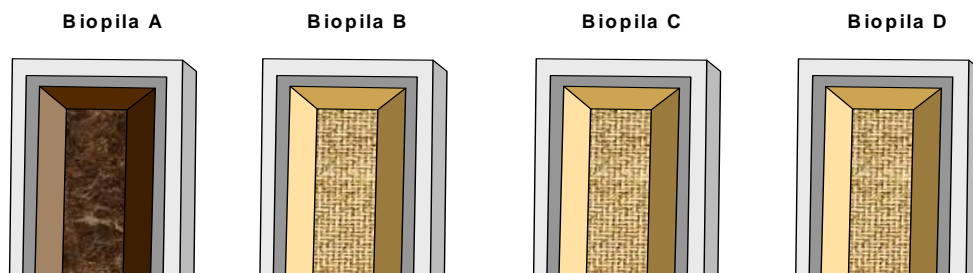




Fig. 5.1 Vista superior de las cuatro biopilas.

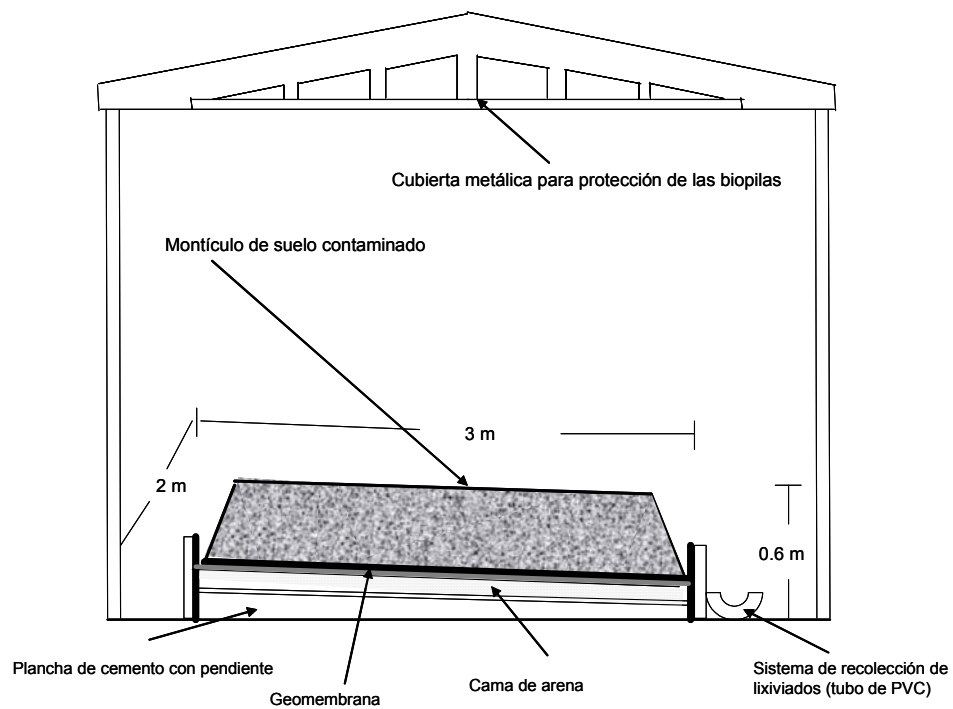


Fig. 5.2 Establecimiento del sistema de biopilas.

Una vez colocados los cuatro montículos de tierra sobre las geomembranas, se procedió a cubrirlos con la misma geomembrana para ser descubiertos durante su mezclado o bien durante la toma de muestras (Fig. 5.3).

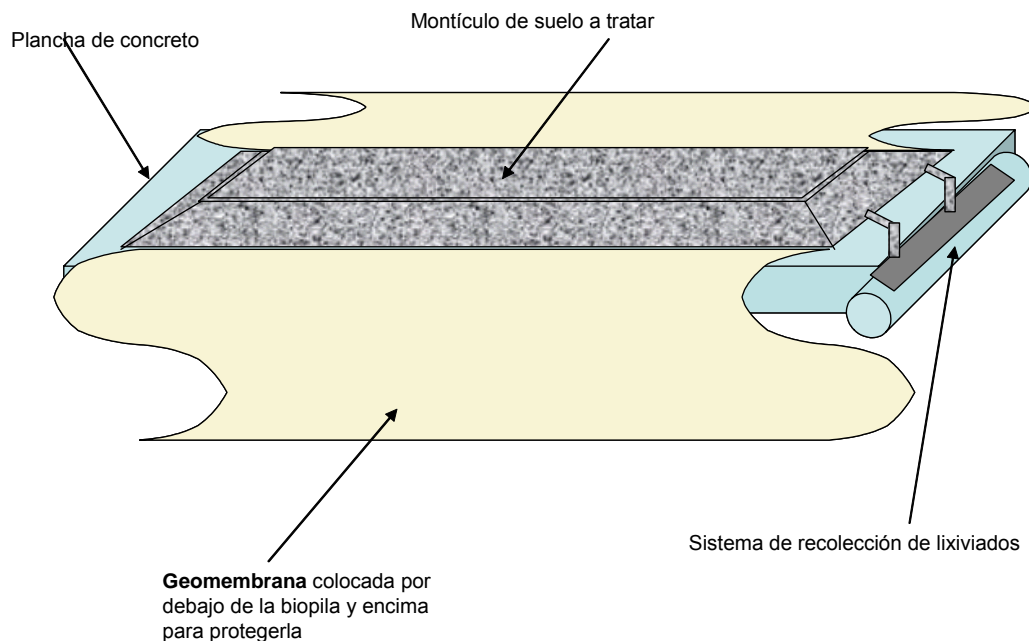


Fig. 5.3 Vista a detalle de una biopila con su geomembrana y su sistema recolector de lixiviados.

### 5.7.1 Mezclado y aireación de las pilas

Las pilas fueron aireadas y mezcladas mecánicamente utilizando palas. Durante los 6 meses que duró el tratamiento, las pilas fueron mezcladas y/o aireadas cada cuarto día, además, cada 24 h se realizó la medición de humedad y temperatura siendo éstos, dos parámetros importantes durante el proceso de composteo. El tratamiento inició en abril y finalizó en septiembre. Debido a la pérdida de agua durante el tratamiento, al mismo tiempo del mezclado se adicionó agua con un rociador manual para evitar la sequedad de las biopilas. Después del mezclado y ajuste de humedad, las pilas fueron cubiertas nuevamente.

### 5.7.2 Muestreo

Cada 30-35 días se tomaron 30 muestras de cada una de las pilas para evaluar los parámetros físico-químicos (pH, humedad, carbono orgánico total, fósforo, nitrógeno e hidrocarburos totales), y los microbiológicos (bacterias heterotrofas, bacterias hidrocarbonoclastas y hongos totales), que dieron la información sobre la eficiencia del tratamiento.

### 5.7.3 Adición de nutrientes y residuo agrícola

Durante el periodo de tratamiento (6 meses) y, con la finalidad de mantener la relación C/N/P a un valor de 100/3/0.5, fue necesario adicionar las fuentes de nitrógeno y fósforo en dos ocasiones a los 30 y 90 días del tratamiento.

En el caso del residuo agrícola (paja), éste se adicionó en dos ocasiones, para evitar la compactación del suelo, a los 90 y 150 días de tratamiento. La cantidad de material adicionado correspondió a 14 kg y 4.5 kg, respectivamente.

La Fig. 5.4 muestra el diagrama completo del establecimiento del sistema de biopilas a nivel de campo.

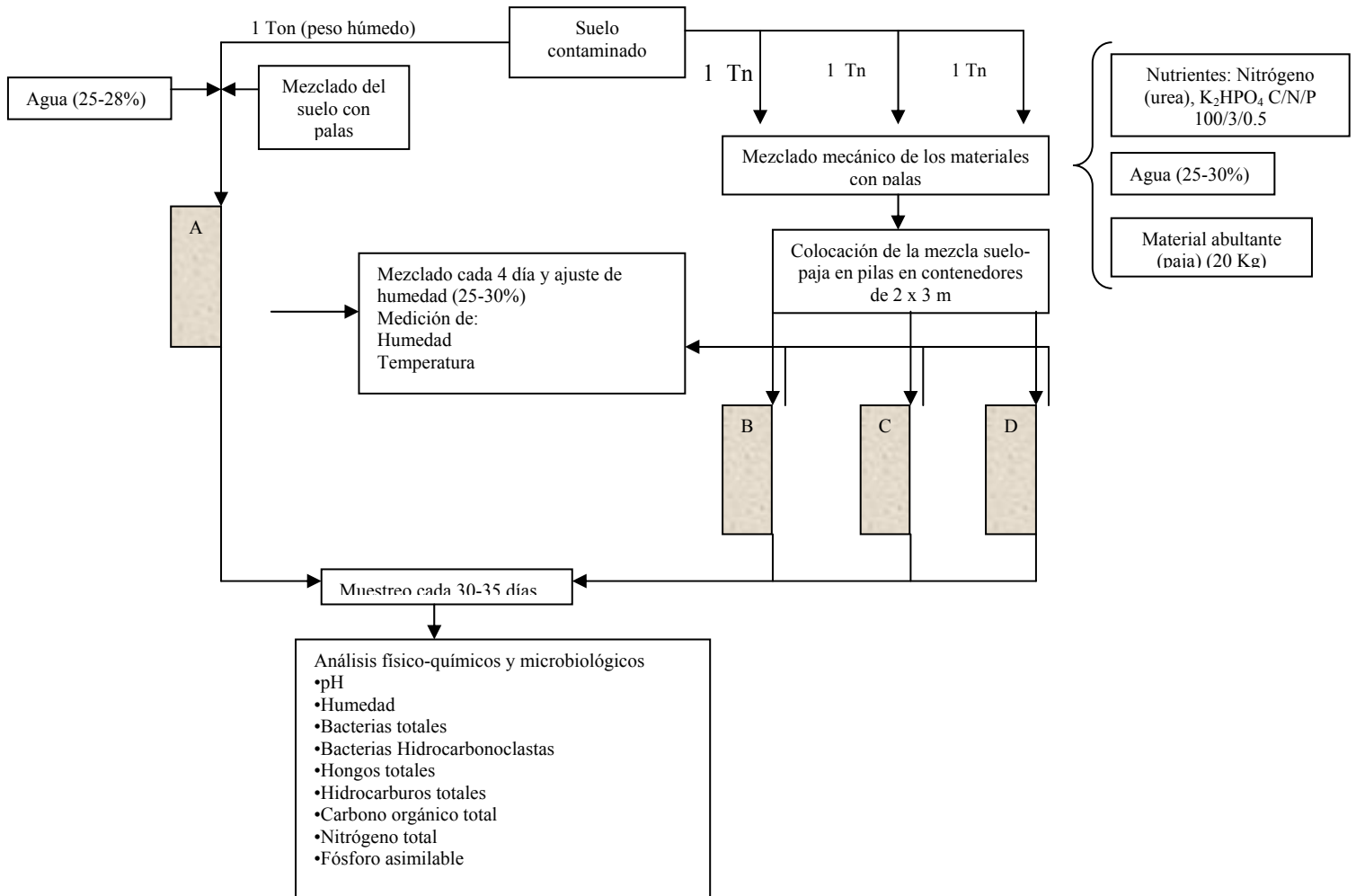


Fig. 5.4 Diagrama que muestra la metodología seguida para el establecimiento de un sistema de biopilas a nivel campo

## 5.8 Conclusiones

En este capítulo se mencionan a *grosso modo* las características de funcionamiento que se presentan en los sistemas biológicos y se explican las razones de la dificultad para obtener

sus modelos, destacando la incapacidad para monitorear la actividad molecular de los microorganismos de una manera sencilla y barata.

En este capítulo se describe el bioproceso construido para el saneamiento de un suelo que ha sido contaminado con Hidrocarburos. Las características mencionadas describen un proceso altamente complejo con altas interacciones de sus variables. Se justifica también que a la fecha la construcción y operación de este tipo de procesos sigue una especie de receta, establecida a través de la experiencia adquirida. Por lo que es de notarse el alto grado de conocimiento heurístico que incluye dicha descripción. Esto incrementa la conveniencia de tener un modelo matemático enmarcado en la teoría de los conjuntos difusos y de la lógica difusa.

Así, la metodología propuesta en los primeros cuatro capítulos se aplicará en el modelado de este proceso de bioremediación.

## CAPITULO 6

### MODELADO DIFUSO DEL BIOPROCESO

#### 6.1 Introducción

De la descripción del bioproceso realizada en el capítulo anterior, se deduce que el sistema es altamente complejo, y que su desempeño depende de una serie de condiciones no necesariamente bien definidas. Para establecer las condiciones adecuadas para una buena bioremediación de los contaminantes, se sigue una especie de *receta* basada en la experiencia, es decir, de los experimentos realizados previamente, se determinaron los rangos más adecuados de cada parámetro para establecer las mejores condiciones de su funcionamiento. Estas condiciones establecen un buen reto para resolver exitosamente la obtención del modelo que permita caracterizar el sistema de manera global. Para resolver este problema, planteamos una estrategia de tipo jerárquica que consiste en *aproximar por partes* el conjunto de variables significativas del sistema. Se considera este esquema para plasmar, de manera fehaciente, la estrategia que siguen los expertos en la resolución de este problema. Este esquema puede incluir las fases de arranque y/o paro del sistema, pues su caracterización se incluye en un supervisor difuso.

Dado que esta tesis se enfoca en el aspecto operativo del proceso, consideramos que el modelo obtenido se puede usar no sólo para obtener información del proceso, sino que también puede hacerlo más eficiente. Además, la metodología propuesta permite que los modelos obtenidos puedan usarse para procesos similares, logrando así acelerar su desarrollo.

En las disciplinas de la ingeniería eléctrica, mecánica y química, la optimización y el control se pueden desarrollar basándose en un modelo del sistema, pero en el caso de los bioprocursos esto no es así. Entre las principales razones está la dificultad para desarrollar un modelo preciso del sistema debido a la complejidad de los procesos biológicos subyacentes y a la carencia de sensores confiables, precisos y baratos.

En lo particular, para el desarrollo del modelo sólo disponemos de un pequeño número de datos experimentales y el conocimiento del proceso.

De manera genérica podemos establecer que la ruta para el desarrollo de un bioproceso industrial se inicia con el planteamiento de un objetivo específico para realizar una cierta actividad bajo un conjunto de especificaciones dadas. El siguiente paso es reducir los costos. En ingeniería de bioprocesos, existen dos métodos para minimizar estos costos de producción: Uno, es encontrar una ruta de control o una estrategia de operación que utilice las dinámicas del proceso de manera óptima en tanto resuelve las restricciones de producción, el otro método consiste en eliminar o disminuir las restricciones mediante mejoras constructivas en el equipo y/o en los microorganismos.

En nuestro caso, nos enfocamos en el aspecto de encontrar la mejor estrategia de operación dados los resultados experimentales, lo que se busca es que el modelo difuso facilite el estudio de las interacciones de los diversos parámetros del proceso con el fin de contar con una herramienta que permita posteriormente diseñar y controlar procesos similares.

## **6.2 Descripción del proceso de modelado propuesto para la biopila estudiada.**

Nuestro ejemplo es un proceso biológico y como tal, lo que ocurre en él resulta complejo de describir matemáticamente; así, nuestro intento por formalizarlo debe ser una tarea que se debe cumplir acorde con las herramientas matemáticas disponibles. En una breve revisión bibliográfica es posible encontrar trabajos de modelado y control de bioprocesos, los cuales presentan situaciones perfectamente caracterizables (debido a la experiencia adquirida a través del tiempo) pues en general, los experimentos se acotan a un sustrato conocido, con condiciones bien caracterizadas y que en general son sistemas del tipo *batch* o *feed batch*, que se desarrollan en un bioreactor, en el cual las condiciones de funcionamiento se pueden establecer con mayor certeza [31], [32].

En nuestro ejemplo, el suelo utilizado para crear el sistema de biopilas contiene, de manera natural, a los microorganismos que utilizan a los hidrocarburos como alimento, esto es gracias a que estos organismos han aprendido a sobrevivir en presencia del contaminante por lo que han adquirido la capacidad para procesarlo. Desafortunadamente, no es posible caracterizar ni cuantificar con certeza qué tipos de microorganismos están presentes, sólo se les cuantifica de manera global y se les denomina como bacterias hidrocarbonoclastas (capaces de procesar las moléculas de los hidrocarburos). El bioproceso que nos ocupa, es una forma de acelerar la biodegradación de los contaminantes mediante la búsqueda de las condiciones óptimas en la biopila.

Considerando jerárquicamente el nivel de detalle respecto de la caracterización del proceso, establecemos que en los niveles superiores se encuentran los análisis a *grosso modo* y en los niveles inferiores los análisis más detallados (en un proceso biológico, a este nivel correspondería una caracterización a escala molecular). En estas condiciones podemos establecer que el trabajo de modelado se puede situar en los niveles superiores, debido a que la información que se requiere para tener una caracterización en los niveles inferiores debe ser más amplia y precisa.

De esta premisa, podemos decir que la información que se tiene es escasa y poco precisa pues las mediciones se realizan manualmente con equipos portátiles y en condiciones reales. Gran parte de la información que se tiene para caracterizar el proceso es la que se obtuvo fuera de línea, en el laboratorio y después de dos días de tomada la muestra. Esto es así, pues el tipo de análisis no se puede realizar en línea, además de que algunas pruebas requieren de tiempo de incubación del orden de días inclusive (para determinar la UFC, por ejemplo).

Esta es la principal justificación del uso de la Lógica Difusa en la caracterización de este tipo de sistemas, pues permite obtener un marco teórico matemático que representa al sistema aún en condiciones de información escasa, incierta, ambigua e incompleta, con el fin de reproducir y mejorar la eficiencia del proceso. Una gran cualidad del marco teórico difuso, es que el modelo obtenido puede ser validado por el experto y en caso de requerirse, el modelo puede ser complementado para caracterizar aquellas zonas del sistema que no estuvieron contenidas en los datos, pero que el experto sabe que son situaciones que se presentan ocasionalmente en el sistema.

A continuación, se mencionan las propiedades más importantes que se utilizan para caracterizar el proceso:

- el suministro de oxígeno,
- la disponibilidad de nutrientes,
- el contenido de humedad,
- el pH,
- la salinidad,
- la estructura del suelo y el contenido orgánico,
- la temperatura,
- la solubilidad del contaminante,
- la concentración de compuestos tóxicos y
- la presencia de microorganismos degradadores del contaminante.

En este trabajo sin embargo, no se incluyen todas éstas pues algunas de ellas no se tienen registradas como variables medidas durante la operación. Esto se debe a varias razones, entre otras, se pueden mencionar el tipo de experimento realizado, el cual fue del tipo exploratorio y no permanente, a la falta de sensores o de equipo, pero principalmente, a la falta de un modelo que estableciera que variables se deben monitorear a fin de supervisar el funcionamiento del bioproceso.

Considerando la dificultad para caracterizar con detalle todos los fenómenos biológicos que se llevan a cabo en el interior de la biopila (principalmente a escala molecular) y mediante los cuales los microorganismos indígenas transforman a los hidrocarburos presentes en el suelo en materia inerte, se consideró conveniente representar el bioproceso como una caja con las siguientes propiedades consideradas como entradas y salidas, como se muestra en la figura 6.1.

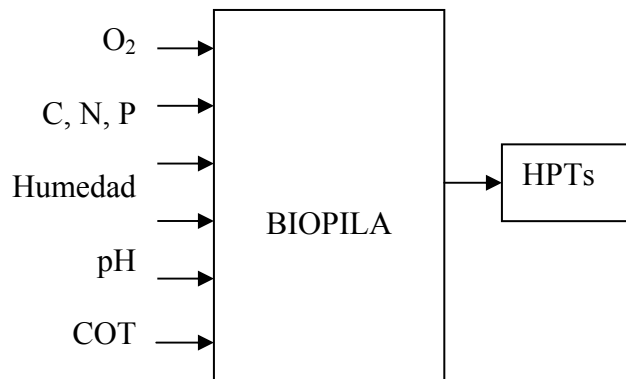


Figura 6.1. Esquema de la Biopila que muestra las variables de entrada y de salida que se utilizan para caracterizar el Bioproceso

Es de hacer notar que se restringe el análisis y modelado del bioproceso a una búsqueda del modelo con los datos obtenidos principalmente en laboratorio, pues aún cuando los datos medidos en campo son útiles, la mayor cantidad de información es aquella obtenida en el laboratorio.

Nuestra propuesta para modelar este sistema es hacer uso de la metodología creada y aplicarla al conjunto de variables que caracterizan el bioproceso. La metodología creada servirá para poder identificar las relaciones entre los nutrientes, los contaminantes y los microorganismos. Se demostrará que los resultados obtenidos son viables de ser implementados en un programa de software de computadora pues contiene un nivel de generalización adecuado para su uso en tiempo real.

Para iniciar el trabajo de análisis nos apoyamos en el conocimiento del experto, el cual establece que las variables que más influyen en la biodegradación son el pH y la temperatura. El análisis se realiza siguiendo el diagrama de flujo mostrado en la figura 6.2.

Un beneficio importante de poder modelar este tipo de bioprocesos es eliminar el contacto directo entre el operador y la biopila, evitando así la posibilidad de contaminación por materiales peligrosos.



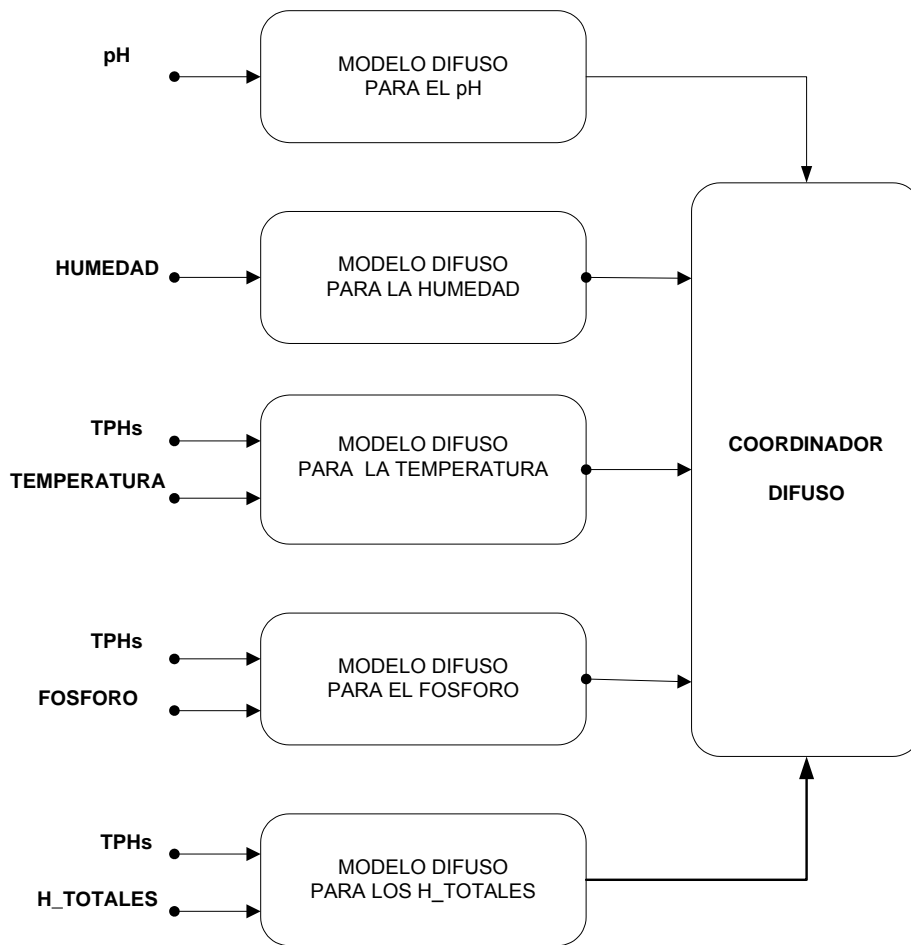


Figura 6.2. Diagrama de flujo con la propuesta para el análisis de la Biopila

Además de usar nuestro marco teórico desarrollado, algunas de las variables se modelan mediante el método directo, es decir, cuando se tiene la información suficiente para caracterizar el comportamiento de un parámetro, éste se puede describir usando una plantilla, donde los conjuntos difusos y las reglas se pueden establecer de manera *a priori* con la posibilidad de crear sistemas difusos tanto del tipo Mamdani así como TSK.

### 6.3 Análisis del pH.

Caracterizamos el rango de valores del parámetro pH del suelo mediante el método directo. De los antecedentes en la literatura, se sabe que un valor adecuado para el pH del suelo está en el intervalo de  $6 < \text{pH} < 8$ . Sin embargo, como valor óptimo se considera un valor neutral cercano a 7, ni ácido ni alcalino, así en nuestra caracterización aproximaremos este valor con el conjunto difuso *alrededor de 7*. De lo anterior, se plantea caracterizar con los conjuntos difusos definidos en la Figura 6.3, el rango de valores para el pH del suelo.

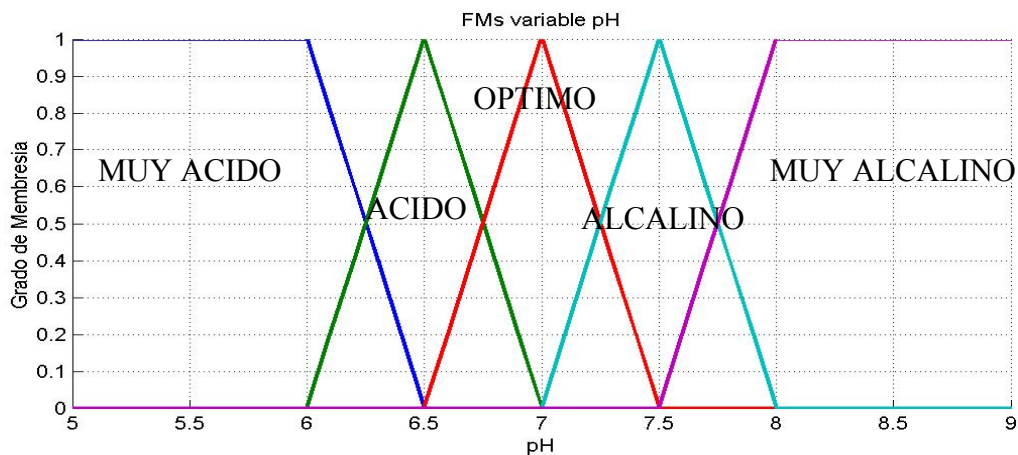


Figura 6.3. Funciones de Membresía utilizadas para caracterizar el Universo de Discurso de los valores del pH.

Los valores del pH del suelo se caracterizan mediante los 5 conjuntos difusos, uniformes y distribuidos equitativamente en el universo de discurso. Para un buen crecimiento de los microorganismos, el pH del suelo debe estar en el rango de 6 a 8; siendo el óptimo un valor cercano a 7 (neutro) por lo que suelos con un valor de pH fuera de este rango, deberán ser ajustados. El pH del suelo se puede incrementar agregándole cal y se puede disminuir si se agrega sulfuro elemental. Se utilizan cinco conjuntos difusos, pues dado que este parámetro cambia muy lentamente, las acciones correctivas también son tardadas. El objetivo del sistema es mantener un valor óptimo para el pH del suelo (alrededor de 7), pues aún cuando un valor de pH entre 6 y 8 se considera como adecuado para el crecimiento microbiano, el sistema difuso buscará alcanzar el valor óptimo para el proceso de bioremediación.

Se crea un sistema difuso estático con dos variables de entrada, *pH* y *cambio pH*, la variable de salida se denomina *cal - sulfuro*, en ésta, se tiene en un extremo el componente cal y en el otro el componente sulfuro, pues para hacer la corrección del pH sólo se agrega un componente a la vez y, cuando el valor del pH se encuentra en el conjunto óptimo, no se requiere de ninguna corrección.

El sistema difuso queda expresado por el siguiente conjunto de reglas creadas para regular el pH según el experto:

Las reglas que contienen la parte proporcional:

Si *ph* es MUY ACIDO entonces *cal sulfuro* es CAL ALTA;  
Si *ph* es ACIDO entonces *cal sulfuro* es CAL;  
Si *ph* es ALCALINO entonces *cal sulfuro* es SULFURO;  
Si *ph* es MUY ALCALINO entonces *cal sulfuro* es SULFURO ALTO;

Las reglas con la parte derivativa:

Si *cambio ph* es MUY ACIDO entonces *cal sulfuro* es CAL ALTA;  
Si *cambio ph* es ACIDO entonces *cal sulfuro* es CAL;  
Si *cambio ph* es ALCALINO entonces *cal sulfuro* es SULFURO;  
Si *cambio ph* es MUY ALCALINO entonces *cal sulfuro* es SULFURO ALTO;

Las reglas que contienen el estado estable

Si *ph* es OPTIMO entonces *cal sulfuro* es CERO;  
Si *cambio ph* es SIN CAMBIO entonces *cal sulfuro* es CERO;

La respuesta del modelo difuso se muestra en la Figura 6.4. En ésta, se puede apreciar la suavidad con que el sistema modela los cambios en los parámetros de entrada.

Los cuatro vértices de la gráfica muestran un comportamiento concreto del estado actual del sistema y de la acción correspondiente para este estado, pues representan valores extremos tanto del valor del pH como de la razón de cambio. En la parte central de la gráfica se observa un comportamiento suave pues recordemos que los valores del pH se deben mantener en el intervalo de 6 a 8. Este comportamiento *suave* es característico de los sistemas difusos.

El sistema difuso empleado en el modelado del pH es es del tipo Mamdani y se evalúa usando los operadores difusos min-max y el método de inferencia difuso de Mamdani.

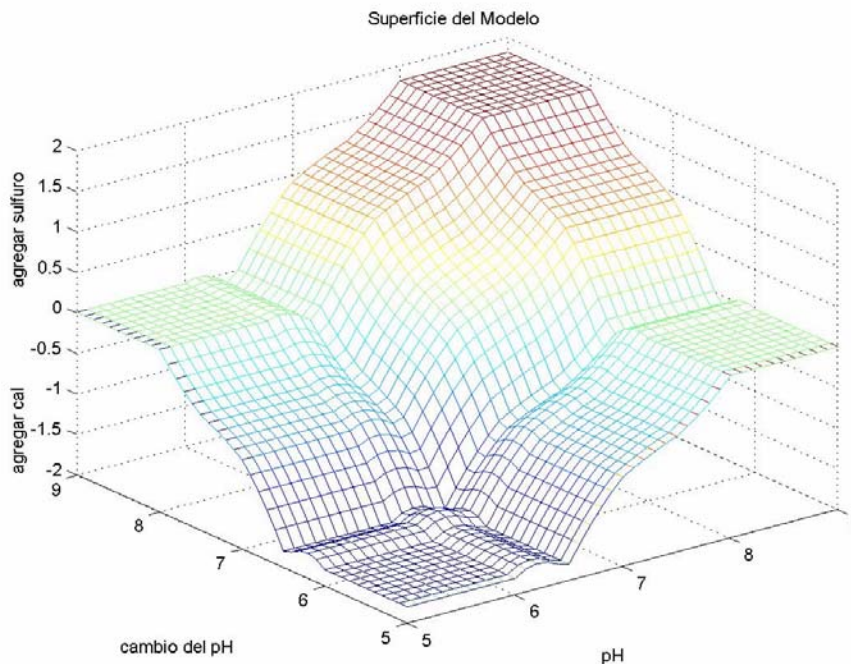


Figura 6.4. Superficie de respuesta para las variables de entrada pH y cambio pH

## 6.4 Humedad.

Para un crecimiento óptimo de los microorganismos contenidos en el suelo, se requieren condiciones de humedad del suelo de alrededor del 30%. Este valor es el que se mantuvo durante todo el experimento aunque al principio, y debido a las condiciones climatológicas del lugar, dicho valor fue mucho menor. El valor de 30% representa el equilibrio entre un suelo seco y una humedad excesiva en el suelo, pues ambas condiciones representan condiciones adversas para el crecimiento de los microorganismos, ya que un suelo muy húmedo puede restringir una correcta aireación de la biopila restringiendo con ello la disponibilidad de oxígeno, el cual es esencial para los procesos metabólicos de los microorganismos aeróbicos.

En general, el suelo debe estar húmedo pero no mojado. El rango ideal para la humedad del suelo está entre el 15 y el 30% del peso. Para mantener este porcentaje de humedad en el suelo, se decidió agregar agua en forma periódica para que el suelo mantuviera un porcentaje de humedad cercano al 30%, pues dadas las condiciones naturales del lugar donde se construyó la biopila, el suelo se seca rápidamente como resultado de la evaporación natural además de la evaporación que se produce durante el proceso de aireación (acción manual).

La irrigación de la biopila es una actividad no regulada para seguir un comportamiento predeterminado, pues dadas las características del sitio, se optó por hacerlo de manera periódica cada 48 horas, es decir, el personal a cargo irrigaba el bioproceso de manera manual. Esta acción se realiza hasta asegurar un valor de humedad de aproximadamente el 30%.

Aquí hay que mencionar que estas acciones se regularon mediante la percepción visual del operador encargado, es decir, no se disponía de equipo para hacer estas actividades de manera precisa y automática, por el contrario, fueron acciones reguladas bajo un criterio completamente heurístico, que los operadores fueron perfeccionando cada vez más en cuanto el proceso avanzaba.

De los datos de la humedad registrados en la biopila y del conocimiento adquirido durante el proceso, se plantea caracterizar esta variable siguiendo el método de modelado directo. Para caracterizar el rango de valores de esta variable, se crean los conjuntos difusos que se muestran en la figura 6.5. Como ya se argumentó, se considera como óptimo el valor del 30% de humedad así, valores inferiores o mayores se requieren ajustar hacia este valor óptimo.

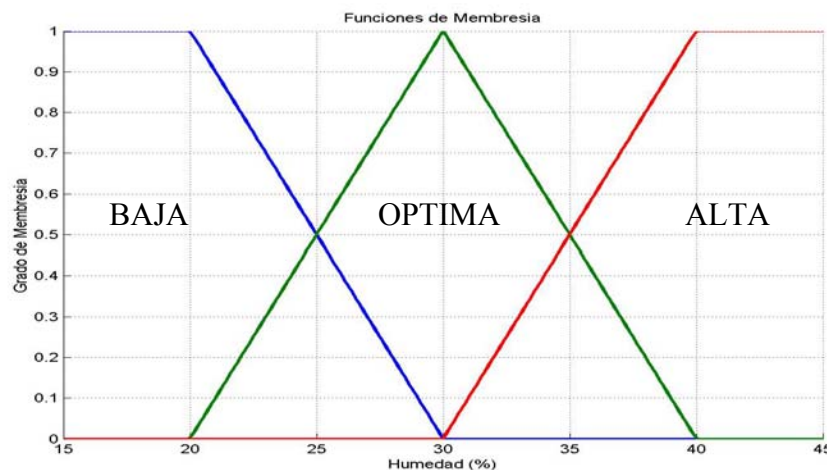


Figura 6.5 Caracterización del Universo de discurso para la variable Humedad mediante Conjuntos Difusos

Siguiendo el esquema para este tipo de modelos y usando una caracterización similar para la variable *cambio\_humedad*, la cual caracteriza el cambio de valores en la humedad, se crea el sistema difuso tipo Mamdani. Para la variable de salida *Irigar* se crean los conjuntos difusos que se muestran en la figura 6.6. Estos conjuntos, se distribuyen equitativamente en el universo de discurso y las reglas se crean para plasmar lo más fehacientemente el conocimiento del experto.

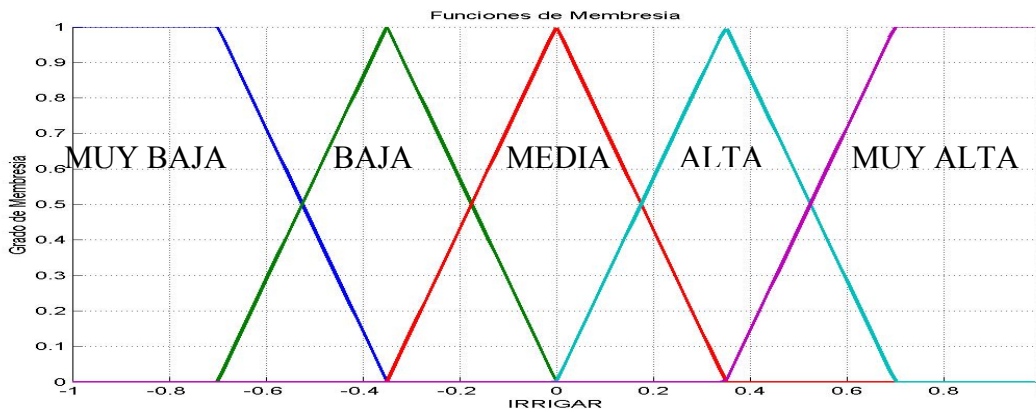


Figura 6.6 Caracterización del Universo de discurso para la variable Humedad mediante Conjuntos Difusos

Las reglas estipuladas son:

Si humedad es BAJA entonces irrigar es ALTA;  
 Si humedad es OPTIMA entonces irrigar es MEDIA;  
 Si humedad es ALTA entonces irrigar es BAJA;

Si cambio\_humedad es LENTO entonces irrigar es MUY ALTA;  
 Si cambio\_humedad es MEDIO entonces irrigar es MEDIA;  
 Si cambio\_humedad es RAPIDO entonces irrigar es MUY BAJA

El análisis de este sistema se puede realizar mediante la superficie de respuesta que se muestra en la figura 6.7.

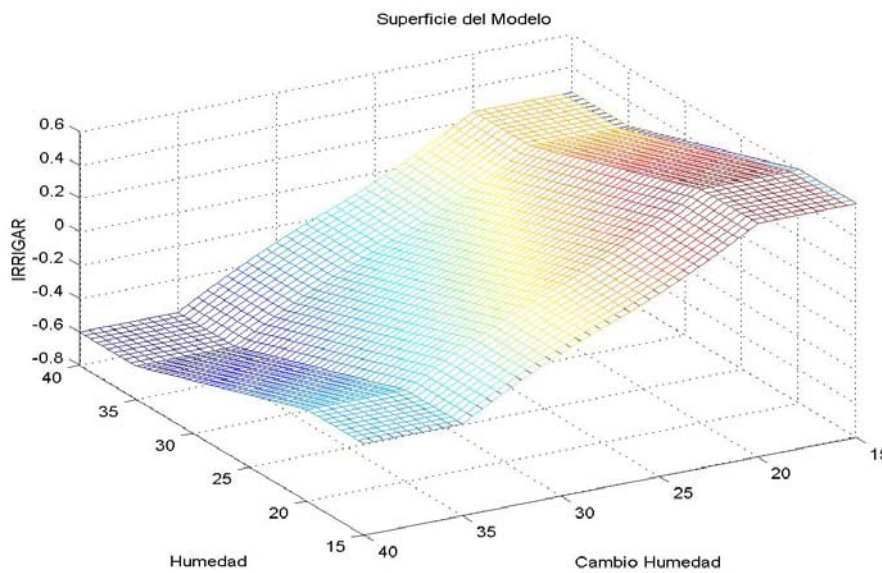


Figura 6.7 Superficie de respuesta del Sistema Difuso creado para caracterizar la variable Humedad

## 6.5 Oxigenación

En la Biopila, la actividad microbiológica aeróbica se estimula mediante el suministro de aire (aireación) y/o, añadiendo minerales y nutrientes. Las Biopilas son sistemas que usan el oxígeno, generalmente del aire, para estimular el crecimiento y reproducción de las bacterias aeróbicas encargadas de degradar los hidrocarburos del petróleo presentes en el suelo.

La acción de airear el suelo se efectúa de manera manual, moviendo mecánicamente mediante palas el suelo de la biopila. Esto permite asegurar el suministro de oxígeno a todo el suelo. Nuevamente, se debe mencionar que esta acción se regula bajo un criterio completamente heurístico, que los operadores fueron perfeccionando cada vez más en cuanto el proceso avanzaba.

## 6.6 Temperatura de la Biopila

Como se mencionó, este parámetro es muy importante para el proceso de bioremediación, pues el crecimiento de la población microbiana es una función de la temperatura. La actividad microbiológica de la mayoría de los microorganismos degradadores de hidrocarburos desaparece a temperaturas menores a 5°C y mayores a 45°C. Por el contrario, en el intervalo de los 10°C hasta los 45°C, esta actividad se duplica con cada 10° de incremento en la temperatura.

La Figura 6.8, muestra el registro del comportamiento de la temperatura de la biopila durante el proceso de bioremediación. También, se debe considerar la temperatura promedio del lugar donde se construyó la biopila (Tabasco tiene un clima cálido húmedo y una temperatura promedio de 26°C). De esta manera se puede prever el rango de temperaturas que se puede presentar en la biopila. En caso de que se requiera regular la temperatura hacia un mejor valor, esta se puede regular si la biopila se confina en una estructura del tipo invernadero o bien calentando el aire que se aplica.

En la gráfica de la figura 6.8, las altas temperaturas iniciales se pueden atribuir a una gran actividad microbiana durante la metabolización de los contaminantes, pues al inicio del proceso la concentración de hidrocarburos del petróleo totales (TPHs) en el suelo corresponde a 99,000 mg TPHs/kg., como se aprecia en la gráfica de la figura 6.9, donde se aprecia también la disminución del contenido de TPHs al final del proceso, correspondiente a una disminución de casi el 95%.

(°C)

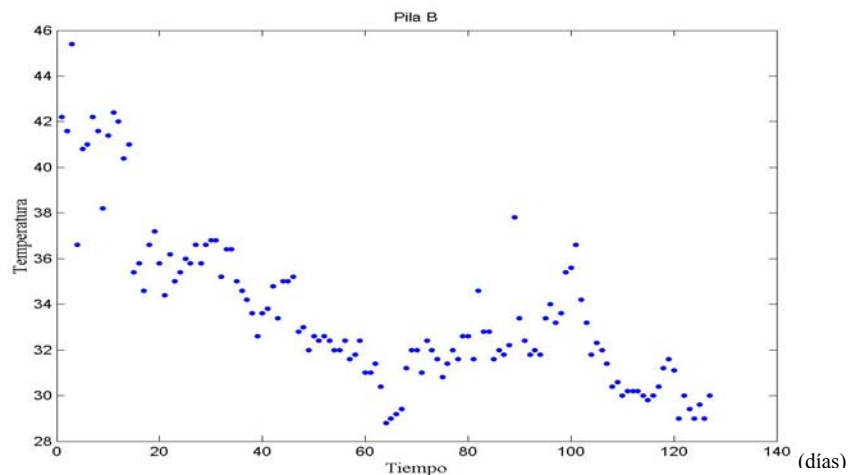


Figura 6.8. Gráfica de los datos de la temperatura de la Biopila

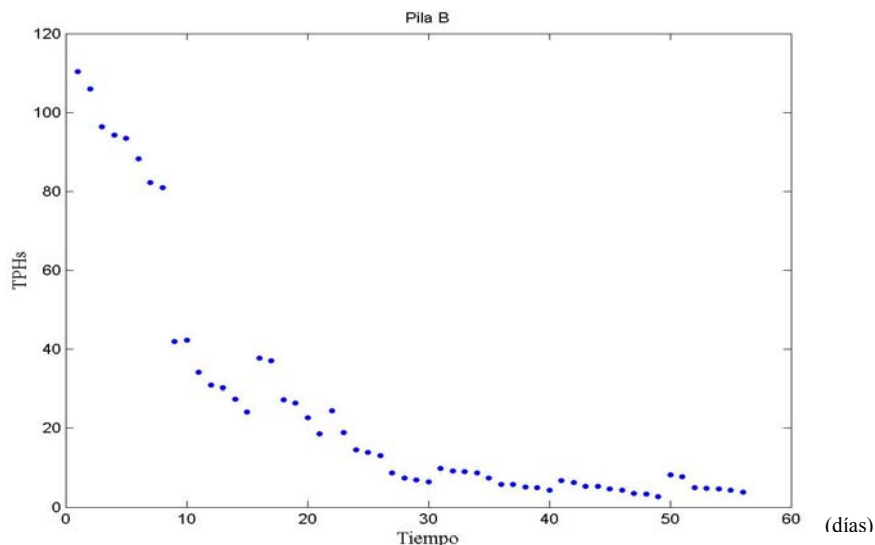


Figura 6.9. Gráfica de los datos de los TPHs de la Biopila

Considerando que la temperatura es uno de los factores más importantes en el proceso de biodegradación, se usa la metodología creada para obtener las relaciones entre la temperatura de la biopila y la cantidad de hidrocarburos totales del petróleo. En la figura 6.10 se pueden ver los datos de ambas variables, los centros obtenidos mediante el proceso de clustering así como las curvas de nivel, las cuales muestran la orientación de cada cluster. Esta orientación muestra la propiedad del algoritmo de clustering Gustafson-Kessel empleado para realizar el proceso de clustering, el cual adapta cada cluster a la forma que siguen los datos a analizar. Aquí, lo importante para aplicar esta metodología es determinar el número adecuado de clusters para hacer la partición, pues el modelo se obtiene a partir



de la información proporcionada por los clusters. Así, si el número de clusters no es el adecuado, el modelo final no podrá reproducir con el mínimo error el comportamiento del sistema.

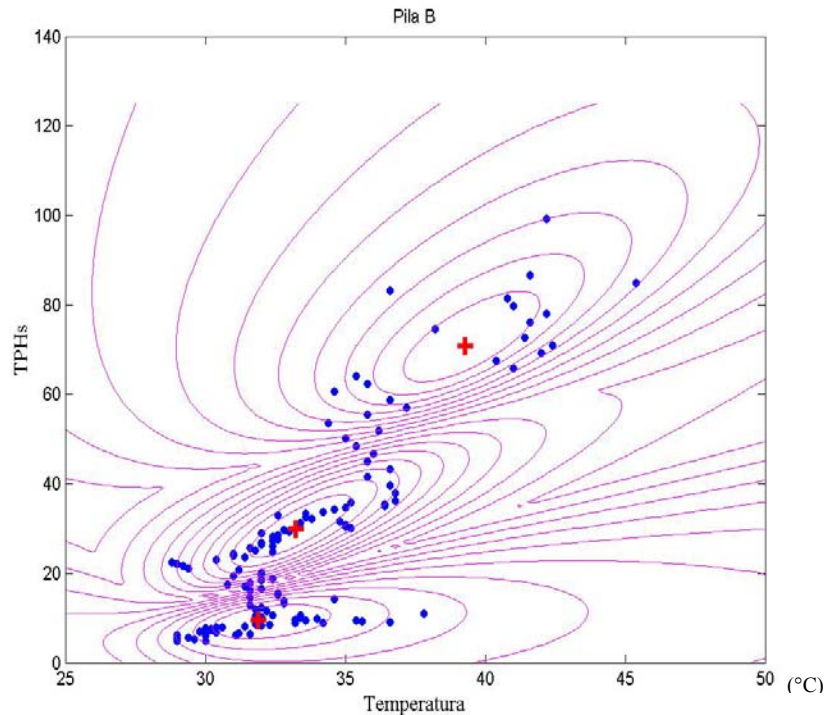


Figura 6.10. Gráfica de los datos de Temperatura Vs los valores para el TPHs. Los centros de cada cluster están indicados con el signo + y se incluyen las curvas de nivel para cada cluster.

Con la información obtenida para cada cluster, se procede a calcular las funciones de membresía para cada variable mediante la operación de proyección. En este caso sobre el eje de la variable temperatura, pues ésta es la variable definida como entrada al sistema difuso. Las funciones de membresía se muestran en la figura 6.11.

Dado que el sistema difuso que se construirá es del tipo Takagi-Sugeno, el conjunto de reglas es el siguiente:

- Si temperatura es BAJA entonces  $TPHs = (0.52)BAJA - 9.35$
- Si temperatura es MEDIA entonces  $TPHs = (2.33)MEDIA - 14.46$
- Si temperatura es ALTA entonces  $TPHs = (2.23)ALTA - 44.97$

Los consecuentes se determinan por el algoritmo de mínimos cuadrados. La salida del sistema se calcula utilizando el método de defusificación de la media ponderada. En la figura 6.11, se muestra el resultado global de la evaluación del sistema difuso cuyo error mínimo alcanzado es de 0.6464.

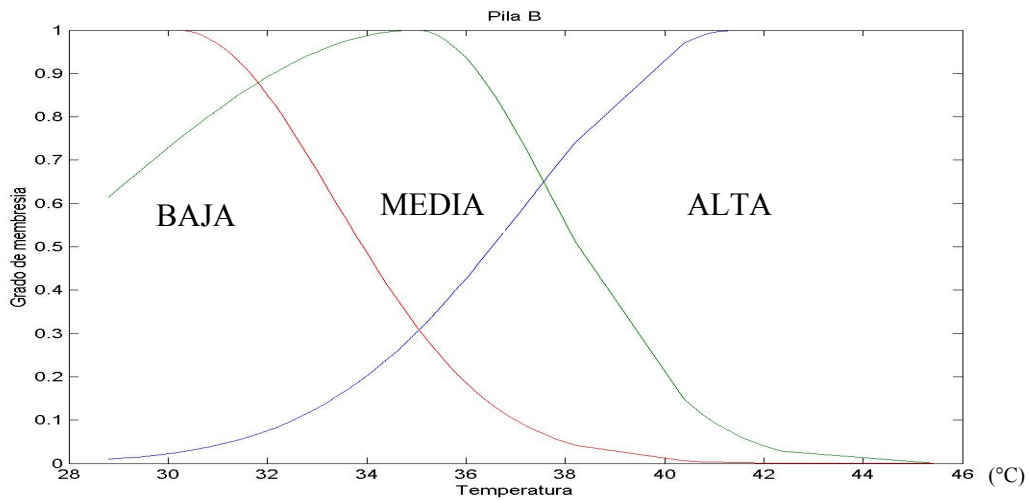


Figura 6.11. Funciones de Membresía obtenidas mediante la proyección de los clusters sobre el eje de la variable Temperatura.

En la gráfica de la salida global del sistema difuso, figura 6.12, se puede apreciar la ventaja de particionar el modelo total, pues aún cuando los datos no siguen un patrón regular además de estar demasiado dispersos, el sistema difuso es capaz de proveer una salida para cada condición de entrada que es muy aproximada a la salida real.

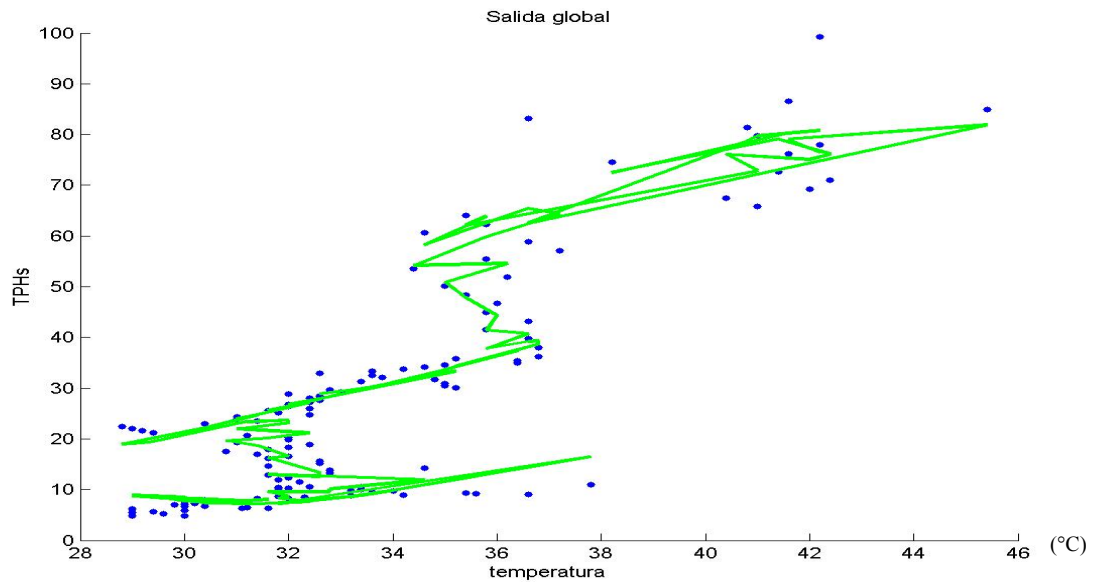


Figura 6.12. La línea continua es la salida del sistema difuso del tipo T-S cuyo error global es de 0.6464.

Debemos recordar que esta caracterización de la biopila es mediante un modelo estático; éste es factible de utilizar pues el sistema es lento y los análisis se pueden programar para ser realizados en periodos de 12 o 24 horas, según se requiera.

Dado que cada cluster es una partición del modelo, la respuesta de cada regla es una línea recta, es decir, un modelo local (ver la figura 6.13).

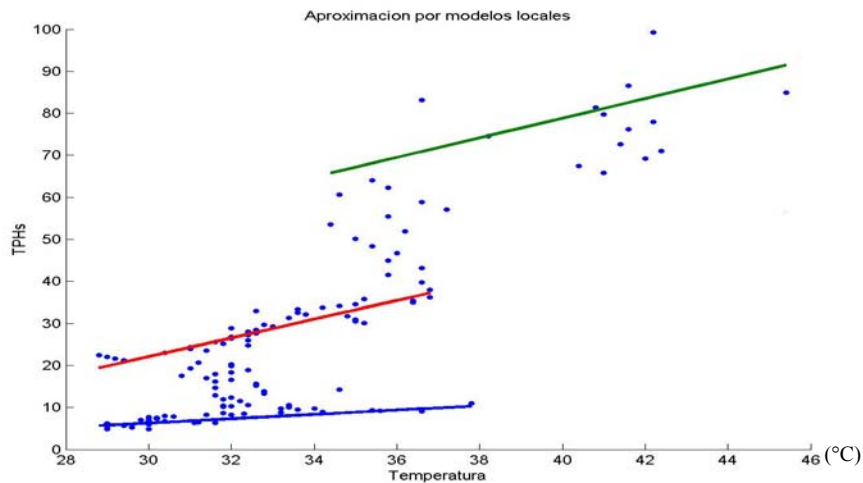


Figura 6.13. Modelos locales obtenidos para cada regla.

## 6.7 H Totales

A continuación, se muestra el proceso de modelado directo usado para caracterizar la variable H Totales y su relación con la variable TPHs. Los datos para este par de variables se muestran en la figura 6.14. En esta figura se puede observar el comportamiento tan irregular de ambas variables, razón por la cual se considera conveniente utilizar este tipo de modelado.

El proceso de modelado se inicia creando un grupo de conjuntos difusos para caracterizar la variable de salida TPHs. Estos conjuntos difusos se muestran en la figura 6.15, los conjuntos son regulares y distribuidos equitativamente en el universo de discurso para la variable TPHs. La definición final del número y tipo de estos conjuntos se obtuvo luego de varias pruebas con diferentes tipos y números de conjuntos difusos esto, hasta alcanzar la aproximación a los datos de salida con el menor error posible.

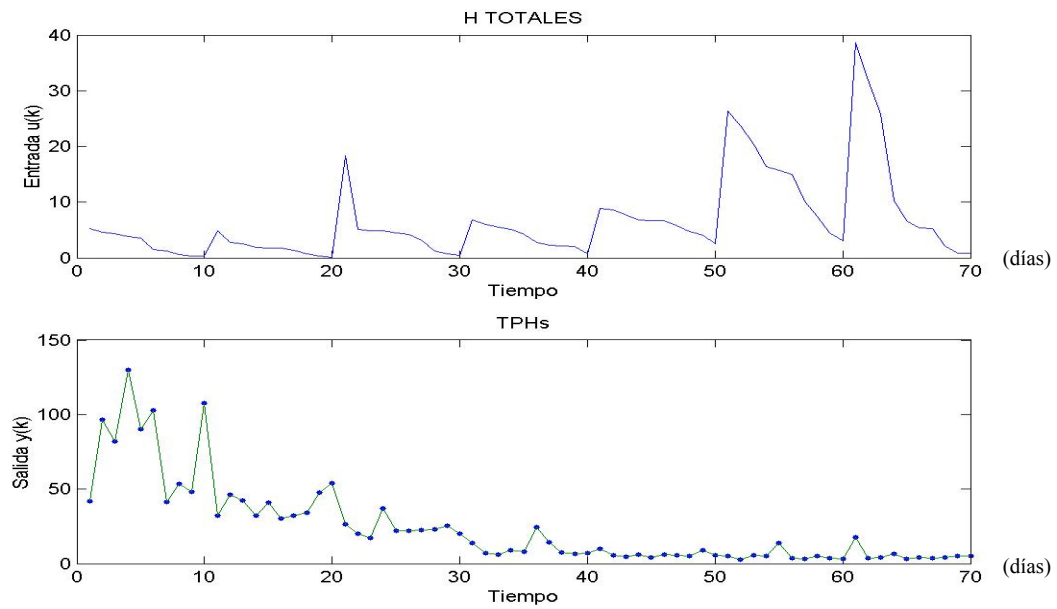


Figura 6.14 Datos para las variables a modelar. La gráfica superior muestra los Datos para la variable H Totales y la gráfica inferior los datos para los TPHs

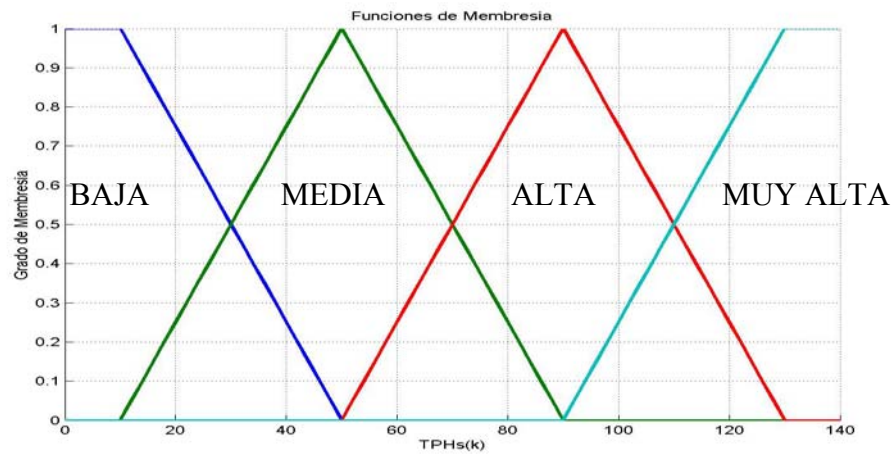


Figura 6.15 Funciones de membresía creadas para caracterizar la variable TPHs

El Sistema Difuso tipo TSK final está formado por las siguientes reglas:

Si TPHs(k) es BAJA entonces  $TPHs(k+1) = 0.9597(TPHs(k)) - 0.0170(H \text{ Totales}(k))$   
Si TPHs(k) es MEDIA entonces  $TPHs(k+1) = 1.2508(TPHs(k)) - 0.6596(H \text{ Totales}(k))$   
Si TPHs(k) es ALTA entonces  $TPHs(k+1) = -0.7339(TPHs(k)) + 43.152(H \text{ Totales}(k))$   
Si TPHs(k) es MUY ALTA entonces  $TPHs(k+1) = 1.7519(TPHs(k)) - 37.0437(H \text{ Totales}(k))$

La salida global de este sistema se puede ver en la gráfica 6.16. En ésta, se puede apreciar cómo el modelo difuso aproxima con un error muy pequeño el comportamiento de la variable TPHs(k).

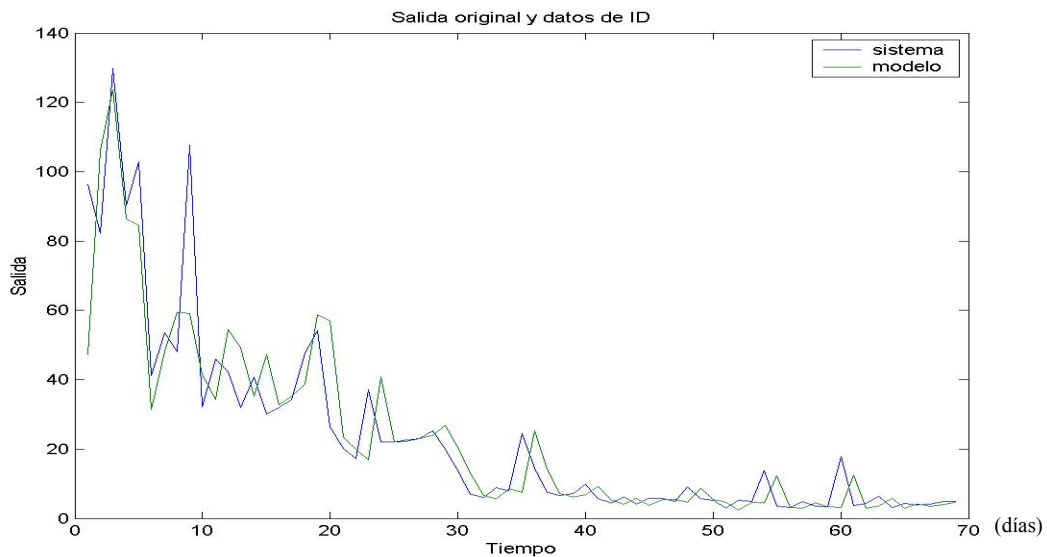


Figura 6.16 Datos correspondientes a la variable TPHs y la salida del modelo difuso.

Nuevamente, el análisis del sistema difuso se puede realizar usando la superficie del modelo en 3D. Para esto, se muestra, mediante una malla, el resultado de la salida del modelo y se sobreponen los datos de la variable TPHs(k+1). Se observa que la malla se adecua con un pequeño error a los datos de la variable. Esta grafica se muestra en la figura 6.17.

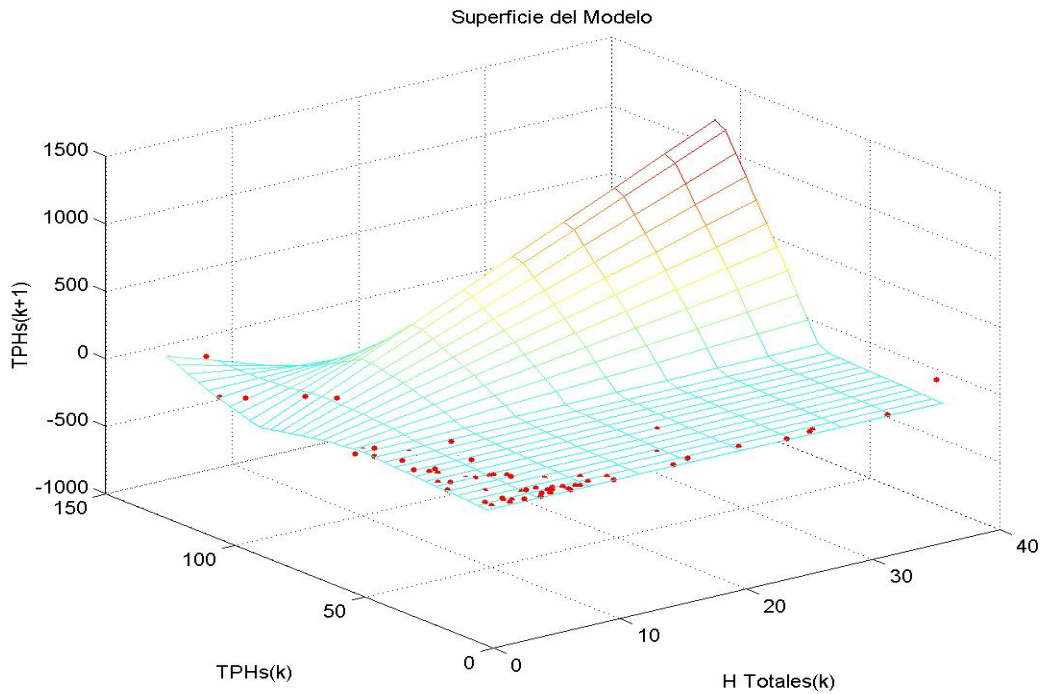


Figura 6.17. Gráfica en 3D de la salida del modelo y los datos de la variable TPHs

## 6.8 Coordinador Difuso

En el caso del coordinador difuso, éste se usa para activar los diferentes modelos difusos según las diferentes etapas clasificadas en sus reglas. Dependiendo de la etapa en proceso, el coordinador difuso activará el correspondiente modelo difuso a fin de determinar el estado presente de una o más variables del sistema. La especificación del conjunto de reglas que constituyen el sistema coordinador está a cargo del experto, quien conoce a detalle la forma en que evoluciona el sistema de la biopila.

Las etapas que se identifican en el funcionamiento de la biopila se muestran en la figura 6.18; cada etapa tiene un conjunto de reglas diferente.

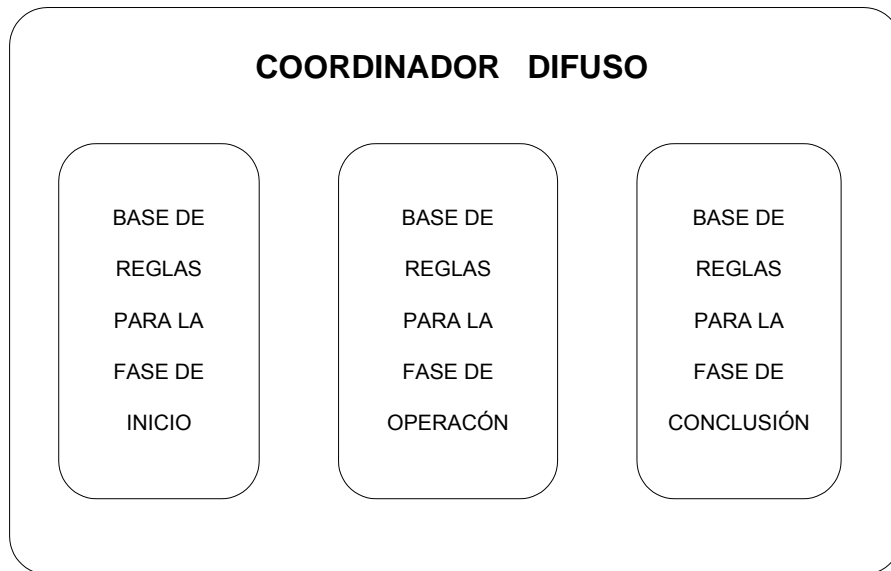


Figura 6.18 Bases de reglas que constituyen el sistema coordinador difuso

El conjunto de reglas para la fase de inicio es:

- Si pH no es óptimo entonces evalúa el Sistema Difuso para el pH
- Si HUMEDAD no es óptima entonces evalúa el Sistema Difuso para la HUMEDAD

El conjunto de reglas para la fase de operación es:

- Si TEMPERATURA no es óptima entonces evalúa el Sistema Difuso para la TEMPERATURA
- Si FOSFORO no es óptimo entonces evalúa el Sistema Difuso para el FOSFORO
- Si H TOTALES no es óptimo entonces evalúa el Sistema Difuso para los H TOTALES

El conjunto de reglas para la fase de conclusión es:

- Si TPHs es MUY BAJO entonces concluye el Sistema de Bioremediación

## 6.9 Conclusiones

En este capítulo se muestra la efectividad de los Sistemas Difusos para modelar Sistemas Complejos. En la resolución del problema abordado, se sigue la estrategia de “*divide y vencerás*”, pues en general, se adopta una metodología jerárquica que caracteriza en tres etapas el funcionamiento de la biopila y, en lo particular, las variables se modelan relacionándolas con la variable más significativa, que es la disminución de los hidrocarburos contaminantes del petróleo en la biopila.

Las diferentes características de cada variable nos permiten utilizar tanto la metodología desarrollada para modelado de sistemas, como el método de modelado directo. Los resultados de las evaluaciones para cada modelo permiten comprobar la gran viabilidad de los Sistemas Difusos para la caracterización de variables con comportamientos altamente no lineales.



## CAPITULO 7.

### Conclusiones Generales

En este trabajo se estableció una alternativa metodológica para construir modelos de sistemas con comportamiento complejo mediante el empleo del marco teórico de los Conjuntos Difusos y de la Lógica Difusa. La metodología obtenida presenta amplias cualidades para obtener el modelo de sistemas no-lineales, cuyo comportamiento es difícil de caracterizar mediante otras metodologías de modelado.

La metodología desarrollada se enmarca dentro de los paradigmas que aplican los métodos de múltiples modelos y regímenes de operación, para el modelado no lineal. Estos paradigmas implementan la estrategia de “divide y vencerás”; estrategia consistente en aproximar por partes el comportamiento global del sistema. De esta manera es posible obtener cada sub modelo mediante el uso de técnicas lineales. Así, el modelo final de un sistema complejo es una sinergia entre varias de las metodologías disponibles, pues cada una de ellas explotan de diferente manera la información disponible.

La búsqueda del Modelo Difuso para un sistema, tiene como propósito final, el explotar la principal característica de los sistemas difusos: la legibilidad o interpretabilidad del modelo final resultante. Esta propiedad permite que la interacción hombre-máquina se realice de una manera familiar para aquellas personas encargadas del funcionamiento del sistema; esto debido a que, el modelo difuso está compuesto de reglas lógicas de la forma canónica *si-entonces*, donde en general los antecedentes son valores lingüísticos y, dependiendo del tipo de sistema difuso empleado, los consecuentes pueden ser igualmente términos lingüísticos o bien, pueden ser una relación lineal de los valores de entrada.

La propiedad de legibilidad o interpretabilidad de los sistemas difusos, les añade una ventaja más, pues permite que el experto, además de validar el conjunto de reglas que conforman el modelo actual, agregue las reglas que considere necesarias para cubrir aquellas zonas no caracterizadas inicialmente por la metodología desarrollada, pero en las que existe la posibilidad de que se presenten en el proceso en un tiempo no previsto; esto se puede considerar como una gran ventaja para los sistemas difusos, pues el modelo final se puede hacer tan general como sea necesario. Esto es, otra ventaja de esta metodología es su flexibilidad.

Además, por la forma en que representan la información, los Sistemas Difusos son considerados como Sistemas Inteligentes, y en el ámbito de control difuso, éstos son considerados como aproximadores universales, ya que pueden aproximar, en un grado arbitrario, una gran gama de sistemas y de manera muy importante, a aquellos sistemas que requieren de la intervención de un operador durante su funcionamiento pues, en general, los

Sistemas Difusos explotan de manera efectiva el conocimiento a priori adquirido por el operador experto a través de la experiencia durante el manejo del sistema. Esto último puede parecer no ortodoxo, sin embargo se presenta en la mayoría de los eventos que ocurren en el mundo real en donde las partes de los sistemas están sometidas a condiciones no previstas por la teoría de diseño original, característica común en lo relacionado a la industria petrolera.

En el transcurso de este trabajo se estuvo en contacto continuo con el experto del sistema, pudiéndose considerar que se cuenta con datos reales en esta colaboración interdisciplinaria.

Para trabajos futuros se presenta un panorama muy extenso no sólo en aplicaciones potenciales, sino también en investigación.

Por una parte, la bioremediación implicando biopilas requiere de condiciones iniciales distintas de acuerdo al lugar contaminado y condiciones del mismo, por ejemplo en algunos lugares secos y fríos, los microorganismos nativos requieren de condiciones de humedad y temperatura distintas para desarrollar su potencial remedial benéfico, pudiéndose adoptar o incorporar esto en lo presentado.

Es posible abordar también el estudio e investigación de nuevas técnicas de minería de datos o reconciliación de datos, en donde se considere incertidumbre en los mismos datos, y de esta manera contar con modelos de procesos de alta complejidad, los cuales sean interpretables, precisos, flexibles y adaptables, para poder realizar labores de optimización y control de procesos o análisis de la fenomenología subyacente en los mismos.

## REFERENCIAS

- [1] Zadeh, L., “Outline of a new approach to the analysis of complex systems and decision processes”, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3, 28-44., 1973.
- [2] Zadeh, L.A. (1965), “Fuzzy sets”, *Information and Control*, vol. 8, 338-353
- [3] Klir G.J. y Yuan B., *Fuzzy Sets and Fuzzy Logic* (1995), Prentice Hall, New Jersey, ISBN 0-13-101171-5.
- [4] Yager, R. y Filev, D., *Essentials of Fuzzy Modeling and Control* (1994), Wiley, New York, ISBN 0-471-01761-2.
- [5] Johansen T.A., “Operating Regime based Process Modeling and Identification” PhD. Thesis, *Norwegian Institute of Technology*, N-7034 Thondheim, Norway, 1994.
- [6] Haykin, S., *Neural Networks*, New York:Mcmillan Maxwell International, 1994.
- [7] Chen,S., C. Cowan and P. Grand (1991), Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Networks* 2(2), 302-309.
- [8] Kosko, B (1994), Fuzzy Systems as universal approximators. *IEEE Trans. Computers* 43, 1329-1333.
- [9] Wang, L.-X (1994), *Adaptive Fuzzy Systems and Control, Design and Stability Analysis*,. New Jersey: Prentice Hall.
- [10] Takagi, T. and M. Sugeno (1985), Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. Systems, Man and Cybernetics* 15(1), 116-132.
- [11] Gaines, B. and J. Boose (1988), *Knowledge Acquisition for Knowledge-Based Systems*, London: Academic Press.
- [12] Mamdani, E. (1977), Application of fuzzy logic to approximate reasoning using linguistic systems. *Fuzzy Sets and Systems* 26, 1182-1191.
- [13] Duda, R. and Hart, P. (1973), *Pattern Classification and Scene Analysis*, John Wiley & Sons.
- [14] Bezdeck, James C. (1981), *Pattern Recognition with Fuzzy Objective Algorithms*, Plenum Press, New York.
- [15] Ruspini. E, (1970), Numerical Methods for Fuzzy Clustering, *Information Science*, vol. 2, 319-350.

- [16] Dunn, J. C., (1974), A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, well Separated Clusters, *Journal of Cybernetics*, vol. 3, 32-57.
- [17] Bezdeck, J., R. Hathaway, R. Howard, C. Wilson and M. Windham (1987). Local convergence analysis of a grouped variable version of coordinate descent, *Journal of Optimization Theory and Applications*, 54(3), 471-477.
- [18] DeSerbo, W. (1982). Gennclus: New models for general nonhierarchical clustering analysis. *Psychometrica*, 47(4), 449-476.
- [19] Babu, G. and M. Marty (1994). Clustering with evolution strategies. *Pattern Recognition*, 27(2), 321-329.
- [20] Gustafson, D. and W. Kessel (1979). Fuzzy Clustering with a fuzzy covariance matrix. In *Proc. IEEE CDC*, San Diego, CA, USA, pp 761-766.
- [21] Gath, I. And A. Geva (1989). Unsupervised optimal Fuzzy Clustering. *IEEE Trans. Pattern Analysis and Machine Intelligence* 7, 773-781.
- [22] Bezdeck, J., C. Coray, R. Gunderson and J. Watson (1981 a). Detection and Characterization of cluster substructure, I Linear structure: Fuzzy c-lines, *SIAM Journal of Appl. Math*, 40(2), 339-357.
- [23] Hathaway, R. and J. Bezdeck. (1993 b). Switching regression models and fuzzy clustering. *IEEE Trans. Fuzzy Systems*, 1(3), 195-204.
- [24] Zimmermann, H.-J. (1987), *Fuzzy Sets, Decision making and Expert Systems*, Boston: Kluwer Academic Publishers.
- [25] Yoshinari, Y, W. Pedrycz and K. Hirota. (1993), Construction of fuzzy models through clustering techniques. *Fuzzy Sets and Systems* 54, 157-165.
- [26] Hunt, K., D. Sbarbaro, R. Zbikowski and P. Gawthrop (1992), Neural Networks for control systems -a survey. *IEEE Trans. Neural Networks* 28, 1083-1112.
- [28] United States Environmental Protection Agency (1996), *How to evaluate Alternative Cleanup Technologies for Underground Storage Tank Sites, A guide for Corrective Action Plan Reviewers*. EPA 510-R-04-002. [www.epa.gov/oust/pubs/tums.htm](http://www.epa.gov/oust/pubs/tums.htm)
- [29] Huesemann, M.H. (1994). Guidelines for land-treating petroleum hydrocarbon-contaminated soils *J. Soil. Contam.* 3(3), pp.1-204
- [30] NORRIS, R.D. (1994). Handbook of bioremediation. CRC Press, Boca Raton, Florida.
- [31] Suteaki, S., Kazuyuki S., and Toshiomi Y. (1999). Knowledge-Based Design and Operation of Bioprocess Systems *J. Biosc. and Bioeng.* 87(3), pp.261-266

- [32] Horiuchi, J-I. (2002). Fuzzy Modeling and Control of Biological processes *J. of Bioscience and Bioengineering*. 94(6), pp. 574-578



## **Apéndice A.**

### Implementación en MATLAB de los Algoritmos Utilizados

Implementación del algoritmo difuso de clustering fuzzy c-means

Implementación del algoritmo difuso de clustering Gustafson-Kessel

Archivo SCRIP con Ejemplo del Algoritmo Fuzzy c-means

Archivo SCRIP para el Modelado de TEMPERATURA Vs TPHs

## Implementación del algoritmo difuso de clustering fuzzy c-means

```

function [U,V] = fcm_fer(Z,U0,m,tol)

%           [U, V] = FCM_FER(Z, U0, m, tol)
%-----
% ENTRADAS: Z ... Matriz de datos N x n ; N = número de muestras n = número de variables
%           U0 ... matriz de partición difusa inicial o bien, el número de agrupamientos
%           m ... exponente de fusividad (m > 1)
%           tol ... valor mínimo o umbral para terminar el proceso
%-----
% SALIDAS: U ... Matriz de partición difusa
%           V ... Valor de los centros
%
[mz,nz] = size(Z); % mz = número de datos; nz = número de variables
c = size(U0,2); % Lee el No de columnas de U0, esto para
% checar si U0 es una matriz o un número
if c == 1, % U0 no es una matriz y solo es el número de clusters
    c = U0; % Asigna el valor de U0 a c (número de clusters)
end;
mZ1 = ones(mz,1); % Vector columna de tamaño mz con 1's
nZ1 = ones(nz,1); % Vector columna de tamaño nz con 1's
V1c = ones(1,c); % Vector renglón de tamaño c con 1's
U = zeros(mz,c); % Matriz de partición difusa de tamaño mz x c
d = U; % Matriz de distancias (mismo tamaño que U0)

%-----Inicialización del algoritmo-----

% Si no se da la matriz de partición difusa inicial, entonces se debe crear
% dicha matriz. Para esto, primero se determinan de manera aleatoria los centros
% iniciales, posteriormente se calculan las distancias a partir de estos centros
% y finalmente se calcula la matriz de partición difusa inicial
%
if size(U0,2) == 1, % Cuando solo se tiene el numero de particiones requeridas
    % se debe calcular una matriz de partición difusa inicial
    minZ = V1c*min(Z); % produce una matriz de tamaño c x nz, que contiene
    % c renglones con los valores mínimos de los datos
    maxZ = V1c*max(Z); % Igual que la anterior pero para los valores máximos

V = minZ + (maxZ-minZ).*rand(c,nz); % Centros iniciales aleatorios pero
%dentro del rango de los valores de Z
for j = 1 : c, % Para cada c partición
    ZV = Z - mZ1*V(j,:); % Crea una matriz (mz x nz) con el j-avo centro
    % y lo resta a la matriz de datos
    d(:,j) = sum((ZV.^2)'); % Eleva al cuadrado cada elemento de la resta y
    % suma por renglón. El resultado es la distancia de
    % cada elemento de los datos al j-avo centro
end;

d = (d+1e-100).^(-1/(m-1)); % Se hace este calculo para evitar la singularidad que
% se puede cuando existe un dato con distancia CERO
U0 = (d ./ (sum(d')*V1c)); % Calculo de la Matriz de partición difusa.
% Suma los renglones y los repite c veces (mz x c). Divide cada
% elemento de d entre la suma de su renglón (con esto normaliza
% la suma de las membresías en 1).
end;

```



```

%-----Proceso Iterativo-----
%
while max(max(abs(U0-U))) > tol % Mientras no se alcance el valor de
                                % tolerancia especificada como mínima
    U = U0; % Actualiza la matriz de partición difusa
    Um = U.^m; % Eleva cada elemento al exponente m
    sumU = sum(Um); % Suma cada columna de la matriz de partición (1 x c)

    V = (Um'*Z)./(nZ1*sumU)'; % Calcula los nuevos centros (nz x c)
                                % Multiplica la matriz de membresías por la
                                % matriz de datos y la divide entre la suma
                                % de las membresías
    for j = 1 : c, % Para cada c cluster
        ZV = Z - mZ1*V(j,:); % Variable auxiliar, contiene la resta de
                                % los datos con el j-avo centro
        d(:,j) = sum((ZV.^2)'); % Actualiza las distancias
    end;
    d = (d+1e-100).^(-1/(m-1)); % asegura que ningún dato tenga distancia CERO

    U0 = (d ./ (sum(d')'*V1c)); % Nueva matriz de partición difusa
end

U = U0; % Matriz de partición difusa final

```

## Implementación del algoritmo difuso de clustering Gustafson-Kessel

```

function [f,v,P,V,D,J,M] = gkfastfer(x,f0,m,e,s,rho)
%
% Entradas:
% X ... M by N data matriz,
% M es el número de puntos de datos y N la dimensión de los mismos
% F0 ... ya sea una matriz de partición difusa, o el número de clusters.
% En el último caso, por default se genera una matriz de partición
% m ... parámetro opcional m > 1, determina la fusividad del
% clustering, para m cercano a 1 los cluster llegan a ser certeros,
% el valor de default es 2
% e ... tolerancia opcional para la conclusión, el algoritmo se detiene
% cuando max(max(F(k-1) - F(k))) <= e, tolerancia de default 1e-3
% s ... parámetro opcional para graficar resultados intermedios
% (solo para datos 2D), default 0 - i.e. no graficar, poner a 1
% para mostrar el proceso de clustering on-line,
% poner a 2 para acelerar las gráficas para datos ordenados
% rho .. vector de 1xM de los volúmenes esperados de los clusters
%
% Salidas:
% F ... Matriz de partición difusa
% C ... matriz de la media de los clusters
% P ... matrices de la covarianza de los clusters concatenados en una
% matriz P = [P1;P2;...Pk], donde k es el número de clusters
% V ... eigenvectores de las matrices de covarianza, correspondientes
% a los eigenvalues más pequeños
% D ... eigenvalues de las matrices de covarianza
% J ... Historia del criterio de clustering J
% M ... matrices induciendo la norma distancia, calculada como
% M = det(P).^(1/N)*inv(P), M = [M1;M2;...Mk]
%
% % % % % %
% ADECUAMIENTO E INICIALIZACION DEL ALGORITMO
% % % % % %

[mx,nx] = size(x); %contienen el tamaño y dimensión de datos de entrada
[mf0,nf0] = size(f0); %contiene tamaño y dimensión de matriz inicial de partición
x1 = ones(mx,1); %variable auxiliar
inx = 1/nx; %variable auxiliar

if max(mf0,nf0) == 1, % Solo se dio el No de clusters, hay que generar la matriz de partición
    c = f0; %Asigna a c el No de clusters
    mm = mean(x); %vector renglón con el valor medio de cada columna de x
    aa = max(abs(x - ones(mx,1)*mm)); %resta a cada elem de x el valor medio de su columna y toma
    %el valor máximo de cada columna
    v = 2*(ones(c,1)*aa).*(rand(c,nx)-0.5) + ones(c,1)*mm; %Se generan centros aleatorios pero dentro
    %del universo de valores de x
elseif nf0 == nx, % Matriz de centros proveída
    c = mf0; % se asigna a c el número de centros
    v = f0; % se asigna a v la matriz de centros proveída
end;

if mf0 ~= mx, % ¿Número de centros diferente al número de datos?
    % Calculo de f0, la matriz de partición difusa inicial

```

```

for j = 1 : c, % Para todos los clusters
    xv = x - x1*v(j,:); % Resta a cada elemento de x el c_avo centro
    d(:,j) = sum((xv.^2)'); % eleva al cuadrado cada resultado de la resta y suma por renglón
    % generando una columna con las distancias de los elementos a los
    % centros
end;

d = (d+1e-10).^(-1/(m-1)); %Esta operación se realiza para prever la existencia de valores en d
    %(distancia) de CERO, evitando que el algoritmo caiga en una singularidad
f0 = (d ./ (sum(d')*ones(1,c))); %MATRIZ INICIAL DE PARTICION DIFUSA
else % Matriz F0 proveída inicialmente
    c = size(f0,2); % Asigna a c el No de clusters
    fm = f0.^m; % Cada elemento se eleva al valor de fusividad "m"
    sumf = sum(fm); % Se suman las columnas
    v = (fm*x)./(sumf*ones(1,nx)); % Se generan los c centros
end

f = zeros(mx,c); % tamaño de la matriz de partición
iter = 0; % contador del número de iteraciones

if (nargin < 6), % Comprueba que el dato de entrada exista y de no ser
    rho = ones(1,c); % así, lo complementa
elseif isempty(rho), % o bien, chequea si es un espacio vacío
    rho = ones(1,c); % en cuyo caso, también escribe el valor de default
end;

% Inicializa graficas
if (s ~= 0) & (nx < 3), %
    subplot(211);
    lines = [v(:,1)*x(:,1)+v(:,2)*ones(1,mx)];
    mask = find(f0 < 0.2);
    lines(mask) = NaN*ones(size(mask));
    H1 = plot(x(:,1),x(:,2),'b.',v(:,1),v(:,2),'r*',x(:,1),lines,'EraseMode','xor','linewidth',2,'markersize',12);
    title('a) Funcion a aproximar (..), Centros de los clusters(*), Modelos Locales (-)');
    xlabel('x'); ylabel('y');
    minx = min(x(:,1)); maxx = max(x(:,1));
    miny = min(x(:,2)); maxy = max(x(:,2));
    ma = 0.3*max(abs(x(:,2)));
    axis([minx maxx miny-miny ma maxy+ma]);
    subplot(212);
    if s == 1,
        H2 = plot(x(:,1),f0,'EraseMode','xor');
    else
        H2 = plot(x(:,1),f0,'EraseMode','xor','linewidth',2);
    end;
    title('b) Funciones de Membresia Parametrizadas');
    xlabel('x'); ylabel('Grado de Membresía');
    set(gcf,'UserData',[H1;H2]);
end;

while max(max(f0-f)) > e
    iter = iter + 1;
    f = f0;
    fm = f.^m;
    sumf = sum(fm);

```

```

% Calculo de los centros

v = (fm'*x)/(sumf*ones(1,nx));
for j = 1 : c,
    xv = x - x1*v(j,:);

% Calculo de la matrix de covarianza

p = ones(nx,1)*fm(:,j)'.*xv'*xv/sumf(j); %matriz de covarianza del c centro

% Calculo de las diatancias
M = (det(p)/rho(j))^inx*inv(p);           %M es la matriz norma distancia A del c cluster
d(:,j) = sum((xv*M.*xv)');               %Calculo de las distancias con la matriz A

% Calculo de los eigen vectors y los cluster prototipos (líneas)
if s ~= 0 & nx < 3,                       %Restringe el calculo a datos 2D y solo cuando se
                                            %habilita la opción de ver resultados intermedios
    [ev,ed] = eig(p);                      %ev, full matrix cuyas columnas son los eigenvectors
                                            %ed, matriz diagonal de eigenvalues
    ed = diag(ed);                          %ed queda como un vector columna con los elementos
                                            %de la diagonal de ed
    ev = ev(:,ed == min(ed));               %"deja" el eigenvector correspondiente al mínimo de
                                            % los eigenvalues
    lines(:,j) = -x(:,1)*ev(1)/ev(2) + v(j,:)*ev/ev(2);

    mask = find(f0 < 0.2);                 % encuentra los grados de membresía < 0.2
    lines(mask) = NaN*ones(size(mask));     % los enmascara con NaN's para las gráficas
end;                                       %End del if s ~= 0 & nx < 3

end;                                       %End del for j = 1 : c

J(iter) = sum(sum(f0.*d));                 %la suma del producto de las membresías por las d's

d = (d+1e-10).^(-1/(m-1));
f0 = (d ./ (sum(d')*ones(1,c)));          %la nueva matriz de partición difusa

% Grafica de resultados intermedios...
if (s ~= 0), %fprintf('Iteration count = %d, J = %f\n\n',iter,max(max(f-f0))),
    if (nx < 3),
        H = get(gcf,'UserData');
        set(H(2),'xdata',v(:,1),'ydata',v(:,2));
        for i = 1 : c,
            set(H(2+i),'ydata',lines(:,i));
            set(H(2+c+i),'ydata',f0(:,i));
        end;
        drawnow;
        end;
end;

if iter == 1,
    pause (2)
else
    pause(1);
end;
end
                                %END del ciclo While

```

```

% CALCULO RESULTADOS FINALES
%

fm = f.^m;
sumf = sum(fm);

P = zeros(c*nx,nx);      % ARREGLO MATRICES DE COVARIANZA
M = P;                  % MATRIZ NORMA-INDUCTORA
V = zeros(c,nx);       % VECTORES PROPIOS
D = V;                  % VALORES PROPIOS

% calculo de P,V,D,M

for j = 1 : c,
    xv = x - ones(mx,1)*v(j,:);

% Calculo de la Matriz de Covarianza
    p = ones(nx,1)*fm(:,j)'.*xv'*xv/sumf(j);

% Calculo de los valores propios y vectores propios
    [ev,ed] = eig(p); ed = diag(ed)';
    ev = ev(:,ed == min(ed));

% Concatena la información de los clusters en una matriz
    P((j-1)*nx+(1:nx),:) = p;
    M((j-1)*nx+(1:nx),:) = (det(p)/rho(j)).^(1/nx)*inv(p);
    V(j,:) = ev';
    D(j,:) = ed;
end;

```

## Archivo SCRIP con Ejemplo del Algoritmo Fuzzy c-means

% para este ejemplo, se crea un conjunto de datos bidimensionales aleatorios.

```
N = 30;           % numero de datos
R = 0.4;         % varianza de los datos
randn('seed',1);
d = [-2-.2*randn(N,1) -0.5+R*randn(N,1)
      2+R*randn(N,1) -0.5+.1*randn(N,1)
      .5+R*randn(N,1) +0.5+R*randn(N,1)];
```

```
figure(1); clf;
plot(d(:,1),d(:,2),'b','markersize',12);
title('Fuzzy C-means');
axis([-4 4 -2 2]);
set(gca,'xticklabel',[],'yticklabel',[])
set(gca,'xtick',[],'ytick',[])
xlabel('x','fontname','times','fontsize',10);
ylabel('y','fontname','times','fontsize',10);
```

```
%-----
[f,c] = fcm_fer(d,rand(size(d,1),3));
```

```
figure(2);
```

```
[ds,ind] = sort(d(:,1));
```

```
subplot(221);
plot(d(:,1),d(:,2),'b','markersize',10); hold on
plot(c(:,1),c(:,2),'r+','linewidth',2,'markersize',7); hold off
axis([-4 4 -2 2])
title('Fuzzy C-means');
set(gca,'xticklabel',[],'yticklabel',[])
set(gca,'xtick',[],'ytick',[])
xlabel('x','fontname','times','fontsize',12);
ylabel('y','fontname','times','fontsize',12);
```

```
subplot(222);
[xx,yy]=meshgrid(-4:0.05:4, -2:0.05:2);
fm1 = xx;
```

```
fg = fgrade([xx(:) yy(:)],c,2);           % calcula los grados de membresía
```

```
fm1(:) = max(fg'); %(81 x 161)
surf(xx,yy,fm1);
view(332,55)
shading interp
```

```
set(gca,'xticklabel',[],'yticklabel',[])
set(gca,'xtick',[],'ytick',[])
title('Conjuntos difusos en 3D');
xlabel('x','fontname','times','fontsize',12);
ylabel('y','fontname','times','fontsize',12);
zlabel('\mu','fontname','times','fontsize',12);
```

```
ff = xx(1,:); %(161 x 1)
```

```

for i = 1 : 3,
    ff(:,i) = max(reshape(fg(:,i),81,161));
end;

% ff = (161 x 3)

subplot(223);
plot(d(:,1),f,'.',markersize,7)
axis([-4 4 0 1])
title('Conjuntos Difusos en 2D');
xlabel('x','fontname','times','fontsize',14);
ylabel('\mu','fontname','times','fontsize',14);

subplot(224);
plot(xx(1,:),ff)
axis([-4 4 0 1])
title('Conjuntos Difusos en 2D con lineas');
xlabel('x','fontname','times','fontsize',14);
ylabel('\mu','fontname','times','fontsize',14);

```

## Archivo SCRIP para el Modelado de TEMPERATURA Vs TPHs

```
% Usa el algoritmo de clustering difuso Gustafson-Kessel
%
NC = 3;          % número de clusters
m = 2;          % exponente de fusividad
eps = 0.01;     % diferencia mínima entre particiones subsecuentes
s = 2;
Ncont = 10;     % Numero de líneas de contorno

Dat_pil_B      % contiene 126 registros diarios de TEMPERATURA y TPHs

x(:,1) = Dat_B(:,1);
x(:,2) = Dat_B(:,2);

figure(1); clf;
plot(x(:,1), x(:,2), 'b', 'markersize', 14);
title('Pila B');
xlabel('Temperatura', 'fontname', 'times', 'fontsize', 12);
ylabel('TPHs', 'fontname', 'times', 'fontsize', 12);

[mx, nx] = size(x);
[mf0, nf0] = size(NC);
x1 = ones(mx, 1);
inx = 1/nx;

c = NC;
mm = mean(x);
aa = max(abs(x - ones(mx, 1) * mm));
v = 2 * (ones(c, 1) * aa) .* (rand(c, nx) - 0.5) + ones(c, 1)*mm;

for j= 1:c,
    xv = x - x1 * v(j,:);
    d(:,j) = sum((xv.^2)');
end;
d = (d+1e-10).^(1/(m-1));

% Calculo de la matriz de membresías inicial

f0 = (d ./ (sum(d')*ones(1,c)));

% Variables auxiliares
f = zeros(mx, c);
iter = 0;

% Asignación del valor de volumen para cada cluster

rho = ones(1,c);

% Proceso iterativo. En tanto la diferencia entre la nueva partición y
% la inmediata anterior no sea menor al valor asignado para epsilon

P = zeros(c*nx,nx);      % p/matriz de covarianza
V = zeros(c,nx);        % p/eigenvectores
```



```

while max(max(f0 - f)) > eps
    iter = iter + 1;
    f = f0;
    fm = f.^m;
    sumf = sum(fm);

%calculo de los nuevos centros

    v = (fm' * x) ./ (sumf' * ones(1, nx));

% Iteración sobre cada "c" columna para los nuevos cálculos

    for j = 1:c,
        xv = x - x1 * v(j,:);

%calculo de la matriz de covarianza

        p = ones(nx,1)*fm(:,j)'.*xv'*xv/sumf(j);
        P((j-1)*nx+(1:nx),:) = p;

%calculo de las distancias

        M = (det(p) / rho(j)) ^ inx * inv(p);
        d(:,j) = sum((xv * M .* xv)');

%calculo de los vectores propios

        if s ~= 0 & nx < 3,
            [ev,ed] = eig(p);
            ed = diag (ed);
            ev = ev(:,ed==min(ed));
            V(j,:) = ev'; %%% Matriz de Eigenvectores
            EV((j-1) * nx + (1:nx), :) = ev;
            ED((j-1) * nx + (1:nx), :) = ed;
        end;
    end;

% actualiza la matriz de particion difusa

    J(iter) = sum(sum(f0 .* d));
    d = (d + 1e-10) .^ (-1/(m-1));
    f0 = (d ./ (sum(d)' * ones(1,c)));

%grafica resultados intermedios

    [xx, yy] = meshgrid(25:0.20:50, 0:125);
    fg = fgrade( [xx(:) yy(:)], v, m, P);
    fm1 = xx;
    fm1(:) = max(fg');

end; % Fin del ciclo WHILE. Indica que ya se alcanzo la mínima diferencia.

hold on;

```

```

plot(x(:,1),x(:,2),'b.','markersize',15,'EraseMode','none');
plot(v(:,1),v(:,2),'r+', 'linewidth',3,'markersize',10);
contour(xx,yy,fm1,Ncont,'m');
hold off;

cutoff = 0.1;
mt = 2 ; % Tipo de FM(1,trapezoide, 2 Exponencial; 3 Gausiana)
OPT = foptions;
OPT(14) = 1000;

% -----
% Calculo de las FMs convexas y sus parámetros
%-----

figure(2); clf
[ds,fs]=smooth(x(:,1),f0,cutoff);
mf = mffit(ds,fs,mt,OPT,[1 1 1]);
lim = mf(:,3) == min(mf(:,3));
mf(lim,2:3) = ones(sum(lim),1)*[min(x(:,1)) min(x(:,1))];
if mf(lim,4) < min(x(:,1)),
    mf(lim,4) = min(x(:,1));
end;
lim = mf(:,4) == max(mf(:,4));
mf(lim,4:5) = ones(sum(lim),1)*[max(x(:,1)) max(x(:,1))];
if mf(lim,3) > max(x(:,1)),
    mf(lim,3) = max(x(:,1));
end;

figure(3); clf
mfplot(ds',mf);
mfs = [mf];
title('Pila B');
xlabel('Temperatura');
ylabel('Grado de membresía');

figure(4); clf

[ds2,fs2]=smooth(x(:,2),f0,cutoff);
mf2 = mffit(ds2,fs2,mt,OPT,[1 1 1]);
lim = mf2(:,3) == min(mf2(:,3));
mf2(lim,2:3) = ones(sum(lim),1)*[min(x(:,2)) min(x(:,2))];
if mf2(lim,4) < min(x(:,2)),
    mf2(lim,4) = min(x(:,2));
end;
lim = mf2(:,4) == max(mf2(:,4));
mf2(lim,4:5) = ones(sum(lim),1)*[max(x(:,2)) max(x(:,2))];
if mf2(lim,3) > max(x(:,2)),
    mf2(lim,3) = max(x(:,2));
end;

figure(5); clf
mfplot(ds2',mf2);
mfs2= [mf2];
title('Pila B');
xlabel('TPHs');
ylabel('Grado de membresía');

```

```
[p1,ym,yl,ylm] = suglms([x(:,1) ones(size(x),1)],x(:,2),f0);
```

```
% -----  
% gráfica de los resultados de la función ajustada  
% -----
```

```
figure(5);clf  
plot(x(:,1),x(:,2),'.b', 'markersize', 14);  
set(gca,'box','off');  
hold on;  
title('Salida global');  
xlabel('temperatura');  
ylabel('TPHs');  
plot(x(:,1),ym,'g','linewidth',2);  
hold off;
```

```
% Modelos Locales  
figure(6);clf  
plot(x(:,1),x(:,2),'.b', 'markersize', 14);  
set(gca,'box','off');  
hold on;  
title('Aproximacion por modelos locales')
```

```
xlabel('Temperatura');  
ylabel('TPHs');  
plot(x(:,1),ylm,'linewidth',2);  
hold off;
```