

# **UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**

*INSTITUTO DE INGENIERÍA*

*MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA*



***“DESARROLLO DE MODELOS PARA ESTIMAR LA GENERACIÓN DE LIXIVIADOS Y BIOGÁS EN EL PROCESO DE BIODEGRADACIÓN ANAEROBIA MEDIANTE HERRAMIENTAS DE SOFTCOMPUTING”***

TESIS PARA OBTENER EL GRADO DE:

**DOCTOR EN INGENIERÍA**

**PRESENTA**

MARÍA DEL CONSUELO MAÑÓN SALAS

**DIRECTOR**

DRA. LILIANA MÁRQUEZ BENAVIDES

Mexicali, B. C.

Enero de 2012

## AGRADECIMIENTOS

### *A Díos,*

*Por darme conciencia sobre una existencia llena de bendiciones por las cuales que agradecer.*

*Por poner en mi camino a seres excepcionales que con su ejemplo, cariño y dedicación, me motivan día a día a no desistir a pesar de las dificultades.*

*Por permitirme dar un paso más en una trayectoria profesional.*

### *A la Familia,*

*Por ser el pilar sobre el cual he podido construir una vida plena.*

### *A mis Asesores y revisores,*

*Por sus comentarios, consejos y tiempo invertido, imprescindible para enriquecer el presente trabajo.*

### *A todos mis Amigos,*

*Por acompañarme en este camino lleno de aventuras, incluyendo a aquellos que por diferentes razones ya no están presentes y a los que por azares de la vida he reencontrado.*

### *A los integrantes de REDISA*

*Por ser una fuente de inspiración.*

### *Y Especialmente a:*

*La Dra. Gabriela<sup>†</sup> por motivar el desarrollo de este trabajo, y que hoy se ha convertido en una realidad.*

*La Dra. Lilitiana por compartir su experiencia y aceptarme como tesista.*

*La Dra. Sara por todo su apoyo y cariño.*

*La Dra. Consuelo por toda su paciencia, apoyo, compañerismo y amistad incondicional.*

## RESUMEN

En este trabajo de tesis se desarrollaron y evaluaron favorablemente modelos para estimar lixiviados y biogás en el proceso de biodegradación anaerobia, mediante herramientas de *Softcomputing* (Redes Neuronales Artificiales-RNA y Sistemas Adaptativos de Inferencia Neurodifusa-ANFIS) y datos experimentales.

El sistema bajo estudio consistió de un conjunto de biorreactores a escala laboratorio cargados con residuos sólidos urbanos del relleno sanitario de Pátzcuaro, Michoacán, los cuales operaron a seis tasas de recirculación de lixiviados (15, 30, 40, 60, 80 y 120 % Volumen). Los lixiviados se recircularon dos veces por semana durante 201 días. Las entradas del sistema quedaron determinadas por las características de los lixiviados recirculados (tasa de recirculación, pH, conductividad eléctrica, ácido acético, ácido propiónico, demanda química de oxígeno, sólidos totales, sólidos volátiles totales, nitrógeno amoniacal) y las salidas por la composición del metano en el biogás generado y los lixiviados producidos (volumen y características).

El proceso inició con la verificación de la validez de los datos, seguido por el análisis exploratorio, el cual fue integrado por el análisis univariado y multivariado. Considerando la importancia de la selección de variables, en el proceso de modelado, se compararon las técnicas: búsqueda exhaustiva (BE), *stepwise* (SW) y el conocimiento del experto (SE) mediante un conjunto de variables propuestas por él.

Con el diseño y empleo de tres funciones en código de MATLAB se facilitó la implementación de la búsqueda exhaustiva, la selección de una ANN y la configuración de ANFIS. Debido a que fueron tres métodos los empleados en la selección de variables (BE, SW y SE) y dos las salidas (metano y lixiviado producido), se evaluaron seis modelos. El criterio de selección se vio justificado con la presencia de valores bajos en el error de entrenamiento y validación, lo que representó una mayor bondad de ajuste con valores de  $R^2$  muy cercanos a uno. Una vez seleccionados los modelos con el menor RMSE se les realizó un análisis de sensibilidad, con el fin de determinar las variables de entrada con mayor influencia en la salida de los modelos.

## ABSTRACT

In this thesis were developed and evaluated favorably models to estimate leachate and biogas in anaerobic biodegradation process, through *Softcomputing* tools (Artificial Neural Networks-ANN and Adaptive Neural Fuzzy Inference Systems-ANFIS) and experimental data.

The system under study was integrated by a set of laboratory-scale bioreactors loaded with municipal solid waste of landfill Patzcuaro Michoacan, which operated at six leachate recirculation rates (15, 30, 40, 60, 80 and 120 % Volume). The leachate was recirculated twice a week by 201 days. The inputs of the system were determined by the characteristics of the recirculated leachate (recirculation rate, pH, electrical conductivity, acetic acid, propionic acid, chemical oxygen demand, total solids, total volatile solids, ammoniacal nitrogen) and the outputs by methane composition in biogas and the produced leachate (volume and characteristics).

The process began with the verification of the validity of the data, followed by an exploratory analysis, which was composed of univariate and multivariate analysis. Considering the importance of variable selection in the process modeling techniques were compared: exhaustive search (BE), *stepwise* (SW) and expert knowledge (SE) through a set of variables proposed by him.

With the design and use of three functions in MATLAB code were facilitated the implementation of the exhaustive search, the selection of an ANN and ANFIS. Because the three methods were used in the selection of variables (BE, SW, SE) and two outputs (methane and produced leachate) were evaluated six models. The selection criterion was justified by the presence of low values in the training error and validation, which represented a higher goodness of fit with  $R^2$  values very close to one. Once selected the models with the lowest RMSE were evaluated by a sensitivity analysis, to determine the input variables with greatest influence on the outputs of the models.

# INDICE

AGRADECIMIENTOS	I
RESUMEN	II
ABSTRACT	III
ÍNDICE	IV
INDICE DE FIGURAS	VII
INDICE DE TABLAS	IX
ABREVIATURAS	XI
<b>1. INTRODUCCION</b>	
1.1. Antecedentes.	1
1.2. Planteamiento del problema.	6
1.3 Justificación.	7
1.4 Hipótesis.	8
1.5 Objetivos.	8
<b>2. MARCO TEÓRICO</b>	9
2.1 Una problemática ambiental.	9
2.2 El proceso de biodegradación y sus productos.	10
2.2.1 Factores que influyen en el proceso de biodegradación.	10
2.2.2 Proceso de biodegradación.	12
2.2.3 Lixiviados.	14
2.2.4 Biogás.	15
2.3 Modelado de sistemas ambientales	15
2.3.1 Sistema.	15
2.3.2 Modelos ambientales.	16
2.3.3 Construcción del modelo.	17
	IV

2.3.3.1	Formulación del problema.	18
2.3.3.2	Recolección y procesamiento de los datos.	18
2.3.3.3	Selección de variables.	21
2.3.3.4	Construcción del modelo.	22
2.3.3.5	Verificación del modelo.	23
2.3.3.6	Análisis de sensibilidad.	24
2.4	Técnicas de <i>Softcomputing</i>	25
2.4.1	Red neuronales artificiales.	26
2.4.2	Lógica Difusa.	29
2.4.2.1	Conjuntos Difusos.	30
2.4.2.2	Reglas Difusas.	31
2.4.2.3	Subclustering Difuso.	31
2.4.2.4	Operaciones Difusas.	32
2.4.3	Sistemas de inferencia difusos.	33
2.4.4	Sistemas Neurodifusos.	34
2.5	Modelando con Herramientas de <i>Softcomputing</i> .	36
<b>3.</b>	<b>METODOLOGÍA</b>	39
3.1	Formulación del problema.	40
3.2	Análisis y procesamiento de datos.	42
3.2.1	Análisis exploratorio de los datos.	42
3.2.1.1	Análisis Univariado.	43
3.2.1.2	Análisis Multivariado.	43
3.3	Selección de variables.	44
3.3.1	Técnica de selección de variables <i>stepwise</i> .	44
3.3.2	Búsqueda exhaustiva.	45

3.3.3 Selección por el experto.	46
3.4 Formulación de los modelos.	46
3.4.1 Selección de las herramientas de modelado.	47
3.4.2 Construcción del modelo con una red neuronal.	47
3.4.3 Construcción de modelos neurodifusos ANFIS.	53
3.5 Validación de modelos.	59
3.6 Análisis de sensibilidad.	60
<b>4. RESULTADOS</b>	
4.1 Análisis exploratorio de los datos.	61
4.2 Selección de variables.	69
4.3 Modelo con una red neuronal.	72
4.3.1 Construcción del modelo.	75
4.3.2 Verificación y validación del modelo.	77
4.4 Modelo ANFIS	79
4.4.1 Construcción del modelo.	82
4.4.2 Elaboración del sistema de inferencia difuso.	83
4.4.3 Verificación y validación del modelo.	89
4.5 Validación de los modelos.	95
4.6 Análisis de sensibilidad.	97
<b>5. DISCUSIÓN</b>	99
<b>6. CONCLUSIONES</b>	103
<b>RECOMENDACIONES</b>	104
<b>APÉNDICES</b>	105
<b>BIBLIOGRAFÍA</b>	129

# ÍNDICE DE FIGURAS

<b><u>Figura</u></b>	<b><u>Página</u></b>
2.1. Fases de degradación a) biogás y b) lixiviados	14
2.2. Metodología para el desarrollo de un modelo	18
2.3. Esquema de una neurona artificial	27
2.4. Modelo de una red neuronal tipo <i>Feedforward</i>	29
2.5. Formas de las funciones de membresía	30
2.6. Arquitectura del sistema ANFIS	35
3.1. Metodología desarrollada en el diseño de los modelos	39
3.2. Sistema a escala y elementos que lo integran a) biogás, b)lixiviados	40
3.3. Distribución de los datos	47
3.4. Diagrama de flujo para la formulación de modelos mediante una RNA	48
3.5. Estructura de la matriz de datos para generar el modelo RNA	49
3.6. Ventana de entrenamiento de la RNA	51
3.7. Gráfico de regresión entre datos generados por el modelo y datos reales	52
3.8. Diagrama de flujo para la formulación de los modelos por ANFIS	53
3.9. Interfaz gráfica del editor ANFIS	57
3.10. Comparación datos de entrenamiento vs datos de validación	58
3.11. Estructura del modelo ANFIS	58
3.12. Reglas difusas del modelo ANFIS	59
3.13. Diagrama de flujo para análisis de sensibilidad	61
4.1. Diagramas de caja de las variables de lixiviados recirculados (LR)	63
4.2. Matriz de dispersión de variables LR	64
4.3. Gráfico de probabilidad normal	65
4.4. Diagramas de caja de variables de de lixiviados producidos (LP)	66

4.5	Matriz de dispersión de variables LP	67
4.6	Análisis de ANOVA para a) lixiviados producidos y b) CH <sub>4</sub>	68
4.7	Estructura de red modelo a) MBE_CH <sub>4</sub> y b) MSE_LP	76
4.8	Validación del modelo a) MBE_CH <sub>4</sub> y b) MSE_LP	76
4.9	Análisis de regresión del modelo a) MBE_CH <sub>4</sub> , b) MSE_LP	77
4.10	Validación de los modelos a) MBE_CH <sub>4</sub> y b) MSE_LP	78
4.11	Interfaz gráfica de MATLAB activada con la función <i>mfedit</i>	84
4.12	Funciones de membresía variables independientes MSE_CH <sub>4</sub>	86
4.13	Visor de reglas difusas del modelo MSE_CH <sub>4</sub>	87
4.14	Funciones de membresía variables independientes MSE_LP	88
4.15	Reglas difusas del modelo MSE_LP	88
4.16	Ventana del Editor ANFIS	90
4.17	Modelo MSE_CH <sub>4</sub> con 100 épocas de entrenamiento	91
4.18	Modelo MSE_CH <sub>4</sub> con 50 épocas de entrenamiento	91
4.19	Modelo MSE_CH <sub>4</sub> con 20 épocas de entrenamiento	92
4.20	Estructura del modelo ANFIS para los modelos MSE_CH <sub>4</sub>	92
4.21	Modelo MSE_LP con 20 épocas de entrenamiento	93
4.22	Modelo MSE_LP con 100 épocas de entrenamiento	94
4.23	Modelo MSE_LP con 40 épocas de entrenamiento	94
4.24	Modelo MSE_LP con 60 épocas de entrenamiento	94
4.25	Estructura del modelo ANFIS para los modelos MSE_LP	95
4.26	Validación del modelo MSE_CH <sub>4</sub> y MSE_LP	96
4.27	Distribución de los datos para el análisis de sensibilidad	97

## ÍNDICE DE TABLAS

<b><u>Tabla</u></b>		<b><u>Página</u></b>
3.1	Definición de variables	41
4.1	Resumen de estadísticos del análisis univariado LR	62
4.2	Matriz de correlación de variables LR	63
4.3	Análisis de ANOVA de variables LR	64
4.4	Resumen de estadísticos del análisis univariado LP	65
4.5	Matriz de correlación de variables de LP	67
4.6	Análisis de ANOVA para las variables de lixiviados producidos (LP)	68
4.7	Selección de variables para estimar CH <sub>4</sub> por método <i>stepwise</i>	69
4.8	Selección de variables para estimar LP por método <i>stepwise</i>	69
4.9	Evaluación de modelos para CH <sub>4</sub> con cuatro variables por BE	70
4.10	Evaluación de modelos para LP con cuatro variables por BE	71
4.11	Resumen de variables seleccionadas por SW, BE y SE	71
4.12	Evaluación del modelo MSW_CH <sub>4</sub>	72
4.13	Evaluación del modelo MBE_CH <sub>4</sub>	73
4.14	Evaluación del modelo MSE_CH <sub>4</sub>	73
4.15	Evaluación de los modelos de RNA para estimar CH <sub>4</sub>	73
4.16	Evaluación del modelo MSW_LP	74
4.17	Evaluación del modelo MBE_LP	74
4.18	Evaluación del modelo MSE_LP	75
4.19	Evaluación de los modelos para estimar LP	75
4.20	Evaluación modelo MSW_CH <sub>4</sub> para estimar metano	79
4.21	Evaluación modelo MBE_CH <sub>4</sub> para estimar metano	80
4.22	Evaluación del modelo MSE_CH <sub>4</sub> para estimar metano	80

4.23	Evaluación de modelos ANFIS para estimar CH <sub>4</sub>	81
4.24	Modelos MSW_LP desarrollado para estimar lixiviados producidos	81
4.25	Modelos MBE_LP desarrollado para estimar lixiviados producidos	81
4.26	Modelos MSE_LP desarrollado para estimar lixiviados producidos	82
4.27	Evaluación de modelos ANFIS para estimar LP	82
4.28	Reglas difusas del modelo MSE_CH <sub>4</sub>	87
4.29	Reglas difusas del modelo MSE_LP	88
4.30	Resultados obtenidos al evaluar los modelos para CH <sub>4</sub> y LP	96
4.31	Definición de variables en el análisis de sensibilidad	97
4.32	Resultado del análisis de sensibilidad	98

## ABREVIATURAS

<b>Símbolo</b>	<b>Descripción [Unidades]</b>
AA	Ácido Acético [mg/L]
AB	Ácido Butírico [mg/L]
AP	Ácido Propiónico [mg/L]
AOV	Ácidos Orgánicos Volátiles
CE	Conductividad Eléctrica [mS/cm]
DQO	Demanda Química de Oxígeno [g/L]
pH	Potencial de hidrógeno
SVT	Sólidos Volátiles Totales [mg/L]
ST	Sólidos Totales [mg/L]
CH <sub>4</sub>	Metano [umoles/kg]
CO <sub>2</sub>	Bióxido de carbono
NH <sub>3</sub> -N	Nitrógeno Amoniacal
SO <sub>4</sub> <sup>-2</sup>	Sulfatos

## Acrónimos

ANFIS	Adaptive Neural Fuzzy Inference Systems
BE	Búsqueda Exhaustiva
BLAB	Biodigestores a escala laboratorio
FIS	Fuzzy Inference System
INEGI	Instituto Nacional de Estadística y Geografía
LPGIR	Ley General para la Prevención y Gestión Integral de los Residuos
RNA	Red Neuronal Artificial
RESA	Relleno Sanitario
RSU	Residuos Sólidos Urbanos
SERMANAT	Secretaría de Medio Ambiente y Recursos Naturales
SE	Selección del Experto
SOFTCOMPUTING	Inteligencia computacional
SW	Stepwise

# 1. INTRODUCCIÓN

## 1.1 Antecedentes

Hoy en día, los residuos generados por la actividad humana se han convertido en un grave problema mundial. Al hablar del medio ambiente, es inevitable mencionar la generación de residuos y los daños que causan en la sociedad, tanto por su aspecto como por los productos de su descomposición. La necesidad por desprenderse de los residuos provenientes de cualquier actividad ha dado origen a la creación de múltiples acciones de saneamiento, una de estas soluciones son los sitios de disposición final o rellenos sanitarios. Estos deben cumplir con especificaciones de protección ambiental que van desde la selección del sitio, diseño, construcción y operación, hasta el monitoreo después de la clausura, tal como lo menciona la norma oficial Mexicana NOM-083-SEMARNAT-2003 (SERMARNAT, 2010).

Una vez depositados los Residuos Sólidos Urbanos (RSU), en el Relleno Sanitario (RESA), su impacto en el ambiente se debe a que la degradación de la materia orgánica (MO) contenida en los residuos, genera como productos lixiviados y biogás, ambos son potencialmente contaminantes en los rellenos sanitarios, debido a sus altas concentraciones de compuestos orgánicos e inorgánicos. Los primeros se originan por la humedad interna de los RSU y el agua infiltrada en el relleno, su generación y composición depende de factores propios de los RSU (composición, densidad, contenido de humedad y capacidad de campo); de factores climáticos y de la misma operación del RESA (Griffith and Trois, 2006; Kjeldsen, 1999). El biogás es resultado de la degradación de la MO de los RSU bajo condiciones anaeróbicas, contiene principalmente  $\text{CH}_4$  y  $\text{CO}_2$ . Las emisiones de  $\text{CH}_4$  desde el RESA es uno de los principales contribuyentes a los gases de efecto invernadero. Sin embargo, el  $\text{CH}_4$  puede recuperarse y aprovechar su poder calórico como carburante en automotores, o en calderas y turbinas para generar electricidad, siempre y cuando sea económicamente viable como ocurre en los grandes rellenos sanitarios (Nayla, 2005).

Para mitigar la contaminación al ambiente por la disposición de los RSU, resulta imprescindible conocer los procesos de descomposición y lixiviación que conllevan a la generación de biogás y lixiviados. En la última década se han desarrollado una serie de teorías dirigidas a explicar la degradación de los RSU y las dinámicas que presentan las características de sus lixiviados y biogás en diferentes condiciones dentro del RESA, a

pesar de que en ocasiones no pueden ser probadas por la heterogeneidad de sus componentes, estacionalidad, costos, seguridad, tiempo u otras (Kylefords, 2003; Lobo *et al.*, 2002a, 2002b). Es aquí donde la modelización ha permitido representar de manera simplificada la realidad bajo estudio, proponiendo una serie de modelos que integran distintos efectos y manifestaciones en el proceso de biodegradación de los RSU, así como en la estimación de sus productos (lixiviados y biogás).

Entre los trabajos revisados se encuentran aquellos dirigidos a estudiar el proceso de biodegradación de los RSU (El-Fadel *et al.*, 1997; Haarstrick *et al.*, 2001; Vavilin *et al.*, 2002; White *et al.*, 2004), conocer las características de los lixiviados y las variaciones que sufren tanto por su calidad como por la cantidad (Demirekler *et al.*, 1999) y aquellos empleados para estimar el flujo a partir de los elementos del sistema, sus interacciones y características del entorno (Papadopoulou *et al.*, 2007). Bajo este contexto, se encuentra el modelo desarrollado por Schroeder (1994), el cual opera principalmente con datos meteorológicos y características de los residuos, del suelo y diseño del RESA, excluye reacciones bioquímicas. En el caso de Safaria y Baronian (2002) su modelo considera patrones de precipitación, evaporación y capacidad de campo de los RSU, los valores son introducidos en función del tiempo y de las entradas/salidas del relleno, debido a la infiltración y evaporación de agua, por medio de un balance hídrico. Lobo *et al.*, (2002a) simula las reacciones biológicas de hidrólisis de sólidos y la gasificación del material biodegradable disuelto, estima los lixiviados producidos en cierto período, así como su contaminación orgánica y los gases generados en el proceso de biodegradación. En el trabajo de Vadillo y Carrasco (2005) estiman la cantidad de lixiviados generados en un relleno no-controlado, aplicando la ecuación de balance hídrico y la caracterización de los diferentes componentes que integran el RESA. Paralelamente, se han venido desarrollando trabajos sobre modelos para estimar biogás, entre estos se encuentra el trabajo de Gurijala *et al.*, (1997), que a través de un análisis de regresión múltiple, evalúan efectos simultáneos de 10 factores ambientales sobre la tasa de producción de  $\text{CH}_4$ , obteniendo como resultado, un modelo capaz de explicar el 95.85% de las variaciones en la producción de  $\text{CH}_4$ . Martin *et al.*, (2001) desarrollan un modelo de flujo de diferencias finitas en dos dimensiones, analizan las leyes que rigen el desplazamiento de biogás. Vavilin *et al.*, (2002) emplean ecuaciones diferenciales parciales y un esquema cinético simplificado de primer orden,

estudian el equilibrio entre la tasa de hidrólisis/acidogénesis y acetogénesis/metanogénesis durante la digestión anaerobia de los RSU. El modelo de la EPA (2005) emplea una ecuación de biodegradación de primer grado, a partir de las cantidades de residuos dispuestos acumulados a través del tiempo, estima las tasas de emisión de gases del relleno como: CH<sub>4</sub>, CO<sub>2</sub>, compuestos orgánicos y contaminantes en el aire.

De los trabajos anteriores, se observa que los modelos matemáticos han permitido analizar el comportamiento de sistemas tan complejos como un RESA, diseñar nuevos procesos y controlarlos. Con el fin de generar predicciones sobre el sistema a mayor o menor escala, a pesar de lo difícil que algunas veces puede llegar a ser su interpretación, debido a la cantidad de simplificaciones que tienen que hacerse con el propósito de representar los procesos del sistema. Aunado a lo anterior, es necesario conocer los principios físicos que reflejan las propiedades del sistema y la interpretación de dichos parámetros.

Al considerar que gran parte de los sistemas biológicos involucran serias dificultades al modelarse, dada la complejidad que algunas veces presentan en los diferentes niveles de abstracción, ha sido necesario utilizar una serie de metodologías y aproximaciones para poder comprender su verdadera existencia relacional. La importancia de la modelización no podría entenderse sin la presencia de medios y métodos computacionales, cada vez más potentes, que hacen posible abordar problemas que por su grado de dificultad hasta ahora era imposible resolver, no solo por la potencia de cálculo que poseen, sino también por la capacidad de representación formal de procesos y sistemas de gran complejidad, que no pueden ser descritos “fíelmente” por la matemática tradicional.

Por lo anteriormente expuesto, el empleo de herramientas de inteligencia computacional (*Softcomputing*) ha resultado ser exitoso, debido a que tienen un paralelo notable a la capacidad de la mente humana de razonar y de aprender en un ambiente de incertidumbre e imprecisión (Zadeh, 1996). Entre los paradigmas más relevantemente empleados, se encuentran la lógica difusa o borrosa (Fuzzy Logic), las redes neuronales artificiales y los algoritmos genéticos, o bien el empleo conjunto de dos de ellos, tal es el caso de los sistemas neurodifusos (combinación entre redes neuronales artificiales y sistemas difusos). De acuerdo a Jang (1995), las principales características del procesamiento del *Softcomputing* son: que utilizan la experticia humana, son modelos

computacionales inspirados en la biología, aplican nuevas técnicas de optimización, involucran cálculo numérico, tienen nuevos campos de aplicación, poseen un modelo de aprendizaje libre, son tolerantes a fallos y tienen aplicaciones en el mundo real.

La lógica difusa, es una metodología que aprovecha la experiencia humana junto con el lenguaje natural en un camino simple de aprendizaje, además de su capacidad para modelar complejas funciones no lineales. En este contexto, se encuentran trabajos como el de Abdallah *et al.*, (2009), quienes desarrollaron un sistema capaz de simular la generación de biogás en un biorreactor anaerobio, buscando controlar y manipular la influencia del contenido de humedad y nutrientes para acelerar la biodegradación de RSU. Garg *et al.*, (2006) estimaron el índice de generación de metano ( $k$ ), considerando el promedio anual de precipitación, promedio diario de temperatura, fracción biodegradable de residuos y profundidad del relleno. Polit *et al.*, (2002) construyeron un modelo dinámico difuso para el proceso de digestión anaerobia, con valores provenientes de la literatura. Obteniendo como salida, la tasa de flujo de los gases. Asignan un coeficiente difuso para tener en cuenta la influencia de la temperatura y pH, cuyos valores no están presentes explícitamente. El modelo fue probado con un conjunto de datos experimentales y con un modelo diferente, la comparación es buena y cumple con los objetivos del modelo.

Existen aplicaciones con modelos de redes neuronales en el área ambiental que tratan de modelar el conocimiento de forma conectiva y adaptativa, tal es el caso de Karaka y Ozcaya (2006), cuyo trabajo consistió en desarrollar un método para modelar el caudal de lixiviados en el relleno de Estambul/Odayeri con redes neuronales artificiales, en conjunción con un sistema para controlar la cantidad de lixiviados dentro del RESA. Con respecto a la estimación de biogás se tienen trabajos reportados como el de Ozcaya *et al.*, (2007), cuyo modelo predice la producción de CH<sub>4</sub> en el biogás procedente de dos biorreactores a escala campo, operados con y sin recirculación de lixiviados. Se consideran factores como la adición de agua y lixiviados además de la temperatura de los residuos en un periodo de 34 meses. En este estudio Ogwueleka y Ogwueleka, (2010), emplean una red neuronal para predecir el valor calorífico inferior de los residuos sólidos urbanos, a partir de plástico, papel, vidrio, textiles y alimentos, el modelo proporciona el mejor ajuste y la evolución prevista de los datos observados. Por su parte (Noori *et al.*, 2009) emplean redes neuronales para estimar la generación de residuos sólidos urbanos, comparan los resultados

del modelo de red contra un modelo de de regresión basado en análisis de componentes principales, determinan que el modelo de red es mejor con respecto al de regresión. Ozcan *et al.*, (2008) emplean algoritmos genéticos para estimar las concentraciones de CH<sub>4</sub>, CO<sub>2</sub> y O<sub>2</sub> a partir de datos reales del relleno, obtenidos durante 3.5 años con el fin de monitorear variaciones a corto y largo plazo. Con respecto al empleo de sistemas neurodifusos se tiene reportado el trabajo de Cakmakci (2007), el cual consistió en modelar un sistema de digestión anaerobia de lodos, en una planta tratadora de aguas residuales. Chan and Chan, (2006) utilizan un enfoque híbrido para construir un sistema de predicción del nivel del agua durante los períodos de inundaciones, para el manejo de los yacimientos. Además, verificaron que el modelo al incluir como variable de entrada la decisión humana, tiene un rendimiento superior con respecto a todos los índices utilizados en el modelo sin dicha entrada.

Como resultado de estos trabajos, se deduce que el modelado de un sistema biológico no es una tarea fácil, debido a la cantidad de procesos involucrados, sin embargo el empleo de herramientas de *Softcomputing* mostró una serie de características positivas en el desarrollo de modelos ambientales. En su caso, los modelos difusos presentaron gran capacidad para incorporar la incertidumbre en la entrada del modelo, permitieron simular perfectamente los datos experimentales, lo cual fue confirmado estadísticamente. Los parámetros del modelo pueden ser fácilmente adaptados en diferentes etapas del proceso, y solo se requiere un periodo corto de adaptación. Por lo anterior el empleo de la lógica difusa en la modelización del proceso de biodegradación de RSU se considera exitoso, debido a que es capaz de describir un gran número de complejos procesos físicos y bioquímicos que se producen en un relleno-biorreactor (Abdallah *et al.*, 2009; Polit *et al.*, 2002; Garg *et al.*, 2007).

Con respecto a las redes neuronales, resalta la viabilidad y versatilidad en la predicción del caudal de lixiviados y biogás en un área de relleno sanitario, su predicción es precisa y eficaz, además que permite un control efectivo y de gran alcance en la gestión del sistema (Karaca y Ozcaya, 2006; Ozkaya *et al.*, 2007). Los resultados obtenidos, señalan que el modelo de redes neuronales artificiales tiene más ventajas en comparación con métodos tradicionales (análisis multivariable), en la predicción de la generación de residuos

sólidos urbanos y en la predicción de contenido energético (Jalili and Noori, 2008; Noori, *et al.*, 2009; Ogwueleka and Ogwueleka, 2010).

Dada la estructura altamente no lineal del modelo neurodifuso, se demostró que ciertos sistemas complejos, como la biodegradación anaerobia, pueden ser fácilmente modelados, su aplicabilidad es sencilla y no es necesario definir complejas reacciones, ni ecuaciones químicas, además su fiabilidad puede verse incrementada al adicionar el conocimiento experto (Cakmakci, 2007; Chan y Chan, 2006).

## 1.2 Planteamiento del problema

Hoy en día, los modelos matemáticos ayudan a simular sistemas complejos en áreas disímiles tales como las ciencias ambientales, la física y la economía, entre otros, además estos modelos juegan un papel clave en la estimación y predicción (Demirekler *et al.*, 1999; El-Fadel *et al.*, 1997; Griffith and Trois 2006; Kiss y Encarnación, 2006; Maris 2006). Normalmente en el área ambiental estos modelos se fundamentan en técnicas tradicionales como regresión lineal, ecuaciones diferenciales o balances de masa, que no son del todo adecuadas cuando los sistemas presentan un comportamiento no-lineal, poseen cierto grado de incertidumbre ó no se dispone de más conocimiento sobre el sistema, como es común en los sistemas biológicos. Aunado a lo anterior, durante el proceso de modelado los parámetros se convierten en los principales obstáculos para obtener predicciones fiables, cuando los datos presentan una elevada variabilidad (Abebe *et al.*, 2000; Hanns, 2000; Scenna, 1999; Zacharof and Butle, 2004).

En la búsqueda de modelos genéricos, simples y aplicables, las herramientas de *Softcomputing* han demostrado ser de las más apropiadas, debido a que reproducen las características del ser humano, como su flexibilidad y la tolerancia a la incertidumbre, además de que facilitan la interpretación de los resultados (Jang 1993; Zadeh, 1996).

Si bien existen modelos ambientales con herramientas de *Softcomputing* (Abdallah *et al.*, 2009; Karaka and Ozcaya, 2006; Ozcan *et al.*, 2008; Ozcaya *et al.*, 2007; Jalili and Noori, 2008; Noori, *et al.*, 2009; Ogwueleka and Ogwueleka, 2010), a la fecha son escasos los trabajos donde se ha empleado más de una técnica en la resolución de un problema en particular (Abu Qdais *et al.*, 2009), en el área de los residuos sólidos se reduce más aún el

número de aplicaciones, que de manera conjunta empleen técnicas de *Softcomputig* en la estimación de la producción de CH<sub>4</sub> y lixiviados. Por lo que resulta de especial interés extender su estudio y verificar con que técnica es posible generar modelos con mayor bondad de ajuste.

### **1.3 Justificación**

Los modelos han demostrado ser una herramienta primordial tanto en el diseño de los experimentos como en la evaluación de los resultados, debido a que brindan soluciones sencillas y de bajo costo. Son el medio a través del cual, es posible reproducir la realidad bajo estudio, para después; obtener valiosa información tanto del funcionamiento del sistema como de la previsible evolución de éste y de los efectos que tienen determinadas acciones. Actualmente existe una variedad de técnicas y paradigmas, donde es posible abordar el desarrollo de modelos predictivos (Kleinbaum *et al.*, 1988; Safari and Baronian, 2002). Bajo este contexto una serie de trabajos han sido desarrollados, cuyo objetivo ha sido establecer relaciones causa-efecto entre una variable independiente (causa) y una variable dependiente (efecto); permitiendo determinar cuantitativamente las propiedades del objeto bajo estudio (Del Real Olvera *et al.*, 2009; Gurijala *et al.*, 1997).

Cuando se tiene un problema de modelado concreto, con un reducido conocimiento sobre los procesos y leyes que gobiernan el sistema, en principio no es sencillo establecer que técnica es más conveniente usar, ni que criterios pueden ayudar a resolver ese problema, lo que supone una limitante importante (Cakmakci, 2007; Garg, 2007). Aún más, para tratar el problema en todas sus vertientes y que sea operable en ambientes de incertidumbre, se precisa de una guía que combine conocimientos, técnicas y metodologías de varias fuentes. Es por ello, que este trabajo pretende establecer una metodología simple para el desarrollo de modelos y la selección de herramientas de *Softcomputing* en la estimación de CH<sub>4</sub> y lixiviados en residuos sólidos urbanos, así como ser un aporte en la solución de problemas prácticos, beneficiando a especialistas e investigadores del área ambiental y a usuarios que deseen obtener los mayores beneficios de sus datos.

## 1.4 Hipótesis

Al comparar distintas herramientas de *Softcomputing* en el desarrollo de modelos en el área de residuos, es posible emplear datos experimentales para estimar la generación de CH<sub>4</sub> y lixiviados con mayor precisión y bondad de ajuste.

## 1.5 Objetivos

### Objetivo general

Determinar qué herramienta de *Softcomputing* es idónea en la construcción de modelos con un elevado nivel de precisión, en la estimación de CH<sub>4</sub> y lixiviados, a partir de datos experimentales.

### Objetivos particulares

- Describir que características presentaron mayores ventajas y desventajas de las herramientas de *Softcomputing* seleccionadas.
- Brindar una serie de pautas que pudieran facilitar la selección de una herramienta de *Softcomputing* en el modelado de los datos.

## **2. MARCO TEÓRICO**

En este capítulo, se describe el marco teórico que sustenta éste trabajo de tesis doctoral, consta de tres secciones. En la primera, se analiza la problemática ambiental en la que contribuyen los residuos sólidos urbanos (RSU) y la creación de sitios de disposición como medio de solución. En la segunda sección, se presenta una visión general de los procesos que se gestan dentro del relleno sanitario (RESA), en especial el proceso de biodegradación y sus productos, explicando la importancia que tiene su cuantificación para el medio ambiente. Por último, se presentan los principios y teorías del modelado, además de las técnicas de *Softcomputing* como una herramienta de solución en el modelado de sistemas ambientales, se hace una breve descripción de sus ventajas.

### **2.1 Una problemática ambiental**

En México, al igual que en muchos otros países del mundo, se enfrentan grandes retos en el manejo de los residuos, debido en gran parte, al elevado índice de crecimiento tanto demográfico como industrial, a las costumbres orientadas al consumo de artículos desechables, los que al tener un tiempo de vida más largo, hacen que su proceso de degradación e integración a la naturaleza sea más complejo. Los RSU de acuerdo a la ley general para la prevención y gestión integral de los residuos, se definen como el material o producto cuyo poseedor desecha, que se encuentra en un estado sólido, líquido o gaseoso contenido en recipientes y que puede ser susceptible de ser valorizado o requiere sujetarse a tratamiento o disposición final (LGPGIR, 2007), su inadecuada disposición provoca contaminación visual, enfermedades y pérdida de recursos naturales, generalmente se componen de residuos como papel, cartón, plásticos, metales, materia orgánica, vidrio, entre otros (SEMARNAT, 2010b).

Debido a las características socioeconómicas del país, el método más adecuado para disponer de los RSU ha sido el RESA, la norma oficial Mexicana NOM-083-SEMARNAT-2003 lo define como una obra de infraestructura que involucra métodos y obras de ingeniería para la disposición final de RSU con el fin de controlar a través de la compactación e infraestructura adicional, los impactos ambientales (SEMARNAT, 2010a).

Actualmente en México, solo el 60.74% de los RSU generados son depositados en un RESA, un 9.62% se dispone en rellenos de tierra controlado y alrededor del 29.64% sigue depositándose en los tiraderos a cielo abierto (INEGI, 2010). Entre los efectos que producen estas acciones se encuentran los riesgos sanitarios, debido a la presencia de roedores, insectos y otros agentes patógenos; la contaminación de las fuentes de abastecimiento de agua subterránea y superficial por la infiltración y escurrimientos de los lixiviados; y el deterioro del entorno particularmente por la gran cantidad de plásticos y desechos artificiales no biodegradables. Finalmente y no menos delicado es el incremento en las emisiones de gases de efecto invernadero, provocado por la liberación de CH<sub>4</sub> y CO<sub>2</sub> a la atmósfera, como resultado de los procesos de degradación de la fracción orgánica en los RSU. México contribuye con alrededor de 1.6 por ciento de las emisiones mundiales de gases de efecto Invernadero, causantes del cambio climático, lo que ubica a la nación en la posición 13 de los países emisores (Sheinbaum y Masera, 2000).

## **2.2 El proceso de biodegradación y sus productos**

La disposición de los RSU dentro del RESA da lugar a la generación de diferentes productos contaminantes, derivados de los procesos de descomposición de los RSU, los cuales puede presentarse en forma sólida como polvo, líquida en lixiviados y gaseosa como biogás (Kiss and Encarnación, 2006). Conocer los procesos de degradación que se realizan en el interior del RESA es de suma importancia, debido a que permiten alcanzar una mayor comprensión de los mecanismos bajo los cuales se producen los lixiviados y el biogás, y con ello, poder evitar posibles riesgos de contaminación ambiental.

### **2.2.1 Factores que influyen en el proceso de biodegradación**

Un RESA puede ser considerado como un enorme reactor bioquímico donde están presentes una gran variedad de reacciones químicas y biológicas, tal como las reacciones de óxido-reducción que afectan la disolución de metales, la disolución de los productos de conversión biológica y otros compuestos principalmente orgánicos en el lixiviado, las reacciones biológicas más importantes que ocurren en el RESA son realizadas por los microorganismos aerobios y anaerobios, y están asociadas con la fracción orgánica contenida en los residuos sometidas a ciertas condiciones de presión y temperatura. Entre

los principales componentes del RESA, se encuentran los RSU depositados, el oxígeno existente en la capa superior de los residuos el agua tanto de los residuos como la que se infiltra y los microorganismos contenidos en los RSU que actúan como catalizadores en las reacciones, y que ocurren en función de factores tales como: la naturaleza de los reactivos depositados, la presencia de oxígeno libre, de sulfatos, de inhibidores, la humedad interior, el pH, la presión parcial de hidrogeno y algunos factores microbiológicos (Orozco *et al.*, 2003; Robles-Martínez, 2008; Tchobanoglous *et al.*, 1993).

La concentración de oxígeno libre ( $O_2$ ) presente en la fase gaseosa al interior de la masa de los residuos, es un factor que determina el tipo de vía metabólica a seguir, es decir si el proceso de biodegradación se desarrollará en forma aerobia o anaerobia. Otro de los factores más importantes para la generación de biogás en un RESA es la humedad, debido a que es necesaria para la actividad microbiana, además de ser un medio útil en el transporte de nutrientes y microorganismos a través de las diferentes capas de residuos en los sitios de disposición final. El pH es un parámetro clave para el control del metabolismo microbiano y optimización del proceso de digestión. En la evolución de los procesos bioquímicos el pH se modifica a lo largo de la vida activa del RESA y sus variaciones pueden tener un gran impacto en el proceso, hasta el punto de inhibir totalmente la actividad microbiana (Pohland and Kim, 1999; White *et al.*, 2004).

Por su parte los ácidos orgánicos volátiles (AOV) representan la mayor parte de la carga orgánica de los lixiviados. Al menos durante las fases de acidogénesis y acetogénesis estos pueden llegar a representar hasta un 90% de la carga orgánica. Se ha demostrado que bajo condiciones desfavorables, los AOV que se acumulan más fácilmente son los ácidos acético, propiónico y butírico, lo anterior demuestra indirectamente que los microorganismos responsables de la degradación de estos compuestos son los más sensibles a los cambios del medio, y su inhibición es la responsable de las disfunciones en el proceso de biodegradación anaerobia de los RSU, su determinación es fundamental para controlar el proceso de digestión anaerobia (Robles-Martínez, 2008). Es posible cuantificar de manera aproximada el contenido total de materia orgánica presente en los lixiviados mediante parámetros como la demanda química de oxígeno (DQO), los sólidos totales (ST) y los sólidos volátiles totales (SVT). La DQO es una medida del  $O_2$  expresado en mg/l, equivalente a la cantidad de materia orgánica contenida en una muestra, que es susceptible

de oxidarse en presencia de un oxidante químico fuerte. Desde el punto de vista del proceso de biodegradación, el cambio de la concentración de la DQO es el más importante indicador de estabilización (Ozkaya *et al.*, 2007). La definición generalizada de ST, se refiere a la suma de los sólidos suspendidos totales, sales disueltas y materia orgánica que queda como residuo después de su evaporación entre los 103°C a 105°C. Esta medida permite conocer el contenido total de sustancias no volátiles presentes en los lixiviados. Los SVT representan la porción de materia orgánica e inorgánica que puede volatilizarse por el efecto de la calcinación a 550°C. Existen otros parámetros como la conductividad, sulfatos y nitrógeno amoniacal en los lixiviados, que permiten conocer la mineralización y la toxicidad que pudiera haber en la matriz de residuos. La conductividad electrolítica es una expresión numérica de la capacidad de una solución para transportar una corriente eléctrica, la cual depende de la presencia de iones, de su concentración total, de su movilidad, valencia y concentraciones relativas, así como de la temperatura. Los sulfatos son la forma más común del azufre en los RESA, es un compuesto que se desplaza con facilidad, está caracterizado por mal olor, y puede ser de gran ayuda para monitorear la migración de lixiviados (SEMARNAT, 2010b). El contenido de nitrógeno amoniacal (NH<sub>3</sub>-N) en lixiviados, deriva del contenido de proteínas de los residuos, la concentración depende de la tasa de solubilidad y/o lixiviación de los RSU. En pruebas para medir la toxicidad de los lixiviados, se ha observado que el NH<sub>3</sub>-N contribuye significativamente a la naturaleza tóxica de éstos (Kjeldsen *et al.*, 2002)

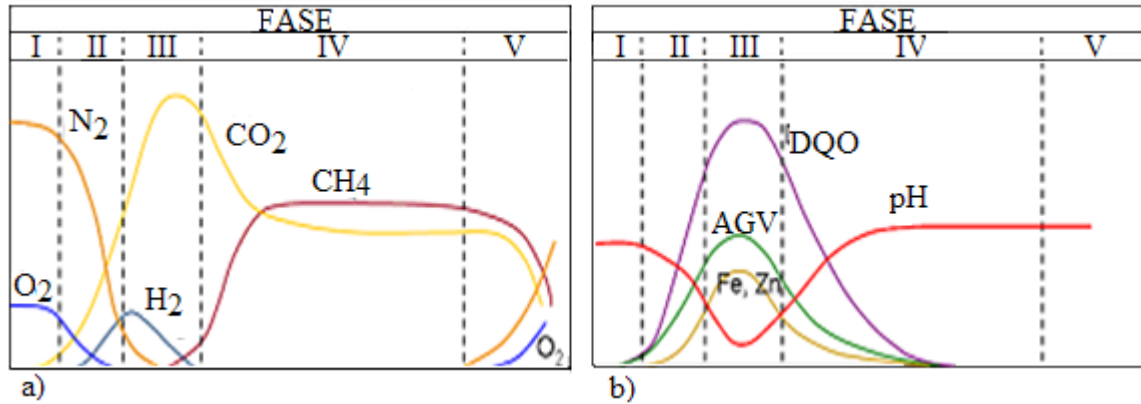
### **2.2.2 Proceso de biodegradación**

La biodegradación es el proceso por el cual una sustancia es degradada por organismos vivos a fragmentos más pequeños. Para que se produzca se requiere de condiciones ambientales, de una población adecuada de microorganismos, de la distribución y disponibilidad de nutrientes, y del contenido de humedad en los RSU.

En la biodegradación se distinguen los procesos aeróbicos y los anaeróbicos. En los primeros se aprovecha la presencia del oxígeno que se encuentra en los espacios vacíos en las capas superiores que están en mayor contacto con el aire; mientras que los procesos anaeróbicos se definen como la acción de microorganismos en ausencia de oxígeno, que estabilizan materiales orgánicos mediante su conversión a CO<sub>2</sub>, CH<sub>4</sub> y otros productos

orgánicos. Cuando la altura de las capas de residuos se incrementa, la transformación por vía anaerobia prevalece. El proceso de biodegradación anaerobia se divide en un conjunto de fases de acuerdo a la actividad metabólica predominante y sus productos finales (Tchobanoglous *et al.*, 1993). Estas se detallan a continuación (Figura 2.1).

- Fase inicial de ajuste (I). Comienza la descomposición biológica de la MO biodegradable en condiciones aerobias. En esta fase se observa una gran concentración de  $N_2$  y una ligera disminución de  $O_2$ .
- Fase de transición (II). Con el decremento de  $O_2$ , comienzan a desarrollarse condiciones anaerobias en el RESA, hay un aumento considerable de  $CO_2$ , un descenso del  $N_2$  y la aparición de  $H_2$ .
- Fase ácida (III). Se acelera la actividad microbiana anaerobia iniciada en la fase anterior, produciendo cantidades importantes de ácidos orgánicos volátiles (AOV). En los lixiviados se aprecia el descenso del pH y el incremento de metales y la demanda química de oxígeno (DQO). Los compuestos de alto peso molecular (lípidos, polisacáridos, ácidos nucleicos y proteínas) son transformados mediante enzimas en compuestos más sencillos y de peso molecular bajo, que son utilizados por los microorganismos como fuentes de energía y carbono celular. El biogás producido en esta etapa está constituido por  $CO_2$ , se reduce el porcentaje de  $H_2$  mientras comienza a aparecer  $CH_4$  en su composición.
- Fase de fermentación metanogénica (IV). El ácido acético obtenido en la fase ácida se transforma en  $CH_4$  y en  $CO_2$  mediante microorganismos metanogénicos, que actúan en condiciones anaerobias estrictas, por lo que el gas generado está constituido principalmente por estos compuestos. En los lixiviados se observa el descenso de los AOV, la DQO y metales, así como el incremento del pH, el cual se estabiliza entre valores de siete a ocho.
- Fase de maduración (V). En esta etapa la velocidad de generación de  $CH_4$  disminuye significativamente, debido a que los componentes que quedan en el RESA son de degradación lenta. Los principales gases que se encuentran en esta fase son el  $CH_4$  y  $CO_2$  procedentes de la fase anterior, y debido a la inclusión de aire también se tienen  $N_2$ ,  $O_2$ . Los lixiviados se estabilizan con valores mínimos de AOV, DQO y metales, y mantienen su pH entre siete y ocho.



**Figura 2.1** Fases de degradación a) biogás y b) lixiviados

Fuente: Tchobanoglous *et al.*, 1993

La digestión anaerobia se sugiere como una alternativa de eliminación de altas concentraciones de residuos orgánicos. En México se reporta que los residuos orgánicos ocupan el primer lugar con un 52.4% del volumen total (INEGI, 2010), por lo que el empleo de esta alternativa parece ser prometedora para el tratamiento de este tipo de residuos, además de que contribuye a disminuir la generación de gases de efecto invernadero, a elevar la producción de biogás, el cual puede emplearse como una fuente no renovable de energía.

### 2.2.3 Lixiviados

El lixiviado se forma con los líquidos que entran al relleno desde el exterior (precipitación, aguas subterráneas, drenajes) y como resultado de la descomposición de la fracción orgánica de los residuos. La composición del lixiviado en un relleno es de gran interés, debido a que proporciona información sobre la actividad química y biológica que se está llevando a cabo en el RESA (Bilgili *et al.*, 2006). Esta composición es variable y depende de factores tales como la naturaleza del residuo, edad del RESA, tecnología del relleno incluido, el grado de compactación (alta, media o baja), frecuencia y forma del material de cobertura y finalmente otros factores como el clima, estación del año, hidrología del lugar, tratamiento previo de los residuos y forma de explotación (recirculación de lixiviados, altura y tipo de relleno). Los compuestos orgánicos presentes en los lixiviados son: Proteínas, carbohidratos, compuestos hidroxiaromáticos, alcoholes y principalmente los ácidos AOV; adicionalmente, los lixiviados contienen gran cantidad de  $NH_3-N$ .

Para evitar que los lixiviados entren en contacto con el agua subterránea y la contaminen, estos son captados y posteriormente tratados; ya sea recirculándolos dentro del mismo RESA, por métodos fisicoquímicos ó mediante una planta depuradora de aguas residuales (Méndez *et al.*, 2004).

#### **2.2.4 Biogás**

En el proceso de digestión anaerobia, se realiza la descomposición de la materia orgánica por las bacterias en ausencia de aire. Durante la digestión los materiales de descomposición pasan por varios procesos: licuefacción, gasificación y mineralización, obteniéndose un producto final inerte con liberación de gas. La digestión pasa por diferentes fases, siendo las principales la fermentación ácida y la fermentación alcalina, de donde resulta la importancia del pH en el control de estas fases. En los RESA el tiempo de la metanogénesis puede durar más de 30 años en su fase productiva, la mayor parte de la producción de gas se efectúa durante los 20 años posteriores a la clausura del relleno, aunque la producción es más intensa durante los primeros cinco años (EPA, 2005). La producción de gas disminuye gradualmente aunque puede continuar durante muchos años. Los principales factores que influyen sobre la producción de biogás en el tiempo son: cantidad, calidad y edad de los residuos (Serrano, 2006).

### **2.3 Modelado de sistemas ambientales**

Hoy en día, es común hablar de modelos, de las diferentes técnicas y tipos que existen, y particularmente de los modelos que se emplean en el área ambiental. Existen prácticamente dos caminos en el cual un modelo puede ser construido; este puede ser deducido empleando las leyes de la naturaleza, llamado modelo mecanicista o puede ser inferido desde una colección de datos resultado de experimentar con un sistema real, llamado modelo de caja negra.

#### **2.3.1 Sistema**

El enfoque de sistemas establece que "*el mundo y cualquiera de sus partes, puede visualizarse como un conjunto de sistemas en interacción dinámica*". Por sistema; se entiende una colección de elementos relacionados cada uno caracterizado por atributos o

propiedades que pueden estar relacionados entre sí. Los objetivos que se persiguen al estudiar uno o varios fenómenos en función de un sistema, son aprender cómo cambian los estados, predecir el cambio y controlarlo.

Si es un sistema de la realidad el que se pretende modelar, los modelos suelen clasificarse en: Modelos estáticos, si omiten ya sea un reconocimiento del tiempo o describen un instante del estado de un sistema en determinado momento. En contraste, un modelo dinámico reconoce explícitamente el transcurso del tiempo, además proporciona una secuencia de instantes del sistema en el tiempo. Son modelos determinísticos, si todas las entidades establecen relaciones matemáticas o lógicas constantes, en cambio son estocásticos, si por lo menos una parte de la variación tienen una naturaleza casual o aleatoria (Naylor *et al.*, 1982). De acuerdo a la cantidad de información disponible acerca del sistema, los modelos se clasifican en: Modelos de caja-negra, si no existe información previa disponible o modelos de caja-blanca (o caja de vidrio) si toda la información necesaria se encuentra disponible. En general, los modelos típicos se encuentran en una posición intermedia entre ambos tipos, y siempre es preferible usar la mayor cantidad de información posible para hacer el modelo más exacto. En un modelo de caja-blanca se pretende conocer las relaciones entre las variables, cuyos valores son conocidos y por ello son considerados más fáciles; mientras que en un modelo de caja negra, tanto las relaciones entre las variables como los valores de los parámetros pretenden ser determinados (Polit *et al.*, 2002).

### **2.3.2 Modelos ambientales**

Los modelos ambientales permiten determinar las concentraciones relativas de un contaminante en el agua, aire, suelo, biota, etc. y cómo los procesos de degradación pueden llegar a controlar su comportamiento. Además, ayudan a identificar el impacto ambiental de las actividades humanas sobre los recursos y, a planificar las mismas o las acciones de mitigación o adaptación o recuperación de los ecosistemas. De acuerdo a la extensión de las aproximaciones hechas se clasifican en fenomenológicos o empíricos, en los primeros cada uno de los términos se basa en leyes fundamentales de la química y la física. Los empíricos en cambio, utilizan las mediciones experimentales de X, P, Y y Z para obtener expresiones analíticas que replacen a la función  $f$  desconocida. En un nivel intermedio se

encuentran los modelos cuasi-empíricos que aprovechan las ventajas de ambos. Desde el punto de vista práctico han demostrado tener mejores resultados los modelos ambientales semi-empíricos estocásticos y dinámicos. Debido a que permiten la utilización de información de campo y/o laboratorio, con lo cual se puede lograr un mejor ajuste entre la teoría, las experiencias en el laboratorio y lo que realmente sucede “*in situ*” (Maris, 1996).

### 2.3.3 Construcción del Modelo

Los modelos son una parte importante del método científico. Como construcciones deductivas esenciales en la teoría de la observación, permiten probar y alcanzar un mayor entendimiento del sistema bajo estudio. El ideal del científico que construye un modelo, es que llegue a ser tan exacto como le sea posible a la realidad que aspira representar (Baird, 1991). De acuerdo al ámbito en el que son empleados los modelos pueden ser:

- Descriptivos, si intentan encontrar relaciones y patrones de comportamiento en el conjunto de datos, para ofrecer conocimiento sobre un problema concreto.
- Predictivos, si el objetivo que persiguen es la precisión, cuantificada por medio del error cuadrático medio (ECM). Es el promedio de los errores de predicción al cuadrado y es un criterio utilizado ampliamente en la literatura para evaluar modelos predictivos. En este sentido, un modelo predictivo será mejor en la medida que tenga un menor ECM, ya que representaría una menor desviación, en términos de magnitud del resultado predicho con respecto al valor observado (Ec. 2.1).

$$ECM = \frac{1}{N} \sum_{i=1}^N (P_p - P_o)^2 \quad (2.1)$$

Donde:

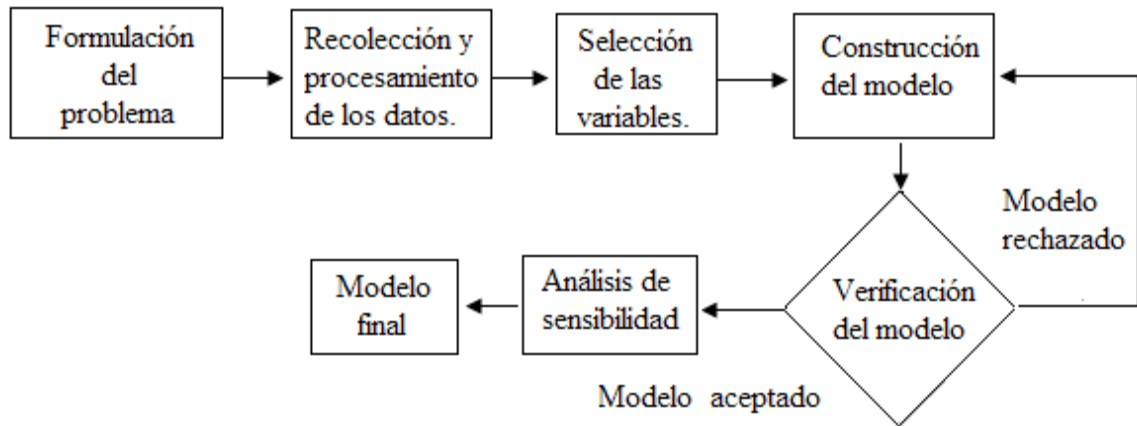
$P_p$ =Valor pronosticado por el modelo.

$P_o$ = Valor observado o medido.

$N$ =Número total de datos.

Es importante aclarar la diferencia entre precisión y exactitud en los modelos. La *precisión* hace referencia a la dispersión del conjunto de valores obtenidos de mediciones repetidas de una magnitud, cuanto menor es la dispersión mayor la precisión. *Exactitud* se refiere a cuan cerca del valor real se encuentra el valor medido, se expresa mediante el error

absoluto. La calidad de un modelo predictivo se evalúa a través de características tales como su simplicidad, interpretabilidad, poder discriminatorio.



**Figura 2.2** Metodología para el desarrollo de un modelo

Naylor *et al.*, (1982) sugiere una serie de pasos en el desarrollo de un modelo en el que pueden realizarse experimentos, para fines de este trabajo se han adecuado seis de los pasos para planear experimentos de simulación (Figura 2.2).

### 2.3.3.1 Formulación del problema

Es el primer paso en el desarrollo del modelo, es donde se establecen claramente los objetivos de la investigación o experimento mediante la especificación de criterios. Esto incluye determinar los objetivos apropiados, las restricciones sobre lo que se puede hacer, las interrelaciones del sistema bajo estudio con otros sistemas y el establecimiento de restricciones. Es un proceso secuencial que generalmente requiere una reformulación continua y progresiva y un refinamiento de los objetivos de experimentación durante su realización.

### 2.3.3.2 Recolección y procesamiento de los datos

Consiste en la recopilación inicial de los datos, integra información tanto descriptiva como cuantitativa. Tiene como propósito la adquisición de información relevante sobre el sistema que se pretende modelar (Beato, 2008).

A menudo los datos recopilados presentan ruido o un bajo nivel de precisión, causado muchas veces porque los valores de estos parámetros fueron sometidos a cierto grado de incertidumbre, o a la falta de puntos de medición, calibración, juicio del experto, o a la falta de información. Durante el proceso de depuración, pueden ser implementadas una serie de técnicas, como el análisis selectivo, que consiste en eliminar datos atípicos o inconsistentes. Las transformaciones consisten principalmente en modificaciones sintácticas llevadas a cabo sobre los datos, sin que supongan un cambio en el significado de los mismos.

Entre las pruebas que se realizan para aceptar o rechazar un dato anómalo se encuentra el Test Q de Dixon, para su aplicación se requiere ordenar los datos en forma creciente, calcular el rango de los resultados, encontrar la diferencia entre el dato sospechoso y su vecino más cercano. Dividir la diferencia obtenida en el paso anterior entre el rango para obtener el coeficiente de eliminación Q. Posteriormente se consulta una tabla de valores Q. Si el valor calculado de Q es mayor que el valor de la tabla; el resultado se puede descartar con un 90% de confianza de que en realidad estuvo sujeto a algún factor que no actuó sobre los otros resultados (Dean y Dixon, 1951). Técnicamente un dato no debe retirarse a menos que se encuentre una causa probable para justificar esta acción. El criterio que se siga es de vital relevancia con respecto a los objetivos que se persigan en la modelación y a limitaciones técnicas. En la práctica se ha encontrado que la limpieza y preparación de los datos toma aproximadamente el 80% del total del esfuerzo de ingeniería, sin embargo, se ve justificada por los resultados.

Otro de los problemas frecuentes durante la depuración, son los datos faltantes, debido a que dificulta la estimación en los modelos. La interpolación es una herramienta importante a la hora de estimar valores intermedios entre valores conocidos. El método comúnmente empleado para este propósito es la interpolación polinomial. Un polinomio de interpolación consiste en determinar el único polinomio de n-ésimo orden que se ajusta a los (n+1) puntos dados. La fórmula más simple de interpolación, es la de conectar dos puntos con una línea recta, este método se conoce como interpolación lineal (Ec. 2.2).

$$f(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0} (x - x_0) \quad (2.2)$$

El análisis exploratorio de los datos, se realiza con el fin de contrastar determinadas hipótesis, modelos o situaciones más complicadas, donde se requiere conocer el tipo de relación existente entre un determinado número de variables (lineal o no lineal) o cuando no existen expectativas claras sobre la naturaleza de estas relaciones. Su propósito es examinar las propiedades más relevantes de los datos adquiridos y reportar los resultados observados. De este modo es posible obtener una estructura explicativa de los datos a través de técnicas que combinan herramientas de la estadística.

El empleo de estadísticos suele estar clasificado de acuerdo al análisis que se realice sobre los datos. Por ejemplo, en el análisis univariado se estudian de manera separada cada una de las variables. Las técnicas más frecuentes en este análisis son la distribución de frecuencias, medidas de tendencia central (media, mediana), de dispersión o variabilidad (desviación estándar, varianza), distribución (curtosis, simetría). En cambio, con el análisis multivariado, las observaciones se ubican dentro de grupos. Su objetivo es examinar la interdependencia de las variables, que abarca desde la independencia total hasta la colinealidad presente, cuando una variable independiente es combinación lineal de algunas de las otras variables independientes. Cuando se duda acerca de su existencia, se emplea una regla empírica que se conoce como factor de inflación de la varianza ( $FIV=1/1-R^2$ ), el cual considera que existen problemas de colinealidad si al calcular el *FIV* se obtiene un valor superior a 10 (Kleinbaum *et al.*, 1988).

Los métodos de análisis multivariado pueden clasificarse de acuerdo a su función en dos tipos:

- Explicativos, si son empleados en contextos de investigación experimental. Consideran a las variables independientes como causas potenciales de las variables dependientes, su objetivo es contrastar hipótesis relativas. Entre las técnicas que pertenecen a esta clasificación se encuentra, el análisis de regresión múltiple, análisis discriminante, análisis multivariado de varianza y covarianza.
- Descriptivos, se aplican en contextos no experimentales, su objetivo es resumir el conjunto de datos conforme a ciertos objetivos establecidos o para detectar la interdependencia entre variables. Entre este tipo de técnicas se encuentra el análisis de datos categóricos, correlación canónica, análisis de clúster, análisis factorial y de componentes principales.

### 2.3.3.3 Selección de variables

La identificación de las variables significativas es una de las tareas más importantes en el modelado y por lo general lo realiza el experto de forma intuitiva, pero cuando se manejan un gran número de variables con cierto nivel de complejidad, conviene recurrir a técnicas matemáticas que aseguren que la selección de éstas, sea la óptima y facilite al mismo tiempo la detección de problemas de multicolinealidad y la eliminación de variables irrelevantes y/o redundantes para un problema dado. Se considera que una variable es irrelevante, cuando el conocimiento del valor de la misma no aporta nada al conjunto original de variables, por otra parte una variable es redundante si su valor puede ser determinado a partir de otras variables (Gestal, 2009). Debe considerarse durante la fase de formulación del modelo, el principio de parsimonia, cuya idea principal menciona, que si existen dos modelos que explican suficientemente bien un conjunto específico de datos, se debe de escoger el modelo más simple de los dos (Gualdrón, 2006).

Existen varias técnicas para seleccionar variables, las cuales se enmarcan bajo dos enfoques. El primero es el análisis elemental, es una técnica estadística que selecciona variables de entrada según su capacidad de explicación de la variable de salida. El segundo enfoque, se refiere al análisis de correlación múltiple, empleado para medir la asociación lineal presente entre las variables. Bajo este enfoque, se emplean algoritmo de búsquedas secuenciales (deterministas) que siguen estrategias que reducen el número de estados a analizar mediante la aplicación de técnicas de búsqueda local. Los métodos más comúnmente usados, son la selección hacia adelante “*Forward*”, que consiste en elegir las variables más representativas del conjunto de variables, hasta llegar a un máximo predeterminado o una condición a cumplir. La selección hacia atrás “*Backward*”, se trata de un algoritmo que selecciona una variable para su eliminación en cada paso, parte de la totalidad del grupo de variables y se van eliminando las menos representativas, y la selección paso a paso “*Stepwise*” (SW) es una combinación de las anteriores y se pueden ejecutar cuantas veces sea necesario. Esto resulta efectivo en el caso de archivos de datos muy complejos en los que existen grupos de variables que dan la solución correcta. El poder eliminar una variable elegida anteriormente, permite re-direccionar el camino de selección. Al repetir este método varias veces, es posible alcanzar una solución óptima, la cual se mide a un determinado nivel de significancia, generalmente designado por  $\alpha$  (alfa),

representa hasta qué punto en términos de probabilidad, se está dispuesto a aceptar un error de tipo I, es decir, a rechazar una hipótesis nula ( $H_0$ ) que sea verdadera. Esto conduce a la aceptación de  $H_0$  como verdadera con una probabilidad  $(1-\alpha)$ . Si se fija  $\alpha=0.05$  o  $p\text{-value}=0.05$  se está preparado a aceptar el error tipo I aproximadamente una de cada 20 veces.

También son empleados como criterios de selección los contrastes t y F. Para valores de  $t > 2$  y  $p\text{-valor} < 0.5$  la variable es significativa, para  $t < 2$  y  $p\text{-valor} > 0.5$  la variable no aporta información y se descarta. Como criterio de evaluación de la bondad de ajuste, se considera el coeficiente de determinación  $R^2$ , el cual proporciona la cantidad de variación total en la variable dependiente explicada por las variables independientes.

Otro método de selección es la búsqueda exhaustiva (BE), que consiste en seleccionar sistemáticamente todas las posibles combinaciones de variables a ser evaluadas y verificar cual es el mejor subconjunto. Este método toma la hipótesis de que un modelo dado con el menor error cuadrático medio (ECM) después de la primera iteración del entrenamiento, tiene una mayor capacidad de estimación aunque el resto de modelos alcance un ECM menor después de varias iteraciones. Esta forma de búsqueda solo puede utilizarse con un número de variables disponibles reducido hasta 4 variables después de realizar  $C(m, n)$  consultas, mediante el cálculo de una serie de combinaciones  $C = [(m!) / (m-n!) * (n!)]$ , en la que no existe posibilidad de repetición de variables.

Por último, el conocimiento del experto (SE) puede aplicarse en la selección de un conjunto de variables, donde es pertinente reflexionar sobre la complejidad del problema en términos de tiempo y recursos. Los expertos hacen sus estimaciones de coste basándose en su experiencia previa.

#### **2.3.3.4 Construcción del modelo**

Una vez definido el problema, la siguiente etapa consiste en reformularlo para su análisis, mediante la construcción de un modelo que represente la esencia del sistema. Para construir un modelo es necesario primero definir las variables en función de las cuales será establecido. Un modelo siempre debe ser menos complejo que el sistema real, es una aproximación abstracta de la realidad con consideraciones y simplificaciones que hacen más

manejable el sistema y permiten evaluar eficientemente las alternativas de solución. En esta etapa es donde son empleadas las herramientas seleccionadas para desarrollar el modelo.

### 2.3.3.5 Verificación del modelo

La validación del modelo tiene dos facetas; la validación interna se conoce como verificación y la validación externa se le denomina validación. En la verificación, se analiza si el modelo es estructuralmente correcto y lógicamente consistente. El mejor método para verificarlo es analizando los resultados generados después de probarlo con datos que posean un amplio rango de valores. Esto implica que las expresiones son dimensionalmente consistentes. Se comprueba que todas las variables utilizadas en el modelo funcionan correctamente, al igual que los datos de entrenamiento elegidos, ya que es fácil incurrir en errores tales como: seleccionar una muestra con un número excesivo de datos de entrenamiento, muestras que no reflejen las características de las variables bajo estudio o muestras con datos faltantes, estos puntos pueden conducir a la creación de modelos inválidos. Por lo tanto, debe cuidarse que el modelo sea siempre una representación válida del problema. El criterio apropiado para juzgar la validez de un modelo es el hecho de si predice o no con suficiente exactitud los efectos relativos a los diferentes cursos de acción, para poder tomar una decisión que tenga sentido. Entonces, lo que se requiere es que exista una alta correlación entre la predicción del modelo y lo que ocurre en la vida real. Para asegurar que este requisito se cumpla, es importante hacer un número considerable de pruebas del modelo y las modificaciones consecuentes. Gran parte del trabajo de verificación del modelo se lleva a cabo durante la etapa de construcción para que sirva de guía en la confección del modelo (Gestal, 2009).

Otra medida directamente derivada del ECM es la raíz cuadrática media, o Root Mean Square (RMSE), se emplea para contrastar en qué medida los valores predichos se acercan a los valores observados, lo que significa que a menor valor de RMSE mayor precisión. Por precisión se entiende el promedio del grado de correspondencia entre pares individuales de valores pronosticados y valores observados o medidos, cuya fórmula es (Ec. 2.3):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_p - P_o)^2}{N}} \quad (2.3)$$

Donde:

$P_p$ =Valor pronosticado por el modelo.

$P_o$ = Valor observado o medido.

$N$ =Número total de datos.

Otra opción es realizar una validación cruzada para obtener una medida de la calidad de predicción del modelo. Para este propósito, los datos obtenidos se dividen en dos partes, una parte se destina al entrenamiento y la otra parte a la comprobación (testeo o validación). Cuando se detecta que el comportamiento del error de entrenamiento (EE) decrece monótonamente durante la fase de entrenamiento, mientras que el error sobre el conjunto de validación (EV) decrece hasta un punto a partir del cual crece, indica que a partir del mismo la red realiza un sobre-ajuste o sobre-entrenamiento sobre los datos de entrenamiento. Por ello, el proceso de entrenamiento debe finalizar cuando se alcance el primer mínimo de la función del error de validación (Demuth *et al.*, 2008).

Dee (1995) señala cuatro aspectos muy importantes asociados con la validación de modelos, los cuales son: *validación conceptual*, se refiere a la pregunta de si el modelo representa con exactitud el sistema que se está estudiando, es de tipo cualitativo y la manera en que se comprueba es cotejándola con la opinión del experto. *Validación de algoritmos*, es la traducción de los conceptos del modelo en formulas matemáticas. *Validación de códigos informáticos*, se refiere a la aplicación de formulas en lenguaje informático.

Finalmente, la *validación funcional* es la verificación del modelo frente a observaciones obtenidas de manera independiente. La evaluación ideal consiste en obtener los datos pertinentes del mundo real y realizar una comparación estadística de los resultados simulados y las observaciones. Así, las distintas fuentes de datos pueden ser útiles para la validación del modelo o proporcionar una base para la generalización.

#### **2.3.3.6 Análisis de sensibilidad**

Según Saltelli (2002), el análisis de sensibilidad es el estudio de las relaciones entre las entradas y las salidas de un modelo. El análisis de sensibilidad trata de medir el efecto que se observa en una salida  $y_k$  debido al cambio que se produce en una entrada  $x_i$ , así, cuanto mayor efecto se observe sobre la salida, mayor sensibilidad podemos deducir que

presenta respecto a la entrada (Montaño *et al.*, 2003). Existen algunas formas de realizarlo, entre estas se encuentran:

- Matriz de sensibilidad Jacobiana, se basa en el cálculo de la derivada parcial del error con respecto a los pesos para averiguar la dirección que se debe tomar para modificar los pesos y reducir el error de forma iterativa.
- Análisis de sensibilidad basado en la salida, consiste en estudiar el efecto que se observa directamente en una variable de salida debido al cambio que se produce en una variable de entrada.
- Análisis de sensibilidad basado en el error, trata de analizar el efecto producido sobre el error debido a cambios en la entrada, emplea el ECM.

La aplicación de este último análisis, consiste en ir variando el valor de una de las variables de entrada a lo largo de todo su rango con pequeños incrementos, mientras que al resto de las variables se les asigna su valor medio el cual queda fijo durante el análisis de la variable (Frost y Karri, 1999). El tamaño del incremento se basa en dividir el rango de la variable bajo análisis entre el número de patrones que se tiene. Una vez aplicados los incrementos a una determinada variable de entrada, se procede a entrenar la red neuronal calculando el valor de ECM. Siguiendo este procedimiento para todas las variables de entrada, se puede establecer una ordenación en cuanto a importancia sobre la salida. Así, la variable de entrada que proporcione el mayor error (ECM) será la variable con más influencia en la variable de respuesta, mientras que la variable de entrada con menor error (ECM) asociado será la que menos contribuya en la predicción de la red (Montaño *et al.*, 2003).

El conocimiento que se obtiene del modelo a partir del análisis de sensibilidad puede ser más valioso que la propia solución. En síntesis, el análisis de sensibilidad consiste en la evaluación sistemática de la respuesta de la solución óptima a las modificaciones de los datos de entrada. En algunos casos, el análisis de sensibilidad requiere volver a resolver el problema para varias combinaciones de los datos de entrada.

## **2.4 Técnicas de *Softcomputing***

Hoy en día, dada la complejidad de los problemas ambientales, se requiere de sistemas inteligentes que combinen conocimiento, técnicas y metodologías de varias

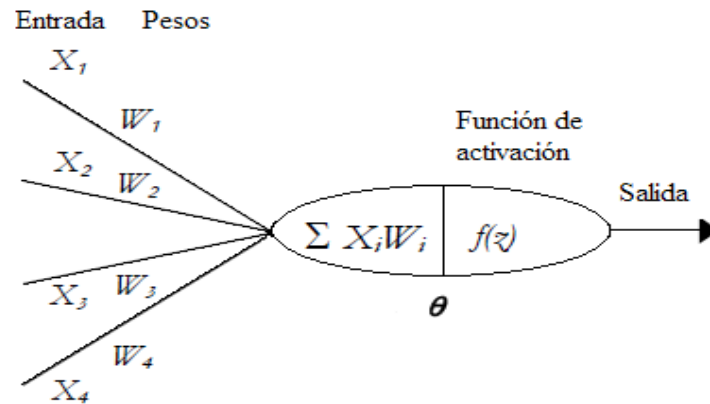
fuentes para tratar el problema en todas sus vertientes. Desde esta perspectiva, las técnicas de *Softcomputing* juegan un papel importante, al ofrecer una solución de forma eficiente y de calidad, además de permitir el diseño de modelos que tomen en cuenta las imprecisiones o la carencia de información en los datos, debido a potenciales errores en las mediciones o conocimiento impreciso. Zadeh (1965) define *Softcomputing* como un sistema que “aprovecha la tolerancia ante la imprecisión, la incertidumbre y la veracidad parcial de los hechos con vistas a lograr un manejo adecuado, robustez, una solución a bajo costo y una mejor afinidad a la realidad”. Bonissone *et al.*, (1999) actualizan la definición de *Softcomputing* como el conjunto de metodologías computacionales que tienen entre sus principales miembros la lógica difusa, la neurocomputación, la computación evolutiva y la computación probabilística. *Softcomputing*, puede observarse como una serie de técnicas y métodos con los que es posible manejar situaciones prácticas reales en la misma manera en que los seres humanos las tratan, es decir, con base a la inteligencia, sentido común, consideración de analogías y aproximaciones. Son técnicas capaces de adaptarse a cada situación.

#### **2.4.1 Redes neuronales artificiales**

Las redes neuronales artificiales (RNA) son un método para resolver problemas de forma individual o combinada con otros métodos, para tareas de clasificación, identificación, optimización o predicción, en las que el balance datos/conocimiento se inclina hacia los datos y en situaciones donde puede haber la necesidad de aprendizaje en tiempo de ejecución y de cierta tolerancia a fallos. Una RNA se compone por un conjunto de procesadores elementales denominados “neuronas”, que intentan reproducir el comportamiento del cerebro. Cada neurona, está caracterizada por entradas de activación  $X_j$ , que son modificadas por un peso  $W_j$ , las señales se combinan entre ellas  $\sum X_i W_i$  y entran a una función de activación  $f(z)$  que determina la salida, con base en una comparación de la suma con el valor de activación  $\Theta$  (Figura 2.3).

La intensidad de interacción entre dos neuronas está dada por el peso sináptico que las relaciona; si es positivo, la neurona pre-sináptica excitará a la post-sináptica, si es negativo la inhibirá. Inicialmente la RNA está conectada aleatoriamente, no dispone de ningún conocimiento útil almacenado. Para que pueda efectuar una determinada tarea

necesita ser entrenada y para ello se emplea una serie de ejemplos, lo que constituye la fase de aprendizaje.



**Figura 2.3** Esquema de una neurona artificial

El conocimiento no se programa directamente en la red neuronal, se obtiene a partir de los patrones que le son suministrados, por ejemplos adecuados de la operación que se pretende que el sistema realice.

El proceso de aprendizaje de la RNA, también conocido como entrenamiento puede ser supervisado o no supervisado. El primero, consiste en entrenar la red a partir de un conjunto de datos o patrones de entrenamiento compuesto por patrones de entrada y salida, cuyo objetivo es ajustar los pesos de la red  $W_j$  de manera tal que la salida generada por la RNA, sea lo más cercano al valor de salida dado, para cierta entrada. En el aprendizaje no supervisado se presenta sólo un conjunto de patrones a la red y, el objetivo del algoritmo de aprendizaje es ajustar los pesos de la red de manera tal que la RNA encuentre alguna estructura o configuración presente en los datos.

Si se ha elegido un modelo de red neuronal conveniente a las necesidades del problema, junto con un adecuado conjunto de patrones representativos de la operación que se desea efectuar, la red tendrá la capacidad de generalización, es decir, a partir de los ejemplos dados abstraerá la regla o reglas subyacentes y responderá adecuadamente en el futuro a nuevos patrones que nunca antes había visto. De acuerdo a la estructura, las arquitecturas neuronales pueden clasificarse en: RNA tipo *feedforward*, compuesta de arreglos de neuronas denominadas capas y RNA recurrente, que se distingue porque

permite la interconexión de neuronas en cualquier sentido y la inclusión de operadores de retardo.

La RNA es una estructura paralela, donde las neuronas están masivamente interconectadas con la capa anterior, de donde reciben la información, y con la capa posterior, hacia donde la transmiten. Una RNA o Perceptron Multicapa, consta de múltiples capas de unidades computacionales (neuronas) interconectadas, son capaces de aproximar cualquier función continua seleccionando los parámetros adecuados, el entrenamiento es del tipo supervisado por medio del algoritmo de retropropagación "BackPropagation".

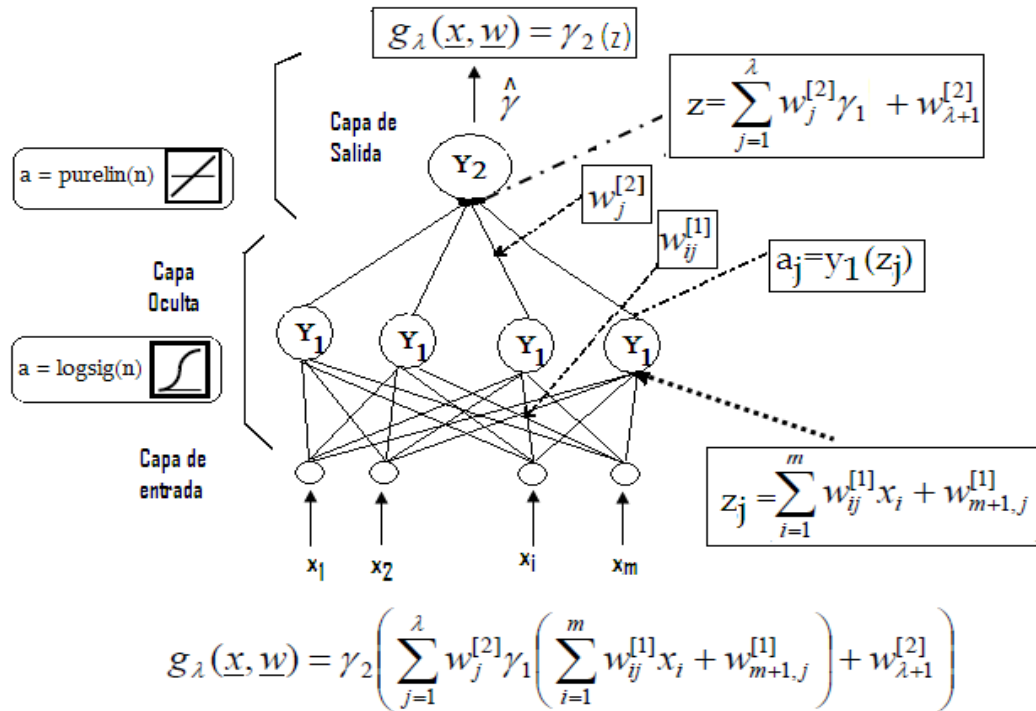
Dado un conjunto de observaciones, la tarea del aprendizaje neuronal es construir un estimador  $g_\lambda(x, w)$  de la función desconocida  $h(x)$  de la que se conoce sólo un conjunto de datos, donde  $w=(w_1, \dots, w_d)^T$  es el vector paramétrico a ser estimado y equivale a las ponderaciones de las conexiones entre las neuronas de la red,  $\gamma_1$  es una función no-lineal acotada y diferenciable con forma de función sigmoidea o de radio basal,  $\gamma_2$  es una función que puede ser lineal o no lineal y  $\lambda$  es el parámetro de control que indica el número de neuronas escondidas. El ajuste de la RNA se produce como resultado de la estimación de los parámetros basado en una muestra de tamaño  $N$ . La estimación es obtenida minimizando una función de costo (Ec. 2.4).

$$w_n^{LS} = \operatorname{argmin}\{L_n(w) : w \in W \subseteq \mathcal{R}^d\} \quad (2.4)$$

Donde  $L_n(w)$  viene dado por la función del promedio de los errores al cuadrado entre el dato estimado por la red y el dato real (Ec. 2.5).

$$L_n = \frac{1}{2n} \sum_{t=1}^n (y_t - g(x^t, w))^2 \quad (2.5)$$

La versatilidad de este tipo de estructuras reside en su capacidad para detectar y establecer relaciones no lineales complejas entre las variables independientes (entradas) y las dependientes (salida) (Figura 2.4).



**Figura 2.4** Modelo de una red neuronal tipo *Feedforward*.

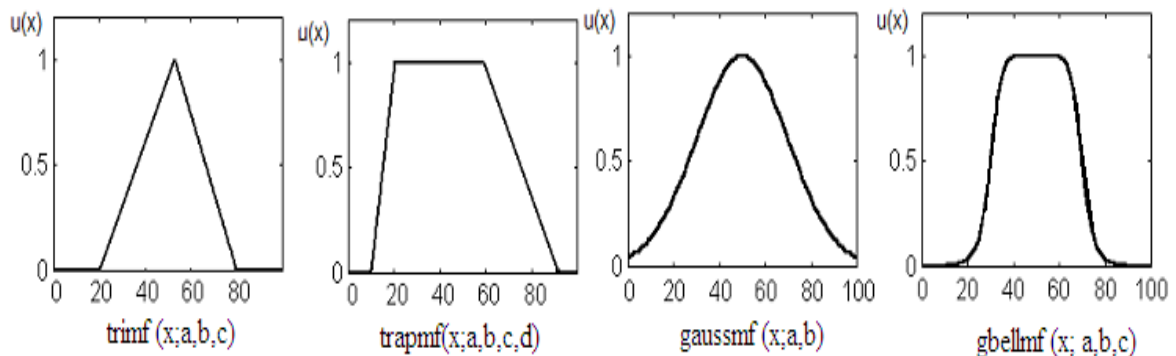
### 2.4.2 Lógica Difusa

La lógica difusa (LD) tiene dos significados diferentes; en el sentido estricto la LD constituye un sistema lógico dedicado a la formalización de modos de razonamiento que son aproximados y no exactos. En el sentido amplio, la LD coexiste con la teoría de conjuntos difusos, como una teoría de clases con fronteras no nítidas. Hoy en día la LD es empleada desde el sentido más amplio (Zadeh, 1996). La LD brinda un lenguaje con la sintaxis y semántica local a través del cual es posible traducir el conocimiento cualitativo sobre el problema a resolver. La LD permite usar variables lingüísticas para modelar sistemas dinámicos, estas variables toman valores difusos que son caracterizados por una etiqueta y un significado, que puede ser interpretado como una restricción de su valor. Estas limitaciones son propagadas por operaciones de inferencia difusa, basadas sobre los conjuntos difusos y la regla composicional de inferencia denominada *Modus Ponens* generalizado. Este mecanismo traslada la regla básica de deducción del cálculo de predicados del *Modus Ponens* al entorno de predicados vagos o difusos, se establece en los siguientes términos: supuesto que la implicación de  $A \rightarrow B$  es cierta y dado que ocurre  $A$ , entonces se ha de concluir que la proposición  $B$  también es cierta (Bonissone *et al.*, 1999).

### 2.4.2.1 Conjuntos difusos

Desde el punto de vista formal, un conjunto difuso  $A$  se define como el conjunto cuyos elementos pueden estar contenidos parcialmente, la frontera del conjunto es difusa, la transición entre “pertener” y “no pertenecer” es gradual. Sea  $X$  una colección de objetos, con un elemento genérico de  $X$  denotado por  $x$ . Un conjunto difuso  $A$  en  $X$ , es un conjunto de pares ordenados:  $A = \{(x, \mu_A(x)) \mid x \in X\}$  caracterizado por una función de membresía  $\mu_A(x)$  que asocia a cada punto en  $X$  un número real en el intervalo  $[0,1]$  con valor de  $\mu_A(x)$ , donde  $x$  representa el “grado de membresía” del elemento  $x$  al conjunto  $A$ . Donde  $X$  se le llama universo de discurso, el cual contiene todos los elementos que pueden ser tomados bajo consideración para asignar valores a las variables del sistema difuso. Si  $\mu_A(x)=1$  indica la pertenencia estricta de  $x$  en  $A$  y  $\mu_A(x)=0$  indica que  $x$  no pertenece a  $X$ . La forma particular de una función de membresía, solo puede ser determinada en el contexto de una aplicación en particular. Entre las funciones más frecuentemente utilizadas, se encuentran la: triangular, trapezoidal, gaussiana y de campana (Figura 2.5).

Las *variables lingüísticas* pueden utilizarse para cuantificar el significado del lenguaje natural, por lo que dicho lenguaje puede manipularse. Se les asignan valores (etiquetas lingüísticas) que son expresiones como palabras, frases o sentencias en un lenguaje natural o artificial. Para estas variables se utilizará un nombre y un valor lingüístico sobre el universo de discurso. Además podrán dar lugar a sentencias generadas por reglas sintácticas, a las que se les podrá dar un significado mediante reglas semánticas.



**Figura 2.5** Formas de las funciones de membresía

Zadeh (1972), considera que las etiquetas lingüísticas pueden clasificarse en dos categorías que informalmente se definen como: Tipo I, si son representadas como operadores que actúan en un conjunto difuso (muy, más, menos, ligeramente, altamente). Tipo II, las que requieren una descripción de cómo actúan en los componentes del conjunto difuso (esencialmente, técnicamente, estrictamente).

La *granularidad* en una variable lingüística, está dada en función del número de valores que puede tomar y normalmente se emplea un conjunto pequeño de valores para una variable lingüística, mientras que un *predicado nítido* es el nombre de una propiedad precisa, exacta de los elementos de un conjunto.

#### **2.4.2.2 Reglas difusas**

Las reglas difusas empleadas en el razonamiento difuso, son la manera que tiene un sistema difuso de guardar el conocimiento lingüístico que le permite resolver el problema para el cual ha sido diseñado. Estas reglas son del tipo IF-THEN y tiene dos partes, el antecedente y la conclusión dada por el consecuente. La estructura de las reglas difiere del sistema difuso utilizado (Mamdani o Takagi-Sugeno). Una regla difusa simple tiene la forma: “*si u es A entonces v es B*”, donde *A* y *B* son conjuntos difusos definidos en los rangos de “*u*” y “*v*” respectivamente. Una regla expresa un tipo de relación entre los conjuntos *A* y *B* cuya función característica sería  $\mu_{A \rightarrow B}(x, y)$  y se representa como implicación lógica. Cada una de las reglas o proposiciones IF-THEN es a su vez un conjunto difuso con su función característica que mide el grado de verdad de la relación de implicación entre *x* e *y*. Sin embargo, tales reglas son difíciles de identificar por inspección manual, por lo tanto son usualmente derivadas desde los datos usando técnicas de clusterización difusas (Medina y Paniagua, 2008; Llano *et al.*, 2007).

#### **2.4.2.3 Subclustering Difuso**

El propósito de la clusterización difusa es dividir el espacio de los datos en conjuntos difusos, cada uno representando una parte específica del comportamiento del sistema. Después de proyectar los conglomerados (clusters) en el espacio de entrada, la parte del antecedente de la regla difusa puede ser encontrado.

El método de agrupamiento substractivo (subtractive clustering) es utilizado para generar sistemas de inferencia difusos (FIS) automáticamente, basados en los

conglomerados creados. Mediante la función *genfis2* se lleva a cabo esta tarea, se genera un FIS tipo Sugeno de un conjunto de datos de entrada-salida usando para ello la función *subclust*, que es la encargada de generar la cantidad de conglomerados y la ubicación de sus centros. La función *genfis2* utiliza la información generada por *subclust* para generar los parámetros de las funciones de pertenencia en la entrada y salida del FIS. Adicionalmente, *genfis2* determina el número de reglas basándose en el número de conglomerados, y establece los antecedentes. Finalmente usa estimación lineal de mínimos cuadrados para determinar los consecuentes de cada regla, el cual puede ser una simple función. En este camino un conglomerado corresponde a una regla del modelo TKS (Medina y Paniagua, 2008).

Los argumentos mínimos requeridos para utilizar la función *genfis2* son: matriz de los valores de entrada, matriz o vector de salida y radio de influencia de los conglomerados. Este último puede ser un vector ó un escalar que varía entre cero y uno, que especifica el rango de influencia del centro de los conglomerados en cada una de las dimensiones de los datos. Valores pequeños para este radio, generan una gran cantidad de grupos, valores aceptables para este parámetro generalmente están entre 0.2 y 0.5 (Chiu, S., 1994).

Cuando se tiene una sola salida, *genfis2* es utilizado para generar un FIS inicial el cual pueda ser mejorado con la técnica de aprendizaje adaptativa neuro-difusa ANFIS. Cuyo propósito es ajustar los valores de los parámetros (funciones de pertenencia, coeficientes lineales) de manera que el modelo pueda describir en forma satisfactoria al sistema real.

#### 2.4.2.4 Operaciones Difusas

El conjunto de operaciones difusas estándar que alcanzan la esencia de la lógica difusa, son la unión (OR), intersección (AND) y complemento aditivo (NOT). Si dos conjuntos difusos *A* y *B* se definen en el universo *X*, para cada elemento *x* perteneciente a *X*.

Intersección, AND:  $\mu_{A \cap B}(x) = \min (m_A(x), m_B(x))$

Union, OR:  $\mu_{A \cup B}(x) = \max (m_A(x), m_B(x))$

Complemento aditivo, NO:  $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$

### 2.4.3 Sistemas de inferencia difusos

Un sistema de inferencia difuso (FIS), o Fuzzy Inference System, por sus siglas en inglés, es una forma de representar conocimientos y datos inexactos en forma similar a como lo hace el pensamiento humano. Un FIS define una correspondencia no lineal entre una o varias variables de entrada y una variable de salida; proporcionando una base sobre la cual se toman decisiones o definen patrones. Los pasos esenciales para el diseño de un sistema difuso son (Jang, 1993):

- Identificar el tipo de problema y el tipo de sistema difuso que mejor se ajusta a los datos.
- Definir variables de entrada y salida, sus valores difusos y sus funciones de membresía.
- Definir la base de conocimiento o reglas difusas.
- Obtener las salidas del sistema mediante la información de las variables de entrada utilizando el sistema de inferencia difuso.
- Trasladar la salida difusa del sistema a un valor nítido o concreto mediante un sistema de “concreción”, finalmente ajustar el sistema validando los resultados.

La estructura básica de un FIS consiste de tres conceptos: Una base de reglas, una base de funciones de pertenencia usadas en las reglas difusas y un mecanismo de razonamiento que ejecuta el proceso de inferencia.

La salida de un FIS es un conjunto difuso, por lo cual se requiere una interfaz de defuzificación que transforme los resultados difusos de la inferencia en salidas nítidas. Los sistemas de inferencia más comunes son el sistema Mamdani y el Takagi-Sugeno, son caracterizados por el consecuente de las reglas difusas que emplean:

- Inferencia de Mamdani: Tanto el antecedente como el consecuente son expresiones difusas. “*Si  $X_1$  es Alto y  $X_2$  es Bajo ENTONCES  $Y$  es Alto*”
- Inferencia de Takagi-Sugeno (TKS): El antecedente son expresiones difusas, pero el consecuente es una función polinómica que depende de las entradas. “*Si  $X_1$  es Alto y  $X_2$  es Bajo ENTONCES  $Y=f(X_1, X_2)$* ”. Si  $f(X_1, X_2)$  es un polinomio de primer orden se dice que el sistema es de primer orden. Si  $f$  es una constante se dice que el sistema es de orden cero.

#### 2.4.4 Sistemas Neurodifusos

Por un lado están las redes neuronales artificiales (RNA) que permiten obtener y ajustar modelos de manera sencilla, la limitante de esta técnica se debe a que el conocimiento aparece de forma implícita, a través de valores numéricos sin un significado fácilmente reconocible. Por otro lado, los modelos basados exclusivamente en conjuntos y lógica difusa tienen problemas de eficiencia cuando la base de reglas es grande y presentan dificultades prácticas a la hora de implementar las funciones de pertenencia. La hibridación de estos dos paradigmas condujo al desarrollo de los sistemas neurodifusos, los cuales permiten aprovechar las ventajas de ambos, consiguiendo la sencillez de ajuste de las RNA y a la vez disponiendo de conocimiento explícito expresado en un conjunto de reglas IF-THEN.

Los modelos neurodifusos son una herramienta que facilita el tratamiento de datos con alta dimensionalidad, no requieren de gran cantidad de datos como con otras técnicas, permite manejar información imprecisa, vaga o de baja la calidad y no requiere conocimiento a priori. La arquitectura ANFIS por sus siglas en inglés (Adaptive-Network-based Fuzzy Inference System) es un sistema de inferencia difuso implementado sobre una red adaptativa. ANFIS puede servir como base para la construcción de un conjunto de reglas difusas con las funciones de membresía asociadas para generar parejas definidas de entrada-salida (Jang, 1993, 1995). Hace uso de reglas IF-THEN del tipo Takagi-Sugeno (Takagi y Sugeno, 1985), en términos de variables lingüísticas para las entradas ( $A_1$ ,  $B_1$ ), y el consecuente de cada regla se obtiene como una combinación lineal de los valores de entrada más un término constante ( $r$ ), tal como se observa en la ecuación (Ec. 2.6).

$$\text{IF } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ THEN } z_1 = p_1x + q_1y + r_1 \quad (2.6)$$

La matriz del sistema de inferencia difuso se genera mediante un entrenamiento inicial, anterior a su construcción, buscando relaciones entre las entradas y las salidas *a priori*, formando agrupaciones “*clúster*” entre los datos que tienen un comportamiento similar, reduciendo así, tanto el número de reglas difusas como el número de clúster encontrados, como consecuencia se tiene el mismo número de reglas. Un factor muy importante es el *radio de influencia*, es un parámetro del algoritmo de entrenamiento cuyo

valor se determina mediante prueba y error: Cuanto menor es el radio de influencia, se obtiene un mayor número de agrupaciones, reglas y funciones de membresía, consiguiendo un modelo más complejo que no generalice, mientras mayor sea el radio de influencia, el número de reglas y funciones disminuye, creando un modelo más grueso (Chiu 1994).

El proceso de inferencia del ANFIS se implementa como una red neuronal de cinco capas, por medio de técnicas de gradiente descendente y mínimos cuadrados, que le permiten ajustar los parámetros del antecedente y del consecuente. Los nodos cuadrados son nodos cuyos parámetros son ajustables, mientras que los nodos en círculos son nodos fijos (Figura 2.6).

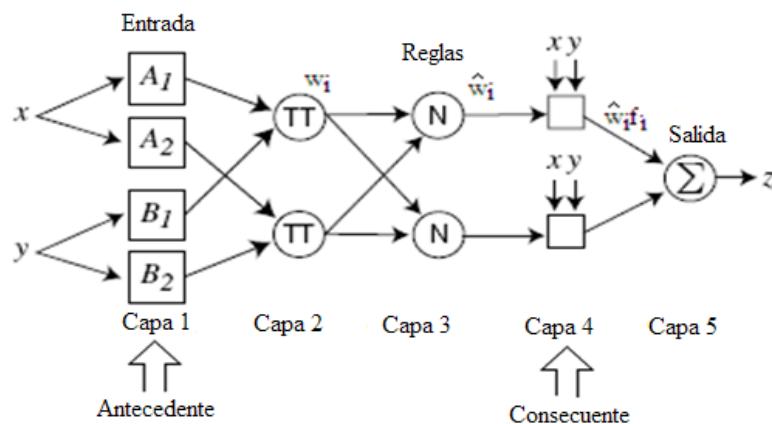
Capa 1, *Pertenencia*. Las entradas a la capa corresponden a los valores  $x$  e  $y$  del *antecedente*, la salida del nodo es el grado de pertenencia para el cual la variable de entrada satisface el término lingüístico asociado a este nodo.

Capa 2, *Grado de activación de la regla*. La función de este nodo es fija y consiste en realizar la t-norma entre los grados de pertenencia dados por la capa 1, dando el peso de la regla en el proceso de obtención del consecuente (Ec. 2.7).

$$w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y) \quad (2.7)$$

Capa 3, *Normalización de los grados de activación*. Asigna el peso de la regla normalizado con respecto a la suma de los grados de activación de una regla (Ec. 2.8).

$$\bar{w}_i = \frac{w_i}{w_1 + w_2} \quad (2.8)$$



**Figura 2.6** Arquitectura del sistema ANFIS (Jang, 1993)

Capa 4, *Calcula la salida* correspondiente al producto entre el grado de activación normalizado por la salida individual de cada regla. Los parámetros de cada capa son conocidos como parámetros del consecuente, para este caso serían  $r_i$ ,  $p_i$  y  $q_i$  (Ec. 2.9).

$$\overline{w}_i f_i = \overline{w}_i (r_i + p_i x + q_i y) \quad (2.9)$$

Capa 5, *Agregación / combinación de las salidas*. El único nodo de esta capa calcula la salida total del FIS, donde  $x$  e  $y$  son las variables del antecedente y del consecuente respectivamente y  $z$  la salida del sistema (Ec. 2.10).

$$f = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (2.10)$$

En el primer paso o “*paso hacia adelante*”, los modelos de entrada son propagados y los consecuentes óptimos son estimados por un procedimiento iterativo de mínimos cuadrados, mientras que los antecedentes permanecen fijos. En el segundo paso o “*paso hacia atrás*” se utiliza el procedimiento de propagación de errores para modificar los antecedentes mientras los consecuentes permanecen constantes. Este procedimiento se repite hasta que alcanza la condición de parada (criterio de error). Cuando los valores de los antecedentes son fijos, la salida general del sistema se expresa como una combinación lineal de los consecuentes (Jang, 1993).

## 2.5 Modelando con Herramientas de *Softcomputing*

La modelización de sistemas biológicos se dificulta, debido a la complejidad de las reacciones y al número de componentes indeterminados, que en muchos de los casos muestran una relación no lineal. Los modelos convencionales basados en técnicas analíticas, que actualmente se emplean para predecir el comportamiento de los RSU dentro del relleno, pueden volverse muy complicados e inflexibles cuando se quiere hacer frente a la imprecisión y la complejidad de los procesos involucrados. La realidad es que si se pretende tratar con estos sistemas, se supone el hecho de enfrentarse a un elevado nivel de incertidumbre, sin olvidarse de la imprecisión. Es en este punto, donde se han desarrollado

una serie de modelos que se benefician de las características de *Softcomputing*, como es la tolerancia a la imprecisión y a la incertidumbre.

En este sentido Cakmakci (2007) emplea ANFIS para desarrollar un modelo que predice el efluente de sólidos volátiles y la producción de metano en un sistema de digestión anaerobia de lodos primarios provenientes de una planta tratadora de aguas. Concluye que la aplicabilidad del modelo ANFIS es muy simple y no fue necesario ni definir reacciones complejas, ni ecuaciones químicas. Debido a la estructura no lineal del modelo ANFIS, demostró que un sistema complejo como la digestión anaerobia puede ser fácilmente modelado. La incertidumbre es una característica importante que está presente en la mayoría de los problemas asociados a los RSU dentro de un RESA.

Garg et al., (2006), presentan un modelo aplicando evaluación sintética difusa, para estimar la constante de la tasa generación de metano ( $k$ ) en un RESA. Emplean cuatro parámetros, precipitación, temperatura, composición de los RSU y asentamientos del relleno. El modelo fue calibrado y verificado con los valores de  $k$  de 42 localidades, los datos de 10 sitios se emplearon para calibrar el modelo y el resto para comprobarlo. El modelo resultó ser razonablemente complejo y preciso, y al mismo tiempo, simple y fácil de ejecutar. Se verifica que la LD ofrece un método barato para la estimación de metano.

Otro elemento positivo que se verifica en las técnicas de *Softcomputing*, es la exactitud en la predicción de la generación de lixiviados y biogás, cuando no se dispone de suficiente información. Ozkaya *et al.*, (2007) comentan que a pesar de la importancia de la degradación de los RSU en los RESA, existe poca información sobre los factores ambientales que actúan simultáneamente en la biodegradación de los RSU. La mayor parte de la información disponible se deriva de experimentos que se diseñan para evaluar los factores individuales que afectan la degradación de los RSU. Desarrollan un modelo con una RNA para predecir la fracción de metano en el biogás procedente de biorreactores a escala campo de un RESA, teniendo en cuenta factores ambientales tales como la adición de agua (recirculación), componentes de temperatura y los lixiviados de los residuos.

Los biorreactores fueron operados con ( $C_2$ ) y sin recirculación de lixiviado ( $C_1$ ). Se monitorearon los lixiviados y los componentes de los gases por 34 meses y con esta información se modeló la fracción de metano en el gas en los biorreactores ( $C_1$  y  $C_2$ ), los componentes de los lixiviados se utilizaron como parámetros de entrada (pH, alcalinidad,

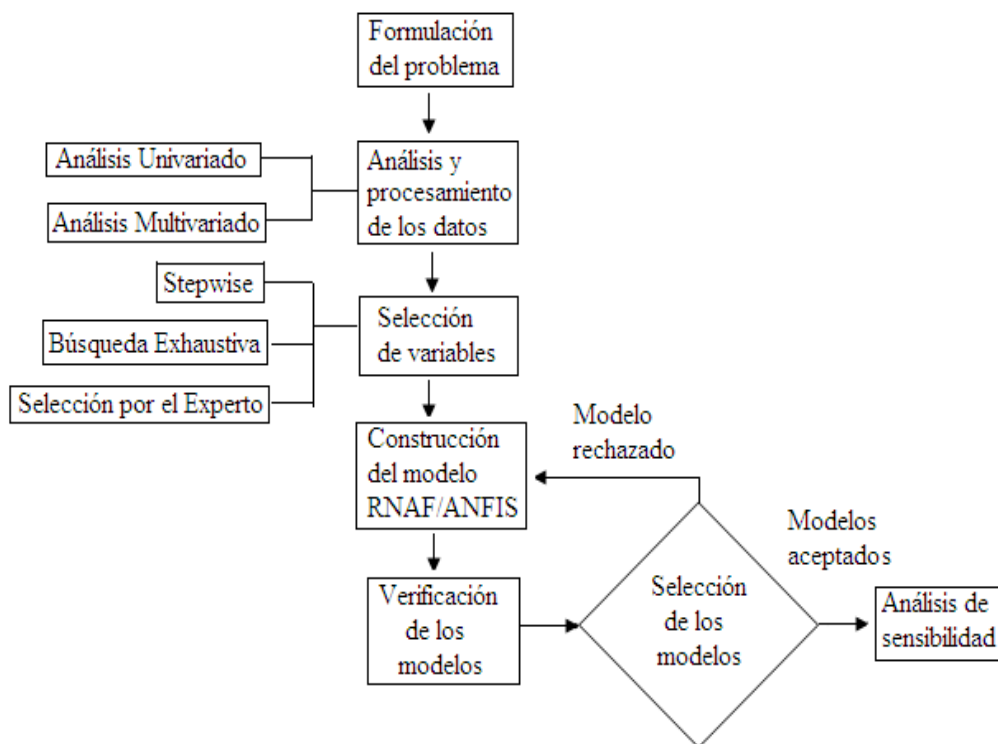
DQO, sulfato, conductividad, cloruro y temperatura de los residuos), para predecir la fracción de metano en el biogás como un producto final de digestión anaerobia. Concluyen que el modelado con redes pueden extenderse y desarrollarse en término de la reducción de metano sobre la base de parámetros de lixiviados y que las estrategias de control puede extenderse a aumentar la producción de metano y a la determinación de estrategias de recirculación de lixiviados y a la reducción de gases de efecto invernadero.

En esta misma dirección Karaca y Ozkaya (2006) proponen un método para modelar el caudal de lixiviados en un relleno sanitario (RESA), empleando RNA y el algoritmo de retropropagación. Después del entrenamiento del algoritmo, el modelo predice la tasa de flujo basado en los datos meteorológicos. Dependiendo del valor de salida, se activan estrategias de control y acciones apropiadas. En la metodología propuesta, se define en primer lugar, la colección de datos, el algoritmo de retropropagación que minimice el error entre el valor de salida de la red y el valor real. En segundo lugar, se establecen las salidas de la red neuronal, utilizando datos meteorológicos para un día determinado, de acuerdo con las mediciones o información meteorológica proporcionada se calcula el caudal del lixiviado. La velocidad diaria del flujo de lixiviado desde el RESA, fue considerada como un parámetro crítico del modelo (salida de la red neuronal).

Los datos fueron divididos en entrenamiento, validación y prueba. Durante el período de estudio, los niveles de flujo superaron el umbral superior un total de 19 ocasiones. Esto reveló la necesidad de un método fiable, rápido y sencillo para predecir estos episodios antes de que se produzcan y tomar las medidas adecuadas para controlar y administrar los niveles de flujo de lixiviados. También se evaluaron trece algoritmos de retropropagación, finalmente se seleccionó el mejor algoritmo Levenberge-Marquardt con 22 neuronas, y una red neuronal de dos capas. Concluyen que las redes neuronales ofrecen un método simple, factible y versátil para predecir los caudales diarios de lixiviados en un RESA, además de que alcanzan una predicción precisa y eficaz de los lixiviados y un control efectivo y de gran alcance en la gestión de los RSU.

### 3. METODOLOGÍA

Para fines de este trabajo, se adecuaron seis de los pasos sugeridos por Naylor et al., (1982) en la planeación de experimentos de simulación los cuales son descritos brevemente (Figura 3.1), En el primer paso, se definieron los elementos de la problemática bajo investigación, se estableció el origen de los datos, variables de entrada y salida, y características del sistema en general. En el segundo paso, los datos se registraron y procesaron, además se verificó tanto su fiabilidad como su validez. Fue en la etapa de procesamiento donde se realizó una descripción individual y en grupo de las variables involucradas. Como tercer paso, se eligieron las variables que integrarían el modelo, de acuerdo a las metodologías propuestas de selección de variables. Durante el cuarto paso se realizó la construcción del modelo, en este trabajo se emplearon y evaluaron dos herramientas de *Softcomputing*, una red neuronal artificial tipo *feed forward* (RNA) y un sistema adaptativo neurodifuso (ANFIS). El quinto paso consistió en la validación y verificación de los modelos desarrollados previamente. Para ello, se calculó la raíz cuadrática media RMSE, entre el valor real y los resultados arrojados por el modelo.

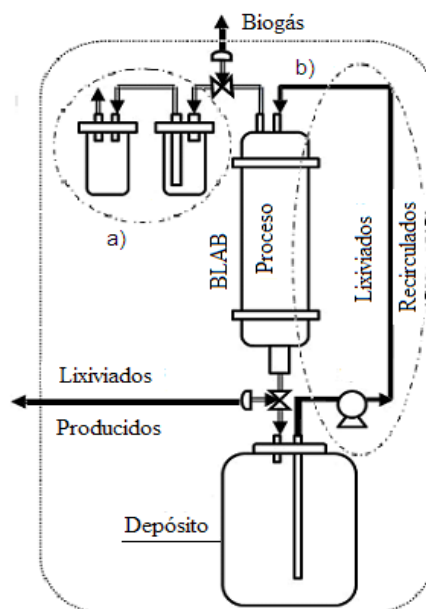


**Figura 3.1** Metodología desarrollada en el diseño de los modelos.

Los dos modelos con el menor valor de RMSE se seleccionaron, uno para estimar lixiviados y el otro para CH<sub>4</sub>, el resto se descartó. Finalmente en el sexto paso, se realizó el análisis de sensibilidad, lo que permitió definir las variables independientes con mayor influencia en el modelo.

### 3.1 Formulación del problema

Los datos empleados en la presente tesis, fueron resultado de un proyecto de investigación realizado en el laboratorio de residuos y medio ambiente del Instituto de Investigaciones Agropecuarias y Forestales (IIAF), perteneciente a la Universidad Michoacana de San Nicolás de Hidalgo (UMSNH). El sistema bajo estudio (Figura 3.2), se realizó en un conjunto de biorreactores a escala laboratorio (BLAB), donde las entradas quedaron determinadas por las características de los lixiviados recirculados y las salidas por la composición del metano en el biogás generado y los lixiviados producidos (volumen y características), cada BLAB fue cargado con residuos sólidos urbanos (RSU) preparados previamente del RESA de Pátzcuaro, Michoacán, cubiertos con suelo de la región y compactados a 600 kg/m<sup>3</sup>, de acuerdo a las condiciones de operación del sitio. Se operaron en tres grupos de siete BLAB, desmontándose un BLAB de cada grupo los días: 15, 45, 75, 105, 135, 165 y 201. Los lixiviados se recircularon dos veces por semana durante 201 días (Hernández-Berriel et al., 2010).



**Figura 3.2** Sistema a escala y elementos que lo integran a) biogás, b) lixiviados.

Conforme a las características del sistema y al conjunto de información registrada, el modelo a desarrollar es del tipo de caja de negra, debido a que fue diseñado a partir de datos de entrada y salida, y con un reducido conocimiento acerca de la interpretación de las variables medidas. Como primer paso se definieron las variables independientes (entrada) y dependientes (salidas), para ello fue necesario establecer una correspondencia entre las entradas y las salidas del BLAB.

Considerando el sistema real, las variables independientes quedaron determinadas por: el tratamiento, referido como tasa de recirculación (15, 30, 40, 60, 80 y 120 %Volumen) asignada a cada grupo de biorreactores a escala laboratorio, cuantificado por el volumen de lixiviado recirculado (62, 124, 165, 248, 330, 496 ml). Además de las características medidas analíticamente (pH, DQO, CE, AA, AP, ST, SVT, NH<sub>3</sub>-N) de los lixiviados recirculados (LR) de cada tratamiento, y como variables dependientes se especificaron: el volumen de lixiviado en (ml) que fue generado en cada grupo de biorreactores por tratamiento y es referido como lixiviado producido (LP), además del porcentaje de metano (CH<sub>4</sub>) cuantificado a partir de la medición del biogás por cromatografía de gases (Tabla 3.1).

**Tabla 3.1** Definición de variables

Nombre de la variable	Simbología	Unidad de Medida	Rango	
			Min.	Máx.
Tratamiento	%V	%	15	120
Potencial de hidrógeno	pH	----	6	8.52
Demanda química de oxígeno	DQO	[g/L]	224	2913
Conductividad Eléctrica	CE	[mS/cm]	0.42	5.83
Ácido Acético	AA	[mg/L]	4	1046
Ácido Propiónico	AP	[mg/L]	3	361
Sólidos Totales	ST	[mg/L]	0	3.07
Sólidos Volátiles Totales	SVT	[mg/L]	0	1.6
Nitrógeno amoniacal	NH <sub>3</sub> -N	[mg/L]	0	64.1
Lixiviados producidos	LP	ml	27	565
Metano	CH <sub>4</sub>	%	0	68.39

### **3.2 Análisis y procesamiento de los datos**

Una vez definido el conjunto de variables de entrada y salida del sistema bajo estudio, las actividades posteriores consistieron; en seleccionar un método de almacenamiento de la información y la preparación de los datos para su análisis. El registro o captura de los datos se realizó inicialmente en hojas de Excel, debido a su facilidad de manipulación. Durante el procesamiento, los datos se organizaron, clasificaron y depuraron por medio de dos controles de calidad. Con el primer control, se efectuó una comparación visual entre el dato de origen con respecto al capturado, mientras que con el segundo, se verificó aleatoriamente la correspondencia entre los datos capturados y las fuentes de origen. Ambos procesos de control permitieron la oportuna detección de anomalías, tales como errores en la recopilación de datos y duplicidad de la información. Posteriormente la información fue exportada a una base de datos (BD) en Microsoft Access 2003®, en la que se aplicaron tres procesos de normalización, el cual consistió en la aplicación de una serie de reglas a las relaciones obtenidas al generar el modelo relacional de la base de datos, para evitar problemas de redundancia o ambigüedad y proteger la integridad de los datos, facilitando su operación y la generación de matrices de datos (Chen, 1976).

#### **3.2.1 Análisis exploratorio de los datos**

Una vez capturados y almacenados los datos, se procedió a realizar el análisis exploratorio de los datos. La población se consideró finita debido a que se tuvo un número específico de variables, con valores de tipo cuantitativo. Estas variables fueron registradas en seis muestras, cada una representada por un tratamiento bajo estudio (15, 30, 40, 60, 80, 120 %V) y replicada tanto para los lixiviados recirculados (LR) como lixiviados producidos (LP), por lo que fue necesario manipular cada conjunto de variables en matrices, que fueron importadas a STATGRAPHICS Plus 5.1 ® donde se realizó el análisis. Se trabajó con un total de 920 datos, constituidos por nueve variables para LR y once variables para LP con 46 valores cada una. Debido a la necesidad de un conocimiento más detallado del comportamiento de los datos, el análisis exploratorio se dividió en análisis univariado y multivariado, se efectuaron de manera independiente, tanto para el conjunto de variables de LP como para el conjuntos de variables de LR, obteniendo así un

entendimiento básico tanto del comportamiento individual como en grupo de las variables de entrada y de salida del sistema bajo estudio.

### **3.2.1.1 Análisis univariado**

Mediante el análisis univariado se evaluó el comportamiento de cada variable por separado, se verificó la calidad de los datos, se investigaron a fondo las inconsistencias en los valores usando los datos originales como puntos de referencia o bien consultando a la persona responsable de dicho experimento (Experto). Después de importar las matrices de datos a STATGRAPHICS Plus 5.1 Plus ®, se generaron una serie de gráficos de dispersión, de tallos y hojas, histogramas y diagramas de caja. Al examinar los gráficos de cada variable, fue posible conocer su distribución, su grado de variación, así como datos faltantes, valores mínimos y máximos. Los valores ausentes detectados se calcularon mediante interpolación lineal (Ec. 2.1), procurando mantener una muestra homogénea capaz de identificar el comportamiento general del sistema. En este caso los diagramas de caja facilitaron la detección de valores extremos, el Test Q de Dixon determinó su consistencia con el promedio de las mediciones, facilitando su permanencia o su eliminación. Posteriormente para conocer las tendencias de los datos, se empleó la media y la mediana, igualmente se usó la desviación estándar, la varianza, el coeficiente de asimetría y curtosis para evaluar la dispersión de los mismos.

### **3.2.1.2 Análisis Multivariado**

Después de analizar las variables individualmente, éstas se agruparon por tratamiento, examinando tanto su relación como dependencia. Se estudio la matriz de correlación y los valores del factor de correlación ( $r$ ), midiendo el grado de relación entre las variables agrupadas. Posteriormente mediante el análisis de regresión y el gráfico de dispersión, se determinó la forma de relación entre las variables. Procurando no degradar la capacidad de predictibilidad del modelo, todas las variables que presentaron valores superiores de  $r > 0.9$  fueron evaluadas para detectar la existencia de colinealidad. Para ello se empleó el diagrama de dispersión y el factor de influencia de la varianza (FIV), con lo que se confirmó la presencia de problemas de colinealidad entre las variables independientes (Kleinbaum *et al.*, 1988).

Para cuantificar la relación entre las variables agrupadas se llevó a cabo el análisis de ANOVA, verificando la hipótesis nula ( $H_0$ ), sobre la existencia de diferencias significativas en las variables de LP debido al tratamiento. Las diferencias encontradas en las variables, se evaluaron mediante la prueba de Tukey HSD con un nivel de confianza del 95% (Salvador y Gargallo, 2003). En resumen, las actividades desarrolladas durante el análisis exploratorio son listadas a continuación:

- Se prepararon y formatearon los datos, haciendolos accesibles a cualquier técnica estadística.
- Se efectuó un examen gráfico y un análisis descriptivo de las variables individualmente.
- Se identificaron y evaluaron datos atípicos en los datos.
- Se valoró el impacto potencial de datos ausentes, en la representatividad de los datos.
- Se realizó un examen gráfico y un análisis descriptivo, que permitiera cuantificar grado de interrelación entre las variables revisadas.
- Se evaluaron supuestos básicos como linealidad y normalidad.
- Se verificaron algunas hipótesis.

### **3.3 Selección de variables**

Con el fin de optimizar el aprendizaje del modelo y facilitar su interpretación, fue necesario reducir el número de variables y seleccionar aquellas con mayor representatividad, por lo que se utilizó tanto la técnica *stepwise* (SW) como la búsqueda exhaustiva (BE).

#### **3.3.1 Técnica de selección de variables *stepwise***

Esta técnica supone el empleo de una ecuación de regresión en la que se ingresan paso a paso variables, que añaden un valor significativo a la varianza ya explicada por las variables previamente seleccionadas. Se verifica en cada caso, dada la nueva situación de las variables introducidas si son adicionadas o excluidas. Se su implementación se usó STATGRAHICS PLUS 5.1® y una matriz de datos, conteniendo en las primeras columnas, los valores de las variables (%V, pH, DQO, CE, AA, AP, ST, SVT y  $\text{NH}_3\text{-N}$ ) de los LR y

en las últimas columnas las variables CH<sub>4</sub> y volumen de LP, se ordenaron ascendentemente por renglones con respecto al tiempo.

El método inició con la selección de la primer variable independiente que integró la ecuación de regresión, ésta se eligió a través de un análisis de regresión simple, donde se evaluaron todas las variables independientes con respecto a las dependientes, se consideró como criterio de selección los contrastes  $t$  y  $F$ , la variable con valores de  $t > 2$  y  $p$ -valor  $< 0.5$  fue seleccionada. A continuación en una ecuación de regresión múltiple, únicamente la variable seleccionada previamente son consideradas, en cada iteración se van agregando una a una las variables independientes, se consideran los mismos criterios de selección  $t$  y  $F$ , las variables que producen la mayor reducción en la suma de cuadrados del error y el mayor incremento en el valor explicativo de la ecuación mediante el factor de determinación ( $R^2$ ) o que posean la correlación parcial más alta con la respuesta deseada, fueron seleccionadas en cada iteración en caso contrario se eliminaron, el proceso de selección de las variables finalizó cuando todas las variables fueron evaluadas.

### 3.3.2 Búsqueda exhaustiva

Posteriormente la consistencia y relevancia de las variables se verificó repetitivamente con la búsqueda exhaustiva de todas las posibles combinaciones de las variables de entrada. Con esta técnica se buscó el conjunto de variables independientes que presentara la mayor influencia con respecto a la estimación de la generación de CH<sub>4</sub> y LP respectivamente. El método se desarrolló mediante la función *exhsrch* de MATLAB®, que consistió en encontrar las mejores cinco variables de nueve de todas las posibles combinaciones  $C(9,5)$ , para ello se evaluaron un total de 126 modelos (MATLAB, 2009). Para correr la función, se requirió construir una matriz conformada por las variables independientes (cinco en cada caso) y la variable dependiente (CH<sub>4</sub> o LP) a partir de los datos de entrenamiento, además de una segunda matriz con la misma estructura que la anterior, sólo que ésta conteniendo datos de validación.

Con el fin de optimizar el proceso de evaluación de los modelos, se programo una función en código de MATLAB para reducir tiempo y recursos en la selección de las variables CH<sub>4</sub> y LP (Apéndice A). Para implementar esta función, fue necesario construir dos pares de matrices, el primer par consistió de una matriz conformada por las nueve

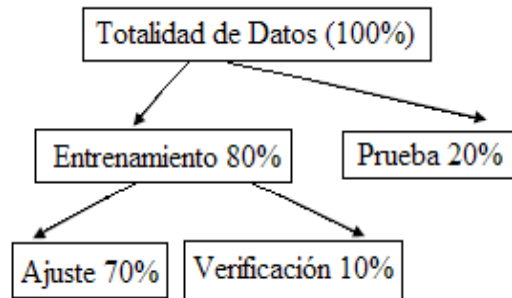
variables independientes y una segunda matriz con la variable dependiente (CH<sub>4</sub> o LP) con datos de entrenamiento. El segundo par, consistió de una matriz conteniendo las nueve variables independientes más la variable dependiente con datos de entrenamiento, y una segunda con la misma estructura a partir de datos de validación. Con el primer par de matrices se construyó el modelo FIS empleando la función *genfis2*, el segundo par de matrices se usaron para generar un modelo ANFIS con una sola época de entrenamiento, a partir del sistema FIS desarrollado previamente, empleando la función *anfis*, lo que fue equivalente a seleccionar las variables empleando la función *exhsrch*. En cada iteración se seleccionó el mínimo Error de Entrenamiento (EE) y Error de Validación (EV), al final se descartaron aquellas combinaciones que presentaron valores nulos o mayores a uno.

### **3.3.3 Selección por el experto**

Finalmente, las variables proporcionadas por el experto fueron evaluadas, se consideró tanto su comportamiento como la importancia en el proceso de biodegradación, además del costo y tiempo en su obtención, como resultado se propusieron seis variables de entrada.

## **3.4 Formulación de los modelos**

Después de seleccionar las variables, el siguiente paso consistió en formular los modelos para su análisis, esto se logró mediante la construcción de un modelo que representará la esencia del problema. Para ello, fue necesario dividir la totalidad de los datos en dos conjuntos. El primer conjunto se integró con el 80% de éstos y fue usado para el entrenamiento del modelo, el segundo se formó con el 20% restante, este conjunto no participó en el entrenamiento del modelo, se usó para proveer una medida de la calidad de la predicción del modelo (Prueba). El conjunto de entrenamiento a su vez se dividió en dos subconjuntos, el primero formado con el 70% los cuales fueron usados para ajustar los parámetros del modelo durante el entrenamiento y el 10% restante se empleó para su verificación, con el fin de evitar el sobre-entrenamiento. El entrenamiento se detiene cuando el error calculado sobre este grupo de datos, comienza a aumentar. Los datos que integraron cada conjunto, se seleccionaron aleatoriamente (Figura 3.3).



**Figura 3.3** Distribución de los datos

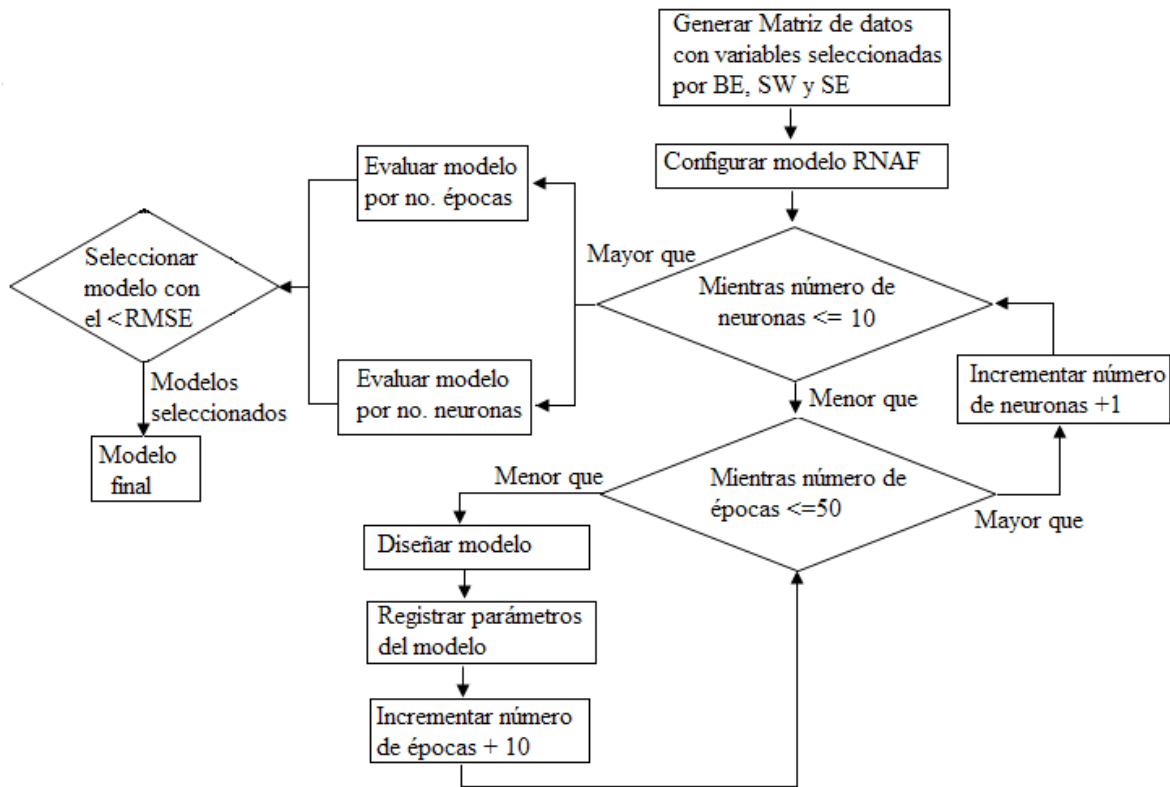
### 3.4.1 Selección de las herramientas de modelado

Antes de elegir la herramienta para el diseño de los modelos, se evaluaron aspectos del propio sistema tales como: el comportamiento no lineal de los datos, presencia de una elevada variabilidad, incumplimiento de supuestos (linealidad, homocedasticidad), reducido conocimiento sobre el comportamiento físico de las variables, partir de un conjunto de datos de entrada y salida para diseñar el modelo.

Por lo anterior, las aproximaciones basadas en técnicas de *Softcomputing* constituyeron una alternativa útil y viable para tratar con los puntos antes mencionados. Con respecto a las redes neuronales las características consideradas fueron: su habilidad para aproximar funciones, la capacidad de aprendizaje y de generalización, su operación tanto con datos cuantitativos como cualitativos, procesamiento de n-entradas y n-salidas, su alta aplicabilidad en sistemas multivariables. El sistema ANFIS, destacó por su aprendizaje a partir de datos entrada-salida, la integración del conocimiento del experto expresado en reglas lingüísticas. Otro punto considerado, fue el disponer de puntos de medición, ya que ambas técnicas (RNA y ANFIS) son útiles para resolver problemas cuando se carece de un conocimiento detallado del sistema; comparten criterios de selección tales como; el número de épocas de entrenamiento, error cuadrático medio (ECM), en el caso de la red el número de neuronas en la capa oculta y para ANFIS el número de funciones de membresía.

### 3.4.2 Construcción del modelo con una red neuronal

El primer modelo construido fue el de la red neuronal tipo *feedforward*, mediante el Neural Network del Toolbox de MATLAB® R2009b (Figure 3.4).



**Figura 3.4** Diagrama de flujo para la formulación de modelos mediante una RNA

Partiendo de un conjunto de variables independientes y una variable dependiente (CH<sub>4</sub>, LP), ordenadas en columnas a partir de los datos de entrenamiento y almacenados en archivos planos. Debido a que fueron tres métodos los empleados en la selección de variables (BE, SW y SE) y dos las salidas (CH<sub>4</sub> y LP), se evaluaron seis matrices de datos, generando un modelo RNA por cada una de ellas.

Con el fin de facilitar el proceso de generación y evaluación de los modelos, fue desarrollada una función en código de MATLAB, que consistió en la asignación de las matrices de entrada y salida con datos de entrenamiento y validación respectivamente. A continuación se desglosan los pasos que se realizaron (Apéndice B).

Por medio de consultas en Access 2003 se generaron matrices de datos. La primera matriz conteniendo las variables independientes con datos de entrenamiento (ebech4.txt) y la segunda los valores de la variable dependiente (SCH4.txt), lo mismo sucedió con los datos de validación (tbech4.txt y tch4.txt), su estructura se ejemplifica en la figura 3.5.

Archivo	Edición	Formato	Ver	Ayuda
120	6	6.35	1944	29.54
120	8	6	2026	26.66
120	13	6.47	2227	21.94
120	17	6.57	2913	17.21
120	24	7.16	2783	14.79
120	27	7.24	2741	9.6
120	34	7.44	2176	8.83
120	37	7.64	2021	3.97
120	41	7.64	1794	4.87
120	45	7.85	1568	16.33
120	48	7.81	1398	21.1
120	55	7.36	1073	30.5
120	59	7.71	943	34.55
120	66	7.73	714	31.61
120	73	7.55	688	34.03
120	80	7.42	662	37.49
120	83	7.67	671	38.87
120	94	7.47	706	45.46

Archivo	Edición	Formato	Ver	Ayuda
15.6				
20.08				
24.56				
28.65				
32.73				
55.72				
57.78				
60.7				
57.83				
57.73				
61.2				
55.59				
63.14				
62.13				
53.66				
52.48				
61.56				
49.74				

**Figura 3.5** Estructura de la matriz de datos para generar el modelo RNA

Cada par de matrices se almacenó en un archivo plano, que fue importado a MATLAB, durante la generación de los modelos. Posteriormente las transpuestas de ambas matrices se asignaron a variables (*pruebaent*, *pruebasal*) empleadas en el diseño del modelo de RNA.

*%asignación de valores*

```
load EBECH4.txt; %variables independientes
load SCH4.txt; %variable dependiente
inputs = EBECH4'; %matriz transpuesta de las variables independientes
targets = SCH4'; %matriz transpuesta de variable dependiente
```

*%asigna datos de validación*

```
load tbech4.txt;
load tch4.txt;
pruebaent = tbech4';
pruebasal = tch4';
```

Se asignaron valores a los parámetros de configuración de la red, como el número de neuronas en la capa oculta, valor que fue determinado por prueba y error, aunque también existió el criterio de considerar el promedio entre el número de entradas y salidas, como un valor referencial del número de neuronas en la capa oculta. Se evaluó un rango de [4 a 10] neuronas en incrementos de 1. La red se construyó mediante la función *newff* (*inputs*, *targets*, *numHiddenNeurons*), se partió de una estructura pequeña y gradualmente se añadieron neuronas hasta que finalizó el proceso de aprendizaje. Se asignó para la capa

oculta una función tangente-sigmoidea (*tansig*) y para la capa de salida una función lineal. Se empleó el algoritmo de optimización del error de la red *trainlm* (entrenamiento por el algoritmo Levenberg Marquardt) por su eficiencia y rapidez en el entrenamiento requerido, la función de actualización de pesos (*learnqdm*), y la función ECM para evaluar el desempeño de la red.

```
% Crea la red neuronal
numHiddenNeurons =4;           %Asigna numero de neuronas capa oculta
net=newff (inputs, targets, numHiddenNeurons); % Construye el modelo con la RNA
net=init(net);                 % Inicializa los pesos de la RNA
```

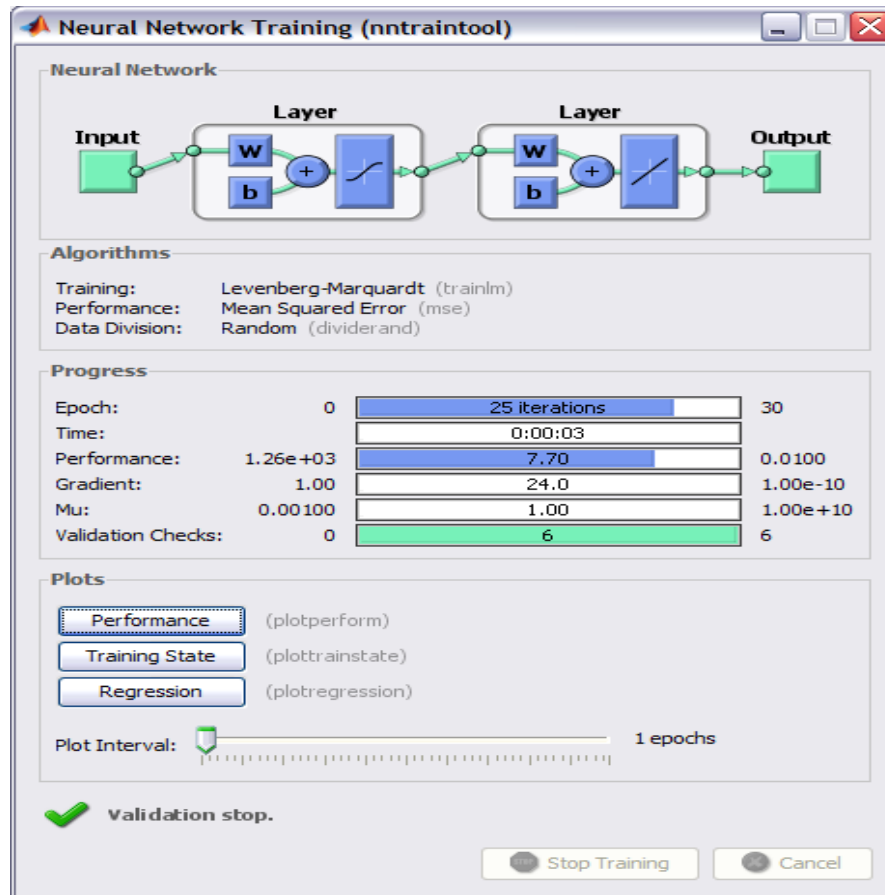
Debido a que el proceso de entrenamiento de la red requirió de un conjunto suficiente de ejemplos (patrones), los datos se asignaron para entrenamiento 70% (*net.divideParam.trainRatio*), 10% para validación (*net.divideParam.valRatio*) y 20% para prueba (*net.divideParam.testRatio*).

```
Asignación de los datos para entrenamiento, validación y prueba
net.divideParam.trainRatio = 70/100;      % Datos de entrenamiento
net.divideParam.valRatio = 10/100;       % Datos de verificación
net.divideParam.testRatio = 20/100;      % Datos de validación o prueba
```

Durante la configuración de los parámetros de la red, se definió el intervalo de visualización de los resultados (40), la tasa de aprendizaje (0.05), para determinar el número óptimo de iteraciones para adquirir convergencia, se evaluó un intervalo de épocas de entrenamiento [10 a 100] con incrementos de cinco, con error máximo de (0.01).

```
%parámetros de configuración de la red
net.trainParam.show = 40;           % Épocas a mostrar
net.trainParam.lr = 0.05;          % tasa de aprendizaje
net.trainParam.epochs =30;        % Máximo número de épocas de entrenamiento
net.trainParam.goal = 0.01;       % Objetivo, rendimiento
```

Una vez creada la red neuronal, y definidos los patrones de entrenamiento con sus respectivos targets, solo resta invocar a la función *train* (*net, inputs, targets*). Con esto MATLAB inicia a entrenar la red neuronal hasta alcanzar el performance deseado, la información de dicho proceso, es presentado a través de una ventana (Figura 3.6).



**Figura 3.6** Ventana de entrenamiento de la RNA.

Una vez que el proceso de entrenamiento terminó, el modelo es evaluado mediante la función *sim*, la cual calcula las salidas generadas por el modelo.

*% entrenamiento de la red*

```
[net, tr] = train (net, inputs, targets);
outputs = sim (net, inputs);
```

```
% Levenberg-Marquardt
% Realiza la simulación del modelo
```

La precisión del modelo se evalúa a través del error cuadrático medio (ECM) y la raíz cuadrada del error (RMSE). Eligiendo aquellos modelos con los menores errores de entrenamiento y validación. Como criterio adicional de selección, se consideraron diferencias entre los errores de dos modelos menores al 5%, considerando el modelo con el menor número de neuronas y épocas de entrenamiento. A cada modelo se le calculó el error de entrenamiento (ET) y el error de validación (EV).

*%calcula el error de la predicción*

```

error = outputs-targets;           % Salida obtenida x la red menos los datos entrada
RMSE = sqrt(mse(error))           % Calculó de la raíz del error cuadrático medio
ECM = mse(error)                   % Calculó del error cuadrático medio
%calcula el error de validación
outputs2 = sim(net, pruebaent);    % Corre el modelo con datos de prueba
error2=outputs2-pruebasal;         % Diferencia de datos de prueba y salida del modelo
RMSE2=sqrt(mse(error2))           % Calcula la raíz del error cuadrático medio

```

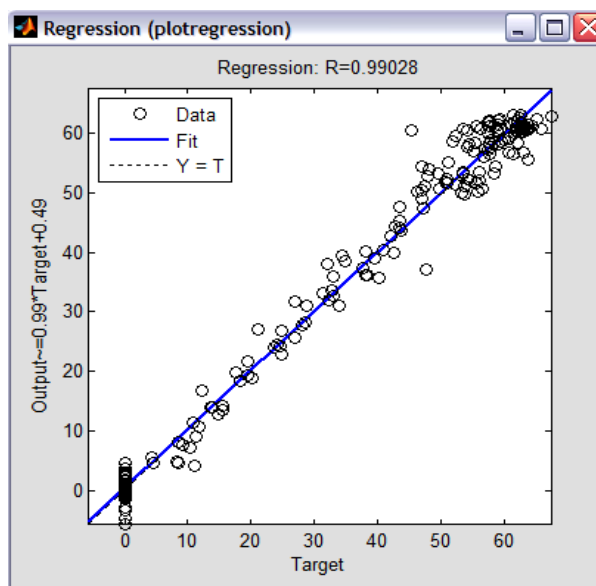
Finalmente con la función *net*, el modelo generado se corre con datos de prueba, calculando el RMSE. Por medio del análisis de regresión entre datos de salida proporcionados por el modelo y los reales, se calcula la bondad de ajuste del modelo representado por el factor de correlación R (Figura 3.7).

```

%Evaluación de la RNA
resultado = net (datos);           % Prueba la RNA
% Despliega resultados de la red generada
plotregression (targets,outputs);  % Realiza análisis de regresión con datos E/S
view (net);                         % Muestra el diseño de la red
plotperf(tr);                       % Muestra el progreso del entrenamiento

```

Este proceso se realizó tanto en los modelos de CH<sub>4</sub> como para LP, por cada técnica empleada (S, BE, SE). El criterio de selección de los modelos se vio justificado con la presencia de valores bajos en el error de entrenamiento y validación, lo que representó una mayor bondad de ajuste con valores de R<sup>2</sup> muy cercanos a uno.



**Figura 3.7** Gráfico de regresión entre datos generados por el modelo y datos reales

### 3.4.3 Construcción del modelo neurodifuso ANFIS

Se generaron los modelos empleando el Fuzzy Logic del Toolbox de MATLAB® R2009b. Se partió de la definición del conjunto de variables independientes y dependientes (CH<sub>4</sub> y LP) a emplear. A través de consultas diseñadas en Access se generaron matrices de datos, presentando ordenadas por columnas las variables dependientes, e independiente, estas se almacenaron en archivos de texto. Uno de ellos con datos de entrenamiento y el otro con datos de verificación. Se evaluaron seis archivos de datos, debido a que fueron tres los métodos empleados en la selección de variables (BE, SW y SE) y dos las salidas esperadas (CH<sub>4</sub> y LP), generando un modelo ANFIS para cada uno (Figura 3.8).

Con el fin de facilitar la construcción de los modelos ANFIS, se desarrolló una función en MATLAB, empleó el par archivos generados con las variables independientes y dependientes, con datos tanto de entrenamiento como de validación (Apéndice C). A continuación se detalla el procedimiento realizado.

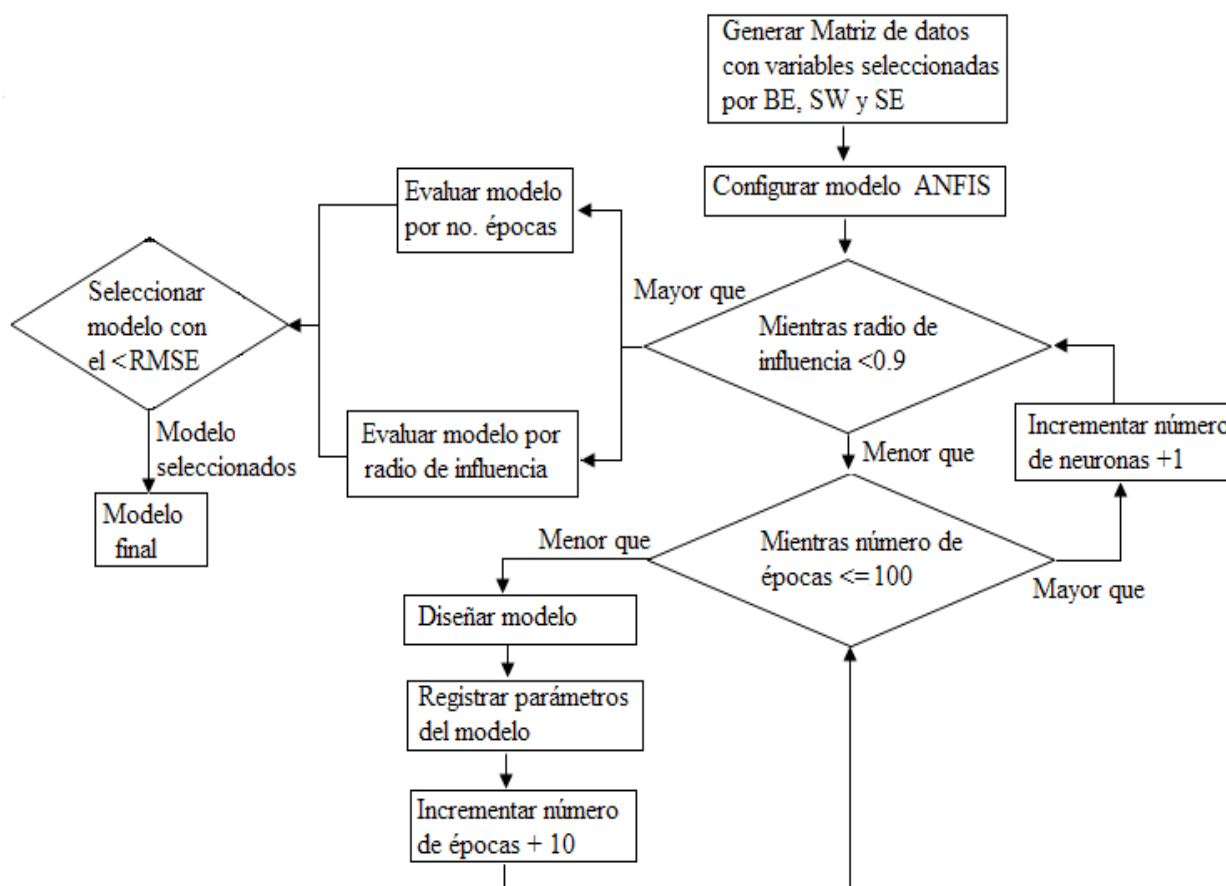


Figura 3.8 Diagrama de flujo para la formulación de los modelos por ANFIS

Se cargaron los archivos con datos de entrenamiento (EBECH4.txt; SCH4.txt), con una estructura similar a la mostrada en la figura 3.5. Los archivos se emplearon para desarrollar el sistema de inferencia difuso. Posteriormente ya unidas todos los valores fueron asignados a una variable temporal (entrena), con la que se creó un archivo (ENTBECH4.txt) usado en el desarrollo del modelo ANFIS. Este mismo procedimiento se realizó para las matrices con de validación.

```
% cargar archivos de datos de entrenamiento
load EBEC4.txt;           %variable independientes
load SCH4.txt;           %variable dependiente
entrada = EBEC4;         %matriz de entrada para FIS
salida=SC4;              %matriz de salida para FIS
entrena=[entrada salida]; %crea matriz de entrenamiento para ANFIS
save ('ENTBECH4.txt','entrena','-ascii'); %Almacena matriz con datos de entrenamiento
```

```
% cargar archivos datos de validación
load TBECH4.txt;        %variable independientes
load TCH4.txt;         %variable dependientes
entval=TBECH4;        %matriz de entrada para FIS
salval=TCH4;          %matriz de salida para FIS
valida=[entval salval]; %crea matriz de validación para ANFIS
save ('VALBECH4.txt','valida','-ascii'); %Almacena matriz con datos de entrenamiento
```

Una vez generadas las matrices de datos, con valores de las variables independientes y dependientes, se empleó la función *genfis2* para generar la estructura del sistema de inferencia difuso inicial. Mediante la implementación del agrupamiento substractivo en los datos, se logró extraer el conjunto de reglas que modelaban su comportamiento. El método de extracción de reglas, uso primero la función de agrupamiento substractivo (*subtractive clustering*) para determinar el número de reglas y las funciones de pertenencia (antecedentes) y luego usa la estimación de mínimos cuadrados para determinar las ecuaciones consecuentes de cada regla. Esta función devuelve una estructura FIS (*fismat*) que contiene un conjunto de reglas difusas tipo *Takagi-Sugeno* para cubrir el espacio de salidas. Como parámetros de configuración se tiene la matriz de entrada, salida y el radio de influencia, el cual se evaluó en un rango de [0.3 a 0.9] con incrementos de 0.1.

```
%crea el sistema de inferencia inicial
fismat = genfis2 (entrada, salida, 0.6);
```

Anterior al empleo del FIS inicial del modelo ANFIS, se evalúa la salida que producida con los datos de entrenamiento mediante la función *evalfis* de MATLAB, obteniendo una estructura del modelo ya entrenado (*fis\_entrena*).

```
%evalua la salida que produce el modelo respecto al entrenamiento
fis_entrena=evalfis (entrada, fismat);      %evaluación del modelo FIS
RMSE_entrena=norm (fis_entrena - salida) / sqrt (length(fis_entrena))  %cálculo del
                                           RMSE de entrenamiento
```

Posteriormente se generó el modelo ANFIS a partir del FIS previamente entrenado, se le asignaron como parámetros, los datos de entrenamiento, tanto variables independientes como dependientes (*entrena*), el sistema FIS (*fismat*), un vector con opciones de entrenamiento (numero de épocas, objetivo de entrenamiento, tamaño del paso inicial), un vector con opciones de mensajes, nombre del archivo con datos de validación. Con respecto al número de épocas de entrenamiento, se aplico un número prolongado de iteraciones que se encontró en un rango de [10-100] épocas con incrementos de 10, con el fin de obtener un modelo con el menor ECM sin perder interpretabilidad. Por cada época de entrenamiento el ECM fue calculado y registrado. Como puntos de parada adicionales en el entrenamiento del modelo, se consideró un error de tolerancia de 0.1 o bien cuando el número de épocas era alcanzado. Las funciones de membresía fueron de tipo Gauss (*gaussmf*), esta función depende de dos parámetros  $c$  y  $\sigma$ , el ancho y el centro respectivamente y responde a la expresión matemática (Ec. 3.1):

$$f(x; c, \sigma) = e^{-\frac{1}{2} \left[ \frac{x-c}{\sigma} \right]^2} \quad (3.1)$$

El número de las funciones de membresía fue elegido por prueba y error, en un rango entre [3-9], de manera que se mantuviera el número total de parámetros a ajustar al mínimo. El tipo de función de salida fue *constante*. El método de entrenamiento seleccionado fue el híbrido.

```
%crea modelo ANFIS
[fis,error_ent,stepsize,fismat2,error_prue] = anfis (entrena,fismat,[20 0 0.1],[],valida)
```

Una vez generado el modelo ANFIS, se evaluó la salida obtenida con respecto a los datos de prueba, obteniendo en cada caso una medida de la bondad de ajuste del modelo.

```
%evalúa la salida que produce el modelo respecto al entrenamiento  
fuzsal2=evalfis(entrada,fismat2);  
trnRMSE2=norm(fuzsal2-salida)/sqrt(length(fuzsal2))
```

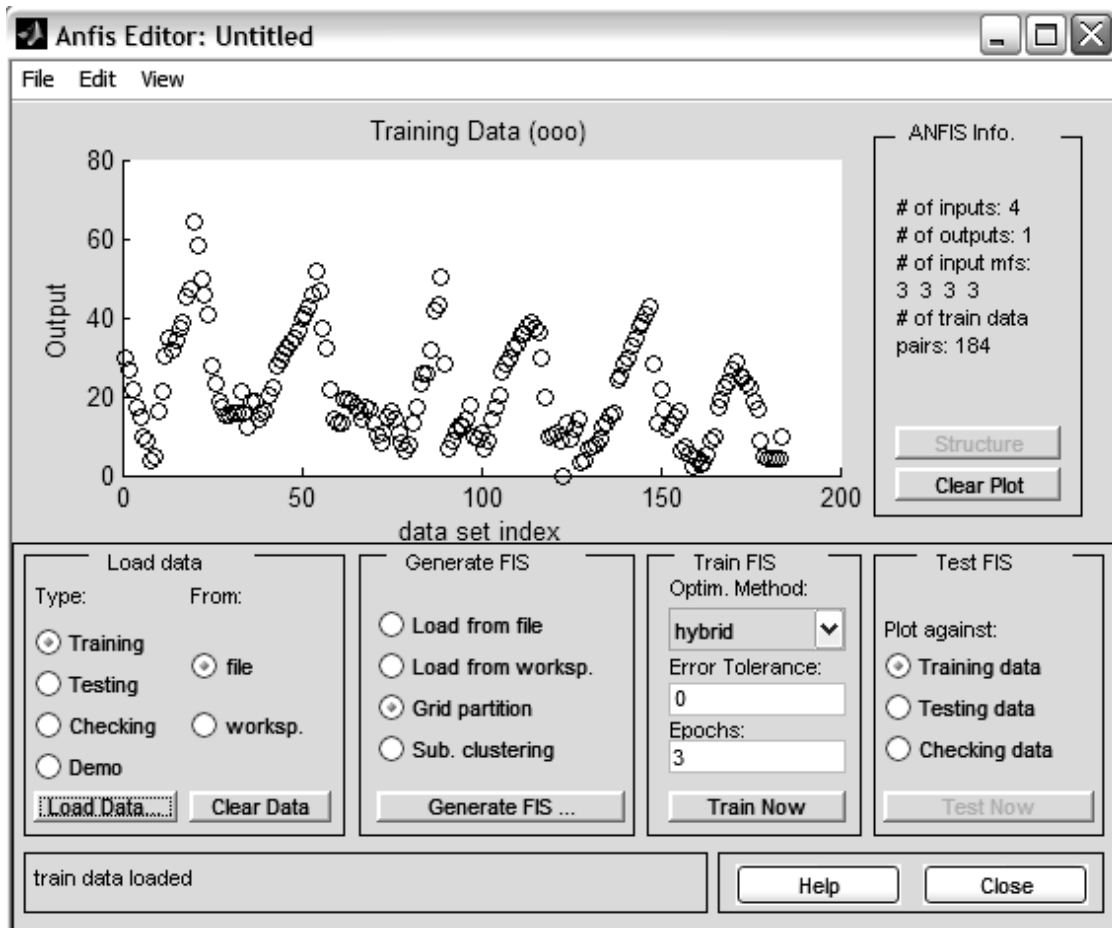
```
%error de validación con datos de entrenamiento  
bondad = evalfis(entrada,fis);  
error = salida - bondad;  
valRmse2 = sqrt(mse(error))
```

```
%error de validación mide bondad de ajuste de la recta estimada con datos de validación  
bondad = evalfis (entval,fis);  
error = salval - bondad;  
pruebRmse = sqrt (mse (error) )
```

Finalmente la salida del modelo se compara con los datos de validación mediante un análisis de regresión.

```
%compara salida del modelo vs datos de validación  
Plotregression (salval, bondad) %genera grafico de dispersión y calcula R2
```

La función anteriormente descrita fue diseñada para optimizar el proceso de selección del modelo, sin embargo este también puede ser desarrollado empleando la interfaz gráfica que ofrece MATLAB. Mediante la instrucción *anfisedit* se llama al editor ANFIS (Figura 3.9). Primeramente en la sección *load data*, se selecciona la opción *training* y se proporciona la ubicación del archivo conteniendo los datos de entrenamiento y después el botón *Load*.



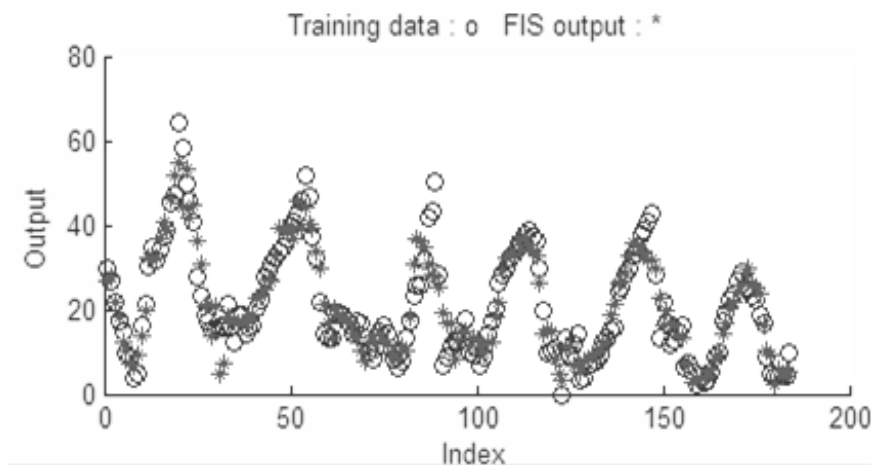
**Figura 3.9** Interfaz gráfica del editor ANFIS

Posteriormente se diseñó el sistema FIS inicial, eligiendo la opción *sub. clustering* y el botón *generate FIS*, se abre una ventana en la que se proporcionan los parámetros de configuración del modelo como son: el rango de influencia, rango de aceptación y rango de rechazo. Seguidamente el método de entrenamiento (*Backpropagation, Hybrid*) es seleccionado, en este caso es el método híbrido, seguido por el error de tolerancia (0.1) y número de épocas de entrenamiento. Para confirmar los cambios a los parámetros, se oprime el botón *Train Now*, se obtiene como resultado el modelo entrenado, en la parte inferior izquierda de la ventana se muestra el número de épocas de entrenamiento y el error generado por el modelo. Estos valores son seleccionados a prueba y error hasta que el modelo alcance el menor RMSE o antes de que ocurra un sobre-entrenamiento.

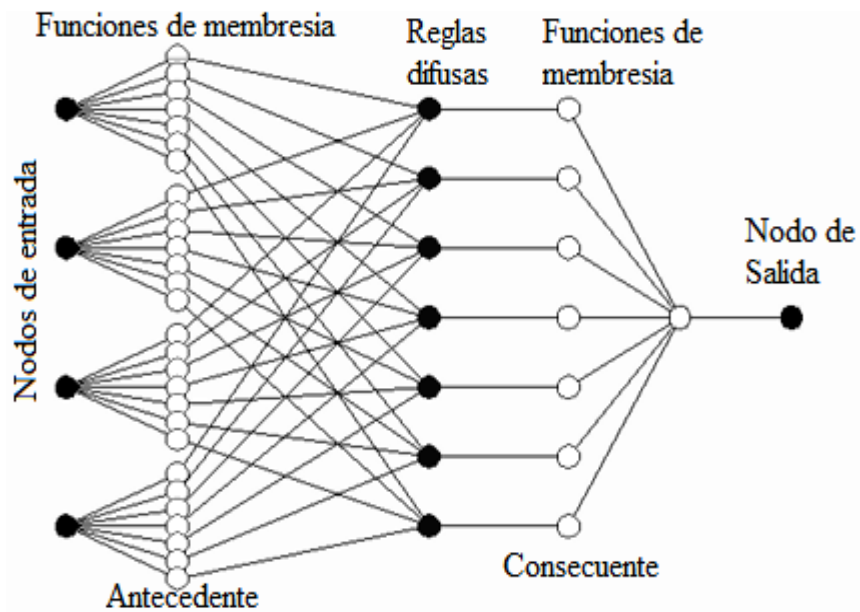
Después de generar el modelo éste debe ser validado, por lo que en la sección *load data*, se selecciona la opción *testing* para verificar el modelo y *Checking* para validarlo, en

la ventana se muestra el ajuste de los datos reales con respecto a los proporcionados por el modelo (Figura 3.10), además del valor del RMSE calculado.

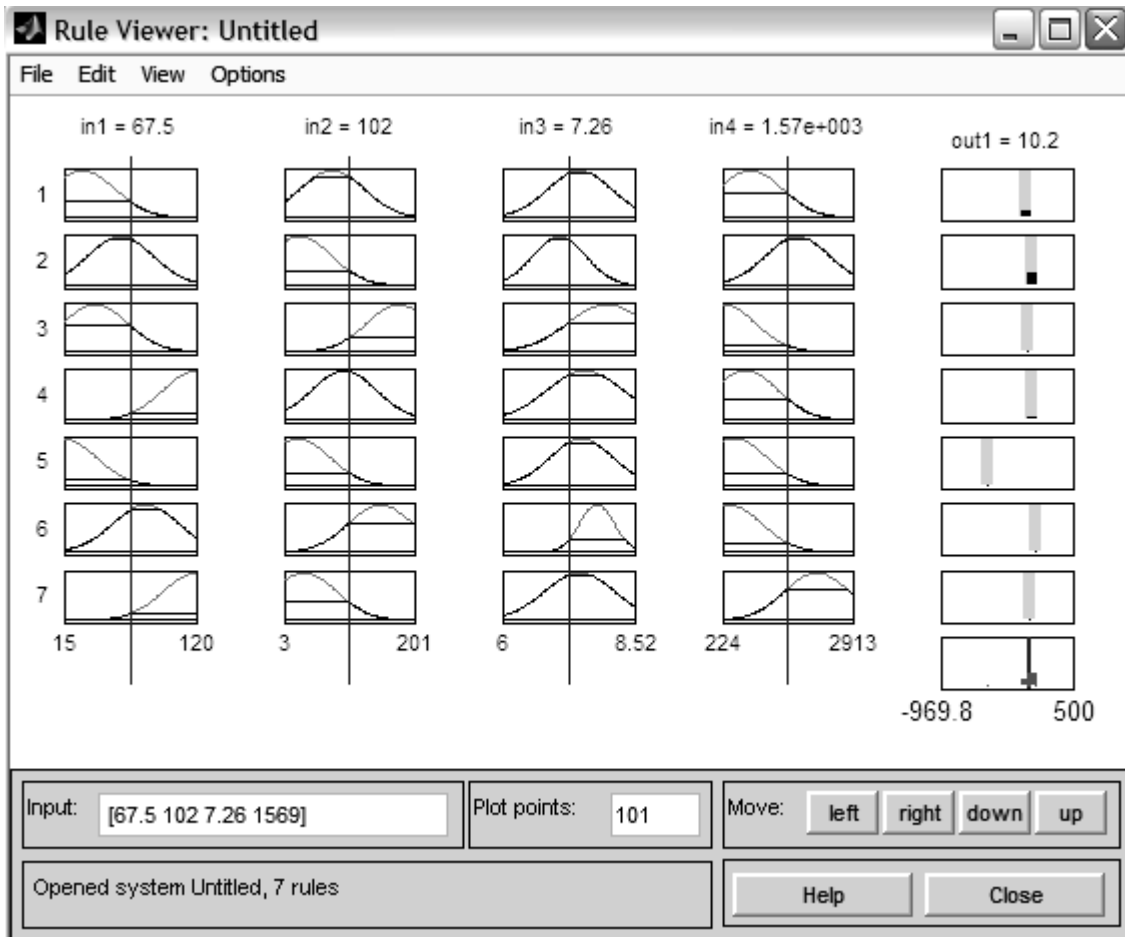
La estructura del modelo ANFIS se genera con el botón *structure* (Figura 3.11), en él se observan los nodos de entrada relacionados con las variables de independientes, las funciones de membresía antecedentes, las reglas difusas, las funciones de membresía consecuentes y el nodo de salida.



**Figura 3.10** Comparación datos de entrenamiento vs datos de validación



**Figura 3.11** Estructura del modelo ANFIS



**Figura 3.12** Reglas difusas del modelo ANFIS

Las reglas difusas relacionadas con el modelo se generan al seleccionar en la barra de menú la opción View/reglas (Figura 3.12). Se presentan ordenadas por columnas las variables independientes se observa que son tipo gaussianas, se pueden realizar simulaciones con el modelo, proporcionando una serie de datos de entrada en la parte inferior izquierda de la ventana.

### 3.5 Validación de los modelos

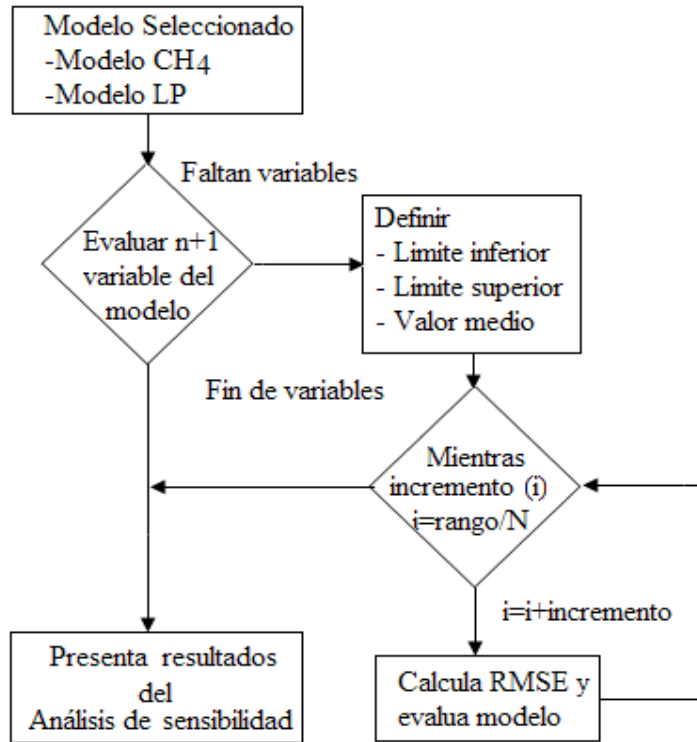
Una vez seleccionados los modelos de RNA y ANFIS para lixiviados y CH<sub>4</sub>, fue necesario evaluar su rendimiento y capacidad predictiva. Para ello se realizó una validación interna del modelo, con datos separados para su verificación (20%). La validación externa permitió medir el nivel de generalización obtenido por el modelo con respecto al grupo de datos de validación totalmente independiente de los datos de entrenamiento.

Los errores se calcularon mediante la interfaz gráfica de MATLAB. Se calculó el error cuadrático medio (ECM), cuya función es dependiente del error entre la salida real del sistema que se pretende modelar y la salida del núcleo estimador. Mediante un análisis de regresión, se midió la bondad de ajuste del modelo y los datos de prueba proporcionados. Se calculó el factor de correlación R entre los datos generados por el modelo después del entrenamiento vs los datos de validación. Al final de este paso se tiene un modelo para estimar CH<sub>4</sub> y otro para estimar LP, los cuales poseen el menor valor de RMSE y en consecuencia una mayor bondad de ajuste.

### **3.6 Análisis de sensibilidad**

Una vez seleccionados los modelos con el menor RMSE (CH<sub>4</sub> y LP) se realizó el análisis de sensibilidad a cada uno, con el fin de determinar las variables de entrada con mayor influencia en la salida del modelo. Dadas las características de los datos y los modelos construidos se eligió el análisis de sensibilidad basado en el error. Como primer paso se calculó el límite inferior y superior de cada variable, así como su valor promedio. De manera iterativa se fue variando el valor de cada variable a lo largo de su rango [límite inferior-límite superior], el incremento fue calculado dividiendo el rango entre el tamaño de la matriz de validación (n=número de datos por variable), debido a que algunas variables presentaron incrementos muy pequeños lo que daría variaciones poco significativas en el RMSE, fue necesario dividir el rango entre (n/2).

El orden de evaluación de las variables fue el mismo al asignado por el modelo. Por cada iteración fue calculado tanto el ECM como el RMSE. Se realizó un análisis fijando la variable independiente a su valor promedio, en un segundo análisis se consideró el valor real de la variable dependiente. Una vez finalizado el análisis, los resultados fueron evaluados con el fin de determinar el nivel de influencia de cada variable de entrada sobre los resultados del modelo (Figura 3.13).



**Figura 3.13** Diagrama de flujo para análisis de sensibilidad

## 4. RESULTADOS

A continuación se describen los resultados obtenidos al desarrollar la metodología definida en el capítulo 3.

### 4.1 Análisis exploratorio de los datos

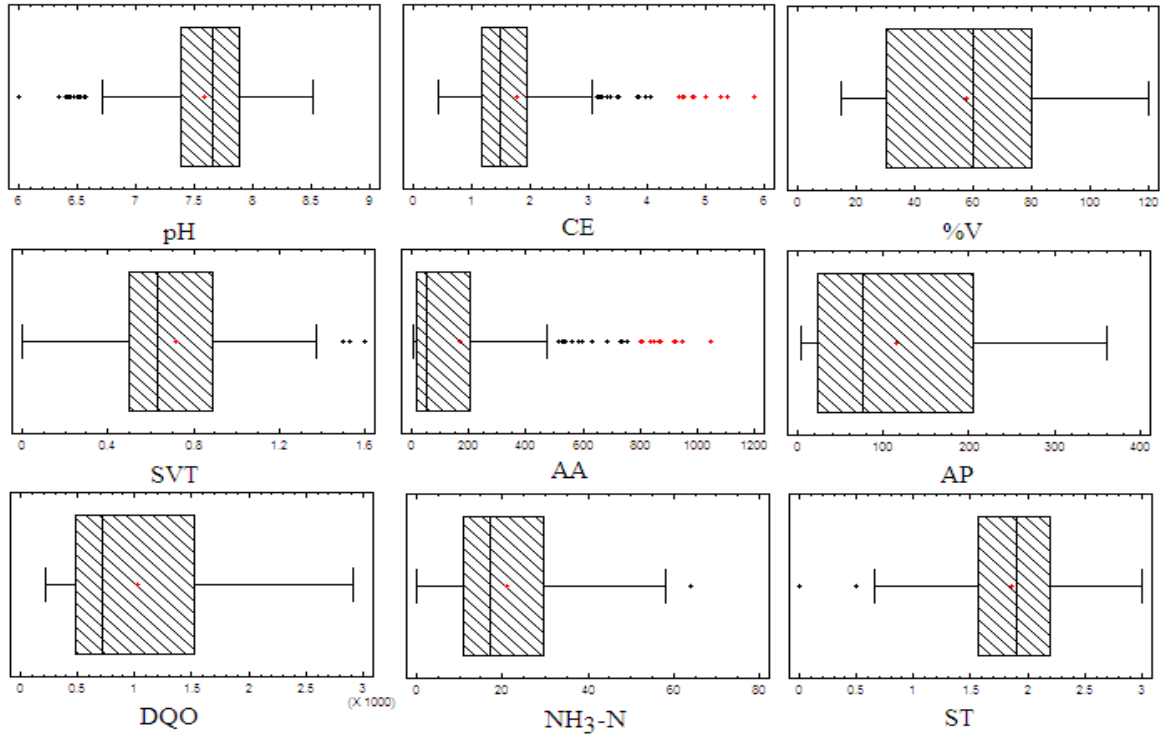
El primer grupo de variables analizadas, fueron las relacionadas con lixiviados recirculados (LR), representan la entrada al sistema. Con el análisis exploratorio se observó, que pH y ST presentaron un ligero sesgo negativo con una distribución muy cercana a la normalidad (Tabla 4.1). Las variables (CE, AA, AP, DQO, NH<sub>3</sub>-N y SVT) presentaron un sesgo positivo lo que demuestra que los datos más extremos se encuentran por encima de la media. La variable NH<sub>3</sub>-N presentó un comportamiento parecido a la curva normal, debido a que presentó un valor de curtosis muy cercano a cero (0.04), el resto de las variables presentaron un elevado grado de concentración alrededor de sus valores centrales revelando una distribución leptocúrtica. Entre las variables con menor variabilidad se encontraron pH, CE, ST y SVT. Por otra parte las variables %V, AP y DQO presentaron una elevada dispersión en sus datos. ST presentó una alta concentración en sus datos, lo contrario a AA cuyos datos mostraron una elevada variabilidad.

Por medio de los diagramas de caja (figura 4.1), se detectó la presencia de datos atípicos en las variables (pH, CE, AA, NH<sub>3</sub>-N, ST y SVT), observando que CE, AA, NH<sub>3</sub>-N y SVT los exhibieron en el límite superior mientras que pH y ST en el límite inferior.

**Tabla 4.1** Resumen de estadísticos del análisis univariado LR

<b>N=268</b>	<b>pH</b>	<b>CE</b>	<b>%V</b>	<b>AA</b>	<b>AP</b>	<b>DQO</b>	<b>NH<sub>3</sub>-N</b>	<b>ST</b>	<b>SVT</b>
<b>Media</b>	7.5	1.7	57.8	172	114	1020	20.9	1.85	0.71
<b>Mediana</b>	7.6	1.4	60	54	76	710	17.14	1.9	0.63
<b>Varianza</b>	0.2	1.1	1239	63701	11096	472636	164	0.26	0.08
<b>Desv. Std.</b>	0.4	1.0	35.2	252	105	687	12.8	0.51	0.28
<b>Asimetría</b>	-0.7	1.7	0.6	1.86	0.82	0.90	0.77	-0.3	0.86
<b>Curtosis</b>	0.5	2.8	-0.8	2.37	-0.63	-0.40	0.04	0.41	0.19
<b>CV</b>	67	59	61	146	92	67	61	27	40

Desv. Std = Desviación estándar; CV=Coeficiente de variación

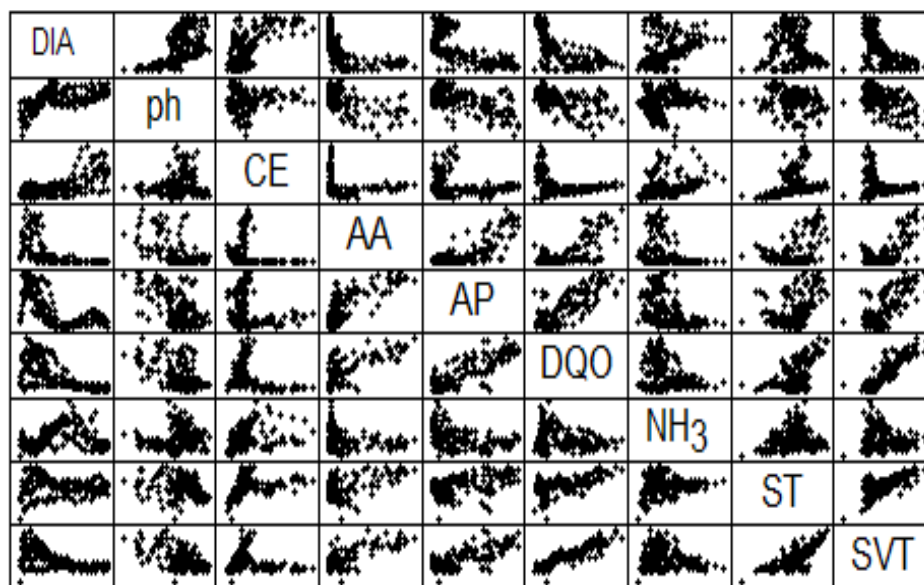


**Figura 4.1** Diagramas de caja de las variables de lixiviados recirculados (LR)

Mediante la matriz de correlación se midió la intensidad de relación entre las variables independientes (Tabla 4.2), se observó la existencia de relaciones fuertes entre DQO y SVT con un coeficiente de correlación  $r=0.91$ , las variables AA, AP y DQO mostraron una relación moderada con un  $r=0.81$  y relaciones débiles entre las variables pH y CE. A través de la matriz de dispersión (Figura 4.2) se verificó la relación lineal existente entre ST y SVT con respecto a DQO, además de una ligera relación entre AA y AP. El resto de las variables presentaron relaciones no lineales.

**Tabla 4.2** Matriz de correlación de variables de lixiviados recirculados (LR)

N=268	%V	pH	CE	AA	AP	DQO	NH <sub>3</sub> -N	ST
<b>pH</b>	-0.27							
<b>CE</b>	0.54	0.11						
<b>AA</b>	0.09	-0.57	-0.21					
<b>AP</b>	0.07	-0.56	-0.21	0.81				
<b>DQO</b>	0.26	-0.51	-0.27	0.75	0.81			
<b>NH<sub>3</sub>-N</b>	0.35	0.05	0.23	-0.28	-0.44	-0.26		
<b>ST</b>	0.63	-0.24	0.31	0.40	0.42	0.64	0.15	
<b>SVT</b>	0.28	-0.50	-0.22	0.73	0.74	<b>0.91</b>	-0.16	0.72



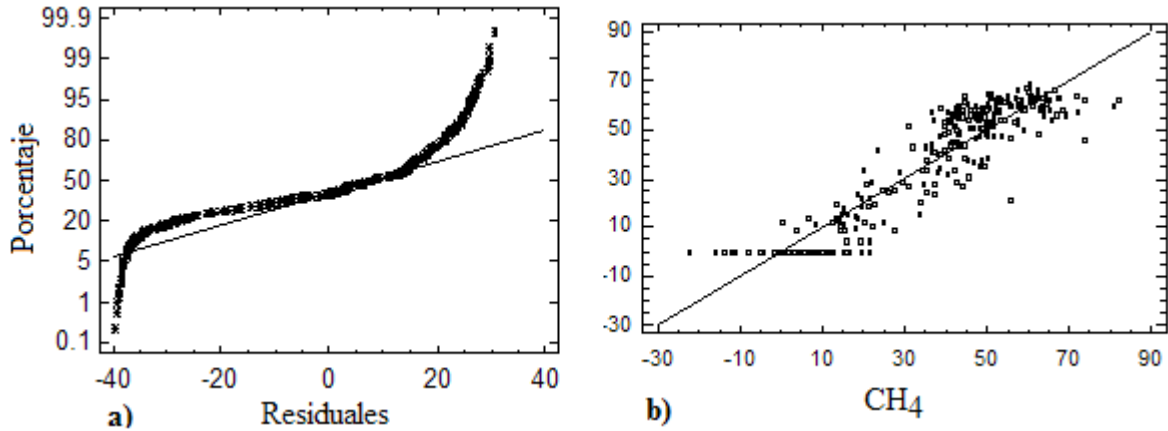
**Figura 4.2** Matriz de dispersión de variables LR

Se realizó el análisis de ANOVA con las variables de lixiviados recirculados (Tabla 4.3). Observando que la variable ST presentó el mayor número de diferencias significativas debido al tratamiento, seguida por la variable CE y NH<sub>3</sub>-N. Por otra parte AA y AP solo presentaron diferencias significativas en un tratamiento.

Posteriormente se verificó el supuesto de normalidad por medio del gráfico de probabilidad normal (figura 4.3a), se observó que los puntos seguían una tendencia lineal, especialmente en el centro, no así al inicio y al final, donde destaca la presencia de datos atípicos, por lo que se confirma que los datos cumplen los supuestos de linealidad y normalidad.

**Tabla 4.3** Análisis de ANOVA de variables LR

pH	CE	AA	AP	DQO	NH <sub>3</sub> -N	ST	SVT
15-30	15-30	30-80	15-60	15-40	15-30,	15-30, 15-40	15-30
15-40	15-40			15-60	15-40,	15- 60, 15-80	15-40
15-60	15- 60			15-80	15-80	15-120	15- 60
15-80	15- 80			15-120	15-120	30-60, 30-80	30-80
15-120	15-120			30-60	60-80	30-120	30-120
	30-120			30-80	60-120	40-80, 40-120	
	30-40				30-120	60-120	
	30-60						
	30-80						



**Figura 4.3** Gráfico de probabilidad normal

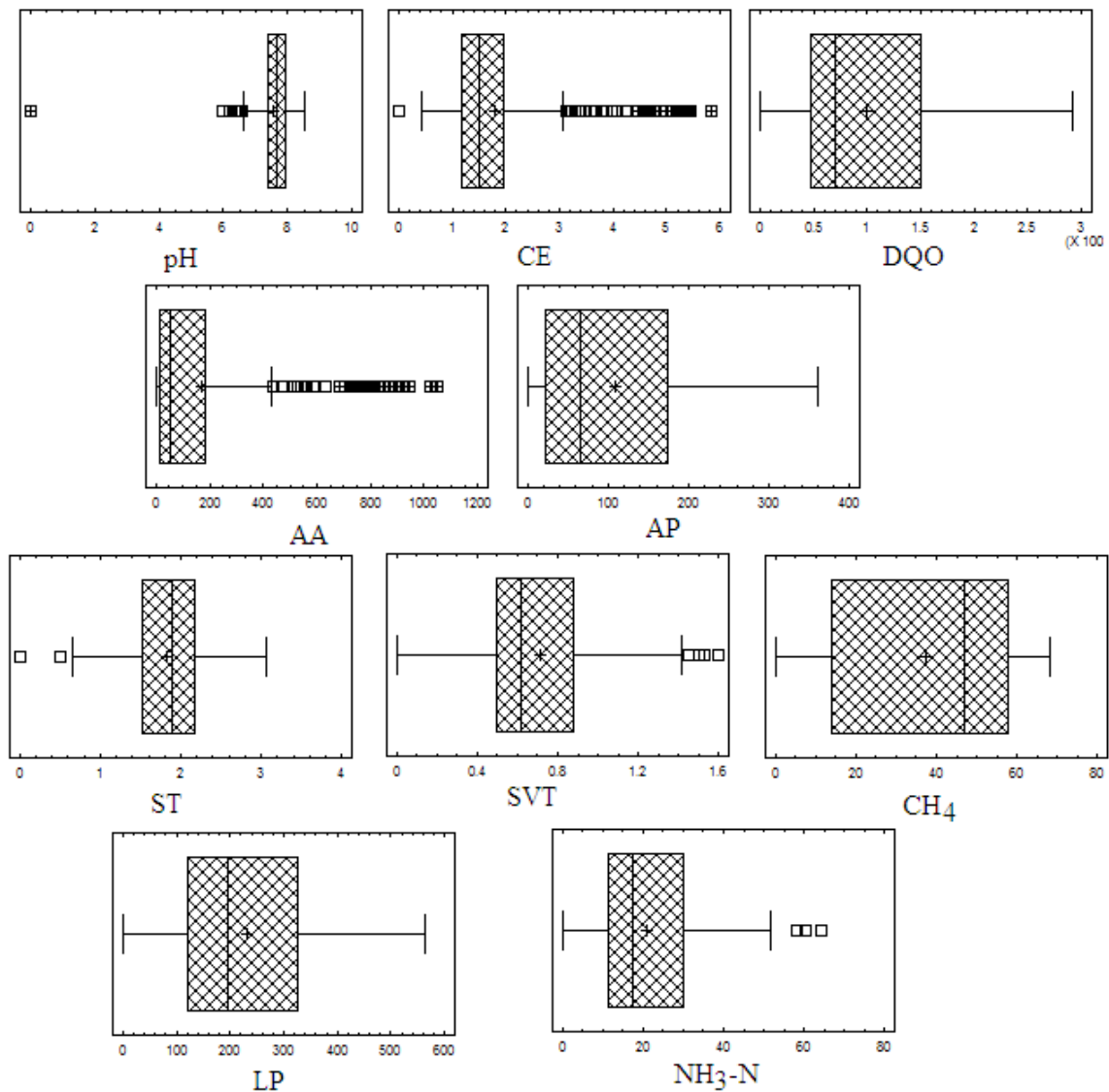
Después de analizar las variables de LR, se examinaron las variables correspondientes a lixiviados producidos (LP). Las variables DQO, AP, ST, SVT, NH<sub>3</sub>-N, LP y CE presentaron una marcada dispersión, mientras que las variables pH y CH<sub>4</sub> mostraron una reducida variabilidad. Sólo pH exhibió una asimetría negativa, el resto de las variables presentaron una marcada aglomeración de sus datos hacia la izquierda, con lo que se verifica una asimetría negativa. En este caso las variables LP y CH<sub>4</sub> presentaron un reducido grado de concentración alrededor de la media, CH<sub>4</sub> presentó mayor variabilidad en sus datos que LP (Tabla 4.4).

Con los diagramas de caja se dedujo que variables no presentaron datos atípicos como fue el caso de DQO, AP, LP y CH<sub>4</sub>, y aquellas que si los presentaron (CE, AA, SVT y NH<sub>3</sub>-N) sobre todo en el límite superior de su rango, en este caso las variables CE y AA presentaron el mayor número de datos atípicos (Figura 4.4).

**Tabla 4.4** Resumen de estadísticos del análisis univariado LP

N=268	pH	CE	AA	AP	DQO	NH <sub>3</sub> -N	ST	SVT	LP	CH <sub>4</sub>
<b>Media</b>	6.2	4.1	1027.8	231.8	6081.8	90.3	6.1	3.2	230.8	37.3
<b>Mediana</b>	6.5	2.6	174.5	91.5	1932	31.3	2.6	1.1	196.5	47.1
<b>Varianza</b>	1.1	16.3	2.8	68268	1.41E+08	43436	133.6	48.6	20853	553.1
<b>Desv. Std.</b>	1	4.1	1667	261.2	11864	208	11.6	6.9	144	23.5
<b>Asimetría</b>	-1.6	2.5	2.6	1.3	4.1	4.8	4.5	4.6	0.7	0.5
<b>Curtosis</b>	7.6	6.3	8	0.9	20.1	25.1	22.3	23.6	-0.6	-1.3
<b>CV</b>	16	100	162	112	195	230	190	218	62	63

Desv. Std = Desviación estándar CV=Coeficiente de variación



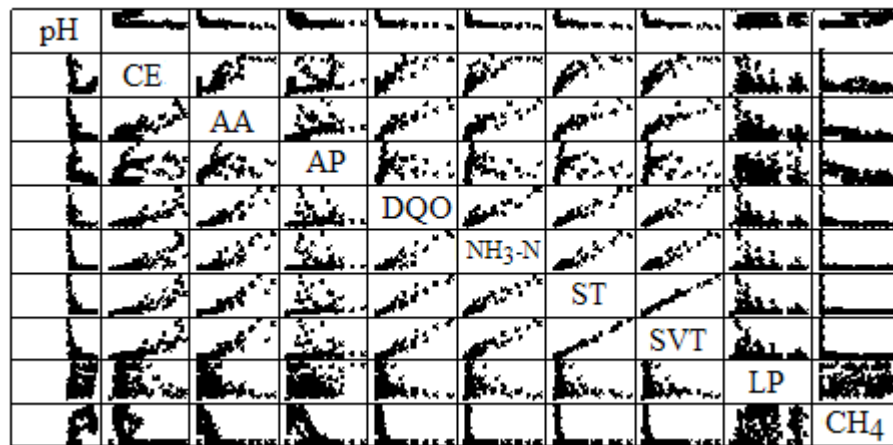
**Figura 4.4** Diagramas de caja de variables de lixiviados producidos (LP)

Las variables CE, DQO, AA, AP, NH<sub>3</sub>-N, ST, SVT y LP se caracterizaron por estar sesgadas a la derecha con valores menores a la media, caso contrario las variables CH<sub>4</sub> y pH exhibieron un sesgo a la izquierda. Se observó que las variables DQO, AP, LP y CH<sub>4</sub> presentaron datos con mayor grado de heterogeneidad que el resto de las variables. Se verificó la existencia de relaciones fuertes entre DQO, NH<sub>3</sub>-N, ST y SVT con valores de  $r > 0.90$  y relaciones moderadas entre AA y CE. Las variables con relaciones débiles fueron AP, pH y CH<sub>4</sub>. En este caso LP, CH<sub>4</sub> y pH sustentaron una correlación inversa con respecto a la mayoría de variables (Tabla 4.5).

**Tabla 4.5** Matriz de correlación de variables de LP

N=268	pH	CE	AA	AP	DQO	NH <sub>3</sub> -N	ST	SVT	LP
CE	-0.26								
AA	-0.56	0.81							
AP	-0.62	0.43	0.67						
DQO	-0.44	0.84	0.92	0.45					
NH <sub>3</sub> -N	-0.31	0.81	0.83	0.25	0.92				
ST	-0.35	0.86	0.88	0.34	0.96	0.96			
SVT	-0.37	0.84	0.88	0.35	0.96	0.96	0.99		
LP	-0.34	-0.08	-0.25	-0.13	-0.27	-0.15	-0.21	-0.24	
CH <sub>4</sub>	0.71	-0.6	-0.74	-0.8	-0.58	-0.46	-0.51	-0.51	-0.49

Con la matriz de dispersión se verificó que las variables ST, SVT, DQO y NH<sub>3</sub>-N presentaron relaciones fuertes con un alto grado de dependencia lineal (figura 4.5). Se detectó la existencia de colinealidad entre DQO y SVT debido a que se calculó un factor de correlación  $R=83.68$ , sin embargo quedó descartada al calcular un valor de  $FIV=5.83$ . Puede observarse que a excepción de estas variables, el resto de ellas no presentó una relación lineal con las variables dependientes LP y CH<sub>4</sub>.



**Figura 4.5** Matriz de dispersión de variables LP

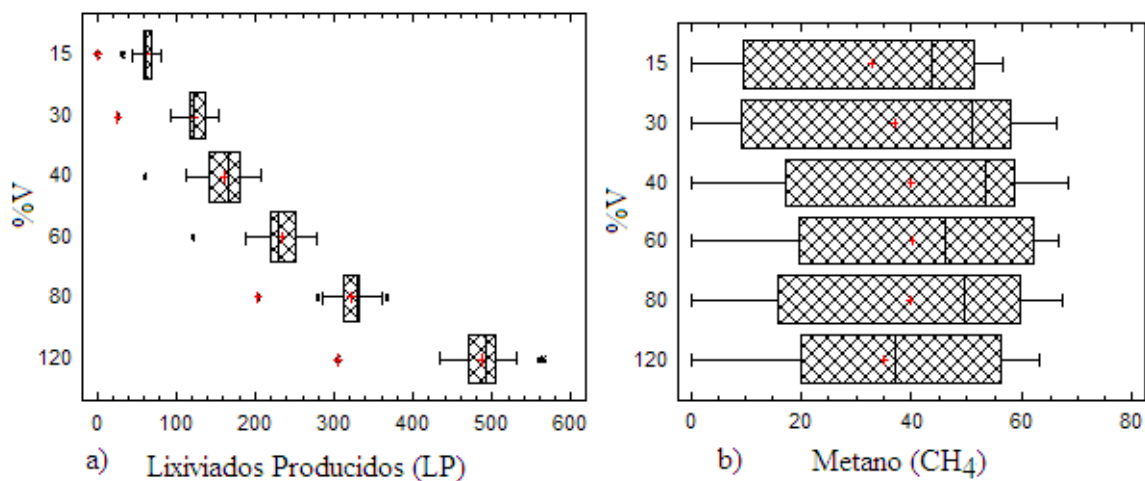
Las diferencias significativas encontradas en el análisis de ANOVA, se verificaron mediante la prueba de Tukey HDS. Observando en este caso, que pH presentó

diferencias significativas debido al tratamiento con un nivel de confianza del 95%, en los tratamientos al 15%V y 30%V, mientras que la variable LP presentó diferencias en todos los tratamientos, el resto de las variables no mostraron diferencias significativas (Tabla 4.6).

Al examinar el diagrama de caja de la variable LP, todos los tratamientos presentan diferencias significativas en sus medias, debido al tratamiento (Figura 4.6a). En los tratamientos 15, 30 y 60 %V los datos tienden a concentrarse hacia la parte superior de la distribución y se extienden más hacia la izquierda de la mediana, lo que indica que los datos están sesgados a la derecha, lo contrario a los tratamientos 40, 80 y 120 %V con un sesgo a la izquierda. La dispersión de los datos es baja y no existen valores máximos y mínimos muy alejados de la tendencia, esto indica una buena reproducibilidad puesto que los datos son muy cercanos entre sí. Con respecto al CH<sub>4</sub> (Figura 4.6b) es posible confirmar que los grupos de datos se encuentran mayormente dispersos y en su mayoría sesgados a la izquierda lo que significa que presentan mayores concentraciones hacia la parte inferior de la distribución.

**Tabla 4.6** Análisis de ANOVA para las variables de lixiviados producidos (LP)

pH	CE	AA	AP	DQO	NH <sub>3</sub> -N	ST	SVT	LP	CH <sub>4</sub>
15-60,80,120	No	No	No	No	No	No	No	Todos	No
30-120	≠	≠	≠	≠	≠	≠	≠	€ ≠	≠



**Figura 4.6** Análisis de ANOVA para a) lixiviados producidos y b) CH<sub>4</sub>

## 4.2 Selección de variables

Concluido el análisis exploratorio, se determinaron como variables independientes (entrada) las características de los lixiviados recirculados ( $\text{NH}_3\text{-N}$ , pH, AA, AP, DQO y CE) además del tratamiento representado por %V, y como variables dependientes  $\text{CH}_4$  y LP. La primera técnica usada para la selección de variables fue *stepwise*, considerando  $\text{CH}_4$  como la variable dependiente (Tabla 4.7). En la primera evaluación AP es seleccionado al obtener el mayor factor de determinación ( $R^2$ ) con un valor de 44.24, seguido por ST con un  $R^2=56.99$ , pH con un  $R^2=65.82$ ,  $\text{NH}_3\text{-N}$  con un  $R^2=72.87$  y %V=74.14. Con respecto al principio de parsimonia, no se tenía sentido incluir otra variable, si su aporte no revelaba un efecto sustancial en la variable dependiente  $\text{CH}_4$ . Posteriormente se consideró como variable dependiente a LP (Tabla 4.8). La variable %V se eligió en la primera evaluación, al presentar el mayor valor de  $R^2=97.08$ , procedida por la variable pH con  $R^2=97.23$  y la variable ST, estas tres variables explican el 97.41% del comportamiento de LP.

**Tabla 4.7** Selección de variables para estimar  $\text{CH}_4$  por método *stepwise*

%V	pH	CE	AA	AP	DQO	$\text{NH}_3\text{-N}$	ST	SVT
-0.35	38.17	6.39	33.13	44.24	15.75	32.43	-0.4	14.98
44.86	52.89	45.46	44.3		50.05	53.23	56.99	46.55
60.46	65.82	56.97	57.72		57.06	59.43		57.99
66.35		65.82	65.7		66.57	72.87		65.85
<b>74.14</b>		72.78	72.9		74.12			72.79

**Tabla 4.8** Selección de variables para estimar LP por método *stepwise*

%V	pH	CE	AA	AP	DQO	$\text{NH}_3\text{-N}$	ST	SVT
97.08	5.06	30.26	0.24	-0.17	5.96	14.1	42.3	7.95
	97.23	97.09	97.06	97.08	97.06	97.16	97.22	97.07
		97.22	97.27	97.22	97.24	97.28	97.41	97.33
		97.22	97.4	97.43	97.43	97.48		97.4

La segunda técnica usada fue *búsqueda exhaustiva*, considerando como variable dependiente CH<sub>4</sub>. Se corrió la función diseñada previamente (Apéndice A), con el fin de seleccionar cuatro variables dependientes de una combinación de nueve C (4,9) evaluando un total de 126 combinaciones. De los modelos generados se eliminaron las que presentaron un error de entrenamiento (EE) y/o error de validación (EV) con valores nulos o donde el error de validación era mayor a 100 (Apéndice D). Como primer criterio de selección, se consideraron modelos con el menor valor de EE y de EV, un segundo criterio fue descartar los modelos cuyas variables representaban un costo elevado para el experto. En la tabla 4.9, se presentan los modelos seleccionados de un total de 126, tres de ellos a pesar de que el error de entrenamiento es superior a cero, tienen un error de validación menor.

Con el fin de comparar el número de variables en los modelos, se corrió la función de búsqueda exhaustiva ahora considerando 5 variables, escogiendo los modelos con un EE=0 y el menor valor de EV. Sin embargo se observó que al agregar una variable en el modelo no había una reducción significativa en el EV, razón por la cual se seleccionó el modelo con cuatro variables (%V, pH, CE y ST). Hay que considerar que el modelo 52 tiene tanto a las variables ST como SVT, lo que no es factible dado el problema de colinealidad que presentaron en el análisis exploratorio, los modelos 25, 4, 15 poseen variables costosas.

El mismo procedimiento se desarrolló para seleccionar las variables representativas en la estimación de lixiviados producidos (LP). Se evaluaron todas las combinaciones de cuatro variables escogiendo los modelos con el menor valor de EE y EV (Apéndice E). En la tabla 4.10 se presenta el número de modelo seleccionado, las variables asignadas, el EE y EV.

**Tabla 4.9** Evaluación de modelos para CH<sub>4</sub> con cuatro variables por BE

No.	Variables	Error Entrenamiento	Error Validación
15	%V pH AP SVT	0.000	1.883
33	%V CE DQO SVT	0.000	1.883
52	%V AP ST SVT	0.000	1.883
5	%V pH CE ST	0.029	0.127
25	%V CE AA ST	0.029	0.127
4	%V pH CE NH <sub>3</sub> -N	0.095	0.179

**Tabla 4.10** Evaluación de modelos para LP con cuatro variables por BE

No.	Variables	Error Entrenamiento	Error Validación
33	%V CE DQO SVT	0	1.883
52	%V AP ST SVT	0	1.883
5	%V pH CE ST	0.0343	0.1262
25	%V CE AA ST	0.0343	0.1262
4	%V pH CE NH <sub>3</sub> -N	0.1069	0.1988

Los dos primeros modelos tienen un EE=0 sin embargo el valor de EV es alto en comparación al resto de los modelos, si se considera el costo que representa obtener ciertas variables sería descartado el último modelo (4), por lo que es seleccionado el tercer modelo integrado por las variables %V, pH, CE y ST. Con respecto a los modelos con cinco variables, se observó que al incrementar una variable, el EV no se reducía.

Finalmente se eligieron un conjunto de variables por cada una de las técnicas implementadas (Tabla 4.11). Se observó que la metodología *stepwise* (SW) fue más fácil de implementar que la búsqueda exhaustiva (BE), el proceso de selección de BE fue lento debido a la cantidad de modelos que se debieron evaluar, sin embargo la función desarrollada facilitó dicho proceso. Otro aspecto que hay que mencionar, es que con BE se seleccionaron las mismas variables (%V, pH, CE, ST) tanto para CH<sub>4</sub> como para LP, y que además poseen una gran similitud con las variables seleccionadas por el experto (%V, pH, CE, DQO). Con respecto a la técnica SW se observan discrepancias tanto en el número de variables seleccionadas, como en las variables, lo anterior se podría justificar dado que la variable LP presentó menos dispersión en sus datos en comparación a CH<sub>4</sub>.

**Tabla 4.11** Resumen de variables seleccionadas por SW, BE y SE

Modelo	Técnica	Variable dependiente	Variables independientes
MSW_CH <sub>4</sub>	SW	CH <sub>4</sub>	AP, ST, pH, NH <sub>3</sub> -N, %V
MBE_CH <sub>4</sub>	BE	CH <sub>4</sub>	%V, pH, CE, ST
MSW_LP	SW	LP	%V, pH, ST
MBE_LP	BE	LP	%V, pH, CE, ST
MSE	SE	LP / CH <sub>4</sub>	%V, pH, CE, DQO

### 4.3 Modelos con una red neuronal

Los primeros modelos desarrollados fueron empleando una red neuronal tipo feedforward), se evaluaron tres modelos para la estimación de CH<sub>4</sub> y tres para LP. En cada caso se asignó un número inicial de neuronas igual al número de variables involucradas en el modelo, se evaluó un rango desde 10 hasta 50 épocas con incrementos de 10, así sucesivamente hasta alcanzar un número máximo de 10 neuronas.

Por cada modelo se calculó el error cuadrático medio (ECM) y la raíz del error cuadrático medio (RMSE) tanto de entrenamiento como de validación. El primer modelo que se evaluó (MSW\_CH<sub>4</sub>), se integró por las variables independientes (AP, ST, pH, NH<sub>3</sub>-N, %V) seleccionada por la técnica *stepwise* y la variable dependiente CH<sub>4</sub> (Tabla 4.12), se observó que el incremento en el número de épocas de entrenamiento no fue un factor determinante en la reducción del ECM. Dado que el modelo con los menores valores en ECM (45.83) y RMSE (6.77) quedo conformada por 9 neuronas y 30 épocas entrenamiento (Apéndice F).

Posteriormente se evaluó el modelo con variables independientes %V, pH, CE y ST seleccionadas por búsqueda exhaustiva (MBE\_CH<sub>4</sub>) con un rango de neuronas [4-10], y por cada una se evaluó el intervalo de épocas [10-50] (Tabla 4.13), finalmente se seleccionaron los modelos con los menores errores de validación (ECM y RMSE), el modelo con 7 neuronas y 10 épocas de entrenamiento presentó el menor RMSE (7.37) Con esta técnica se observa una reducción sustancias tanto en el número de épocas como de neuronas (Apéndice G).

**Tabla 4.12** Evaluación del modelo MSW\_CH<sub>4</sub>

No Neuronas	Épocas	ECM_E	RMSE_E	ECM_V	RMSE_V
5	30	82.24	9.07	94.87	9.74
6	30	35.99	6.00	57.43	7.58
7	40	79.85	8.94	98.13	9.91
8	10	42.74	6.54	68.78	8.29
<b>9</b>	<b>30</b>	<b>22.15</b>	<b>4.71</b>	<b>45.83</b>	<b>6.77</b>
10	20	42.95	6.55	83.40	9.13

**Tabla 4.13** Evaluación del modelo MBE\_CH<sub>4</sub>

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	10	93.13	9.65	110.05	10.49
5	40	65.08	8.07	77.22	8.79
6	20	83.25	9.12	106.32	10.31
<b>7</b>	<b>10</b>	<b>59.02</b>	<b>7.68</b>	<b>54.33</b>	<b>7.37</b>
8	50	77.4	8.8	75.24	8.67
9	30	58.4	7.64	102.41	10.12
10	30	46.26	6.8	71.78	8.47

Posteriormente se realizó la evaluación del modelo MSE\_CH<sub>4</sub> integrado con las variables %V, pH, CE, DQO (Tabla 4.14). Los valores en RMSE presentan un ligero descenso, con respecto al anterior modelo, en este caso el modelo con ocho neuronas y 50 épocas de entrenamiento presentó el menor RMSE con un valor de 7.09 (Apéndice H).

Al comparar los tres modelos para estimar CH<sub>4</sub> (Tabla 4.15), el modelo MSE\_CH<sub>4</sub> requiere mayor entrenamiento (50 épocas) para reducir el RMSE, aunque el incremento de neuronas no es sustancial, es necesario para reducir el error, lo que puede inferir una demanda mayor procesamiento (9 neuronas) para alcanzar un menor ECM. Por lo anteriormente expuesto, es el modelo MBE\_CH<sub>4</sub> el que demanda menor cantidad de recursos.

**Tabla 4.14** Evaluación del modelo MSE\_CH<sub>4</sub>

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	50	82.8	9.1	77.7	8.81
5	10	87.22	9.34	108.75	10.43
6	20	48.07	6.93	65.87	8.12
7	10	54.28	7.37	75.78	8.71
<b>8</b>	<b>50</b>	<b>25.84</b>	<b>5.08</b>	<b>50.29</b>	<b>7.09</b>
9	30	35.25	5.94	61.07	7.81
10	50	25.09	5.01	55.08	7.42

**Tabla 4.15** Evaluación de los modelos de RNA para estimar CH<sub>4</sub>

Modelo	Neuronas	Épocas	ECM_V	RMSE_V
MSW_CH <sub>4</sub>	9	30	45.83	6.77
<b>MBE_CH<sub>4</sub></b>	<b>7</b>	<b>10</b>	<b>54.33</b>	<b>7.37</b>
MSE_CH <sub>4</sub>	8	50	50.29	7.09

Con respecto a la estimación de LP, el primer modelo evaluado fue el MSW\_LP (Apéndice I), compuesto por las variables %V, pH, ST. Se observó que se requirieron 40 épocas de entrenamiento y cinco neuronas para alcanzar un RMSE=28.08 (Tabla 4.16). En el caso del modelo MBE\_LP (%V, pH, CE y ST) se observó un descenso en el número de épocas de entrenamiento a diez (Tabla 4.17), sin embargo se incrementó el número de neuronas a ocho en la capa oculta (Apéndice J).

Por último, se evaluó el modelo MSE\_LP (HUM, pH, CE y DQO), se observó que requirió el mismo número de épocas que el modelo anterior (Apéndice K), además presentó una reducción en el número de neuronas, a solo ocho con un valor de RMSE=25.34 (Tabla 4.18).

**Tabla 4.16** Evaluación del modelo MSW\_LP

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	10	467.07	21.61	935.54	30.59
<b>5</b>	<b>40</b>	<b>459.18</b>	<b>21.43</b>	<b>788.43</b>	<b>28.08</b>
6	40	440.44	20.99	943.15	30.71
7	20	474.53	21.78	1104.81	33.24
8	10	355.67	18.86	1030.14	32.10
9	10	382.09	19.55	1000.2	31.63
10	10	366.09	19.13	897.15	29.95

**Tabla 4.17** Evaluación del modelo MBE\_LP

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	10	524.48	22.9	752.18	27.43
5	10	479.87	21.91	1114.36	33.38
6	30	407.94	20.2	855.44	29.25
7	10	418.14	20.45	909.58	30.16
<b>8</b>	<b>10</b>	<b>419.9</b>	<b>20.49</b>	<b>706.22</b>	<b>26.57</b>
9	10	326.95	18.08	1089.64	33.01
10	10	395.38	19.88	842.63	29.03

**Tabla 4.18** Evaluación del modelo MSE\_LP

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	20	402.54	20.06	795.12	28.20
5	30	438.49	20.94	859.54	29.32
<b>6</b>	<b>10</b>	<b>456.50</b>	<b>21.37</b>	<b>641.98</b>	<b>25.34</b>
7	10	418.75	20.46	960.71	31.00
8	10	379.40	19.48	883.88	29.73
9	10	469.27	21.66	808.33	28.43
10	50	426.7	20.66	789.81	28.10

**Tabla 4.19** Evaluación de los modelos para estimar LP

Modelo	Neuronas	Épocas	ECM_V	RMSE_V
MSW_LP	5	40	788.43	28.08
MBE_LP	8	10	<b>706.22</b>	<b>26.57</b>
<b>MSE_LP</b>	<b>6</b>	<b>10</b>	<b>641.98</b>	<b>25.34</b>

De los tres modelos evaluados para estimar LP, el modelo MSE\_LP integrado por las variables asignadas por el experto, fue seleccionado debido a que presentó un menor valor de ECM (641.98) y RMSE (25.34), constituido por cuatro variables, seis neuronas y 10 épocas de entrenamiento (Tabla 4.19). Se observa en este caso que a pesar de incrementar las épocas de entrenamiento y disminuir el número de neuronas no se alcanza una reducción significativa del RMSE.

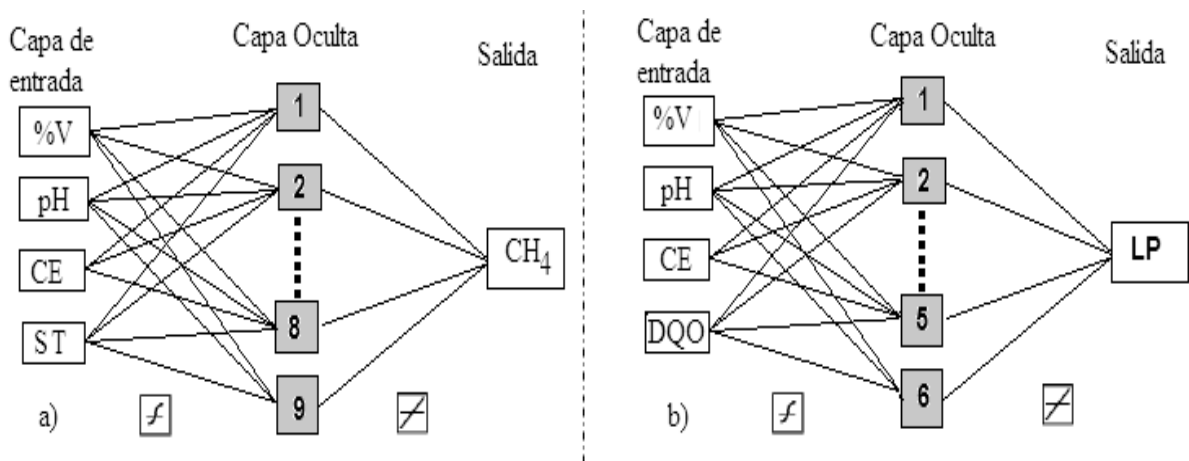
#### 4.3.1 Construcción del modelo

Una vez seleccionados los modelos para estimar CH<sub>4</sub> y LP, así como los parámetros de configuración (épocas de entrenamiento, número de neuronas y variables) y un menor valor de ECM y RMSE. El modelo para estimar CH<sub>4</sub>, quedó definido por las variables (%V, pH, CE y ST) seleccionadas mediante búsqueda exhaustiva (MBE\_CH<sub>4</sub>). Con respecto al modelo para estimar lixiviados (MSE\_LP), las variables que lo constituyen son %V, pH, CE y DQO, ésta fueron propuestas por el experto (Hernández-Berriel *et al.*, 2010).

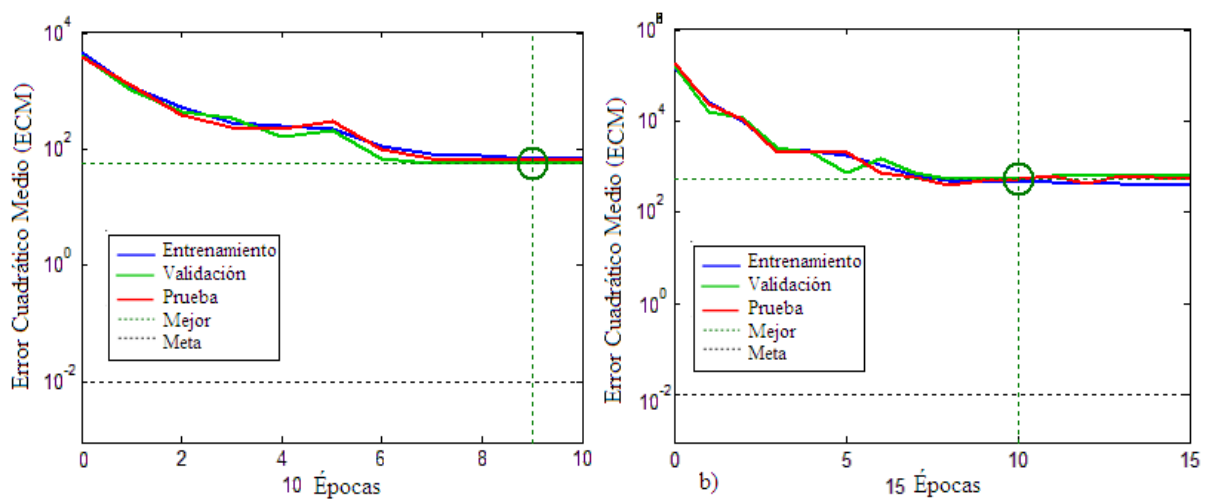
La estructura óptima del modelo de red para MBE\_CH<sub>4</sub> quedó definida por cuatro entradas, siete neuronas en la capa oculta y una en la capa de salida, con 10 épocas de entrenamiento (figura 4.7a). Con respecto al modelo MSE\_LP fue construido a partir de cuatro entradas y seis neuronas en la capa oculta, una neurona capa de salida y 10 épocas de

entrenamiento (Figura 4.7b), ambos modelos requirieron una función de transferencia tangente sigmoidea para la capa oculta y una función de transferencia lineal en la capa de salida, con el fin de verificar la calidad de predicción de los modelos se realizó una comparación entre los errores de entrenamiento y de validación (Figura 4.8).

La prueba evidenció que no existía sobre entrenamiento de la red, dado que la curva del error de prueba no sufría un incremento significativo antes que la curva del error de validación. El modelo MBE\_CH<sub>4</sub> alcanza su mayor rendimiento en la época nueve y el MSE\_LP en la época diez.



**Figura 4.7** Estructura de red modelo a) MBE\_CH<sub>4</sub> y b) MSE\_LP



**Figura 4.8** Validación del modelo a) MBE\_CH<sub>4</sub> y b) MSE\_LP

### 4.3.2 Verificación y validación del modelo

Se realizó un análisis de regresión entre la salida proporcionada por la red con respecto a los datos de entrenamiento (70%), validación (15%) y prueba (15%). Se observó la presencia de fuertes relaciones entre la salida de la red y los datos empleados en su entrenamiento, verificación y prueba (Figura 4.9).

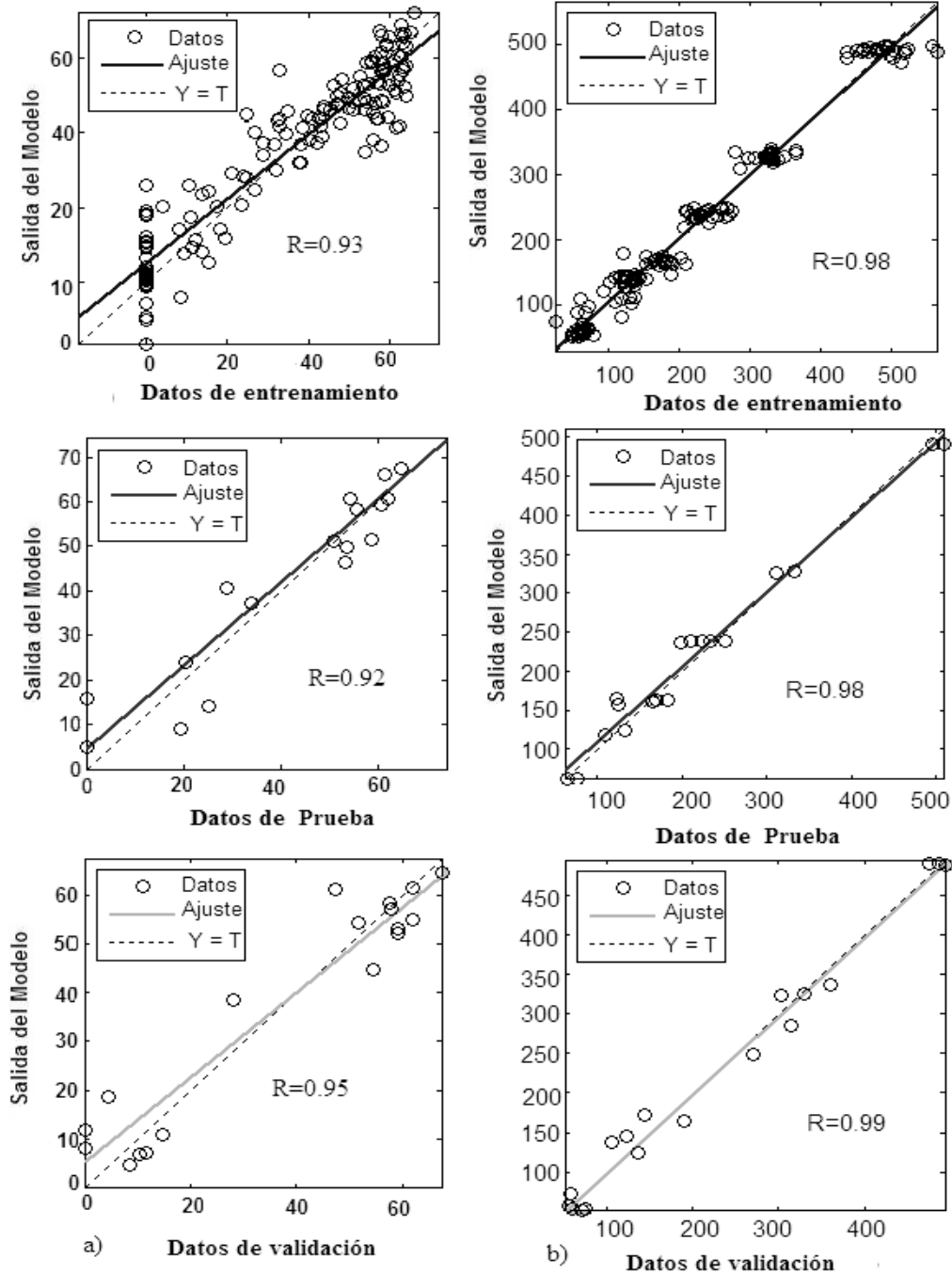
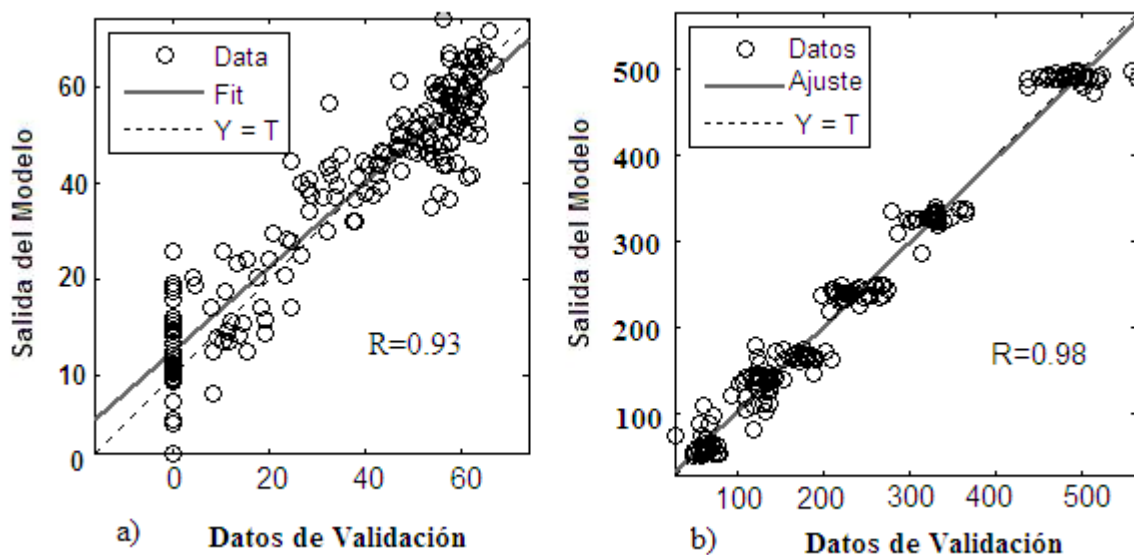


Figura 4.9 Análisis de regresión del modelo a) MBE\_CH4, b) MSE\_LP

La línea de puntos representa el resultado de la validación, la línea sólida representa el mejor ajuste de la línea de regresión entre los resultados de la red y los datos de validación. Se observa que el valor del factor de correlación del modelo MBE\_CH<sub>4</sub> es menor tanto en el entrenamiento como en la verificación, sin embargo asciende en la etapa de validación. En cuanto al modelo MSE\_LP los valores calculados en el factor de correlación se acercan a uno, lo que indica que el modelo presenta una mayor bondad de ajuste.

Dado lo anterior fue evaluada su validez, para ello se empleó el 20% de los datos de validación, por medio de un análisis de regresión entre la salida de la red y los datos de validación (figura 4.10), se calculó un factor de determinación de  $R^2=0.93$  y un RMSE=9.62 para CH<sub>4</sub>. Con respecto al MSE\_LP, presentó un  $R^2=0.98$  y un RMSE=30.14. Significa que la red explica el 93% del comportamiento de los datos para CH<sub>4</sub> y un 98% para LP.



**Figura 4.10** Validación de los modelos a) MBE\_CH<sub>4</sub> y b) MSE\_LP

## 4.4 Modelo ANFIS

El modelo para CH<sub>4</sub> fue elegido considerando criterios como el error de entrenamiento y validación, y ciertos parámetros de configuración, como el número de reglas difusas y radio de influencia, se evaluó la bondad de ajuste del modelo mediante el cálculo del ECM y RMSE para ello se desarrolló una función que facilitó su generación (Apéndice C). Se evaluó el rango del radio de influencia (RI) entre [0.5 a 0.8] con incrementos de 0.05, así como del número de épocas de entrenamiento en un rango de [10 a 50] con incrementos de 10, en cada caso se eligió el modelo con el menor RMSE.

El primer modelo ANFIS desarrollado fue el MSW\_CH<sub>4</sub>, integrado por las variables AP, ST, pH, NH<sub>3</sub>-N y %V (Apéndice L). El modelo presentó el menor RMSE (4.35) con un radio de influencia de (0.6), con 20 épocas de entrenamiento y nueve reglas difusas (Tabla 4.20). Se verificó que cuando el RI disminuye, el número de reglas difusas (RD) tiende a incrementarse. Posteriormente se analizó el modelo MBE\_CH<sub>4</sub> integrado por las variables %V, pH, CE y ST (Apéndice M). El modelo con mejores características, estuvo constituido por seis RD, un RI=0.55 y 30 épocas de entrenamiento, con un RMSE=10.47. Se observó una notoria reducción en el número de RD, pero no así en el número de épocas de entrenamiento con respecto al modelo anterior (Tabla 4.21).

Por último se evaluó el modelo MSE\_CH<sub>4</sub>, configurado por las variables %V, pH, CE, DQO (Apéndice N). No se observó una reducción en los errores, sin embargo el modelo con el menor RMSE (7.23) estuvo configurado por seis RD y un RI=0.6 con 30 épocas de entrenamiento (Tabla 4.22).

**Tabla 4.20** Evaluación modelo MSW\_CH<sub>4</sub> para estimar metano

No Reglas	RI	No Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
13	0.50	40	2.41	3.96	7.19	6.50
10	0.55	50	3.32	5.61	7.19	6.81
<b>9</b>	<b>0.60</b>	<b>20</b>	<b>4.28</b>	<b>6.07</b>	<b>6.79</b>	<b>4.35</b>
9	0.65	40	3.73	6.07	7.19	6.67
6	0.70	30	5.61	8.03	7.19	6.79
5	0.75	50	5.82	8.43	7.19	7.08
4	0.80	40	6.94	8.86	7.19	7.49

**Tabla 4.21** Evaluación modelo MBE\_CH<sub>4</sub> para estimar metano

No Reglas	RI	No Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
7	0.5	50	8.46	10.63	10.38	10.63
<b>6</b>	<b>0.55</b>	<b>30</b>	<b>9.12</b>	<b>11.74</b>	<b>10.47</b>	<b>10.47</b>
5	0.6	40	10.14	11.61	11.62	11.65
4	0.65	50	10.65	11.79	13.23	13.23
4	0.7	50	10.68	11.78	13.14	13.21
4	0.75	50	10.69	11.77	13.01	13.01
4	0.8	50	10.72	11.76	12.92	12.92

**Tabla 4.22** Evaluación del modelo MSE\_CH<sub>4</sub> para estimar metano

No Reglas	Radio	No Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
9	0.50	10	5.82	6.75	7.19	7.47
7	0.55	10	6.21	7.31	7.19	7.39
<b>6</b>	<b>0.60</b>	<b>30</b>	<b>6.36</b>	<b>8.49</b>	<b>7.19</b>	<b>7.23</b>
6	0.65	40	5.80	8.59	7.19	7.45
5	0.70	20	6.96	9.77	7.19	7.57
5	0.75	40	6.71	9.48	7.19	7.98
4	0.80	50	7.34	10.59	7.19	8.19

De los tres modelos revisados, el modelo MSW\_CH<sub>4</sub> presentó el menor RMSE (4.35), con 20 épocas, cinco variables y un RI=0.60, sin embargo posee un número elevado de RD (9) en comparación a los otros modelos, considerando el principio de parsimonia, este modelo no es sencillo ya que consta de cinco variables, es por ello que se descarta y en su lugar se selecciona el modelo MSE\_CH<sub>4</sub> configurado por seis RD, un RI=0.60 y cuatro variables de entrada (Tabla 4.23).

Posteriormente fueron evaluados los modelos para estimar lixiviados producidos (LP), el primer modelo que se revisó fue MSW\_LP configurado con las variables %V, pH, ST (Apéndice O). Presentó un RMSE de 29.04 con un RI=0.75 y dos RD (Tabla 4.24), sin embargo presento un número muy reducido de reglas difusas (2) lo que pudiera representar una limitación en la interpretabilidad del modelo, por lo que fue descartado y en su lugar se eligió el modelo con tres RD.

**Tabla 4.23** Evaluación de modelos ANFIS para estimar CH<sub>4</sub>

Modelo	No Reglas	Radio	No Épocas	ECM_V	RMSE_V
MSW_CH <sub>4</sub>	9	0.6	20	<b>6.79</b>	<b>4.35</b>
MBE_CH <sub>4</sub>	6	0.55	30	10.47	10.47
MSE_CH <sub>4</sub>	6	0.6	30	<b>7.19</b>	<b>7.23</b>

**Tabla 4.24** Modelos MSW\_LP desarrollado para estimar lixiviados producidos

No Reglas	Radio	No Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	0.50	10	18.68	21.18	29.85	31.60
4	0.55	10	18.80	21.18	30.29	31.76
4	0.60	50	18.29	21.20	29.46	31.37
<b>3</b>	<b>0.65</b>	<b>30</b>	<b>19.61</b>	<b>21.79</b>	<b>29.51</b>	<b>30.57</b>
2	0.70	50	20.84	22.19	28.54	29.06
2	0.75	50	20.58	22.22	28.50	29.04
2	0.80	30	20.91	22.26	29.10	29.10

El modelo MBE\_LP se construyó con las variables %V, pH, CE, ST (Apéndice P), cuya estructura estuvo conformada por cinco RD, un RI=0.65 y 10 épocas de entrenamiento. Dado su elevado número de épocas de entrenamiento y número de reglas difusas tuvo que ser descartado y sustituido por el modelo con un RI=0.65 y cinco RD (Tabla 4.25). Por último se seleccionó el modelo MSE\_LP integrado con las variables %V, pH, CE, DQO (Apéndice Q), diseñado con una estructura de cuatro RD, 20 épocas de entrenamiento, un RI=0.65 y un RMSE=28.14 (Tabla 4.26).

**Tabla 4.25** Modelos MBE\_LP desarrollado para estimar lixiviados producidos

No Reglas	Radio	No Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
6	0.5	20	18.31	20.62	31.84	31.84
6	0.55	10	19.42	21.07	31.78	32.18
5	0.6	10	19.94	20.98	33.1	33.63
<b>5</b>	<b>0.65</b>	<b>10</b>	<b>21.23</b>	<b>22.52</b>	<b>29.18</b>	<b>29.61</b>
4	0.7	10	21.43	22.48	29.28	29.82
4	0.75	30	20.58	22.68	30.35	30.48
3	0.8	10	22.18	22.92	29.41	29.8

**Tabla 4.26** Modelos MSE\_LP desarrollado para estimar lixiviados producidos

No Reglas	Radio	No Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
7	0.50	10	18.61	19.96	30.77	32.42
6	0.55	10	18.89	20.14	30.25	30.76
5	0.60	20	19.61	20.97	28.72	28.96
<b>4</b>	<b>0.65</b>	<b>20</b>	<b>19.99</b>	<b>21.83</b>	<b>27.96</b>	<b>28.14</b>
4	0.70	10	20.85	21.57	29.08	29.72
3	0.75	20	21.11	22.32	28.05	28.16
3	0.80	10	21.74	22.35	29.06	29.52

**Tabla 4.27** Evaluación de modelos ANFIS para estimar LP

Modelo	No Reglas	Radio	No Épocas	ECM_V	RMSE_V
MSW_LP	3	0.65	30	29.51	30.57
MBE_LP	5	0.65	10	29.18	29.61
<b>MSE_LP</b>	<b>4</b>	<b>0.65</b>	<b>20</b>	<b>27.96</b>	<b>28.14</b>

Los tres modelos tienen un RI=0.65, el modelo MSW\_LP requiere mayor cantidad de épocas de entrenamiento, en comparación al resto de los modelos, el modelo MBE\_LP no se encuentra muy alejado del MSE\_LP, solo se diferencian por una RD más. Se puede inferir que a menor cantidad de RD mayor número de épocas de entrenamiento (Tabla 4.27). Después de valorar los tres modelos, se eligió al modelo MSE\_LP, debido a que presentó el menor error de validación RMSE=28.14, con 20 épocas de entrenamiento, cuatro reglas difusas, además de que las variables que lo integran no representan un costo elevado para el experto.

#### 4.4.1 Construcción del modelo

Como primer paso en la construcción de los modelos ANFIS, se realizó el diseño del sistema de inferencia difuso (FIS), partiendo de los datos de configuración obtenidos en la etapa anterior. Para ello fue necesario proporcionar los valores de las variables independientes (entrada) y dependientes (salida) por separado para cada uno de los modelos (CH<sub>4</sub>, LP).

#### 4.4.2 Elaboración del sistema de inferencia difuso

Una vez determinados los parámetros de configuración del sistema de inferencia difuso tipo Sugeno, cuyos componentes son: variables de entrada y salida, número de funciones de pertenencia y reglas difusas. Se construyó el FIS, almacenándolo en un archivo .fis, resultado de una serie de instrucciones en la ventana de comandos de MATLAB, el proceso se detalla a continuación junto con el código empleado para el modelo MSE\_CH<sub>4</sub>.

Se cargan las variables independientes y la variable dependiente previamente almacenadas en un archivo, en el espacio de trabajo de MATLAB, éstas se asignan a variables temporales.

```
% Código en MATLAB para desarrollo modelo ANFIS  
load eSECH4.txt;           %variables independientes  
load sCH4.txt;            %variables dependientes  
entrada=eSECH4;          %variables temporal con datos de entrada  
salida=sCH4;              %variables temporal con datos de salida
```

Se asigna a una variable temporal (*fismat*) la estructura del sistema FIS, creada mediante la función *genfis2*, se proporcionó como entrada: los datos de las variables independientes, la variable dependiente y el radio de influencia previamente seleccionado.

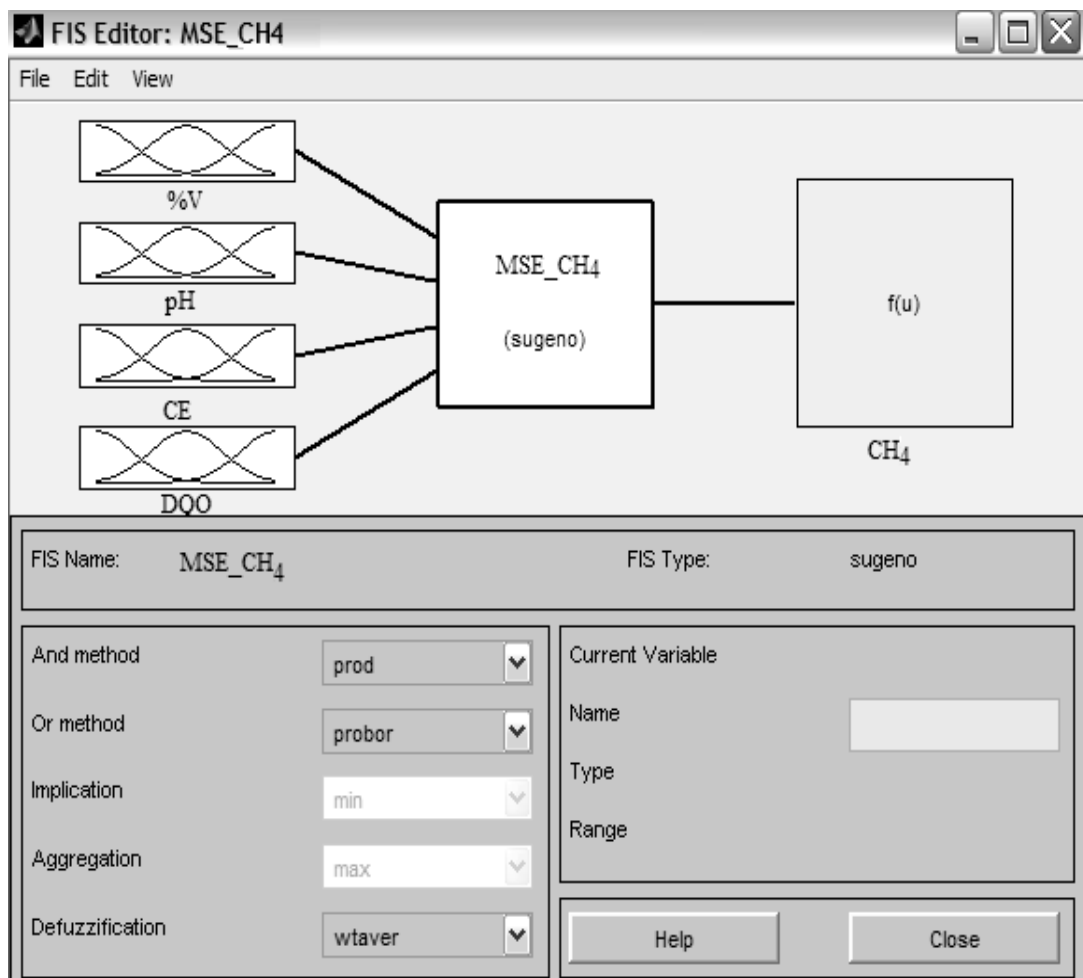
```
fismat = genfis2(Ent_CH4, Sal_CH4, 0.60);           %Asigna a fismat el modelo difuso inicial.
```

Se consultó la estructura recién creada del FIS, escribiendo el nombre de la variable (*fismat*) en la línea de comandos. La información que presentó fue el nombre del sistema, el tipo, los métodos de comparación (and, or e implicación) a los que se les asignó el producto (prod), la probabilidad (probor), el máximo (max), el método de defusificación fue *wtaver* (promedio de pesos), usado en sistemas tipo sugeno, el cual promedia los pesos de todos los resultados parciales obtenidos a partir de la aplicación de las reglas difusas. Verificando que la estructura quedó definida por cuatro entradas, una salida y seis reglas difusas.

```
name: 'sug41'  
type: 'sugeno'  
andMethod: 'prod'  
orMethod: 'probor'  
impMethod: 'prod'  
aggMethod: 'max'  
defuzzMethod: 'wtaver'
```

input: [1x4 struct]  
output: [1x1 struct]  
rule: [1x6 struct]

Debido a que se partió de un conjunto de datos para la creación del sistema FIS, fue necesario realizar algunos ajustes, como el cambio del nombre al sistema FIS y la asignación de etiquetas a cada una de las variables de entrada (input). Este proceso se realizó mediante el uso del comando *mfedit(fismat)*, el cual activa la interfaz gráfica de MATLAB. Desde esta ventana se modificaron las funciones de pertenencia del archivo .fis (figura 4.11), el proceso también puede realizarse desde la línea de comandos, con las siguientes instrucciones.



**Figura 4.11** Interfaz gráfica de MATLAB activada con la función *mfedit*

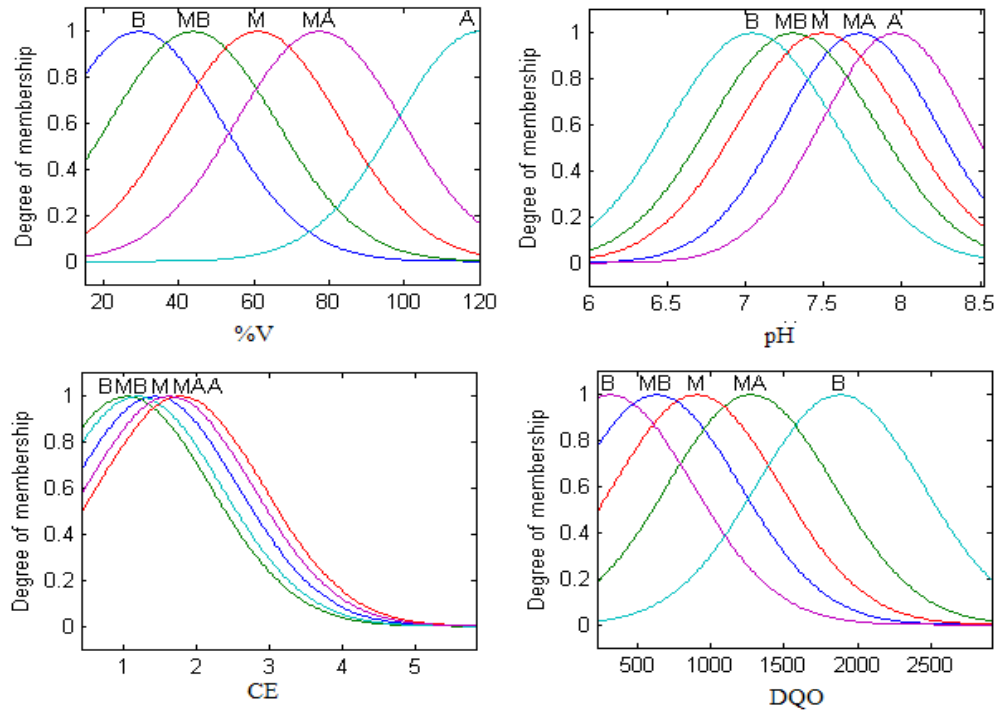
```
fismat = setfis(fismat,'name','MSE_CH4');           % cambio de nombre al FIS
fismat=setfis(fismat,'input',1,'name','V');        %cambio de nombre a variables
fismat=setfis(fismat,'input',2,'name','pH');
fismat=setfis(fismat,'input',3,'name','CE');
fismat=setfis(fismat,'input',4,'name','DQO');
fismat=setfis(fismat,'output',1,'name','CH4');
```

Posteriormente se asignaron las etiquetas a las funciones de membresía [1-6], encargadas de definir, cómo cada punto en el espacio de entrada representa un valor de pertenencia entre cero y uno. Se observó que en algunas variables, el FIS generó un número menor de conglomerados (clusters), por lo que se optimizó el número de funciones de membresía y se redujo a cinco en el caso del modelo MSE\_CH<sub>4</sub> y tres con el modelo MSE\_LP. Se etiquetaron de mayor a menor conforme al rango asignado, los valores fueron: Alto (A), Medio Alto (MA), Medio (M), Medio Bajo (MB), Bajo (B) para el modelo MSE\_CH<sub>4</sub> y Alto (A), Medio (M) y Bajo (B) para MSE\_LP.

```
fismat = setfis(fismat,'input',1,'mf',1,'name','B');
fismat = setfis(fismat,'input',1,'mf',2,'name','MB');
fismat = setfis(fismat,'input',1,'mf',3,'name','M');
fismat = setfis(fismat,'input',1,'mf',4,'name','MA');
fismat = setfis(fismat,'input',1,'mf',5,'name','A');
```

Después de configurar las funciones de membresía del FIS, los cambios se guardaron en el archivo MSE\_CH4.fis mediante la instrucción *writefis(fismat, 'MSE\_CH4.fis')*, a continuación se muestran los gráficos de las funciones de membresía, generados con el comando *plotmf (fismat, 'input', 1)* (Figura 4.12).

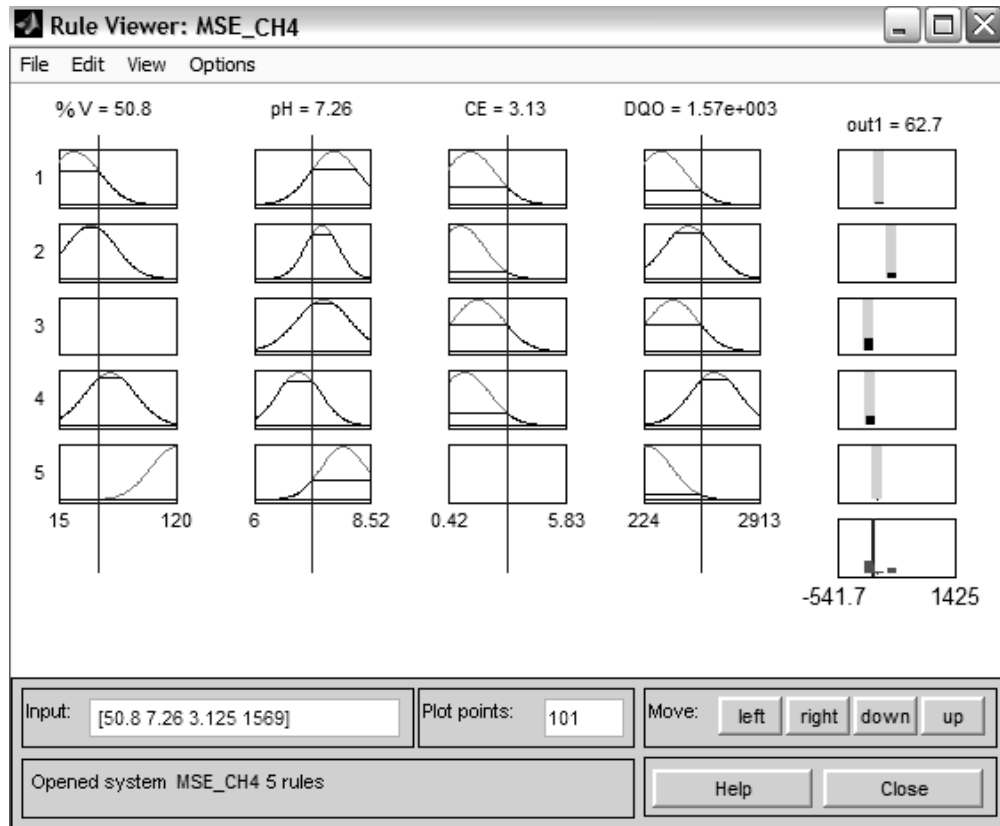
Desde la interfaz gráfica del editor de reglas difusas *ruleview(fismat)*, se evaluaron las reglas difusas del FIS del modelo MSE\_CH<sub>4</sub> (Figura 4.13). Se muestra el diagrama del proceso de inferencia difusa, cada regla está representada por una línea de gráficos. Cada columna representa una variable: las cuatro columnas muestran las funciones de pertenencia del antecedente de cada regla.



**Figura 4.12** Funciones de membresía variables independientes MSE\_CH<sub>4</sub>

La última columna muestra las funciones de pertenencia del consecuente de cada regla. Cada una de las reglas se encuentra numerada al principio de la línea. Es posible mover la línea que atraviesa los gráficos de la variable de entrada, seleccionándola y arrastrándola con el ratón. De este modo se visualiza cómo un valor de entrada forma parte de cada una de las reglas difusas y cómo contribuye a la salida. El valor real de la función de pertenencia se rellena bajo la curva, hasta el grado de pertenencia al que corresponde la entrada. Si se sigue la regla, se puede observar cómo el consecuente es truncado en el mismo grado que el antecedente.

Finalmente el “Proceso de agregación” tiene lugar en la esquina inferior derecha. El valor de salida desfuzificado se muestra bajo la línea gruesa que atraviesa el conjunto difuso resultante de la agregación. El proceso de validación puede realizarse desde la línea de comandos mediante la instrucción `sal_eval= evalfis(entrada, fismat)`, suministrando los datos de validación y el sistema de inferencia difuso (*fismat*), como resultado se obtienen los valores estimados por el modelo (Figura 4.13).

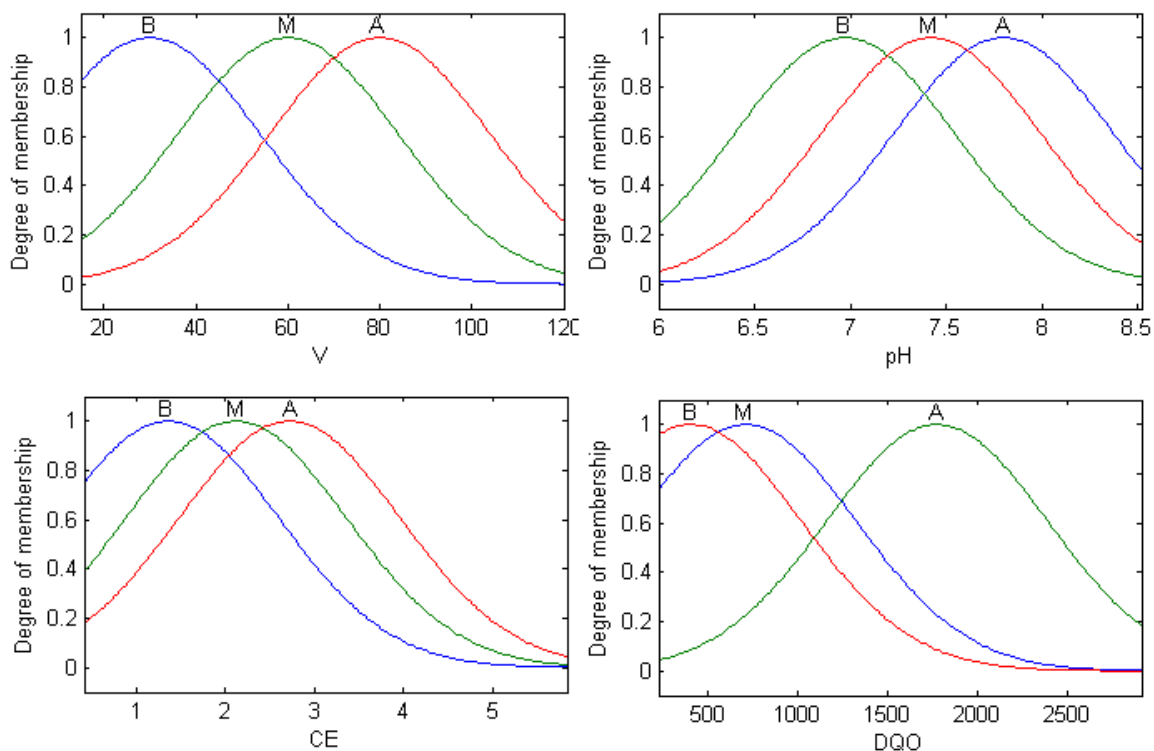


**Figura 4.13** Visor de reglas difusas del modelo MSE\_CH<sub>4</sub>

En la tabla 4.28 se detallan las cinco reglas difusas generadas por el modelo MSE\_CH<sub>4</sub>. El proceso fue generado de igual forma para el modelo MSE\_LP, desde el editor de reglas, se realizó la evaluación de las reglas difusas del FIS, en este modelo se optimizó de cuatro a tres el número de reglas difusas (Figura 4.14).

**Tabla 4.28** Reglas difusas del modelo MSE\_CH<sub>4</sub>

No.	Descripción de las reglas difusas
1	If (V is B) and (pH is MA) and (CE is M) and (DQO is MB) then (CH <sub>4</sub> = 0.4065*V -27.32*pH - 19.59*CE+ 0.04161 *DQO+288.6
2	If (V is MB) and (pH is MB) and (CE is B) and (DQO is MA) then (CH <sub>4</sub> = 0.272*V-60.38 *pH +90.5*CE -0.004736*DQO+ 294.7
3	If (pH is M) and (CE is A) and (DQO is M) then (CH <sub>4</sub> = -0.4099*V -32.08 *pH -31.85*CE - 0.034*DQO+423
4	If (V is M) and (pH is B) and (CE is MB) and (DQO is B) then (CH <sub>4</sub> = -0.2579*V -6.765 *pH +23.45*CE+ 0.0224*DQO -26.35
5	If (V is A) and (pH is A) and (DQO is B) then (CH <sub>4</sub> = -0.9154*V -5.895*pH -2.957*CE+ 0.2009*DQO+ 136.8

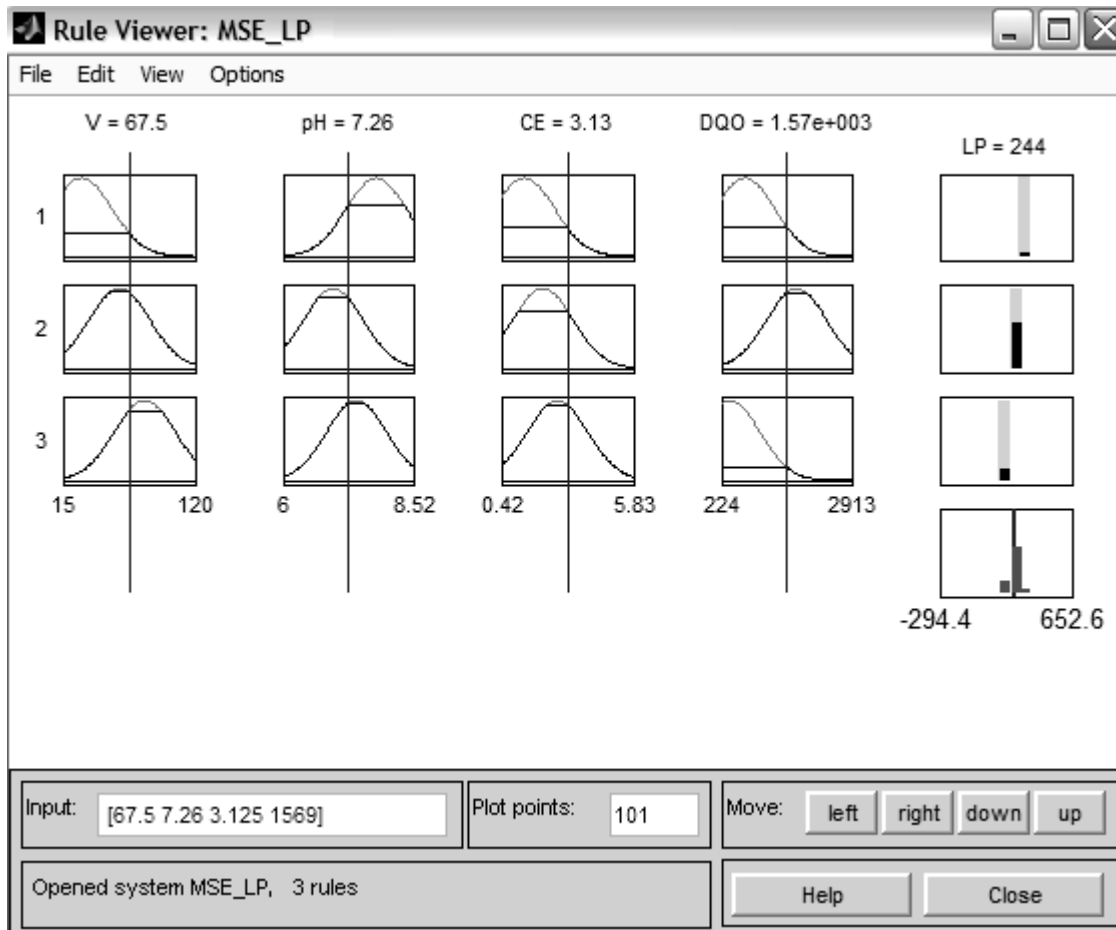


**Figura 4.14** Funciones de membresía variables independientes MSE\_LP

Finalmente el proceso de validación del sistema FIS del modelo MSE\_LP, se realizó con el visualizador de reglas difusas. En la tabla 4.29 se detallan las reglas difusas de este modelo. El valor de salida se muestra en la parte superior de la última columna, donde se encuentra la variable dependiente LP (Figura 4.15).

**Tabla 4.29** Reglas difusas del modelo MSE\_LP

No.	Descripción
1	If (V is B) and (pH is A) and (CE is B) and (DQO is M) then LP= 3.918*V -2.133 *pH+ 15.5 *CE+0.007*DQO+ 4.906
2	If (V is M) and (pH is B) and (CE is M) and (DQO is A) then (LP = 3.679*V+ 7.413 *pH+ 8.73 *CE+0.046 *DQO-144.3
3	If (V is A) and (pH is M) and (CE is A) and (DQO is B) then (LP =3.941*V+ 64.74*pH -15.13 *CE-0.03989*DQO -458.6



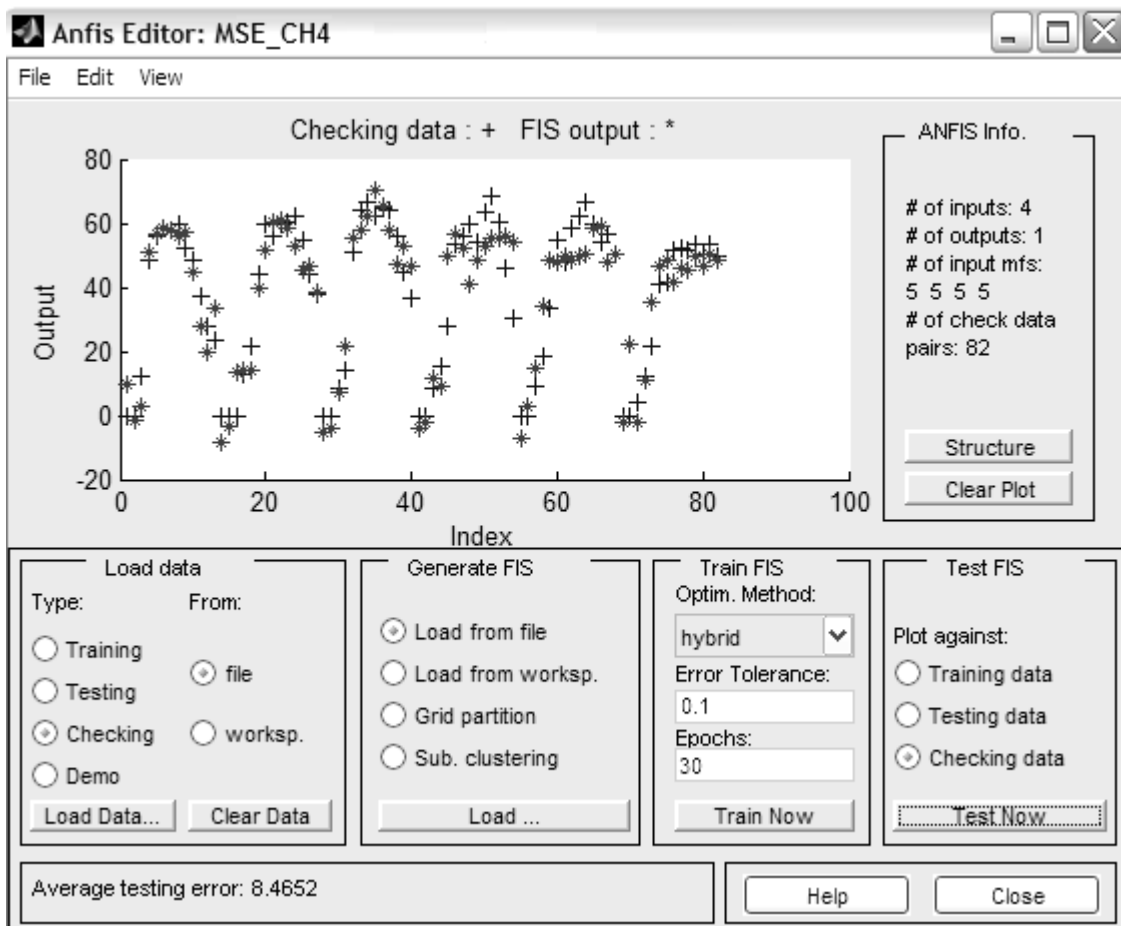
**Figura 4.15** Reglas difusas del modelo MSE\_LP

#### 4.4.3 Verificación y validación del modelo

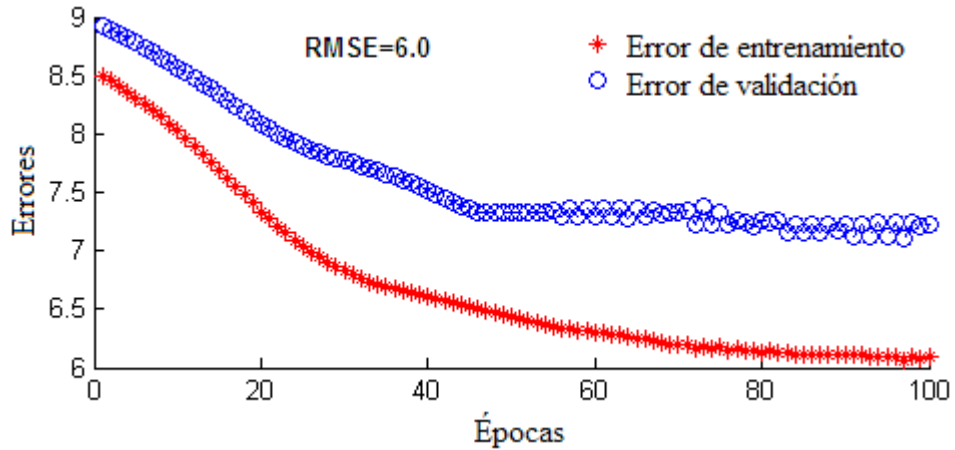
Para la validación del modelo MSE\_CH<sub>4</sub>, se partió de la estructura del FIS optimizada en la etapa anterior, definida por cinco reglas difusas, cuatro variables independientes y una variable dependiente. Teniendo presente que el número de datos de entrenamiento debía ser mayor que el número de parámetros a determinar, se verificó que la razón entre datos y variables fuera alta, alrededor de (298/4). El proceso inició cargando la interfaz gráfica de ANFIS, con el comando *anfisedit*, a continuación se despliega una ventana (Figura 4.16), en la parte inferior izquierda de ésta, se encuentra el panel (*load data*) donde se cargan los datos de entrenamiento (*training*) y prueba (*checking*). Después de cargar los datos, se lee la estructura del FIS, en el panel *Generate FIS* se elige la opción *load from file* y el botón *Load*, en este caso el archivo que se carga es el 'MSE\_CH4.fis'. En

el panel *training FIS* se selecciona el método híbrido, un error de tolerancia de 0.1 y 30 épocas de entrenamiento. Para observar el ajuste que hace ANFIS se oprime el botón *test now*, a continuación se muestra una gráfica de los errores de entrenamiento (asteriscos) con respecto a los errores de validación (círculos).

Como se observa se calcula un error de prueba=8.46, por lo que se evaluó un número mayor de épocas de entrenamiento. Se generó un modelo con 100 épocas de entrenamiento para determinar la época en la que se presentaba un sobre-ajuste del modelo (Figura 4.17). Se observó que a partir de la época 16, el error de entrenamiento tiende a disminuir en mayor proporción que el error de validación, y después de la época 45 el error de validación no ya no disminuye, lo que es indicio de un sobre-ajuste del modelo.

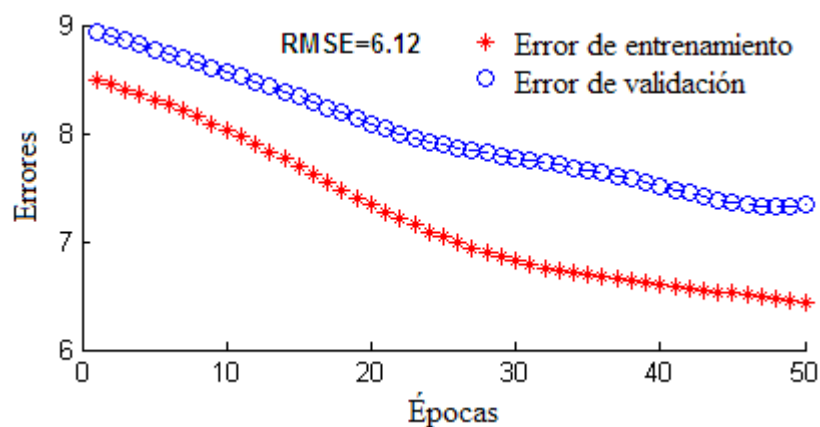


**Figura 4.16** Ventana del Editor ANFIS

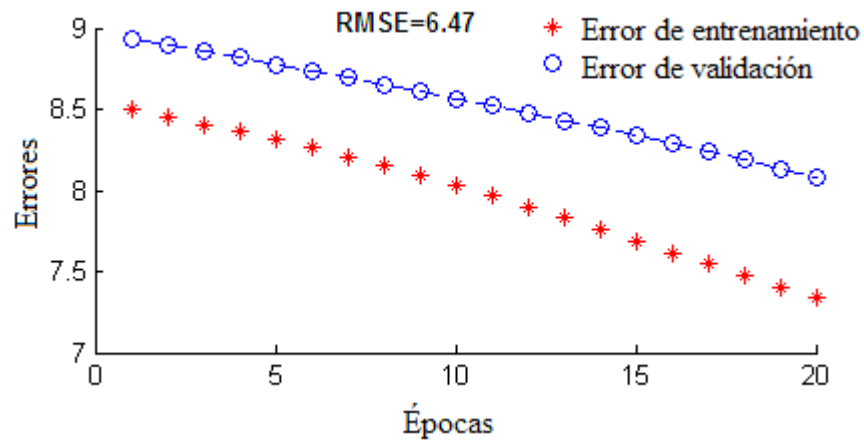


**Figura 4.17** Modelo MSE\_CH<sub>4</sub> con 100 épocas de entrenamiento

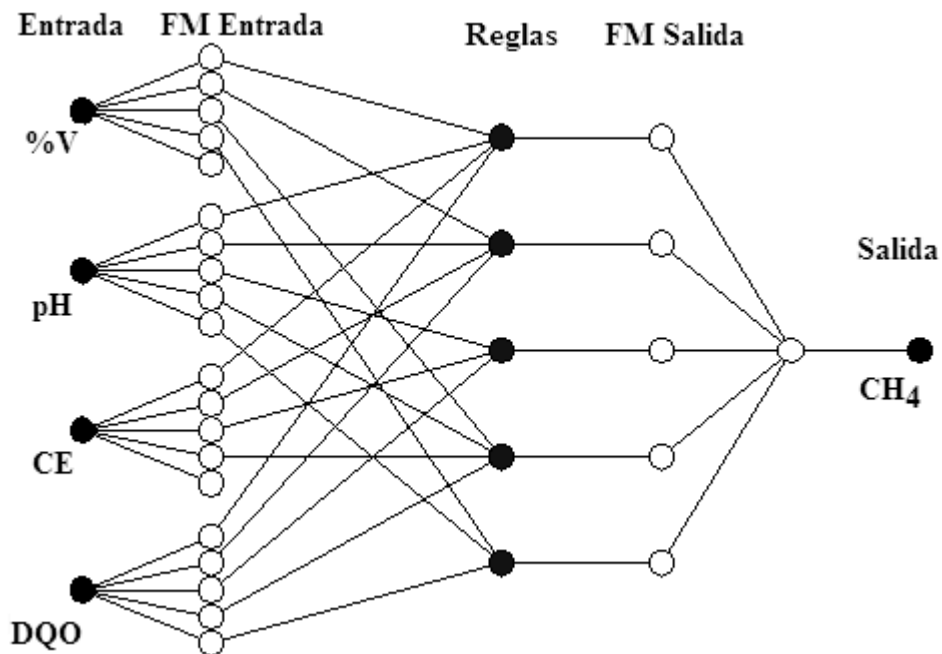
Por lo anterior, se generó un nuevo modelo reduciendo a la mitad el número de épocas de entrenamiento. En la gráfica (Figura 4.18) se observó que el error de validación se reduce en menor proporción que el de entrenamiento alrededor de la época 20. No se encontraron diferencias significativas entre el RMSE calculado con 100 y con 50 épocas (Figura 4.19). Se generó un modelo con 20 épocas de entrenamiento, con un RMSE de 6.47, nuevamente las diferencias no son significativas, al reducir el número de épocas, lo que se deduce es que el modelo puede entrenarse con un número menor de épocas y reducir el tiempo de procesamiento en su construcción.



**Figura 4.18** Modelo MSE\_CH<sub>4</sub> con 50 épocas de entrenamiento



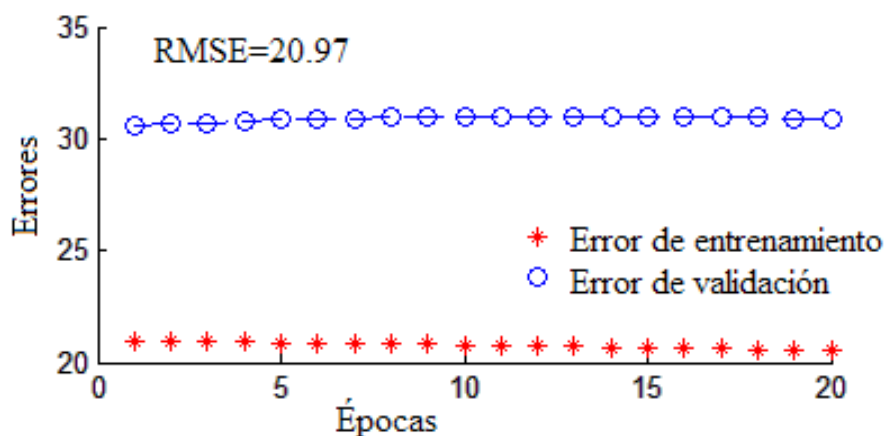
**Figura 4.19** Modelo MSE\_CH<sub>4</sub> con 20 épocas de entrenamiento



**Figura 4.20** Estructura del modelo ANFIS para los modelos MSE\_CH<sub>4</sub>

Desde la ventana de edición de ANFIS, se generó la estructura final del modelo, al oprimir el botón *Structure* (Figura 4.20), los nodos en negro presentan las variables de entrada, cada una consta de cinco funciones de membresía, las cuales convergen en cinco reglas difusas, con su respectivo consecuente y finalmente la salida representada por la variable dependiente buscada.

Al verificar el modelo MSE\_LP, se tomó la estructura FIS inicial, almacenada en el archivo MSE\_LP.fis, cuyo parámetros de configuración consistían de tres reglas difusas, cuatro variables independientes (%V, pH, CE y DQO) y una variable dependiente (LP). Mediante la interfaz gráfica de ANFIS, se evaluó el FIS, se seleccionó el método híbrido de entrenamiento, un error de tolerancia de 0.1 y 20 épocas de entrenamiento, con el fin de observar el ajuste y calcular el RMSE, se generaron los gráficos de los errores de entrenamiento (asteriscos) con respecto a los errores de validación (círculos). En el primer gráfico con 20 épocas de entrenamiento (Figura 4.21) se observó que la distancia entre el error de entrenamiento y de validación es constante, por lo que se generó un nuevo modelo con 100 épocas de entrenamiento para determinar si existía un cambio en los errores (Figura 4.22), observando que a partir de la época 40 había un descenso en el error de validación, se mantenía estable hasta la época 75, sufriendo un incremento después de ésta época. Por lo anterior se generó un modelo con 40 épocas de entrenamiento (Figura 4.23) y otro con 60 (Figura 4.24), no se observaron grandes diferencias en los valores del RMSE, sin embargo presentó un menor RMSE en comparación del modelo de 20 épocas, y con respecto al modelo de 100 épocas fue mínimo el incremento, por lo que el modelo seleccionado para estimar lixiviados es el que tiene 40 épocas de entrenamiento.



**Figura 4.21** Modelo MSE\_LP con 20 épocas de entrenamiento

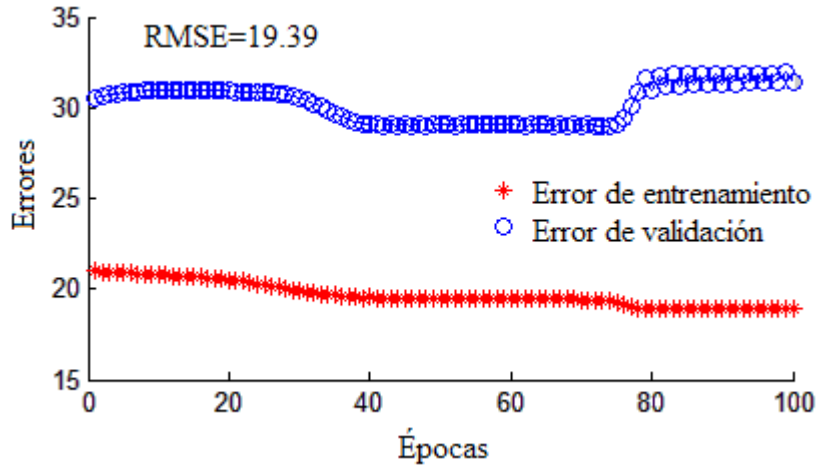


Figura 4.22 Modelo MSE\_LP con 100 épocas de entrenamiento

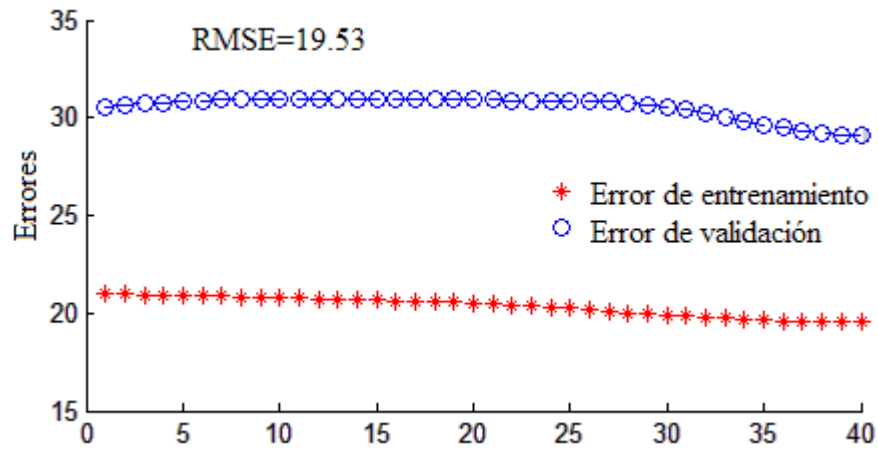


Figura 4.23 Modelo MSE\_LP con 40 épocas de entrenamiento

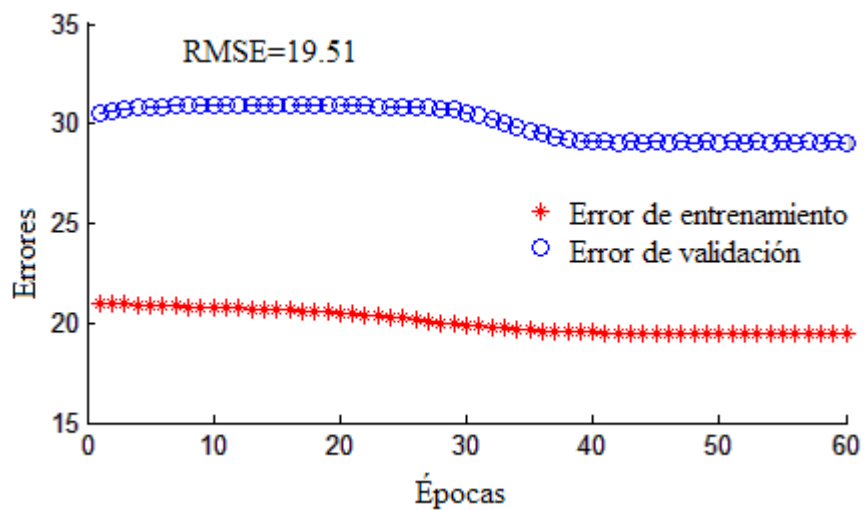
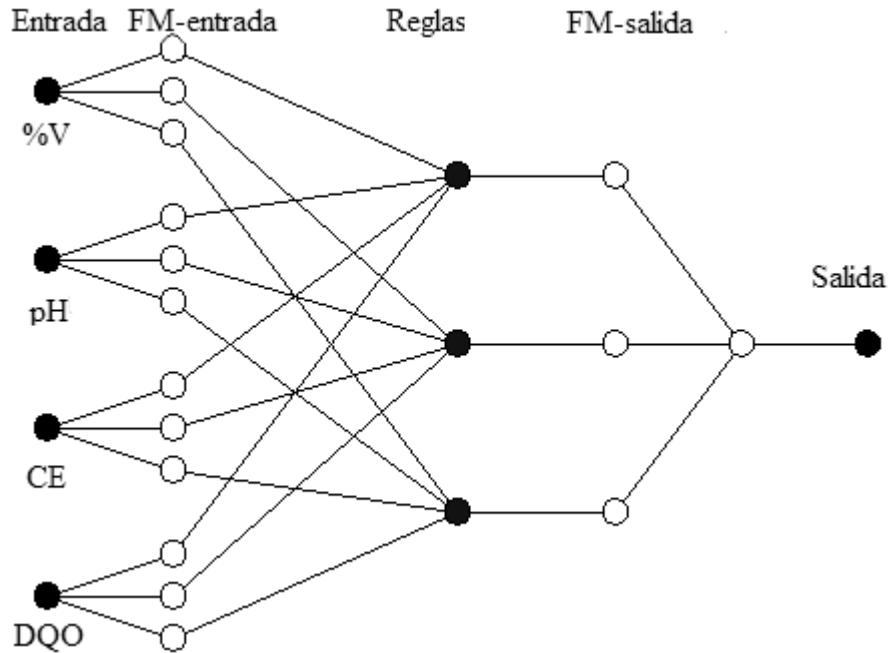


Figura 4.24 Modelo MSE\_LP con 60 épocas de entrenamiento

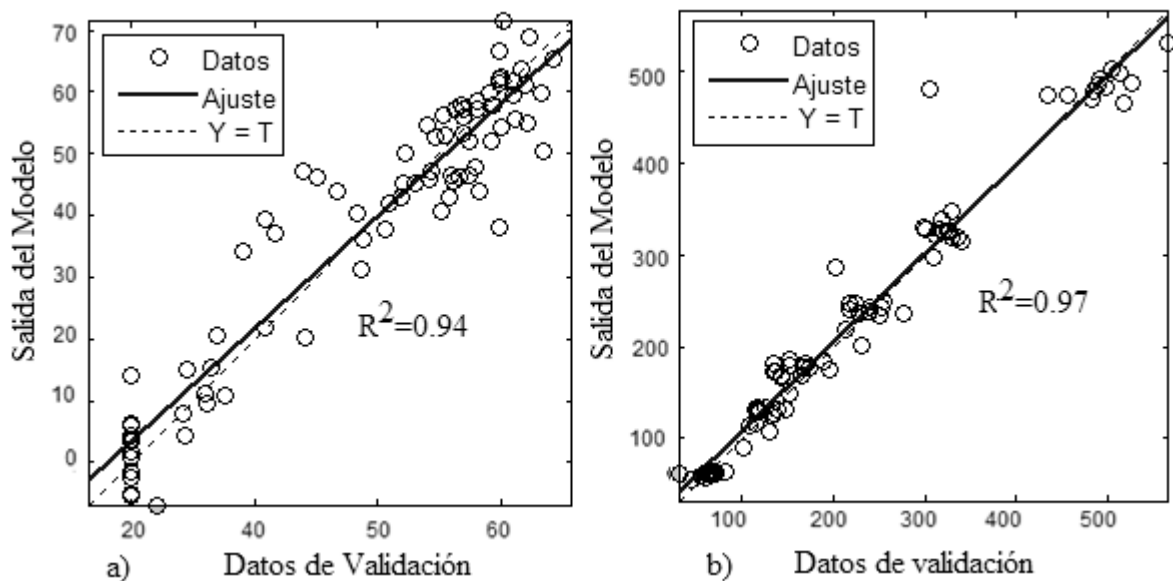


**Figura 4.25** Estructura del modelo ANFIS para los modelos MSE\_LP.

Finalmente se generó la estructura del modelo MSE\_LP (Figura 4.25), en este caso es muy similar al del modelo para CH<sub>4</sub>, las diferencias radican en el número de funciones de membresía y reglas difusas.

#### 4.5 Validación de los modelos

En la validación de los modelos ANFIS, se emplearon datos que no participaron en la etapa de entrenamiento (20% de los datos), con los resultados obtenidos al correr el modelo con éstos datos, se realizó un análisis de regresión para medir la fuerza de relación entre la salida de la red y los datos de validación para los dos modelos (Figura 4.26). Para el modelo MSE\_CH<sub>4</sub> se obtuvo un factor de determinación de  $R^2=0.94$  y para MSE\_LP de  $R^2=0.97$ , lo que indica que los modelos son capaces de explicar el 94 y 97% del comportamiento de CH<sub>4</sub> y LP respectivamente.



**Figura 4.26** Validación del modelo MSE\_CH<sub>4</sub> y MSE\_LP

Finalmente se compararon los resultados de los cuatro modelos desarrollados (Tabla 4.30). Se observó que los modelos con RNA requirieron un menor número de épocas de entrenamiento (10), lo contrario que los modelos ANFIS. En tres de los cuatro modelos, la combinación de las variables %V, pH, CE y DQO se hizo presente, lo que demuestra que la contribución de la experiencia del experto fue significativa en la selección de variables.

Debido a que las diferencias presentadas en  $R^2$  en los modelos de RNA y ANFIS fueron mínimas, se consideró como criterio de selección aquel modelo con el mejor desempeño, es decir el modelo con el menor RMSE. Por lo anterior son seleccionados los modelos desarrollados con ANFIS, tanto para CH<sub>4</sub> como para LP.

**Tabla 4.30** Resultados obtenidos al evaluar los modelos para CH<sub>4</sub> y LP

Técnica	Modelo	VARIABLES	Épocas	RMSE	$R^2$	Neuronas	FM
RNA	MBE_CH <sub>4</sub>	%V, pH, CE, ST	10	9.62	0.93	7	
	MSE_LP	%V, pH, CE, DQO	10	30.14	0.98	6	
ANFIS	<b>MSE_CH<sub>4</sub></b>	%V, pH, CE, DQO	30	<b>6.47</b>	0.94		5
	<b>MSE_LP</b>	%V, pH, CE, DQO	20	<b>19.82</b>	0.97		3

## 4.6 Análisis de sensibilidad

Esta etapa inicia con la selección previa de los modelos MSE\_CH<sub>4</sub> y MSE\_LP. El análisis de sensibilidad se realizó con el fin de medir el impacto de las variables de entrada sobre los cambios en la salida (CH<sub>4</sub> ó LP). Por lo que fue necesario establecer los rangos de cada una de las variables [mínimo máximo], su valor promedio y el incremento calculado para cada variable (Tabla 4.31).

Se inició analizando el modelo MSE\_CH<sub>4</sub>, las variables fueron evaluadas en el orden de asignación en el modelo (%V, pH, CE, DQO), con respecto a las unidades de medida de las variables, no hay cambios dado que las variaciones en sus valores quedaron establecidas a lo largo de su rango. Como se observa en la figura 4.27, la variable %V es la que está bajo estudio, es por ello que sus valores varían a lo largo de su rango, en relación a los incrementos calculados previamente, mientras que el resto de las variables adopta el valor medio. Esta matriz de datos se importa a un archivo de texto, se separan las variables dependientes e independiente, posteriormente se leen desde la ventana de comandos de MATLAB, previamente se cargó el modelo ANFIS, ya que es a través de éste que se realiza la evaluación de los datos y el cálculo del RMSE.

```
fis=readfis('MSE_CH4.fis');
load Valv.txt;           %variables dependientes
load Vch4.txt;          %variables independientes
bondad=evalfis(Valv,fis); %calcula la salida del modelo

%calcula el RMSE entre datos del modelo y los de validación
error=(Vch4-bondad);
rmse=sqrt(mse(error))
```

**Tabla 4.31** Definición de variables en el análisis de sensibilidad

<b>Variable</b>	<b>Min</b>	<b>Max</b>	<b>Valor promedio</b>	<b>Incremento</b>
%V	15	120	56.70	1.28
pH	6	8.52	7.58	0.03
CE	0.42	5.83	1.82	0.06
DQO	224	2913	1003	32.79
CH <sub>4</sub>	0	68.39	37.67	0.83
LP	27	565	224.58	6.56

H	I	J	K	L	M
<b>%V</b>	<b>pH</b>	<b>CE</b>	<b>DQO</b>	<b>CH4</b>	<b>LP</b>
118.7	7.58	1.82	1004	37.67	224.6
117.4	7.58	1.82	1004	37.67	224.6
116.1	7.58	1.82	1004	37.67	224.6
114.8	7.58	1.82	1004	37.67	224.6
113.6	7.58	1.82	1004	37.67	224.6
112.3	7.58	1.82	1004	37.67	224.6
111	7.58	1.82	1004	37.67	224.6
109.7	7.58	1.82	1004	37.67	224.6
108.4	7.58	1.82	1004	37.67	224.6
107.2	7.58	1.82	1004	37.67	224.6
105.9	7.58	1.82	1004	37.67	224.6
104.6	7.58	1.82	1004	37.67	224.6
103.3	7.58	1.82	1004	37.67	224.6

**Figura 4.27** Distribución de los datos para el análisis de sensibilidad

Se evaluó cada una de las variables del modelo MSE\_CH<sub>4</sub>, observando que CE es la variable con mayor sensibilidad a los cambios en sus valores, seguida por DQO y pH. Con respecto a la variable %V, la variación en sus valores no afecta en gran medida a la estimación de CH<sub>4</sub>, como las anteriormente mencionadas. Con respecto al modelo MSE\_LP, la variable con mayor sensibilidad es %V, debido a que el volumen generado de lixiviados depende en gran medida, del volumen recirculado de lixiviados. En segundo término se encuentra la variable pH seguida por CE y finalmente DQO (Tabla 4.32).

**Tabla 4.32** Resultado del análisis de sensibilidad

<b>Variable</b>	<b>MSE_CH<sub>4</sub> RMSE</b>	<b>MSE_LP RMSE</b>
%V	19.3	137.11
pH	21.13	39.65
CE	35.75	14.12
DQO	23.30	10.31

## 5. DISCUSIÓN

Uno de los objetivos planteados en el estudio de un sistema, es desarrollar modelos útiles, que guarden un equilibrio entre complejidad y aplicabilidad. La complejidad del modelo depende de la precisión requerida en la estimación, de la información disponible y de las herramientas de modelado. Con respecto a éstas últimas, las herramientas RNA y ANFIS empleadas durante esta investigación, demostraron una vez más, que sí es posible desarrollar un modelo para estimar la generación de CH<sub>4</sub> y lixiviados, a partir de una colección de datos experimentales de entrada / salida, sin una solución analítica previa, ni un conocimiento detallado de las propiedades físicas del proceso de biodegradación (Chang y Chang, 2006; Jang, 1993; Cakmakci, 2007; Abdallah *et al.*, 2009). No obstante, para su correcta aplicación fue necesario disponer de conocimiento sobre su operación, ya que en la medida en que se dominó la metodología se alcanzó un proceso de modelado eficiente, y con mejores resultados en el desarrollo de los modelos.

Previo a la construcción del modelo, diversos autores sugieren realizar un análisis exploratorio de los datos, lo cual tiene sentido si se considera que existen situaciones en las que solo se parte de una idea preliminar sobre el sistema bajo estudio y de su contexto (Noori *et al.*, 2009; Morales *et al.*, 2000, Naylor *et al.* 1982). Durante la segunda fase de la metodología, el empleo de técnicas estadísticas, reveló un método eficaz para comprender rápidamente la distribución de los datos y sus características. No obstante, se debe considerar que es un proceso que requiere tiempo, y que el cálculo de algunos estadísticos aunque puedan parecer exhaustivos y sin consecuencias, al final proporcionan una explicación viable del fenómeno estudiado.

Durante el análisis exploratorio de los datos, se observó la existencia de dos tipos de incertidumbre. La primera, suministrada a los valores durante su medición, debido a imprecisiones por la sensibilidad de los aparatos utilizados, y por la habilidad mostrada por el experimentador. La segunda, debido a la ambigüedad aportada por datos atípicos o extremos y valores faltantes. La primera fue tratada mediante la teoría de la lógica difusa ya

que no fue posible definir de manera precisa el valor de una medición, en el caso de la segunda, se trató mediante técnicas estadísticas (Laviolette M. y Seaman, 1994).

En este trabajo, se desarrollaron tres enfoques para elegir las técnicas de selección de variables, desde el punto de vista estadístico se consideró la técnica *stepwise*, por la parte de inteligencia artificial, la técnica de búsqueda exhaustiva, y desde el conocimiento empírico fue la selección de variables por el experto. Los resultados demostraron que el conocimiento del experto contribuyó en la construcción de modelos parsimoniosos y con una apropiada bondad de ajuste, lo que también es cierto, es que cuando no se tiene la experiencia de un experto en el área bajo estudio, la técnica de búsqueda exhaustiva es una buena opción.

Entre los beneficios que se alcanzan al reducir el número de variables, es facilitar su visualización, alcanzar una mayor comprensión de los datos, reducir el entrenamiento y el tiempo de ejecución del modelo, además de que se deja de lado el problema de la dimensionalidad (Guyon and Elisseeff, 2003; Chang and Chang, 2006; Gestal, 2009). Los resultados revelaron que al reducir el número de variables, la bondad de ajuste del modelo tendía a incrementarse, lo que fue confirmado en los métodos de selección *stepwise* y búsqueda exhaustiva, lo anterior demuestra que en diseños experimentales similares, podrá monitorearse un número menor de variables. Lo que a su vez incrementa la utilidad del modelo, ya que puede ser empleado en un número mayor de casos de estudio.

Uno de los inconvenientes observados durante la determinación de las variables con búsqueda exhaustiva, fue el tiempo invertido para este fin, el cual resultó ser proporcional al número de variables de entrada que integran el modelo. En cuanto a la técnica *stepwise*, su principal limitante fue el número de variables explicativas que es necesario seleccionar.

Actualmente se han desarrollado numerosas técnicas para el diseño de modelos de predicción, una de estas técnicas es la regresión lineal múltiple y no lineal (Gurijala *et al.* 1997). Sin embargo, si el fenómeno tiene un comportamiento no lineal y es significativo en alguna medida dentro de los datos investigados, el desempeño del modelo desarrollado se

ve limitado. Al respecto se han realizado análisis comparativos entre técnicas multivariantes y herramientas de *Softcomputing*. Donde se ha reportado, que un modelo difuso es de mayor utilidad que uno de análisis multivariable debido a la habilidad que tiene el modelo difuso de incorporar la incertidumbre en la entrada del modelo (Noori et al., 2009; Garg, 2007; Abebe *et al.*, 2000; Hanns 2000, Pitarque *et al.*, 1998). Ahora bien, si se consideran las características de los datos empleados en este trabajo, el modelo desarrollado con una RNA sería más preciso que uno de regresión, debido a que los datos no cumplen los supuestos (linealidad y normalidad). Con respecto a la incertidumbre presentada por los datos, se destaca la habilidad ANFIS para integrarla a través de las reglas difusas y funciones de membresía.

Desde una perspectiva ambiental, la recuperación del CH<sub>4</sub> ofrece beneficios en la economía con la producción de energía, que son aprovechados por los administradores de los rellenos y consumidores (Ogwueleka and Ogwueleka, 2010; Ozkaya *et al.*, 2007). Al disponer de herramientas que permitan estimar volumen de lixiviados y/o la tasa de generación de CH<sub>4</sub> a través de modelos parsimoniosos que requieren un mínimo de datos de entrada, se favorece la determinación de estrategias de recirculación de lixiviados y de paso se reduce la emisión de gases de efecto invernadero.

A pesar de que el uso de herramientas de *Softcomputing* en el diseño de modelos data de 1965, cuando Lotfi A. Zadeh (1965) introdujera el concepto de conjunto difuso. A la fecha son reducidos los trabajos que han desarrollado modelos para estimar la generación de CH<sub>4</sub> y el volumen de lixiviados a partir de residuos sólidos urbanos (Abdallah *et al.*, 2009; Karaka y Ozcaya, 2006), este número se reduce cuando se hace referencia a modelos que han sido implementados exitosamente dentro de un relleno sanitario, muchas veces limitado por la misma infraestructura del relleno sanitario (Kurtulus et al., 2008; Ozkaya et al., 2006).

Los resultados presentados en este trabajo, comprueban la factibilidad para desarrollar modelos con herramientas de *Softcomputing*, sin embargo existen limitaciones que hay que conocer antes de usarlas. Una de ellas, es el nivel de complejidad al que se

enfrenta el modelador cuando desarrolla un modelo de caja negra, debido a que tiene que modelar un sistema de contenido desconocido, con entradas y salidas cuya estructura y procesos pueden estudiarse únicamente por inferencia. Por otro lado, la falta de experiencia en el empleo de las herramientas, puede generar resultados inadecuados o bien puede retardar la obtención de los modelos esperados. No existen estándares en las herramientas empleadas, que determinen la estructura óptima del modelo, dadas las características de los datos. El proceso debe realizarse a prueba y error, lo que implica mayor consumo de recursos. Los modelos que se generan se encuentran limitados a los datos que sirvieron de sustento para su construcción. Se observa una resistencia a la explotación de estos modelos por parte de los expertos en el área ambiental, debido a la falta de una interfaz sencilla que facilite tanto su uso como su comprensión.

## 6. CONCLUSIONES

Es de suma importancia establecer el objetivo que se persigue (exactitud o precisión) en el desarrollo del modelo. Si es la precisión el objetivo principal, el empleo del conocimiento experto representa una buena opción, seguido por las herramientas de *Softcomputing*.

A pesar de la importancia que tiene la selección de variables en el desarrollo de los modelos, se observó que es en el análisis exploratorio, donde se alcanza el entendimiento global del sistema bajo estudio, además de que ofrece una vía para asegurar la fiabilidad y validez de los datos, sobre todo cuando se trata de un modelo de caja negra, el cual la mayoría de la veces se deduce de un conjunto de datos recolectados, durante una experiencia práctica con el sistema.

En la búsqueda por aumentar la precisión del modelo y reducir la dimensionalidad en los datos, fueron tres los métodos evaluados para seleccionar las variables con mayor significancia en los modelos. Observando que la técnica de búsqueda exhaustiva es efectiva para los modelos ANFIS o bien cuando el número de variables es reducido, debido a que requiere una mayor cantidad de recursos en comparación a la técnica *stepwise*.

El empleo de técnicas estadísticas en el tratamiento de la incertidumbre debido a datos faltantes y datos atípicos, permitió asegurar la fiabilidad de las estimaciones al impedir reducir el tamaño de la muestra.

Valores relativamente bajos del error cuadrático medio y  $R^2$  muy cercanos a uno, además de un comportamiento evidentemente no lineal en las variables, demostraron que tanto las redes neuronales, como el sistema neurodifuso son una herramienta útil y eficaz para modelar sistemas complejos como los de un sistema de degradación anaeróbica.

Se verificó que las técnicas de *Softcomputing* mostraron ser de gran utilidad, a la hora de aprovechar datos experimentales, sin embargo el proceso de implementación no resulta ser sencillo, debido a que se requiere tener nociones elementales sobre su arquitectura y construcción, y la forma de cómo interpretar los resultados.

## **RECOMENDACIONES**

- Integrar las funciones diseñadas en código MATLAB, a través de un lenguaje de programación, para llevarlas a un nivel de simulación.
- Investigar y evaluar el uso de Algoritmos Genéticos, en el proceso de selección de variables.
- Elaborar y comparar modelos con una estructura MIMO tanto para el caso de RNA como de ANFIS.

## APÉNDICE A

### Función para búsqueda exhaustiva en MATLAB

```
clear all;

% arreglos para realizar las combinaciones
grup1 = [1 2 3 4 5 6 7 8 9];
grup2 = [1 2 3 4 5 6 7 8 9];
grup3 = [1 2 3 4 5 6 7 8 9];
grup4 = [1 2 3 4 5 6 7 8 9];

%datos para construcción del sistema de inferencia difuso inicial
load dat1.txt;      %variable independientes (10)
load sal1ch.txt; %variable dependiente (1)
trn_dato=dat1;      %datos de entrenamiento
chk_dato=sal1ch;    %datos de validación

%datos para la construcción modelo ANFIS
load trn_dat1.txt;
load chk_dat1.txt;

%contadores de los ciclos
i=0;
j=0;
k=0;

%arreglo para valores temporales
trn_error=(500);
chk_error=(500);

%nombre de las variables independientes
input_name = str2mat('%V','pH','CE','AA','AP','DQO','NH3','ST','SVT');
model=1;
cont=1;

%evaluación de los 4 grupos de variables
for m=1:length(grup1),
    for i=1+m:length(grup2),
        for j=1+i:length(grup3),
            for k=1+j:length(grup4),

                %obtiene nombre de variables
                in0 = input_name(grup1(m), :);
                in1 = input_name(grup2(i), :);
                in2 = input_name(grup3(j), :);
                in3 = input_name(grup4(k), :);
                %asigna posición de la variable a evaluar
                pos0=grup1(m);
                pos1=grup2(i);
                pos2=grup3(j);
                pos3=grup4(k);
```

```

% selecciono las variables para FIS
trn_data=[dat1(:,pos0 :pos1) dat1(:,pos2 :pos3)];    % 4 variables

% asigna var salida modelo FIS
chk_data = sal1ch;

% crea matriz para entrenamiento modelo ANFIS
trn_fis=[trn_dat1(:,pos0 :pos1) trn_dat1(:,pos2 :pos3) trn_dat1(:,10)];
% crea matriz para validacion modelo ANFIS
chk_fis=[chk_dat1(:,pos0 :pos1) chk_dat1(:,pos2 :pos3) chk_dat1(:,10)];

% crea el FIS inicial con un radio de 0.6 valor std
fismat = genfis2(trn_data, chk_data, 0.55);

% crea el ANFIS con 1 epocas de entrenamiento
[trn_out_fis,t_error,stepsize,chk_out_fis,c_error]=anfis(trn_dat1,fismat,1,[0 0 0
0],chk_dat1,1);

% asigna los valores menores de los errores
trn_error(model) = min(t_error);
chk_error(model) = min(c_error);

% Despliega resultados
fprintf('%s %s %s %s', in0, in1, in2, in3);
fprintf('%d %d %d %d %d %d', model, pos0, pos1, pos2, pos3);
fprintf(' --> trn=% .4f',trn_error(model));
fprintf('chk=% .4f', chk_error(model));
fprintf('\n');
model = model+1;
cont=cont+1;
end
end
end
end

```

## APÉNDICE B

### Función para construir un modelo RNA en MATLAB

```
%asignación de valores
load EBECH4.txt; % variables independientes
load SCH4.txt; % variable dependiente
inputs = EBECH4'; % matriz transpuesta de las variables independientes
targets = SCH4'; % matriz transpuesta de variable dependiente

%asigna datos de validación
load tbech4.txt;
load tch4.txt;
pruebaent = tbech4';
pruebasal = tch4';

% Crea la red neuronal
numHiddenNeurons = 4; % asigna numero de neuronas capa oculta
net = newff(inputs,targets,numHiddenNeurons); % construye el modelo con la red neuronal

% divide los datos para entrenamiento y validación
net.divideParam.trainRatio = 70/100; % datos de entrenamiento
net.divideParam.valRatio = 10/100; % datos de verificación
net.divideParam.testRatio = 20/100; % datos de validación o prueba

%parámetros de configuración de la red
net.trainParam.show = 40; % épocas a mostrar
net.trainParam.lr = 0.05; % tasa de aprendizaje
net.trainParam.epochs = 30; % máximo número de épocas de entrenamiento
net.trainParam.goal = 0.01; % objetivo, rendimiento

% entrenamiento de la red
[net, tr] = train (net, inputs, targets); % Levenberg-Marquardt
outputs = sim (net, inputs); % realiza simulación con el modelo

%calcula el error de la predicción
error = outputs-targets; % a la salida obtenida por la red le resta datos de entrada
RMSE = sqrt(mse(error)) % calcula la raíz del error cuadrático medio
ECM = mse(error) % calcula el error cuadrático medio

%calcula el error de validación
outputs2 = sim(net, pruebaent); % corre el modelo con datos de prueba
error2=outputs2-pruebasal; % obtiene diferencia dat_prue/dat_gen_modelo
RMSE2=sqrt(mse(error2)) % calcula raíz del error cuadrático medio

%Evaluar red
resultado = net (datos); % probar la red

% Despliega resultados de la red generada
plotregression (targets,outputs); % realiza análisis de regresión datos de entrada/salida
view (net); % muestra el diseño de la red
plotperf(tr); % muestra el progreso del entrenamiento
ploterrhist(error); % histograma de los errores
```

## APÉNDICE C

### Función para construir modelos ANFIS en MATLAB

```

% cargar archivos de datos de entrenamiento
load eSECH4.txt;           %var independientes
load sCH4.txt;            %var dependientes
entrada=eSECH4;          %matriz de entrada para FIS
salida=sCH4;              %matriz de salida para FIS
entrena=[entrada salida]; %crea matriz de entrenamiento para ANFIS
save ('entSECH4.txt','entrena','-ascii');
%datos de validacion
load tSECH4.txt;          %var independientes
load tCH4.txt;           %var dependientes
entval=tSECH4;
salval=tCH4;
valida=[entval salval];  %crea matriz de validación para ANFIS
save ('valSECH4.txt','valida','-ascii');

%crea el sistema de inferencia inicial
fismat = genfis2(entrada, salida, 0.6);
%evalua la salida que produce el modelo respecto al entrenamiento
fis_entrena=evalfis (entrada, fismat);
RMSE_entrena=norm (fis_entrena - salida) / sqrt (length(fis_entrena));

%evaluar la salida que produce el modelo respecto a datos de evaluacion
fis_valida=evalfis(valida,fismat);
RMSE_valida=norm (fis_valida-salida)/sqrt(length(fis_valida));

%crea modelo ANFIS
[fis,error_ent,stepsize,fismat2,error_prue] = anfis(entrena,fismat,[20 0 0.1],[],valida) ;

%evalua la salida que produce el modelo respecto al entrenamiento
fuzsal2=evalfis(entrada,fismat2);
trnRMSE2=norm(fuzsal2-salida)/sqrt(length(fuzsal2));

%error de validación con datos de entrenamiento
bondad = evalfis(entrada,fis);
error = salida - bondad;
valRmse2 = sqrt(mse(error));

%error de validación mide bondad de ajuste de la recta estimada con datos de validación
bondad = evalfis(entval,fis);
error = salval - bondad;
pruebRmse = sqrt(mse(error));
%compara salida del modelo vs datos de validación
plotregression(salval,bondad); %genera grafico de dispersion y calcula R2
plot(fis_entrena,'x');
hold on;
plot(fuzsal2,'o');
hold off;

```

## APÉNDICE D

### Evaluación con BE, considerando variable dependiente a CH<sub>4</sub>

No.	Variables	Indices	Error Entrenamiento	Error Validación
1	%V pH CE AA	1 2 3 4	44.197	41.818
2	%V pH CE AP	1 2 3 5	147.436	242.340
3	%V pH CE DQO	1 2 3 6	4.485	5.759
4	%V pH CE NH3	1 2 3 7	0.095	0.179
5	%V pH CE ST	1 2 3 8	0.029	0.127
6	%V pH CE SVT	1 2 3 9	1.810	24.438
7	%V pH AA AP	1 2 4 5	39.646	45.312
8	%V pH AA DQO	1 2 4 6	162.076	184.652
9	%V pH AA NH3	1 2 4 7	NaN	NaN
10	%V pH AA ST	1 2 4 8	NaN	NaN
11	%V pH AA SVT	1 2 4 9	NaN	NaN
12	%V pH AP DQO	1 2 5 6	49.055	46.107
13	%V pH AP NH3	1 2 5 7	218.693	223.592
14	%V pH AP ST	1 2 5 8	0.000	24.995
15	%V pH AP SVT	1 2 5 9	0.000	1.883
16	%V pH DQO NH3	1 2 6 7	NaN	NaN
17	%V pH DQO ST	1 2 6 8	NaN	NaN
18	%V pH DQO SVT	1 2 6 9	0.000	24.995
19	%V pH NH3 ST	1 2 7 8	NaN	NaN
20	%V pH NH3 SVT	1 2 7 9	NaN	NaN
21	%V pH ST SVT	1 2 8 9	NaN	NaN
22	%V CE AA AP	1 3 4 5	147.436	242.340
23	%V CE AA DQO	1 3 4 6	4.485	5.759
24	%V CE AA NH3	1 3 4 7	0.095	0.179
25	%V CE AA ST	1 3 4 8	0.029	0.127
26	%V CE AA SVT	1 3 4 9	1.810	24.438
27	%V CE AP DQO	1 3 5 6	135.126	178.054
28	%V CE AP NH3	1 3 5 7	NaN	NaN
29	%V CE AP ST	1 3 5 8	NaN	NaN
30	%V CE AP SVT	1 3 5 9	NaN	NaN
31	%V CE DQO NH3	1 3 6 7	190.867	6610.596
32	%V CE DQO ST	1 3 6 8	0.000	24.995
33	%V CE DQO SVT	1 3 6 9	0.000	1.883
34	%V CE NH3 ST	1 3 7 8	NaN	NaN
35	%V CE NH3 SVT	1 3 7 9	0.000	24.995
36	%V CE ST SVT	1 3 8 9	NaN	NaN
37	%V AA AP DQO	1 4 5 6	4.485	5.759
38	%V AA AP NH3	1 4 5 7	0.095	0.179
39	%V AA AP ST	1 4 5 8	0.029	0.127

<b>No.</b>	<b>Variables</b>	<b>Indices</b>	<b>Error Entrenamiento</b>	<b>Error Validación</b>
40	%V AA AP SVT	1 4 5 9	1.810	24.438
41	%V AA DQO NH3	1 4 6 7	NaN	NaN
42	%V AA DQO ST	1 4 6 8	NaN	NaN
43	%V AA DQO SVT	1 4 6 9	NaN	NaN
44	%V AA NH3 ST	1 4 7 8	0.000	24.995
45	%V AA NH3 SVT	1 4 7 9	0.000	1.883
46	%V AA ST SVT	1 4 8 9	0.000	24.995
47	%V AP DQO NH3	1 5 6 7	0.095	0.179
48	%V AP DQO ST	1 5 6 8	0.029	0.127
49	%V AP DQO SVT	1 5 6 9	1.810	24.438
50	%V AP NH3 ST	1 5 7 8	NaN	NaN
51	%V AP NH3 SVT	1 5 7 9	NaN	NaN
52	%V AP ST SVT	1 5 8 9	0.000	1.883
53	%V DQO NH3 ST	1 6 7 8	0.029	0.127
54	%V DQO NH3 SVT	1 6 7 9	1.810	24.438
55	%V DQO ST SVT	1 6 8 9	NaN	NaN
56	%V NH3 ST SVT	1 7 8 9	1.810	24.438
57	pH CE AA AP	2 3 4 5	NaN	NaN
58	pH CE AA DQO	2 3 4 6	NaN	NaN
59	pH CE AA NH3	2 3 4 7	NaN	NaN
60	pH CE AA ST	2 3 4 8	NaN	NaN
61	pH CE AA SVT	2 3 4 9	NaN	NaN
62	pH CE AP DQO	2 3 5 6	NaN	NaN
63	pH CE AP NH3	2 3 5 7	NaN	NaN
64	pH CE AP ST	2 3 5 8	0.000	24.995
65	pH CE AP SVT	2 3 5 9	0.000	1.883
66	pH CE DQO NH3	2 3 6 7	NaN	NaN
67	pH CE DQO ST	2 3 6 8	NaN	NaN
68	pH CE DQO SVT	2 3 6 9	0.000	24.995
69	pH CE NH3 ST	2 3 7 8	NaN	NaN
70	pH CE NH3 SVT	2 3 7 9	0.000	1156.501
71	pH CE ST SVT	2 3 8 9	NaN	NaN
72	pH AA AP DQO	2 4 5 6	NaN	NaN
73	pH AA AP NH3	2 4 5 7	NaN	NaN
74	pH AA AP ST	2 4 5 8	NaN	NaN
75	pH AA AP SVT	2 4 5 9	NaN	NaN
76	pH AA DQO NH3	2 4 6 7	NaN	NaN
77	pH AA DQO ST	2 4 6 8	0.000	24.995
78	pH AA DQO SVT	2 4 6 9	0.000	1.883
79	pH AA NH3 ST	2 4 7 8	NaN	NaN
80	pH AA NH3 SVT	2 4 7 9	0.000	24.995

No.	Variables	Índices	Error Entrenamiento	Error Validación
81	pH AA ST SVT	2 4 8 9	0.000	1156.501
82	pH AP DQO NH3	2 5 6 7	NaN	NaN
83	pH AP DQO ST	2 5 6 8	NaN	NaN
84	pH AP DQO SVT	2 5 6 9	NaN	NaN
85	pH AP NH3 ST	2 5 7 8	0.000	24.995
86	pH AP NH3 SVT	2 5 7 9	0.000	1.883
87	pH AP ST SVT	2 5 8 9	0.000	24.995
88	pH DQO NH3 ST	2 6 7 8	NaN	NaN
89	pH DQO NH3 SVT	2 6 7 9	NaN	NaN
90	pH DQO ST SVT	2 6 8 9	0.000	1.883
91	pH NH3 ST SVT	2 7 8 9	NaN	NaN
92	CE AA AP DQO	3 4 5 6	NaN	NaN
93	CE AA AP NH3	3 4 5 7	NaN	NaN
94	CE AA AP ST	3 4 5 8	0.000	24.995
95	CE AA AP SVT	3 4 5 9	0.000	1.883
96	CE AA DQO NH3	3 4 6 7	NaN	NaN
97	CE AA DQO ST	3 4 6 8	NaN	NaN
98	CE AA DQO SVT	3 4 6 9	0.000	24.995
99	CE AA NH3 ST	3 4 7 8	NaN	NaN
100	CE AA NH3 SVT	3 4 7 9	NaN	NaN
101	CE AA ST SVT	3 4 8 9	NaN	NaN
102	CE AP DQO NH3	3 5 6 7	NaN	NaN
103	CE AP DQO ST	3 5 6 8	0.000	24.995
104	CE AP DQO SVT	3 5 6 9	0.000	1.883
105	CE AP NH3 ST	3 5 7 8	NaN	NaN
106	CE AP NH3 SVT	3 5 7 9	0.000	24.995
107	CE AP ST SVT	3 5 8 9	NaN	1156.501
108	CE DQO NH3 ST	3 6 7 8	0.000	24.995
109	CE DQO NH3 SVT	3 6 7 9	0.000	1.883
110	CE DQO ST SVT	3 6 8 9	0.000	24.995
111	CE NH3 ST SVT	3 7 8 9	0.000	1.883
112	AA AP DQO NH3	4 5 6 7	NaN	NaN
113	AA AP DQO ST	4 5 6 8	NaN	NaN
114	AA AP DQO SVT	4 5 6 9	0.000	24.995
115	AA AP NH3 ST	4 5 7 8	NaN	NaN
116	AA AP NH3 SVT	4 5 7 9	NaN	NaN
117	AA AP ST SVT	4 5 8 9	NaN	NaN
118	AA DQO NH3 ST	4 6 7 8	NaN	NaN
119	AA DQO NH3 SVT	4 6 7 9	0.000	24.995
120	AA DQO ST SVT	4 6 8 9	NaN	NaN
121	AA NH3 ST SVT	4 7 8 9	0.000	24.995

<b>No.</b>	<b>Variables</b>	<b>Indices</b>	<b>Error Entrenamiento</b>	<b>Error Validación</b>
122	AP DQO NH3 ST	5 6 7 8	NaN	NaN
123	AP DQO NH3 SVT	5 6 7 9	NaN	NaN
124	AP DQO ST SVT	5 6 8 9	NaN	NaN
125	AP NH3 ST SVT	5 7 8 9	NaN	NaN
126	DQO NH3 ST SVT	6 7 8 9	NaN	NaN

**APÉNDICE E**  
**Evaluación BE considerando variable dependiente a LP**

No.	Variables	Índices	Error entrenamiento	Error Validación
1	% V pH CE AA	1 2 3 4	46.901	44.8343
2	% V pH CE AP	1 2 3 5	158.5396	184.8113
3	% V pH CE DQO	1 2 3 6	5.7535	20.9687
4	% V pH CE NH3	1 2 3 7	0.1069	0.1988
5	% V pH CE ST	1 2 3 8	0.0343	0.1262
6	% V pH CE SVT	1 2 3 9	9.5469	30.7011
7	% V pH AA AP	1 2 4 5	45.2639	65.2727
8	% V pH AA DQO	1 2 4 6	173.7058	218.0332
9	% V pH AA NH3	1 2 4 7	NaN	NaN
10	% V pH AA ST	1 2 4 8	NaN	NaN
11	% V pH AA SVT	1 2 4 9	NaN	NaN
12	% V pH AP DQO	1 2 5 6	54.4783	47.615
13	% V pH AP NH3	1 2 5 7	206.1158	207.0694
14	% V pH AP ST	1 2 5 8	0	24.9954
15	% V pH AP SVT	1 2 5 9	0	1.883
16	% V pH DQO NH3	1 2 6 7	NaN	NaN
17	% V pH DQO ST	1 2 6 8	NaN	NaN
18	% V pH DQO SVT	1 2 6 9	0	24.9954
19	% V pH NH3 ST	1 2 7 8	NaN	NaN
20	% V pH NH3 SVT	1 2 7 9	NaN	NaN
21	% V pH ST SVT	1 2 8 9	NaN	NaN
22	% V CE AA AP	1 3 4 5	158.5396	184.8113
23	% V CE AA DQO	1 3 4 6	5.7535	20.9687
24	% V CE AA NH3	1 3 4 7	0.1069	0.1988
25	% V CE AA ST	1 3 4 8	0.0343	0.1262
26	% V CE AA SVT	1 3 4 9	9.5469	30.7011
27	% V CE AP DQO	1 3 5 6	148.9132	175.2398
28	% V CE AP NH3	1 3 5 7	NaN	NaN
29	% V CE AP ST	1 3 5 8	NaN	NaN
30	% V CE AP SVT	1 3 5 9	NaN	NaN
31	% V CE DQO NH3	1 3 6 7	209.2303	13432.6378
32	% V CE DQO ST	1 3 6 8	0	24.9954
33	% V CE DQO SVT	1 3 6 9	0	1.883
34	% V CE NH3 ST	1 3 7 8	NaN	NaN
35	% V CE NH3 SVT	1 3 7 9	0	24.9954
36	% V CE ST SVT	1 3 8 9	NaN	NaN
37	% V AA AP DQO	1 4 5 6	5.7535	20.9687
38	% V AA AP NH3	1 4 5 7	0.1069	0.1988
39	% V AA AP ST	1 4 5 8	0.0343	0.1262

<b>No.</b>	<b>Variables</b>	<b>Índices</b>	<b>Error entrenamiento</b>	<b>Error Validación</b>
40	%V AA AP SVT	1 4 5 9	9.5469	30.7011
41	%V AA DQO NH3	1 4 6 7	NaN	NaN
42	%V AA DQO ST	1 4 6 8	NaN	NaN
43	%V AA DQO SVT	1 4 6 9	NaN	NaN
44	%V AA NH3 ST	1 4 7 8	0	24.9954
45	%V AA NH3 SVT	1 4 7 9	0	1.883
46	%V AA ST SVT	1 4 8 9	0	24.9954
47	%V AP DQO NH3	1 5 6 7	0.1069	0.1988
48	%V AP DQO ST	1 5 6 8	0.0343	0.1262
49	%V AP DQO SVT	1 5 6 9	9.5469	30.7011
50	%V AP NH3 ST	1 5 7 8	NaN	NaN
51	%V AP NH3 SVT	1 5 7 9	NaN	NaN
52	%V AP ST SVT	1 5 8 9	0	1.883
53	%V DQO NH3 ST	1 6 7 8	0.0343	0.1262
54	%V DQO NH3 SVT	1 6 7 9	9.5469	30.7011
55	%V DQO ST SVT	1 6 8 9	NaN	NaN
56	%V NH3 ST SVT	1 7 8 9	9.5469	30.7011
57	pH CE AA AP	2 3 4 5	NaN	NaN
58	pH CE AA DQO	2 3 4 6	NaN	NaN
59	pH CE AA NH3	2 3 4 7	NaN	NaN
60	pH CE AA ST	2 3 4 8	NaN	NaN
61	pH CE AA SVT	2 3 4 9	NaN	NaN
62	pH CE AP DQO	2 3 5 6	NaN	NaN
63	pH CE AP NH3	2 3 5 7	NaN	NaN
64	pH CE AP ST	2 3 5 8	0	24.9954
65	pH CE AP SVT	2 3 5 9	0	1.883
66	pH CE DQO NH3	2 3 6 7	NaN	NaN
67	pH CE DQO ST	2 3 6 8	NaN	NaN
68	pH CE DQO SVT	2 3 6 9	0	24.9954
69	pH CE NH3 ST	2 3 7 8	NaN	NaN
70	pH CE NH3 SVT	2 3 7 9	0	1156.5006
71	pH CE ST SVT	2 3 8 9	NaN	NaN
72	pH AA AP DQO	2 4 5 6	NaN	NaN
73	pH AA AP NH3	2 4 5 7	NaN	NaN
74	pH AA AP ST	2 4 5 8	NaN	NaN
75	pH AA AP SVT	2 4 5 9	NaN	NaN
76	pH AA DQO NH3	2 4 6 7	NaN	NaN
77	pH AA DQO ST	2 4 6 8	0	24.9954
78	pH AA DQO SVT	2 4 6 9	0	1.883
79	pH AA NH3 ST	2 4 7 8	NaN	NaN
80	pH AA NH3 SVT	2 4 7 9	0	24.9954

<b>No.</b>	<b>Variables</b>	<b>Índices</b>	<b>Error entrenamiento</b>	<b>Error Validación</b>
81	pH AA ST SVT	2 4 8 9	0	1156.5006
82	pH AP DQO NH3	2 5 6 7	NaN	NaN
83	pH AP DQO ST	2 5 6 8	NaN	NaN
84	pH AP DQO SVT	2 5 6 9	NaN	NaN
85	pH AP NH3 ST	2 5 7 8	0	24.9954
86	pH AP NH3 SVT	2 5 7 9	0	1.883
87	pH AP ST SVT	2 5 8 9	0	24.9954
88	pH DQO NH3 ST	2 6 7 8	NaN	NaN
89	pH DQO NH3 SVT	2 6 7 9	NaN	NaN
90	pH DQO ST SVT	2 6 8 9	0	1.883
91	pH NH3 ST SVT	2 7 8 9	NaN	NaN
92	CE AA AP DQO	3 4 5 6	NaN	NaN
93	CE AA AP NH3	3 4 5 7	NaN	NaN
94	CE AA AP ST	3 4 5 8	0	24.9954
95	CE AA AP SVT	3 4 5 9	0	1.883
96	CE AA DQO NH3	3 4 6 7	NaN	NaN
97	CE AA DQO ST	3 4 6 8	NaN	NaN
98	CE AA DQO SVT	3 4 6 9	0	24.9954
99	CE AA NH3 ST	3 4 7 8	NaN	NaN
100	CE AA NH3 SVT	3 4 7 9	NaN	NaN
101	CE AA ST SVT	3 4 8 9	NaN	NaN
102	CE AP DQO NH3	3 5 6 7	NaN	NaN
103	CE AP DQO ST	3 5 6 8	0	24.9954
104	CE AP DQO SVT	3 5 6 9	0	1.883
105	CE AP NH3 ST	3 5 7 8	NaN	NaN
106	CE AP NH3 SVT	3 5 7 9	0	24.9954
107	CE AP ST SVT	3 5 8 9	NaN	NaN
108	CE DQO NH3 ST	3 6 7 8	0	24.9954
109	CE DQO NH3 SVT	3 6 7 9	0	1.883
110	CE DQO ST SVT	3 6 8 9	0	24.9954
111	CE NH3 ST SVT	3 7 8 9	0	1.883
112	AA AP DQO NH3	4 5 6 7	NaN	NaN
113	AA AP DQO ST	4 5 6 8	NaN	NaN
114	AA AP DQO SVT	4 5 6 9	0	24.9954
115	AA AP NH3 ST	4 5 7 8	NaN	NaN
116	AA AP NH3 SVT	4 5 7 9	NaN	NaN
117	AA AP ST SVT	4 5 8 9	NaN	NaN
118	AA DQO NH3 ST	4 6 7 8	NaN	NaN
119	AA DQO NH3 SVT	4 6 7 9	0	24.9954
120	AA DQO ST SVT	4 6 8 9	NaN	NaN
121	AA NH3 ST SVT	4 7 8 9	0	24.9954

<b>No.</b>	<b>Variables</b>	<b>Índices</b>	<b>Error entrenamiento</b>	<b>Error Validación</b>
122	AP DQO NH3 ST	5 6 7 8	NaN	NaN
123	AP DQO NH3 SVT	5 6 7 9	NaN	NaN
124	AP DQO ST SVT	5 6 8 9	NaN	NaN
125	AP NH3 ST SVT	5 7 8 9	NaN	NaN
126	DQO NH3 ST SVT	6 7 8 9	NaN	NaN

**APÉNDICE F**  
**Evaluación del modelo MSW\_CH<sub>4</sub> por RNA**

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
5	10	103.02	10.15	99.28	9.96
	20	99.69	9.98	99.25	9.96
	30	82.24	9.07	94.87	9.74
	40	82.24	9.07	94.87	9.74
	50	82.24	9.07	94.87	9.74
6	10	77.47	8.8	77.52	8.8
	20	77.47	8.8	77.52	8.8
	30	35.99	6	57.43	7.58
	40	35.99	6	57.43	7.58
	50	35.99	6	57.43	7.58
7	10	360.74	18.99	330.44	18.18
	20	89.2	9.44	102.89	10.14
	30	88.59	9.41	105.22	10.26
	40	79.85	8.94	98.13	9.91
	50	79.85	8.94	98.13	9.91
8	10	42.74	6.54	68.78	8.29
	20	42.74	6.54	68.78	8.29
	30	42.74	6.54	68.78	8.29
	40	42.74	6.54	68.78	8.29
	50	42.74	6.54	68.78	8.29
9	10	96.27	9.81	94.98	9.75
	20	60.13	7.75	76.04	8.72
	30	22.15	4.71	45.83	6.77
	40	23.5	4.85	65.68	8.1
	50	23.5	4.85	65.68	8.1
10	10	385.65	19.64	375.26	19.37
	20	42.95	6.55	83.4	9.13
	30	42.95	6.55	83.4	9.13
	40	42.95	6.55	83.4	9.13
	50	42.95	6.55	83.4	9.13

**APÉNDICE G**  
**Evaluación del modelo MBE\_CH<sub>4</sub> por RNA**

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	10	93.13	9.65	110.05	10.49
	20	82.62	9.09	116.05	10.77
	30	82.62	9.09	116.05	10.77
	40	82.62	9.09	116.05	10.77
	50	82.62	9.09	116.05	10.77
5	10	96.83	9.84	121.84	11.04
	20	71.38	8.45	90.14	9.49
	30	71.38	8.45	90.14	9.49
	40	65.08	8.07	77.22	8.79
	50	65.08	8.07	77.22	8.79
6	10	98.98	9.95	130.33	11.42
	20	83.25	9.12	106.32	10.31
	30	83.25	9.12	106.32	10.31
	40	83.25	9.12	106.32	10.31
	50	83.25	9.12	106.32	10.31
7	10	59.02	7.68	54.33	7.37
	20	59.02	7.68	54.33	7.37
	30	59.02	7.68	54.33	7.37
	40	59.02	7.68	54.33	7.37
	50	43.01	6.56	60.86	7.8
8	10	79.49	8.92	75.88	8.71
	20	79.49	8.92	75.88	8.71
	30	79.49	8.92	75.88	8.71
	40	79.49	8.92	75.88	8.71
	50	77.4	8.8	75.24	8.67
9	10	84.71	9.2	112.51	10.61
	20	84.71	9.2	112.51	10.61
	30	58.4	7.64	102.41	10.12
	40	58.4	7.64	102.41	10.12
	50	58.4	7.64	102.41	10.12
10	10	64.07	8	80.83	8.99
	20	64.07	8	80.83	8.99
	30	46.26	6.8	71.78	8.47
	40	46.26	6.8	71.78	8.47
	50	46.26	6.8	71.78	8.47

**APÉNDICE H**  
**Evaluación del modelo MSE\_CH<sub>4</sub> por RNA**

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	10	113.3	10.64	115.66	10.75
	20	87.91	9.38	81.16	9.01
	30	87.91	9.38	81.16	9.01
	40	77.66	8.81	79.07	8.89
	50	82.8	9.1	77.7	8.81
5	10	87.22	9.34	108.75	10.43
	20	87.22	9.34	108.75	10.43
	30	87.22	9.34	108.75	10.43
	40	87.22	9.34	108.75	10.43
	50	87.22	9.34	108.75	10.43
6	10	62.92	7.93	71.71	8.47
	20	48.07	6.93	65.87	8.12
	30	48.07	6.93	65.87	8.12
	40	48.07	6.93	65.87	8.12
	50	48.07	6.93	65.87	8.12
7	10	54.28	7.37	75.78	8.71
	20	36.67	6.06	83.93	9.16
	30	36.67	6.06	83.93	9.16
	40	36.67	6.06	83.93	9.16
	50	36.67	6.06	83.93	9.16
8	10	53.29	7.3	67.51	8.22
	20	53.29	7.3	67.51	8.22
	30	44.17	6.65	67.38	8.21
	40	44.17	6.65	67.38	8.21
	50	25.84	5.08	50.29	7.09
9	10	77.34	8.79	73.45	8.57
	20	77.34	8.79	73.45	8.57
	30	35.25	5.94	61.07	7.81
	40	35.25	5.94	61.07	7.81
	50	35.25	5.94	61.07	7.81
10	10	33.36	5.78	64.03	8
	20	33.36	5.78	64.03	8
	30	33.36	5.78	64.03	8
	40	33.36	5.78	64.03	8
	50	25.09	5.01	55.08	7.42

**APÉNDICE I**  
**Evaluación del modelo MSW\_LP con RNA**

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	10	467.07	21.61	935.54	30.59
	20	457.76	21.4	968.97	31.13
	30	457.76	21.4	968.97	31.13
	40	447.08	21.14	964.49	31.06
	50	447.08	21.14	964.49	31.06
5	10	484.35	22.01	925.67	30.42
	20	484.35	22.01	925.67	30.42
	30	484.35	22.01	925.67	30.42
	40	459.18	21.43	788.43	28.08
	50	459.18	21.43	788.43	28.08
6	10	503.64	22.44	235344.72	485.12
	20	503.64	22.44	235344.72	485.12
	30	503.64	22.44	235344.72	485.12
	40	440.44	20.99	943.15	30.71
	50	440.44	20.99	943.15	30.71
7	10	507.17	22.52	1214.91	34.86
	20	474.53	21.78	1104.81	33.24
	30	474.53	21.78	1104.81	33.24
	40	474.53	21.78	1104.81	33.24
	50	474.53	21.78	1104.81	33.24
8	10	355.67	18.86	1030.14	32.1
	20	355.67	18.86	1030.14	32.1
	30	355.67	18.86	1030.14	32.1
	40	355.67	18.86	1030.14	32.1
	50	355.67	18.86	1030.14	32.1
9	10	382.09	19.55	1000.2	31.63
	20	382.09	19.55	1000.2	31.63
	30	345.21	18.58	1247.43	35.32
	40	345.21	18.58	1247.43	35.32
	50	345.21	18.58	1247.43	35.32

**APÉNDICE J**  
**Evaluación del modelo MBE por RNA**

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	10	524.48	22.9	752.18	27.43
	20	465.68	21.58	785.61	28.03
	30	465.68	21.58	785.61	28.03
	40	465.68	21.58	785.61	28.03
	50	465.68	21.58	785.61	28.03
5	10	479.87	21.91	1114.36	33.38
	20	346.35	18.61	1195.87	34.58
	30	346.35	18.61	1195.87	34.58
	40	346.35	18.61	1195.87	34.58
	50	346.35	18.61	1195.87	34.58
6	10	513.56	22.66	945.1	30.74
	20	453.15	21.29	875.92	29.6
	30	407.94	20.2	855.44	29.25
	40	407.94	20.2	855.44	29.25
	50	407.94	20.2	855.44	29.25
7	10	418.14	20.45	909.58	30.16
	20	418.14	20.45	909.58	30.16
	30	418.14	20.45	909.58	30.16
	40	428.2	20.69	1077.82	32.83
	50	428.2	20.69	1077.82	32.83
8	10	419.9	20.49	706.22	26.57
	20	419.9	20.49	706.22	26.57
	30	419.9	20.49	706.22	26.57
	40	419.9	20.49	706.22	26.57
	50	419.9	20.49	706.22	26.57
9	10	326.95	18.08	1089.64	33.01
	20	326.95	18.08	1089.64	33.01
	30	326.95	18.08	1089.64	33.01
	40	326.95	18.08	1089.64	33.01
	50	326.95	18.08	1089.64	33.01
10	10	395.38	19.88	842.63	29.03
	20	395.38	19.88	842.63	29.03
	30	395.38	19.88	842.63	29.03
	40	395.38	19.88	842.63	29.03
	50	395.38	19.88	842.63	29.03

**APÉNDICE K**  
**Evaluación del modelo MSE\_LP por RNA**

No Neuronas	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	10	536.62	23.17	877.52	29.62
	20	402.54	20.06	795.12	28.20
	30	402.54	20.06	795.12	28.20
	40	405.62	20.14	859.2	29.31
	50	405.62	20.14	859.2	29.31
5	10	520.41	22.81	923.35	30.39
	20	520.41	22.81	923.35	30.39
	30	438.49	20.94	859.54	29.32
	40	438.49	20.94	859.54	29.32
	50	438.49	20.94	859.54	29.32
6	10	456.5	21.37	641.98	25.34
	20	456.5	21.37	641.98	25.34
	30	411.09	20.28	705.18	26.56
	40	411.09	20.28	705.18	26.56
	50	411.09	20.28	705.18	26.56
7	10	418.75	20.46	960.71	31
	20	418.75	20.46	960.71	31
	30	418.75	20.46	960.71	31
	40	418.75	20.46	960.71	31
	50	418.75	20.46	960.71	31
8	10	379.40	19.48	883.88	29.73
	20	379.40	19.48	883.88	29.73
	30	379.40	19.48	883.88	29.73
	40	379.40	19.48	883.88	29.73
	50	379.40	19.48	883.88	29.73
9	10	469.27	21.66	808.33	28.43
	20	391.26	19.78	822.03	28.67
	30	391.26	19.78	822.03	28.67
	40	391.26	19.78	822.03	28.67
	50	391.26	19.78	822.03	28.67
10	10	873.21	29.55	1300.81	36.07
	20	873.21	29.55	1300.81	36.07
	30	461.97	21.49	853.94	29.22
	40	461.97	21.49	853.94	29.22
	50	426.70	20.66	789.81	28.1

**APÉNDICE L**  
**Evaluación de modelos MSW\_CH<sub>4</sub> por ANFIS.**

Reglas	Radio	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
13	0.5	10	2.88	3.96	7.19	8.77
		20	2.65	3.96	7.19	7.94
		30	2.50	3.96	7.19	6.53
		40	2.41	3.96	7.19	6.5
		50	2.34	3.96	7.19	6.93
10	0.55	10	4.06	5.61	7.19	7.09
		20	3.58	5.61	7.19	7.17
		30	3.47	5.61	7.19	7.59
		40	3.37	5.61	7.19	7.1
		50	3.32	5.61	7.19	6.81
9	0.6	10	4.70	6.07	7.19	4.79
		20	4.28	6.07	6.79	4.35
		30	4.17	6.07	6.79	4.35
		40	4.10	6.07	6.79	4.35
		50	4.01	6.07	6.79	4.35
9	0.65	10	4.64	6.07	7.19	6.97
		20	4.05	6.07	7.19	7.49
		30	3.86	6.07	7.19	7.11
		40	3.73	6.07	7.19	6.67
		50	3.56	6.07	7.19	14.41
6	0.7	10	6.52	8.03	7.19	7.32
		20	6.03	8.03	7.19	7.06
		30	5.61	8.03	7.19	6.79
		40	5.50	8.03	7.19	6.85
		50	5.40	8.03	7.19	6.88
5	0.75	10	6.65	8.43	7.19	7.18
		20	6.41	8.43	7.19	7.3
		30	6.18	8.43	7.19	7.22
		40	5.92	8.43	7.19	7.1
		50	5.82	8.43	7.19	7.08
4	0.8	10	7.46	8.86	7.19	8.13
		20	7.33	8.86	7.19	7.74
		30	7.10	8.86	7.19	7.57
		40	6.94	8.86	7.19	7.49
		50	6.81	8.86	7.19	7.58

**APÉNDICE M**  
**Evaluación de modelos MBE\_CH<sub>4</sub> por ANFIS**

Reglas	Radio	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
7	0.5	10	9.01	10.63	10.96	10.96
		20	8.65	10.63	10.65	10.94
		30	8.59	10.63	10.38	10.79
		40	8.54	10.63	10.38	10.81
		50	8.46	10.63	10.38	10.63
6	0.55	10	9.94	11.74	11.8	11.83
		20	9.4	11.74	11.07	11.11
		30	9.12	11.74	10.47	10.47
		40	8.91	11.74	10.47	10.55
		50	8.88	11.74	10.45	10.6
5	0.6	10	10.75	11.61	12.3	12.3
		20	10.34	11.61	11.74	11.74
		30	10.19	11.61	11.62	11.7
		40	10.14	11.61	11.62	11.65
		50	10.09	11.61	11.62	11.65
4	0.65	10	11	11.79	13.77	13.81
		20	10.78	11.79	13.45	13.45
		30	10.7	11.79	13.37	13.37
		40	10.68	11.79	13.28	13.28
		50	10.65	11.79	13.23	13.23
4	0.7	10	10.95	11.78	13.3	13.79
		20	10.8	11.78	13.24	13.48
		30	10.73	11.78	13.21	13.36
		40	10.7	11.78	13.18	13.28
		50	10.68	11.78	13.14	13.21
4	0.75	10	10.9	11.77	13.4	13.4
		20	10.79	11.77	13.18	13.18
		30	10.74	11.77	13.14	13.14
		40	10.72	11.77	13.07	13.07
		50	10.69	11.77	13.01	13.01
4	0.8	10	10.94	11.76	13.38	13.38
		20	10.83	11.76	13.11	13.11
		30	10.78	11.76	13.04	13.05
		40	10.75	11.76	12.97	12.97
		50	10.72	11.76	12.92	12.92

**APÉNDICE N**  
**Evaluación de modelos MSE\_CH<sub>4</sub> por ANFIS**

Reglas	Radio	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
9	0.5	10	5.82	6.75	7.19	7.47
		20	5.25	6.75	7.19	8.16
		30	5.13	6.75	7.19	8.07
		40	4.92	6.75	7.19	8.33
		50	4.71	6.75	7.19	8.02
7	0.55	10	6.22	7.31	7.19	7.39
		20	5.94	7.31	7.19	7.61
		30	5.83	7.31	7.19	7.54
		40	5.73	7.31	7.19	7.48
		50	5.64	7.31	7.19	7.44
6	0.6	10	6.66	8.50	7.19	7.50
		20	6.47	8.50	7.19	7.25
		30	6.36	8.50	7.19	7.23
		40	6.23	8.50	7.19	7.28
		50	6.13	8.50	7.19	7.25
6	0.65	10	6.51	8.60	7.19	7.74
		20	6.19	8.60	7.19	7.48
		30	5.92	8.60	7.19	7.47
		40	5.80	8.60	7.19	7.46
		50	5.74	8.60	7.19	7.46
5	0.7	10	7.12	9.77	7.19	7.71
		20	6.97	9.77	7.19	7.57
		30	6.87	9.77	7.19	7.58
		40	6.78	9.77	7.19	7.65
		50	6.71	9.77	7.19	7.69
5	0.75	10	8.19	9.48	7.19	9.16
		20	7.53	9.48	7.19	8.74
		30	7.16	9.48	7.19	8.43
		40	6.88	9.48	7.19	8.16
		50	6.71	9.48	7.19	7.98
4	0.8	10	8.04	10.59	7.19	9.07
		20	7.58	10.59	7.19	8.45
		30	7.51	10.59	7.19	8.43
		40	7.40	10.59	7.19	8.20
		50	7.34	10.59	7.19	8.19

**APÉNDICE O**  
**Evaluación de modelos MSW\_LP por ANFIS**

Reglas	Radio	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
4	0.5	10	18.68	21.18	29.85	31.60
		20	18.48	21.18	29.85	33.95
		30	18.40	21.18	29.85	33.64
		40	18.35	21.18	29.85	34.46
		50	18.29	21.18	29.85	34.08
4	0.55	10	18.80	21.18	30.29	31.76
		20	18.55	21.18	30.29	33.02
		30	18.45	21.18	30.29	33.67
		40	18.37	21.18	30.29	33.48
		50	18.32	21.18	30.29	33.64
4	0.6	10	18.76	21.20	29.46	32.48
		20	18.45	21.20	29.46	33.03
		30	18.35	21.20	29.46	32.52
		40	18.30	21.20	29.46	31.76
		50	18.29	21.20	29.46	31.37
3	0.65	10	20.15	21.79	29.51	31.17
		20	19.77	21.79	29.51	30.72
		30	19.61	21.79	29.51	30.57
		40	19.55	21.79	29.51	30.76
		50	19.51	21.79	29.51	30.85
2	0.7	10	22.02	22.19	29.71	30.04
		20	21.97	22.19	29.71	30.31
		30	21.42	22.19	29.71	30.06
		40	20.93	22.19	28.80	29.59
		50	20.84	22.19	28.54	29.06
2	0.75	10	22.02	22.22	29.57	29.92
		20	21.97	22.22	29.57	30.17
		30	21.02	22.22	29.49	29.49
		40	20.81	22.22	28.50	29.34
		50	20.58	22.22	28.50	29.04
2	0.8	10	22.02	22.26	29.49	29.83
		20	21.96	22.26	29.49	30.02
		30	20.91	22.26	29.10	29.10
		40	20.59	22.26	28.71	28.71
		50	20.49	22.26	27.99	28.58

**APÉNDICE P**  
**Evaluación de modelos MBE\_LP por ANFIS**

<b>Reglas</b>	<b>Radio</b>	<b>Épocas</b>	<b>ECM_T</b>	<b>RMSE_T</b>	<b>ECM_V</b>	<b>RMSE_V</b>
6	0.5	10	19.15	20.62	32.22	32.46
		20	18.31	20.62	31.84	31.84
		30	18.22	20.62	31.84	32.42
		40	18.17	20.62	31.84	32.57
		50	18.13	20.62	31.84	32.68
6	0.55	10	19.42	21.07	31.78	32.18
		20	18.86	21.07	31.78	32.22
		30	18.67	21.07	31.78	32.29
		40	18.62	21.07	31.78	32.53
		50	18.57	21.07	31.78	32.79
5	0.6	10	19.94	20.98	33.1	33.63
		20	18.3	20.98	33.1	38.96
		30	18.21	20.98	33.1	39.69
		40	18.16	20.98	33.1	39.15
		50	18.15	20.98	33.1	38.04
5	0.65	10	21.23	22.52	29.18	29.61
		20	21.02	22.52	29.18	30.48
		30	20.99	22.52	29.18	30.85
		40	20.97	22.52	29.18	30.85
		50	20.96	22.52	29.18	30.88
4	0.7	10	21.43	22.48	29.28	29.82
		20	21.03	22.48	29.28	30.55
		30	20.99	22.48	29.28	30.42
		40	20.96	22.48	29.28	31
		50	20.95	22.48	29.28	31.03
4	0.75	10	21.58	22.68	31.03	31.19
		20	20.79	22.68	30.35	30.64
		30	20.58	22.68	30.35	30.48
		40	20.44	22.68	30.35	30.54
		50	20.35	22.68	30.35	30.64
3	0.8	10	22.18	22.92	29.41	29.8
		20	21.72	22.92	29.41	30.94
		30	21.62	22.92	29.41	31.28
		40	21.59	22.92	29.41	31.26
		50	21.58	22.92	29.41	31.24

**APÉNDICE Q**  
**Evaluación de modelos MSE\_LP por ANFIS**

Reglas	Radio	Épocas	ECM_T	RMSE_T	ECM_V	RMSE_V
7	0.5	10	18.61	19.96	30.78	32.43
		20	17.71	19.96	30.78	32.50
		30	17.71	19.96	30.78	32.50
		40	17.28	19.96	30.78	33.03
		50	16.66	19.96	30.78	33.24
6	0.55	10	18.90	20.15	30.25	30.76
		20	18.77	20.15	30.25	30.82
		30	18.64	20.15	30.25	30.90
		40	18.55	20.15	30.25	30.94
		50	18.46	20.15	30.25	30.98
5	0.6	10	19.70	20.98	28.72	29.87
		20	19.62	20.98	28.72	28.97
		30	18.99	20.98	28.72	32.04
		40	18.96	20.98	28.72	32.36
		50	18.96	20.98	28.72	32.36
4	0.65	10	20.31	21.84	28.84	29.15
		20	20.00	21.84	27.97	28.14
		30	19.89	21.84	27.97	28.54
		40	19.82	21.84	27.97	28.58
		50	19.77	21.84	27.97	28.76
4	0.7	10	20.86	21.58	29.09	29.73
		20	20.57	21.58	29.09	30.96
		30	20.44	21.58	29.09	31.24
		40	20.36	21.58	29.09	31.29
		50	20.32	21.58	29.09	31.31
3	0.75	10	21.20	22.33	28.79	28.80
		20	21.11	22.33	28.06	28.17
		30	20.96	22.33	28.06	28.97
		40	20.89	22.33	28.06	29.41
		50	20.85	22.33	28.06	29.72
3	0.8	10	21.75	22.36	29.07	29.53
		20	21.65	22.36	29.07	29.55
		30	21.59	22.36	29.07	29.55
		40	21.52	22.36	29.07	29.56
		50	21.48	22.36	29.07	29.57

## BIBLIOGRAFÍA

- Abdallah, M S., Fernandes L., Rendra S.(2009) A fuzzy logic model for biogas generation in bioreactor landfills. *Canadian Journal of Civil Engineering*. 36, 701–708.
- Abebe, A.J., Guinot, V., Solomatine, D.P. (2000). Fuzzy Alpha-Cut vs. Monte Carlo techniques in assessing uncertainty in model parameters. *Proceedings of 4th International Conference on Hydroinformatics*. Iowa City, USA.
- Abu Qdais H., Bani Hani K., Shatnawi N. (2009). Modeling and optimization of biogas production from a wastedigester using artificial neural network and genetic algorithm. *Resources, Conservation and Recycling*. 54(6). 359-363.
- Baird, D.C. Experimentación. (1991) Una introducción a la teoría de mediciones y al diseño de experimentos. Segunda Edición. Pearson Educación. Prentice Hall Hispanoamericana S.A. pp.207.
- Beato C.L. (2008). Extensiones de PMML para el preproceso de datos en proyectos de Data Mining. Tesis de Máster. Universidad Politécnica de Madrid, España. 2008.
- Bilgili M.S, Demir A., Ozkaya B. (2006). Quality and Quantity of Leachate in Aerobic Pilot-Scale Landfills. *Environmental Management*. 38(2). 189–196.
- Bonissone P., Chen Y.T., Goebel K. and Khedkar P. S. (1999). Hybrid Soft Computing Systems: Industrial and Commercial Applications. *Proceedings of the IEEE*. 87(9). 1641-1667.
- Cakmakci M. (2007). Adaptive Neuro-fuzzy Modelling of anaerobic digestion of primary sedimentation sludge. *Bioprocess Biosyst Eng*. 30. 349–357.
- Chang Fi-John, Chang Ya-Ting. (2006). Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Elsevier Advances in Water Resources*. 29. 1–10.
- Chen P., (1976). The entity-relationship model: Towards a unified view of data, *ACM Trans. On Database Systems*. 1(1). 9-36.
- Chiu S.L.(1994). Fuzzy Model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems*. 2. 267-278.
- Dean R. B. and Dixon W. J. (1951) Simplified Statistics for Small Numbers of Observations. *Anal. Chem*. 23(4). 636–638.
- Dee D. P. (1995) A pragmatic approach to model validation. *American Geophysical Union. Quantitative skill Assessment for Coastal Ocean Models. Coastal and Estuarine Studies*. 47. 1-13.
- Del Real O. J., Prieto G., Santos L. E., Román G. A. y Gordillo M.A. (2009). Simulación matemática de un digestor anaerobio tipo tanque agitado para el tratamiento de vertidos residuales. *Rev Int Contam Ambie*. 25(1). 33-41.
- Demirekler E., Rowe R.K., Unlu k. (1999). Modeling leachate production from municipal solid waste landfills. *Environmental Sanitary Engineering Centre. Cagliari Italy*. II. 17. 24.
- Demuth, H., Beale, M., Hagan, M. (2008). *Neural Network Toolbox MATLAB*.
- El-Fadel M., Findikakis A.N., and Leckie J. O. (1997). Modelling leachate generation and transport in solid waste landfills. *Environmental Technology*. 18. 669-686.
- EPA (U.S. Environmental Protection Agency). (2005). *Landfill Gas Emissions Model LandGEM Version 3.02 User's Guide*. Office of Research and Development Washington. D.C.

- Frost, F, Karri, V. (1999). Determining the Influence of Input Parameters on BP Neural Network Output Error Using Sensitivity Analysis', Third International Conference on Computational Intelligence and Multimedia Applications. New Deli. India. 45-49.
- Gualdrón G. O. E. (2006). Desarrollo de diferentes métodos de selección de variables para sistemas multisensoriales. Tesis de Grado. Universitat Rovira I Virgili, España.
- Garg, A., Achari G. and Joshi C. R., (2006). A model to estimate the methane generation rate constant in sanitary landfills using fuzzy synthetic evaluation. *Waste Management Research*. 24. 363–375.
- Garg A., Achari G. and Joshi C. R (2007). Application of Fuzzy Logic to Estimate Flow of Methane for Energy Generation at a Sanitary Landfill. *Journal of energy engineering*. 133(4).
- Griffith M. and Trois C. (2006) Long-term emissions from mechanically biologically treated waste: Influence on leachate quality. *Water SA*. 32. 3.
- Gestal P. M. (2009). Computación evolutiva para el proceso de sección de variables en espacios de búsqueda multimodales. Tesis doctoral. Universidad de Coruña.
- Gurijala K. R., Sa P., Robinson J. A. (1997). Statistical Modeling of Methane Production from Landfill Samples. *Appl Environ Microb*. 63(10). 3797–3803.
- Guyon I., Elisseeff A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3. 1157-1182.
- Hanss M. A (2000). Nearly Strict Fuzzy Arithmetic for Solving Problems with Uncertainties. Institute A of Mechanics University of Stuttgart. *IEEE*. 5. 212-220.
- Haarstrick A., Hempel D. C., Ostermann L., Ahrens H., Dinkler D. (2001). Modelling of the biodegradation of organic matter in municipal landfills. *Waste Manage Res*. 19. 320–331.
- Hernández-Berriel, M. C., Márquez-Benavides L. (2010). Biometanización acelerada de residuos sólidos urbanos con recirculación de lixiviados. 6º Congreso Estatal de Ciencia y Tecnología Memorias en extenso.
- INEGI., Instituto Nacional de Estadística Geografía e Informática. Estadísticas ambientales /Residuos. <http://www.inegi.gob.mx/est/default.aspx?c=119> (Documento Web, último acceso Junio 2011).
- Jalili GZM and Noori R. (2008). Prediction of Municipal Solid Waste Generation by Use of Artificial Neural Network: A Case Study of Mashhad Int. *J. Environ Res*. 2(1). 13-22.
- Jang J.S.R., Sun C. T., Mizutani E. (1997). *Neuro-Fuzzy and soft computing. A computational approach to learning and machine intelligence*. London, UK: Prentice-Hall. pp. 614.
- Jang J.R. (1993). ANFIS: Adaptive-Neural-Based Fuzzy Inference System. *IEEE Trans. on Systems, Man and Cybernetics*. 23(3). 665-685.
- Jang J.R. (1995). Neuro-Fuzzy Modeling and Control. Member, *IEEE Proceedings of the IEEE*. 83. 3.
- Karaca F. and Ozkaya B.(2006). NN-LEAP: A neural network-based model for controlling leachate flow-rate in a municipal solid waste landfill site. *Environ Modell Softw*. 21, 1190-1197.
- Kiss G. y Encarnación G. (2006). Los productos y los impactos de la descomposición de residuos sólidos urbanos en los sitios de disposición final Instituto Nacional de Ecología, México. *Gaceta Ecológica*. 79. 39-51.

- Kjeldsen P., Barlaz M.A., Rooker A.P., Baun A., Ledin A., Christensen T.H.(2002). Present and Long-Term Composition of MSW Landfill Leachate: A Review. *Critical Reviews in Environmental Science and Technology*, 32(4). 297–336.
- Kleinbaum D.G., Kupper L.L., Muller K.E. (1988) *Applied Regression Analysis and Other Multivariables Methods*. PWS-KENT Publishing Company.
- Kylefors K., Ecke H., Lagerkvist A. (2003). Accuracy of COD Test for landfill leachates. *Water, Air, and Soil Pollution*, 146. 153–169.
- Laviolette M., Seaman W.J. (1994). The efficacy of fuzzy representations of Uncertainty. *IEEE Transactions on Fuzzy Systems*. 2(1). 4-15.
- Llano L., Zapata G., Ovalle D. (2007). Fuzzy Inference System to Identify Faults Events in Real Time using SOE Registries. *Congreso Colombiano de Computación. Revista Avances en Sistemas e Informática*. 4. 2.
- LPGGIR. Ley General para la Prevención y Gestión Integral de los Residuos. Diario Oficial de la Federación (DOF) Última reforma 19 de junio de 2007. México. <http://www.cddhcu.gob.mx/LeyesBiblio/pdf/263.pdf> (último acceso, diciembre 2007).
- Lobo G.C. A., Herrero L. J., Montero F. O., Tejero M. I. (2002a). Modelling for environmental assessment of municipal solid waste landfills (Part 1: Hydrology). *Waste Management Research* 20. 198–210.
- Lobo G.C. A., Herrero L. J., Montero F. O., Tejero M. I. (2002b). Modelling for environmental assessment of municipal solid waste landfills (Part II: Biodegradation). *Waste Management Research*. 20, 514–528.
- Maris R.A. Evaluaciones de impacto ambiental: utilidad de los modelos matemáticos semiempíricos. *Revista Gerencia Ambiental*. Número 25, Julio. Informe de la 2ª. Jornada de Preservación de Agua, Aire y Suelo en la Industria Petrolera, Neuquén, Argentina. <http://www.ingenieroambiental.com/informes/matematicos.htm> (Documento Web, último acceso noviembre 2006).
- Martín del Brío, B. y Sanz, A. (2007). *Redes Neuronales y Sistemas Borrosos*. Tercera edición. Alfaomega Grupo Editor, México pp. 404.
- Martín S., Marañón E., Sastre H. (2001). Mathematical modelling of landfill gas migration in MSW sanitary landfills. *Waste Management Research*. 19. 425–435.
- MATLAB. The language of technical computing. Version 7.7.0.471 (R2009b). TheMathWorks Inc. U.S. 2009.
- Méndez N.R., Castillo B.E., Sauri R.R., Quintal F.C. (2004). Tratamiento fisicoquímico de los lixiviados de un relleno sanitario. *Ingeniería Revista Académica*. 8(2). 155-163.
- Medina H.S. y Paniagua G.G. (2008). Modelo de inferencia difuso para estudio de Crédito Dyna, Año 75. 154. 215-229.
- Morales O. M., Jurado M. T. y López D. M. L. (2000). Evaluación del sesgo en el análisis exploratorio de datos multivariados. *Psicothema* 12(2). 393-395.
- Montaño J.J., Palmer A. (2003). Numeric sensitivity analysis applied to feedforward neural networks. *Neural Comput Appl*. 12. 119–125.
- Nayla R., Cahuich P., Sauri Riancho M. R., Castillo Borges E. R. y Méndez Novelo R. I. (2005). Monitoreo de emisiones de metano y bióxido de carbono en un relleno sanitario. *Memorias del Primer Congreso Interamericano de Residuos*. Asociación Interamericana de Ingeniería Sanitaria y Ambiental (AIDIS). Federación Mexicana de Ingeniería Sanitaria y Ciencias Ambientales, A. C. (FEMISCA). Sección Mexicana de la AIDIS, Mérida, Yucatán, México.

- Naylor T. H., Balintfy J. L., Burdick D.S. y Kong C. (1982). Quinta reimpression. Técnicas de Simulación en Computadoras. Editorial Limusa, México, Distrito Federal. pp.390.
- Noori R., Abdoli MA, Jalili G. M, Samieifard R. (2009). Comparison of Neural Network and principal Component-Regression Analysis to Predict the Solid Waste Generation in Tehran. *Iranian J Publ Health*. 38(1). 74-84.
- Ogwueleka T.Ch., Ogwueleka F.N. (2010). Modelling energy content of municipal solid waste using artificial neural network. *Iran. J. Environ. Health. Sci. Eng.* 7(3). 259-266.
- Orozco B.C., Perez S.A., Gonzalez D.N.,Rodriguez V.F., Alfayate B. M. (2003). Contaminación Ambiental: Una visión desde la química. Ed. Thomson. España. pp. 680.
- Ozcan H. K, Balkaya N., Bilgili E., Demir G., Nuri O., Bayat C. (2008). Modeling of methane distribution in a landfill using genetic algorithms. *Environ Eng Sci*. 26(4). 1-9.
- Ozkaya, B., Ahmet D., Sinan M.B. (2007). Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. *Environ Modell Softw*. 22. 815-822.
- Papadopoulou M. P., Karatzas G. P., Bougioukou G. G. (2007). Numerical modelling of the environmental impact of landfill leachate leakage on groundwater quality – a field application. *Environ Model Assess* 12. 43–54.
- Pitarque A., Roy J.F. y Ruiz J.C. (1998). Redes neurales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación. *Psicológica*. 19. 387-400.
- Polit M., Estaben M., Labat P. A (2002). fuzzy model for ananaerobic digester, comparison with experimental results. *Engineering Applications of Artificial Intelligence*. 15. 385–390.
- Pohland, F.G., Kim J.C. (1999). “In situ” anaerobic treatment lf leachate in landfill bioreactors. *Water Science Technology*. 40(8). 203-210.
- Robles-Martínez, F. (2008). Generación de Biogás y Lixiviados en Rellenos Sanitarios. 2ª. Ed. Instituto Politécnico Nacional. México. pp. 115.
- Safari E. and Baronian C. (2002). Modeling temporal variations in leachate quantity generated at Kahrizak Landfill. *Proceedings of International Environmental Modeling Software Society, (IEMSS’02), Faculty of Environmental Engineering, University of Tehran, Iran*. 482-487.
- Saltelli A. (2002). Sensitivity analysis for Importance Assessment. *Risk Anal*. 22. 3.
- Schroeder, P. R., Dozier, T.S., Zappi, P. A., McEnroe, B. M., Sjostrom, J. W., and Peyton, R. L. (1994). The Hydrologic Evaluation of Landfill Performance (HELP) Model: Engineering Documentation for Version 3," EPA/600/R-94/168b, September 1994, U.S. Environmental Protection Agency Office of Research and Development, Washington, DC.
- Salvador F. M. y Gargallo P. Análisis Exploratorio de Datos, Estadística <<http://www.5campus.com/leccion/aed>> [Ultimo acceso, noviembre 2010]
- Scenna Nicolás J. (1999). Modelado, Simulación y Optimización de Procesos Químicos. pp. 825.
- Sheinbaum C y Masera O. (2000). Mitigating Carbon Emissions while Advancing National Development Priorities. The Case of Mexico. *Climatic Change*. 47. 259-282.
- SEMARNAT (Secretaria de Medio Ambiente y Recursos Naturales). Normas mexicanas vigentes.(último acceso noviembre 2010a). <http://www.semarnat.gob.mx/leyesy normas/Pages/normasmexicanasvigentes.aspx?&p=1>

- SEMARNAT (Secretaria de Medio Ambiente y Recursos Naturales). Estadísticas ambientales. [http://dgeiawf.semarnat.gob.mx:8080/approot/dgeia\\_mce/html/mce\\_index.html](http://dgeiawf.semarnat.gob.mx:8080/approot/dgeia_mce/html/mce_index.html) (último acceso junio 2010b).
- Serrano C. C. (2006). Alternativas de utilización de biogás de rellenos sanitarios en Colombia, Fundación Universitaria Iberoamericana, Universidad de las Palmas de Gran Canaria (ESPAÑA). Tesis de Grado, Bogota, Colombia.
- Takagi T. and Sugeno M. (1985). Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on fuzzy systems, Man, Cybern.* 15. 116–132.
- Tchobanoglous, G., H. Theisen, S. Vigil. (). *Gestión Integral de Residuos Sólidos. En: Evacuación de residuos sólidos y rechazos.* (1993) McGraw-Hill-Interamericana. España, S.A. México. pp. 1120.
- Vadillo, I. y Carrasco, F. (2005). Estimación del volumen de lixiviado generado en el vertedero de residuos sólidos urbanos de La Mina mediante balance hídrico. *Geogaceta*, 36, 123-126.
- Vavilin, V.A., Rytov, S.V., Lokshina, L.Y., Pavlostathis, S.G., Barlaz, M.A. (2002). Distributed model of solid waste anaerobic digestion: Effects of leachate recirculation and pH adjustment. *Biotechnol. Bioeng.* 81(1). 66-73.
- White J.K., Robinson J.P., Ren Q. (2004). Modelling the biochemical degradation of solid waste in landfills. *Waste Management.* 24. 227–240.
- Zacharof A.I., Butle A.P. (2004). Stochastic modelling of landfill leachate and biogas production incorporating waste heterogeneity. Model formulation and uncertainty analysis. *Waste Management.* 24. 453–462.
- Zadeh L. A. (1996). Nacimiento y evolución de la lógica borrosa, el Softcomputing y la computación con palabras, *Psicothema.* 8(2). 421-429.
- Zadeh L.A. (1972). A Fuzzy-set-teoric Interpretation of linguistic Hedges. *J Cybernetics*, 2. 4-34.
- Zadeh L.A. (1965). Fuzzy Sets. *Information and control* 8. 338-353.