



REPÚBLICA BOLIVARIANA DE VENEZUELA
UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA QUÍMICA

**“ESTRATEGIA DE DOSIFICACIÓN INTELIGENTE PARA EL
PROCESO DE
POTABILIZACIÓN DE AGUA”**

www.bdigital.ula.ve

Trabajo de Grado presentado ante la Escuela de Ingeniería Química como requisito parcial para optar al Título de Ingeniero Químico.

Br. Pablo Alfonso Guerrero Avendaño

Tutores:

Prof. Yohn E. García Z.

Prof. José O. Mayorga

Mérida, 15 de diciembre de 2016.

DERECHO DE AUTOR

Otorgo a la Universidad de Los Andes el derecho de reproducir y difundir el presente Trabajo, con las únicas limitaciones que establece la legislación vigente en materia de Derechos de Autor.

En la ciudad de Mérida a los 15 días del mes de diciembre de dos mil dieciséis.

www.bdigital.ula.ve

Firma

Nombre y Apellido: Pablo Alfonso Guerrero Avendaño.

Cédula de Identidad N° V-20.199.788

Reconocimiento-No comercial-Compartir igual

APROBACION DEL TUTOR

Considero que el Trabajo de Grado titulado:

“ESTRATEGIA DE DOSIFICACIÓN INTELIGENTE PARA EL PROCESO DE
POTABILIZACIÓN DE AGUA”

Realizado por el ciudadano: PABLO ALFONSO GUERRERO AVENDAÑO para optar al título de INGENIERO QUÍMICO reúne los requisitos exigidos por la Escuela de Ingeniería Química de la Universidad de Los Andes, y tiene méritos suficientes para ser sometido a la presentación y revisión exhaustiva por parte del Jurado Examinador que se designe.

www.bdigital.ula.ve

En la ciudad de Mérida, a los 15 días del mes de diciembre de dos mil dieciséis.

Firma:

Prof. Yohn E. García Z.

Co-Tutor

Firma:

Prof. José O. Mayorga

Co-Tutor

ACTA DE VEREDICTO

Nosotros, los abajo firmantes, constituidos como Jurado Examinador, y reunidos en Mérida el día 15 de diciembre de 2016, con el propósito de evaluar el Trabajo de Grado titulado:

“ESTRATEGIA DE DOSIFICACIÓN INTELIGENTE PARA EL PROCESO DE POTABILIZACIÓN DE AGUA”

Presentado por el ciudadano: PABLO ALFONSO GUERRERO AVENDAÑO

Como requisito parcial para optar al título de:

INGENIERO QUÍMICO

Emitimos el siguiente veredicto:

APROBADO:

REPROBADO:

Nota definitiva:

OBSERVACIONES:

Prof. Moira Miranda

Prof. Rubén Montoya

Prof. Yohn García

Prof. José Mayorga

En la ciudad de Mérida, a los 15 días del mes de diciembre de 2016.

AGRADECIMIENTO

Gracias a Dios todopoderoso por haberme dado la paciencia, fuerza y salud para lograr este objetivo.

A mis tutores.

Al Profesor Yohn García por darme la oportunidad y confiar en mí en todo momento. Su apoyo, sabios consejos e invaluable enseñanzas permitieron el desarrollo y culminación de este trabajo de investigación. Los espacios que alcanzamos, me demostraron que el esfuerzo y trabajo duro es el camino correcto hacia el éxito. Muchísimas gracias.

Al profesor José Mayorga, por su asesoramiento y atención constante a lo largo de este trabajo de investigación. Por sus enseñanzas y orientaciones a lo largo de mi formación universitaria. Muchísimas gracias.

A la profesora Silvia Calderón por su orientación durante mi formación universitaria y su ayuda a lo largo de la investigación. A usted, muchas gracias.

Agradezco la colaboración prestada por el personal del Laboratorio de Alimentos y el Laboratorio de Mezclado, Separación y Síntesis Industrial de la Universidad de Los Andes.

Agradezco a mis amigos Rhonald Ortega, Ivana Haack, Néstor Cerrada, Alejandro Villamarin, Nieves Carrero e Ismaela Siso por su apoyo y animo en todas las etapas de este trabajo de investigación.

A Hecmaris Bellorin por su apoyo incondicional, comprensión y atención en todas las etapas de mi formación universitaria. Su presencia, sus palabras y detalles me impulsaron a seguir adelante en los momentos difíciles. A ti, muchísimas gracias.

A mis padres, Blanca y Pablo, mi hermano Juan Pablo, mi sobrina Guadalupe, mi tía Marina, mi prima María Angel, mis abuelos y toda mi familia. Gracias por haber creído en mí y apoyarme siempre, a ustedes les debo todo. Mis logros, son de ustedes también.

INDICE DE CONTENIDO

DERECHO DE AUTOR.....	I
APROBACION DEL TUTOR	II
ACTA DE VEREDICTO	III
AGRADECIMIENTO.....	IV
INDICE DE TABLAS	VII
INDICE DE FIGURAS	IX
RESUMEN	1
I. INTRODUCCIÓN	2
II. MARCO TEÓRICO	5
II.1 POTABILIZACIÓN DE AGUA.....	5
II.1.1 ETAPAS EN EL PROCESO DE POTABILIZACION.....	7
II.2 MODELOS ESTADISTICOS.....	15
II.2.1 REGRESION LINEAL MÚLTIPLE POR MINIMOS CUADRADOS.....	15
II.2.2 REGRESIÓN NO LINEAL POR ANALISIS DE RESIDUOS.....	16
II.2.3 DETECCIÓN DE OBSERVACIONES ATÍPICAS O “OUTLIERS”... ..	18
II.2.4 COEFICIENTE DE CORRELACION DE PEARSON.....	21
II.3 MODELOS DIFUSOS.....	22
II.3.1 LOGICA DIFUSA.....	22
II.3.2 MODELO DIFUSO TIPO MAMDANI.....	32
II.3.3 MODELO DIFUSO TIPO TAKAGI-SUGENO	33
II.3.4 SISTEMAS ADAPTATIVOS DE INFERENCIA NEURO-DIFUSA... ..	36
II.4 ANTECEDENTES DE LA INVESTIGACIÓN	37
II.5 APLICACIONES DE LÓGICA DIFUSA EN INGENIERÍA.....	38
III. METODOLOGÍA.....	40
III.1 RECOPIACIÓN DE LOS DATOS.....	40
III.2 PROCESAMIENTO DE LOS DATOS.....	40
III.3 ESTRATEGIA DEL MODELADO ESTADÍSTICO.....	41
III.3.1 MODELO LINEAL.....	41
III.3.2 MODELO NO-LINEAL.....	42
III.4 ESTRATEGIA DEL MODELADO DIFUSO.....	43
III.4.1 MODELO TIPO MAMDANI.....	43

III.4.2	MODELO TIPO ANFIS.....	45
IV.	RESULTADOS Y DISCUSIÓN	47
IV.1	DOSIFICACIÓN DE COAGULANTE EN LA PLANTA.	47
IV.1.1	SULFATO DE ALUMINIO.....	47
IV.1.2	POLICLORURO DE ALUMINIO, PAC.	69
IV.2	ETAPA DE DESINFECCIÓN.....	73
IV.2.1	COLOR GASEOSO.....	73
IV.3	MODELOS OBTENIDOS EN EL LABORATORIO.....	76
IV.3.1	SULFATO DE ALUMINIO.....	76
V.	CONCLUSIONES Y RECOMENDACIONES	79
VI.	REFERENCIAS	81
VII.	ANEXOS	84
VII.1	ESTRUCTURA DE MODELOS DIFUSOS.	84
VII.1.1	MAMDANI.	84
VII.1.2	ANFIS.....	85
VII.2	GENERADOR DE REGLAS MODELOS MAMDANI.	87
VII.3	TABLAS DE PARAMETROS DE FP DE SALIDA MODELOS ANFIS....	88
VII.4	TABLA DE DATOS DEL LABORATORIO.	90
VII.5	FIGURAS DE LOS MODELOS OBTENIDOS EN PLANTA.	91
VII.5.1	MODELOS “SULFATO DE ALUMINIO”.....	91
VII.5.2	MODELOS “PAC”	94
VII.5.3	MODELOS “COLOR”	98

INDICE DE TABLAS

Tabla 1. Componentes relativos a la Calidad del Agua en Venezuela	6
Tabla 2. Conjuntos difusos para el pH del agua	26
Tabla 3. División de los datos para el modelo de dosificación de Sulfato de Aluminio.....	47
Tabla 4. Conjuntos difusos para la variable “Turbidez”	57
Tabla 5. Conjuntos difusos para la variable “Color”	57
Tabla 6. Conjuntos difusos para la variable “Dosis de sulfato de aluminio”	58
Tabla 7. Reglas difusas generadas para el modelo de dosis de Sulfato de Aluminio.....	59
Tabla 8. Bloque de Reglas base para el modelo Mamdani de dosificación de Sulfato de Aluminio.....	60
Tabla 9. Coeficientes de correlación de Pearson del modelo Mamdani inicial con datos de modelación.....	61
Tabla 10. Nuevos coeficientes de correlación de Pearson modelo Mamdani.	62
Tabla 11. Parámetros de entrada para el Fuzzy Logic Designer Toolbox (MatLab®, 2011).....	64
Tabla 12. Parámetros de entrada para el comando ANFIS (MatLab®, 2011).....	65
Tabla 13. Parámetros de las FP de la variable Sulfato de Aluminio modelo ANFIS.	66
Tabla 14. Coeficientes de correlación de Pearson modelo ANFIS.....	67
Tabla 15. Modelo lineal y no-lineal para la dosificación de PAC.....	70
Tabla 16. Comparación entre los modelos desarrollados con los datos de validación (PAC).....	72
Tabla 17. Modelo lineal y no-lineal para la dosificación de Cloro.....	73
Tabla 18. Comparación entre los modelos desarrollados con los datos de validación (Cloro).....	74
Tabla 19. Modelo lineal y no-lineal para la dosificación de Sulfato de aluminio en el Laboratorio.....	76
Tabla 20. Comparación entre los modelos desarrollados con los datos de validación. Datos de Laboratorio.....	77
Tabla 21. Generador de reglas difusas para el modelo de dosis de PAC en planta.....	87
Tabla 22. Generador de reglas difusas para el modelo de dosis de Cloro en planta.....	87

Tabla 23. Generador de reglas difusas para el modelo de dosis de Sulfato de Aluminio en Laboratorio.....	87
Tabla 24. Parámetros de las FP de la variable PAC en planta.	88
Tabla 25. Parámetros de las FP de la variable Cloro en planta.	89
Tabla 26. Parámetros de las FP de la variable Sulfato de Aluminio en Laboratorio.	89
Tabla 27. Datos para los modelos de laboratorio de dosificación de Sulfato de Aluminio.....	90

www.bdigital.ula.ve

INDICE DE FIGURAS

Figura 1. Diagrama de flujo del proceso de Potabilización de Agua (OSE, 2015).	7
Figura 2. Turbidez del agua (Perlman, 2015).	8
Figura 3. Proceso de Coagulación-Floculación (A.N.C, 2007).	9
Figura 4. Equipo para prueba de jarras.	10
Figura 5. Funciones matemáticas elementales.	17
Figura 6. Análisis de residuos con patrón de x^2 .	17
Figura 7. Análisis de residuos con comportamiento aleatorio.	18
Figura 8. Distribución normal de los residuos del modelo (Probabilidad).	20
Figura 9. Distribución normal de los residuos del modelo (Histograma).	21
Figura 10. Funciones de pertenencia utilizadas en lógica difusa.	24
Figura 11. Funciones de pertenencia para la variable pH.	27
Figura 12. Funciones de pertenencia convencionales para la variable pH.	28
Figura 13. Estrategia para la determinación del número de RB (Piegat, 2001).	30
Figura 14. Arquitectura general de un modelo Difuso tipo Mamdani.	33
Figura 15. Arquitectura general de un modelo Difuso tipo T-S.	34
Figura 16. Estructura de una Red Neuronal Artificial.	36
Figura 17. Neuro-Fuzzy Designer Toolbox (MatLab®, 2011).	46
Figura 18. Neuro-Fuzzy Designer Toolbox "Red Neuronal" (MatLab®, 2011).	46
Figura 19. Distribución normal para los residuos del modelo lineal de Sulfato de Aluminio.	48
Figura 20. Análisis de leverage y Cook del modelo lineal de Sulfato de Aluminio.	49
Figura 21. Desempeño del modelo lineal de Sulfato de Aluminio sobre los datos de validación.	50
Figura 22. Representación de cada variable de entrada vs dosis de Sulfato de Aluminio.	51
Figura 23. Representación del Color vs Residuales del modelo no-lineal (43).	52
Figura 24. Representación de la Alcalinidad vs Residuales del modelo no-lineal (45).	53
Figura 25. Representación de cada variable de entrada con los residuales del modelo (ecuación (47)).	54
Figura 26. Distribución normal para los residuos del modelo no-lineal de Sulfato de Aluminio.	55

Figura 27. Análisis de Leverage y Cook del modelo no-lineal de Sulfato de Aluminio.....	56
Figura 28. Desempeño del modelo no-lineal de Sulfato de Aluminio sobre los datos de validación.....	56
Figura 29. Funciones de pertenencia para el modelo Mamdani de dosificación de Sulfato de Aluminio.....	58
Figura 30. Funciones de pertenencia optimizadas para el modelo Mamdani de dosificación de Sulfato de Aluminio.....	61
Figura 31. Desempeño del modelo Mamdani de Sulfato de Aluminio sobre los datos de validación.....	62
Figura 32. pH vs Sulfato de Aluminio: antes (a) y después (b) de la transformación de los datos.....	63
Figura 33. FP del modelo ANFIS de Sulfato de Aluminio.....	65
Figura 34. Desempeño del modelo ANFIS de Sulfato de Aluminio sobre los datos de validación.....	67
Figura 35. Comparación entre modelos estadísticos con datos de validación (Sulfato de Aluminio).....	68
Figura 36. Comparación entre modelos difusos con datos de validación (Sulfato de Aluminio).....	68
Figura 37. Comparación entre los mejores modelos (Sulfato de Aluminio).....	69
Figura 38. Comparación entre modelos estadísticos con datos de validación (PAC).....	70
Figura 39. Comparación entre modelos difusos con datos de validación (PAC).....	71
Figura 40. Comparación entre mejores modelos (PAC).....	72
Figura 41. Comparación entre modelos estadísticos con datos de validación (Cloro).....	75
Figura 42. Comparación entre modelos difusos con datos de validación (Cloro).....	75
Figura 43. Comparación entre modelos estadísticos con datos de validación. Datos de Laboratorio.....	77
Figura 44. Comparación entre modelos difusos con datos de validación. Datos de Laboratorio.....	78
Figura 45. Comparación entre los mejores modelos. Datos de laboratorio (Sulfato de Aluminio).....	78
Figura 46. Funciones de pertenencia para el modelo Mamdani de dosificación de PAC.....	84
Figura 47. Funciones de pertenencia para el modelo mamdani de dosificación de Cloro.....	84
Figura 48. Funciones de pertenencia para el modelo Mamdani de dosificación de Sulfato de Aluminio (Laboratorio).....	85

Figura 49. FP de las variables de entrada al modelo ANFIS de PAC.	85
Figura 50. FP de las variables de entrada al modelo ANFIS de Cloro.	86
Figura 51. FP de las variables de entrada al modelo ANFIS de Sulfato de Aluminio (Laboratorio).....	86
Figura 52. Desempeño del modelo lineal de Sulfato de Aluminio sobre los datos de modelado.....	91
Figura 53. Desempeño del modelo no-lineal de Sulfato de Aluminio sobre los datos de modelado.	92
Figura 54. Desempeño del modelo Mamdani de Sulfato de Aluminio sobre los datos de modelado.	92
Figura 55. Nuevo desempeño del modelo Mamdani de Sulfato de Aluminio sobre los datos de modelado.....	93
Figura 56. Desempeño del modelo ANFIS de Sulfato de Aluminio sobre los datos de modelado.....	93
Figura 57. Desempeño del modelo lineal de PAC sobre los datos de modelado.	94
Figura 58. Desempeño del modelo lineal de PAC sobre los datos de validación.	94
Figura 59. Desempeño del modelo no-lineal de PAC sobre los datos de modelado.....	95
Figura 60. Desempeño del modelo no-lineal de PAC sobre los datos de validación.....	95
Figura 61. Desempeño del modelo Mamdani de PAC sobre los datos de modelado.....	96
Figura 62. Desempeño del modelo Mamdani de PAC sobre los datos de validación.....	96
Figura 63. Desempeño del modelo ANFIS de PAC sobre los datos de modelado.	97
Figura 64. Desempeño del modelo ANFIS de PAC sobre los datos de validación.	97
Figura 65. Desempeño del modelo lineal de Cloro sobre los datos de modelado.	98
Figura 66. Desempeño del modelo lineal de Cloro sobre los datos de validación.	98
Figura 67. Desempeño del modelo no-lineal de Cloro sobre los datos de modelado.....	99
Figura 68. Desempeño del modelo no-lineal de Cloro sobre los datos de validación.....	99
Figura 69. Desempeño del modelo ANFIS de Cloro sobre los datos de modelado.	100

Figura 70. Desempeño del modelo ANFIS de Cloro sobre los datos de validación.

..... 100

www.bdigital.ula.ve

RESUMEN

La potabilización del agua es un tratamiento que si, por un lado es sumamente vital para la salud pública, por otro lado es un proceso de dosificación complejo, no lineal, y difícil de llevar a cabo de manera óptima mediante el uso de modelos convencionales.

En el presente trabajo, se desarrolló una estrategia de dosificación inteligente, basada en modelos de predicción, con técnicas no convencionales como Lógica Difusa y sistemas híbridos de ésta, con Redes Neuronales Artificiales, a lo que se conocen como Sistemas Adaptativos Neuro-Difusos (del inglés ANFIS). Los modelos difusos implicados en este trabajo de investigación son de tipo Mamdani y ANFIS, los cuales fueron comparados con modelos convencionales lineales y no lineales.

El objetivo principal se enfocó en darle un seguimiento a las diferentes dosificaciones de reactivos en las etapas de coagulación-floculación y desinfección (con cloro gas) en la planta de potabilización ubicada en el Vigía, estado Mérida utilizando la base de datos de los registros de las dosificaciones realizadas en los últimos años.

En la etapa de coagulación-floculación se habían utilizado en la planta dos tipos de coagulante, para los cuales se desarrollaron modelos predictivos. Al primer coagulante, sulfato de aluminio, se le desarrollaron modelos predictivos convencionales (lineales y no lineales) y no convencionales (Mamdani y ANFIS), presentando el modelo neuro-difuso (ANFIS) el mejor ajuste, con un 93.7% de precisión. Con el segundo coagulante (policloruro de aluminio), el modelo ANFIS también reportó el mejor desempeño, con una precisión del 95.1%.

En la etapa de desinfección con cloro no se pudieron establecer las predicciones buscadas, debido a que las variables involucradas no están correlacionadas. El mejor valor obtenido fue para el modelo ANFIS con un 32.1% de precisión.

Finalmente, se procedió a realizar una experiencia de laboratorio que permitiera validar el modelo obtenido en la planta.

I. INTRODUCCIÓN

Como es sabido, el agua es el compuesto que se encuentra en mayor proporción en el planeta, y el más importante para el desarrollo de la vida. Se utiliza para el consumo doméstico, agricultura, ganadería, como medio de calefacción y enfriamiento en procesos industriales, y como materia prima en muchos procesos químicos, entre muchos otros usos.

A medida que transcurre el tiempo, la población y la industrialización crecen a nivel mundial, lo que conlleva a un aumento en el consumo de agua, y de la contaminación de las fuentes naturales de la misma, debido a la gran cantidad de desechos químicos producidos y al vertido de aguas residuales.

El agua presente en la naturaleza tiene ciertas características fisicoquímicas y biológicas que cambian a cada momento, lo que limita su uso para una necesidad específica. Si se requiere agua para el consumo doméstico, es necesario un proceso de potabilización, en donde el efluente cumpla con las normas sanitarias de calidad del agua potable del país en cuestión (OMS, 2008).

El proceso de potabilización del agua tiene como objetivos fundamentales eliminar los microorganismos patógenos que promueven enfermedades y disminuir la turbidez, mediante la dosificación de reactivos químicos en diferentes etapas. En líneas generales, la potabilización incluye la eliminación de compuestos volátiles disueltos, precipitación de impurezas suspendidas, sedimentación, filtración y desinfección.

En el proceso de potabilización es necesario controlar diversos parámetros, como pH, turbidez, sólidos en suspensión, microorganismos, temperatura y dureza (entre otros). Debido al alto número de variables, y a las interacciones que ocurren entre ellas, el proceso es altamente complejo y resulta complicado realizar una estimación que se acerque a la verdadera dosis de reactivos necesarios en un día particular para una planta potabilizadora, lo que se traduce en altos costos de operación. Es por esto, que se plantea el desarrollo de una nueva estrategia, que permita realizar las predicciones correspondientes a las dosis de reactivos en cada una de las etapas del proceso.

La estrategia propuesta se basa en el desarrollo de modelos predictivos que sean capaces de mejorar el ajuste obtenido por los modelos convencionales (lineales y no-lineal). Estos nuevos modelos tendrán una estructura regida por la lógica difusa.

La lógica difusa es una extensión de la lógica convencional, cuya diferencia es la introducción de un grado de imprecisión o vaguedad en los conjuntos numéricos en estudio, con el fin de imitar a la lógica humana, en donde se asigna un peso o grado de importancia a diferentes acciones, para tomar una decisión definitiva.

Con lógica difusa se incluyen variables lingüísticas como Alto, Mediano y Bajo, lo que permite tomar en cuenta otros factores importantes, que con la lógica convencional no son posibles. Con la inclusión de estas variables lingüísticas, es posible plasmar la experiencia por parte de los expertos en el proceso.

El uso de los modelos difusos se justifica cuando el fenómeno en estudio es altamente complejo, es decir, que si con modelos convencionales se obtienen predicciones con un alto porcentaje de precisión, no es necesario el desarrollo de una estrategia inteligente.

En lógica difusa se manejan diversos tipos de estructura, como son Mamdani (Mamdani, 1974) y Takagi-Sugeno (Takagi & Sugeno, 1985). Estas estructuras no tienen una ecuación algebraica definida, y es por esto que, desde su invención, (Zadeh, 1965) muchos matemáticos se han opuesto a estas nuevas técnicas.

Con el paso del tiempo, se han buscado métodos de optimización de modelos difusos, y fueron Jyh-Shing y Jang (Jyh-Shing & Jang, 1991) quienes propusieron el uso de Redes Neuronales Artificiales sobre las estructuras de tipo Takagi-Sugeno, dando lugar a los Sistemas Adaptativos de Inferencia Neuro-Difusa (con sus siglas en inglés ANFIS), tratando de reducir el error entre la respuesta del modelo y el valor real. Esta técnica tiene la particularidad de que permite al modelo entrenar y aprender a partir de datos conocidos.

Con el uso de estructuras difusas y técnicas de optimización, los problemas complejos o que presentan comportamientos altamente no lineales pueden ser solucionados con gran precisión.

Es por esto que se plantean los siguientes objetivos en este trabajo de investigación:

- Desarrollar una estrategia de dosificación inteligente para un proceso de potabilización de agua basada en modelos difusos.
- Desarrollar modelos con estructuras de tipo Mamdani y ANFIS para cada una de las etapas del proceso de potabilización de agua en una planta.
- Desarrollar modelos lineales y no-lineales para cada una de las etapas del proceso de potabilización de agua en una planta.
- Desarrollar modelos con estructuras de tipo Mamdani y ANFIS para el proceso de potabilización, a nivel de laboratorio.
- Desarrollar modelos lineales y no-lineales para para el proceso de potabilización a nivel de laboratorio.
- Comparar y validar los modelos obtenidos.

www.bdigital.ula.ve

II. MARCO TEÓRICO

II.1 POTABILIZACIÓN DE AGUA

Las fuentes de agua natural contienen, microorganismos y compuestos que pueden ser nocivos para la salud humana. Es por esto que resulta necesario un proceso de potabilización, cuyas etapas dependerán de la caracterización del afluente que será tratado.

El proceso de potabilización se aplica con un fin higiénico a cualquier tipo de agua, buscando la disminución de la concentración de sales, compuestos orgánicos, protozoos parásitos y otros microorganismos, para obtener un efluente que cumpla con los índices del agua potable, y no ocasione riesgos significativos en la salud de las personas que la consumen (OMS, 2008).

También hay en el procedimiento una finalidad estética, debido a que se reduce el grado de turbidez y se mejoran las propiedades organolépticas como sabor, olor y color (OPS/CEPIS, 2002).

Con el paso de los años, se han establecido índices de potabilidad en el agua como se pueden observar en la Tabla 1, en donde se presentan los valores deseables y máximos permitidos establecidos en las Normas Sanitarias de Calidad del Agua Potable, NSCAP de Venezuela (Gaceta Oficial de la Republica de Venezuela, 1998).

Tabla 1. Componentes relativos a la Calidad del Agua en Venezuela
(Normas Sanitarias de Calidad del Agua Potable. (**Gaceta Oficial de la Republica de Venezuela, 1998**))

Parámetro	Unidad	Valor deseable menor a	Valor máximo aceptable (a) ¹
Color	UCV (b) ²	5	15 (25)
Turbiedad	UNT (c) ³	1	5 (10)
Sabor y olor	-	Aceptable para la mayoría de los consumidores	
Sólidos disueltos totales	mg/L	600	1000
Dureza total	mg CaCO ₃ /L	250	500
pH	-	6.5-8.5	9
Aluminio	mg/L	0.1	0.2
Cloruro	mg/L	250	300
Cobre	mg/L	1.0	(2.0)
Cloro Residual	mg/L	-	1.0 (3.0) (a)
Coliformes Fecales	NMP/100 ml	-	0

¹ (a) Los valores entre paréntesis son aceptados provisionalmente en casos excepcionales, plenamente justificados ante la autoridad sanitaria.

² (b) UCV: Unidades de color verdadero.

³ (c) UNT: Unidades Nefelométricas de Turbidez.

Para que el proceso cumpla los índices de potabilidad establecidos, se requiere del uso de operaciones unitarias tales como sedimentación y filtración, y de etapas de dosificación de reactivos, las cuales serán diferentes en cada planta de potabilización, ya que estas se deben adecuar a las características del afluente.

II.1.1 ETAPAS EN EL PROCESO DE POTABILIZACION

El proceso de potabilización de agua en líneas generales consta de las siguientes etapas, como se ve en la Figura 1:

- Cribado.
- Coagulación-Floculación.
- Sedimentación.
- Filtración.
- Desinfección.

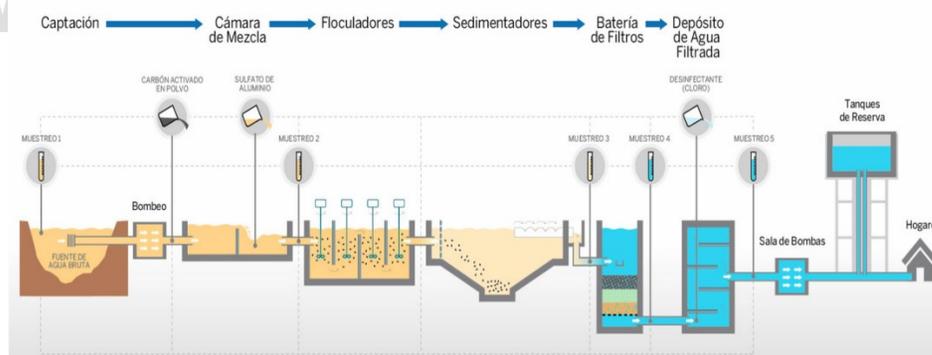


Figura 1. Diagrama de flujo del proceso de Potabilización de Agua (OSE, 2015).

II.1.1.1 CRIBADO

Consiste en hacer pasar el agua a través de un tamiz o rejilla, para retener los sólidos de mayor tamaño como algas, palos, latas, bolsas entre otros.

II.1.1.2 COAGULACIÓN-FLOCULACIÓN

Es una de las etapas más importantes en el proceso. En ella, se reduce el grado de turbidez del agua, lo que es estéticamente primordial, ya que a ninguna persona le sería agradable beber agua con un aspecto sucio.

La turbidez es una propiedad óptica basada en la medida de luz reflejada por las partículas suspendidas en el agua, que no se depositan por simple sedimentación, y este parámetro es directamente proporcional a la concentración de las mismas. Las partículas que generan la turbidez son los sedimentos procedentes de la erosión, sedimentos resuspendidos (debido a los peces que se alimentan del fondo), arcilla, algas, plancton, fango, materia orgánica, microorganismos, etc. Estas partículas suspendidas poseen cargas superficiales que generan una repulsión entre sí ellas, lo que no les permite aglomerarse en partículas más grandes (Parsons & Jefferson, 2006).

Para realizar la medición de la turbidez, se utiliza un turbidímetro que mide la intensidad de la luz dispersada a 90° cuando un haz pasa a través de una muestra de agua (Lenntech, 2015). La unidad de medición es UNT (Unidades Nefelométricas de Turbidez).

El agua con una turbidez menor a 5 UNT suele ser aceptable para los consumidores, aunque para un proceso de desinfección eficaz, es necesaria una turbidez menor a 0.1 UNT (OMS, 2008). En la Figura 2, se puede observar la turbidez para diferentes muestras de agua, y el valor en UNT correspondiente.

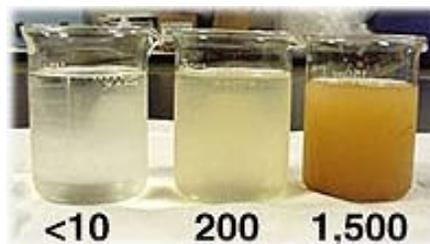


Figura 2. Turbidez del agua (Perlman, 2015).

En la etapa de coagulación-floculación, se adiciona un compuesto que desestabilice las cargas eléctricas superficiales de las partículas, haciendo que ellas se acerquen entre sí, formando “flóculos” más pesados y grandes (ver Figura 3). En este proceso, técnicamente no se destruyen microorganismos patógenos, pero se reducen entre un 60-90% la concentración de los mismos, debido a que son encapsulados en los flóculos formados (Cheremisinoff, 2002).



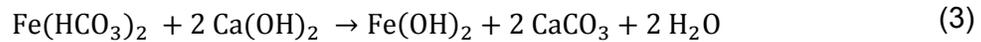
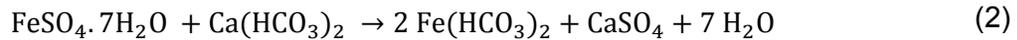
Figura 3. Proceso de Coagulación-Floculación (A.N.C, 2007)

En la Figura 3 se observa que, en el punto de mezcla coagulante-agua, se agita rápidamente para que el compuesto agregado se distribuya de manera eficiente en el agua, y comiencen a formarse los flóculos. Posteriormente, se realiza una agitación lenta, dándole suficiente tiempo de residencia a los flóculos formados para que se aglomeren de forma eficiente, aumenten de tamaño, y puedan ser separados del líquido en un tanque de sedimentación.

Los coagulantes utilizados son compuestos de aluminio o hierro, que producen hidróxidos insolubles en medio alcalino, y que arrastran impurezas suspendidas (OPS/CEPIS, 2002). El coagulante de aluminio más utilizado es el sulfato de aluminio, (alumbre) cuya fórmula química es $\text{Al}_2\text{SO}_4 \cdot 18\text{H}_2\text{O}$. En la ecuación (1), se presenta la reacción que ocurre cuando este compuesto es dosificado en el agua.



El coagulante de hierro que más se emplea es el sulfato de hierro II. Su fórmula química es $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$. En las ecuaciones (2) y (3) se presentan las reacciones que ocurren al emplearlo como coagulante.



Los coagulantes deben cumplir con tres propiedades claves (Toprak, 2006):

- Debe tener, preferiblemente, un catión trivalente, ya que las partículas o coloides que comúnmente se encuentran en las aguas sin tratar, están cargadas negativamente. El hecho de ser trivalente es debido al aumento en la eficiencia de neutralización con la carga del catión.
- La potabilización se lleva a cabo para purificar el agua, por lo tanto, el compuesto que se agregue no debe ser tóxico.
- El coagulante agregado debe precipitar a pH lo más próximo a 7, de manera que beneficie el proceso de remoción de los coloides.

Para conocer la cantidad de coagulante a dosificar, es necesario realizar la prueba de jarras. El equipo consiste en 6 vasos de precipitado, donde se colocan muestras de agua, a la que se añaden cantidades crecientes de coagulante. Tiene paletas que permiten el mezclado a diferentes velocidades, como se ve en la Figura 4.



Figura 4. Equipo para prueba de jarras.

La prueba de jarras permite seleccionar:

- Tipo de coagulante.
- Dosificación óptima de coagulante.
- pH óptimo de operación.
- Tiempo de sedimentación óptimo
- Turbidez deseada
- Secuencia de adición de compuestos químicos.
- Energía óptima y tiempo de mezclado rápido.
- Energía óptima y tiempo de mezclado lento.

Básicamente, el procedimiento consiste en colocar muestras de agua en los vasos de precipitado, y agregar cantidades crecientes de coagulante de una determinada concentración. Las muestras se agitan rápidamente durante 30-60 segundos, a una velocidad que se puede modificar (simulando la mezcla rápida que ocurre en la planta potabilizadora). Luego, se agitan lentamente entre 15-20 minutos para promover la floculación. Se dejan en reposo durante 20-45 minutos. Se observa y anota la apariencia y el tamaño del flóculo, su tiempo de formación y las condiciones en reposo (tamaño final, facilidad de aglomeración). Se puede medir, además la turbidez y el pH final. Con esta información, se puede determinar la dosis óptima de coagulante en función de la muestra de mejor calidad con la menor cantidad de coagulante agregado (Wang, Hung, & Shamma, 2005).

II.1.1.3 SEDIMENTACIÓN.

Los flóculos formados en la etapa de coagulación-floculación tienen una densidad mayor a la del agua tratada, sin embargo permanecen en suspensión aquellos de menor tamaño. Es por esto que la siguiente etapa del proceso de potabilización es la sedimentación, cuyo agente de separación es la gravedad y el gradiente impulsor es la diferencia de densidades entre los flóculos y el agua.

La finalidad de esta etapa es separar los flóculos suspendidos en un estado más calmado, de manera de que comience la precipitación. En el proceso de potabilización de agua, los flóculos precipitan en régimen laminar, por lo tanto se

rigen por la ecuación (4) conocida como la ley de Stokes (Parsons & Jefferson, 2006).

$$v_t = \frac{g * (\rho_f - \rho) * d^2}{18 * \mu} \quad (4)$$

En donde v_t es la velocidad terminal, ρ_f la densidad del floculo, d es el diámetro de la partícula, ρ la densidad del agua y μ la viscosidad del agua.

II.1.1.4 FILTRACIÓN.

La etapa de filtración es utilizada debido a que la sedimentación no proporciona una remoción total de las partículas suspendidas, por lo tanto, se emplea la filtración a través de un lecho poroso, conformado por grava, arena y antracita, donde el agua se mueve por gradiente de presión.

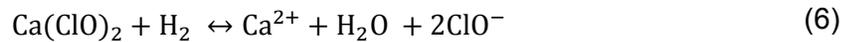
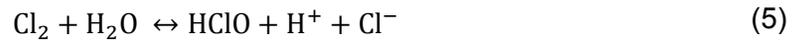
II.1.1.5 DESINFECCIÓN.

El mayor impacto en la salud pública generado por el agua no potable es la enfermedad diarreica, y muchos de los brotes virales están relacionados a la calidad microbiológica del agua potable, ya que sin haber tenido una etapa de desinfección, puede contener microorganismos como bacterias, virus, helmintos y protozoos. La mayoría de estos microorganismos presentes en el agua provienen del contacto con materia orgánica en descomposición y materia fecal. Los microorganismos patógenos son las bacterias, virus y protozoos más peligrosos para la salud. Estos últimos se encuentran comúnmente en las aguas afectadas por contaminación fecal. (Lenntech, 2015). Es por esto que se realiza la etapa de desinfección, mediante la adición de compuestos con propiedades oxidativas, como el cloro.

La fuerte propiedad oxidativa del cloro libre, el cual se encuentra como cloro molecular Cl_2 , ácido hipocloroso $HOCl$ y como ión hipoclorito OCl^- lo hace un excelente biocida, capaz de destruir las bacterias, oxidar enzimas y proteínas (Black & Veatch, 2010), descomponer plaguicidas y oxidar especies disueltas, como el

manganeso (II), formando productos insolubles, que pueden eliminarse mediante filtración (OMS, 2008).

El cloro, al disolverse en el agua, forma ácido hipocloroso HOCl y libera iones hidrógeno H⁺ (como se ve en la ecuación (5)). Los hipocloritos adicionados al agua como sales (de calcio y/o sodio) se disocian, formando iones hipocloritos ClO⁻ (6), que reaccionan con los iones hidrógeno para formar ácido hipocloroso (7).



La acción desinfectante del cloro se debe al ácido hipocloroso, el cual se encuentra en mayor proporción cuando el pH del agua se mantiene entre 4 y 7. La cantidad de cloro necesario para eliminar los microorganismos patógenos se denomina demanda de cloro, y se usa el término cloro residual para designar la cantidad de cloro libre remanente en el agua después del proceso de desinfección (Mayorga, 2009). El cloro tiene la ventaja de tener un efecto residual sobre el agua, ya que se agrega un exceso como medida de seguridad, debido a posibles infecciones posteriores al proceso de desinfección, y el valor máximo permitido de cloro residual es de 1 mg/L.

Además del cloro, existen otros tipos de desinfectantes, tales como el ozono, cloraminación, dióxido de cloro y radiación UV.

Para el análisis microbiológico del agua, debido a las dificultades de analizar microorganismos específicos, y a la necesidad de hacer una evaluación rápida de la presencia de patógenos en el agua, se trabaja con microorganismos indicadores que logran permanecer más tiempo en el agua que los patógenos. Estos microorganismos indicadores están presentes en el tracto gastrointestinal de los animales de sangre caliente y se comportan de igual manera que los patógenos en los sistemas de desinfección, y se analiza entonces este grupo, llamado coliforme. (Campos Pinilla, 2003).

El grupo de microorganismos “coliformes totales” incluye, bacterias de forma bacilar, aeróbicas, o anaeróbicas facultativas (viven en presencia o ausencia de oxígeno) que fermentan la lactosa, formando gas en un rango de 24-48 horas a 35°C. Además son conocidos como “Gram negativos” ya que no cambian el color de la solución de Gram (Mayorga, 2009). Un sub-grupo de los coliformes totales, llamado coliformes fecales o termoresistentes, es el más adecuado como indicador de contaminación fecal. Los coliformes totales están conformados por familias de bacterias como Escherichia, Citrobacter, Enterobacter, Klebsiella y los coliformes fecales están conformados principalmente por la Escherichia Coli (Dufour, y otros, 2003).

La determinación de coliformes se lleva a cabo por diferentes métodos, entre ellos el de tubos múltiples, en donde se determina el Número Más Probable (NMP) de colonias de bacterias presentes en una muestra.

- **NUMERO MÁS PROBABLE (NMP).**

La aplicación de este método bacteriológico consiste en dos fases: la presuntiva y la confirmativa. En la fase presuntiva, el medio de cultivo es caldo lactosado, y los resultados permiten presumir la presencia de coliformes totales al incubar la muestra y sus diluciones a 35 °C. La producción de gas se observa luego de 48 horas, en un pequeño tubo invertido, que se coloca dentro de los tubos donde se vierten las muestras. La determinación de coliformes fecales se realiza a partir de los tubos positivos de la prueba presuntiva, de donde se saca una pequeña porción de líquido y se lleva a un segundo medio de cultivo (caldo de bilis-verde brillante). Luego, se incuban a la misma temperatura durante 48 horas y si se observa formación de cualquier cantidad de gas, el ensayo se considera positivo (Mayorga, 2009).

Para hallar el índice coliforme (NMP/100 mL), se utiliza una tabla, donde aparecen las diluciones utilizadas y los posibles resultados cuando se siembran 3 ó 5 tubos por dilución (3 para aguas residuales y 5 para aguas blancas) (Mayorga, 2009).

II.1.1.6 CONTROL DEL pH.

Consiste en la dosificación de hidróxido de calcio Ca(OH)_2 que en presencia de agua, se comporta como una base fuerte. La finalidad, es ajustar el valor del pH hasta el rango óptimo de la potabilización, es decir, 6,5 a 8,5 (ver Tabla 1).

Una vez definidas las etapas generales del proceso de potabilización de agua, se explicarán los términos utilizados a lo largo del trabajo de investigación.

II.2 MODELOS ESTADÍSTICOS.

Los modelos estadísticos no son más que relaciones matemáticas de variables de entrada(s) y salida(s) vinculadas entre sí. A continuación, se detallan los modelos estadísticos utilizados en el presente trabajo.

II.2.1 REGRESION LINEAL MÚLTIPLE POR MINIMOS CUADRADOS.

Los modelos de regresión lineal por mínimos cuadrados siguen la estructura presentada en la ecuación (8).

$$Y = \beta_1 + \beta_2 * X_1 + \beta_3 * X_2 + \beta_4 * X_3 \dots + \beta_k * X_k + e \quad (8)$$

En donde, y es la respuesta, β_n los parámetros lineales, X_n los predictores y e el error del modelo. Para una muestra de n observaciones, se tiene el siguiente sistema de n ecuaciones lineales:

$$\begin{cases} Y_1 = \beta_1 + \beta_2 * X_{11} + \beta_3 * X_{21} + \beta_4 * X_{31} \dots + \beta_{k1} * X_{k1} + e_1 \\ Y_2 = \beta_1 + \beta_2 * X_{12} + \beta_3 * X_{22} + \beta_4 * X_{32} \dots + \beta_{k2} * X_{k2} + e_2 \\ \vdots \\ Y_n = \beta_1 + \beta_2 * X_{1n} + \beta_3 * X_{2n} + \beta_4 * X_{3n} \dots + \beta_{kn} * X_{kn} + e_n \end{cases}$$

Al expresar este sistema de forma matricial, como se ve en la ecuación (9).

$$Y = X.B + e \quad (9)$$

Se tiene:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & \dots & X_{k1} \\ 1 & X_{12} & \dots & X_{k1} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 1 & X_{1n} & \dots & X_{kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix}, e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

Conociendo las matrices de Y y X , es posible determinar los coeficientes β mediante un cálculo matricial (Chatterjee & Hadi, 1986), como se ve en la ecuación (10).

$$\beta = (X^T * X)^{-1} * X^T * Y \quad (10)$$

Si bien, para la estimación por mínimos cuadrados de los coeficientes de un modelo de regresión, solo es necesario asumir la linealidad y normalidad de los mismos, se debe comprobar que dichas suposiciones se cumplen mediante un análisis de probabilidad normal de los residuos (ver II.2.3.3).

II.2.2 REGRESIÓN NO LINEAL POR ANALISIS DE RESIDUOS.

Los modelos de regresión no-lineal pueden ser obtenidos a partir de la ecuación (10), pero antes, se debe aplicar una transformación a los predictores contenidos en la matriz "X", ó en algunos casos, al vector de respuesta "Y". Para encontrar la transformación adecuada de los datos, se realiza un análisis de residuos tal como muestran Ingels (M. Ingels, 1980).

El análisis de residuos se basa en la representación gráfica de cada una de las variables con respecto a los residuos del modelo encontrado. En estas representaciones, se buscan patrones de comportamiento que puedan ser descritos por funciones elementales conocidas (x , $\log X$, $\exp(X)$, X^2 , etc.), como se observa en la Figura 5.

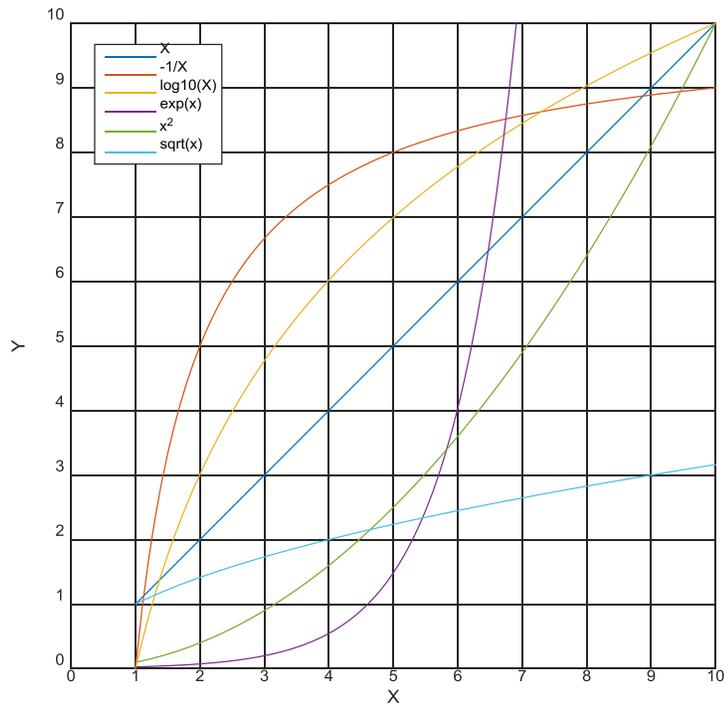


Figura 5. Funciones matemáticas elementales.

Al encontrar un patrón (Figura 6) que sea descrito por una función elemental, se debe agregar este término a la matriz de predictores X , y utilizar la ecuación (10) para generar un nuevo modelo.

Se considera que el modelo está terminado, cuando al representar cada una de las variables con los residuos del modelo, muestre un comportamiento aleatorio (figura 7).

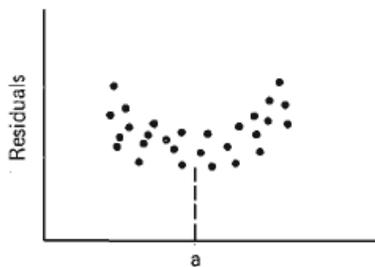


Figura 6. Análisis de residuos con patrón de x^2

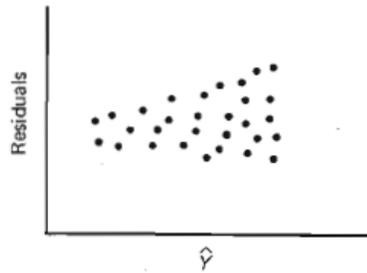


Figura 7. Análisis de residuos con comportamiento aleatorio.

II.2.3 DETECCIÓN DE OBSERVACIONES ATÍPICAS O “OUTLIERS”.

Con la finalidad de determinar si una observación es atípica (outlier) con respecto a los predictores, se estudia la distancia que tiene el valor residual de esta observación con respecto a las demás. Si esta distancia es significativamente grande, la observación se considera atípica o “un outlier”.

Se toman los siguientes criterios:

II.2.3.1 DISTANCIA DE LEVERAGE.

El análisis mediante las distancias de leverage permite determinar si una observación está muy alejada del resto. Por lo tanto, existirá un leverage (h_{ii}) asociado a cada observación, el cual se encuentra en la diagonal principal de la matriz H (Renatas, Ángel, & Perez, 2015), y que se determina con la ecuación (11).

$$H = X * (X^T * X)^{-1} * X^T \quad (11)$$

El leverage tiene las siguientes propiedades:

- Será tanto más grande cuanto sea diferente, en términos de los predictores, del resto de las observaciones.
- Está acotado entre $\frac{1}{n} \leq h_{ii} \leq 1$, donde n es el número de observaciones.

Ahora bien, el análisis mediante las distancias de leverage proporciona una idea de las posibles observaciones atípicas incluidas en el modelo. El criterio para presumir a una observación como atípica, será cuando la distancia de leverage sea mayor a 2 veces la media de todos los leverage, como se ve en la ecuación (12).

$$\text{Observación (posiblemente atípica) si } h_{ii} > 2 * \bar{h} \quad (12)$$

Los resultados de este método (h_{ii}) son necesarios para realizar el análisis mediante las distancias de Cook, el cual comprueba la existencia de observaciones atípicas incluidos en el modelo.

II.2.3.2 DISTANCIA DE COOK.

El análisis mediante la distancia de Cook es una medida compleja que permite detectar la influencia de una observación sobre un modelo, cuando dicha observación no es tomada en cuenta para la construcción del mismo (MatLab®, 2011). La distancia de Cook, para la observación “i”, se calcula con la ecuación (13).

$$d_i = \frac{r_i^2}{p * MSE} * \left(\frac{h_{ii}}{(1 - h_{ii})^2} \right) \quad (13)$$

$$MSE = \sum_{i=1}^n \frac{r_i^2}{(n - p)} \quad (14)$$

En donde r_i^2 es el residuo, h_{ii} es la distancia de leverage, p el número de predictores en el modelo de regresión, MSE la media del cuadrado de los residuos y n el número de observaciones.

El criterio para etiquetar a una observación como atípica será cuando la distancia de Cook sea mayor a 3 veces la media de todas las distancias de Cook, como se ve en la ecuación (15).

$$\text{Observación atípica si } d_i > 3 * \bar{d} \quad (15)$$

Para obtener los valores de las distancias de Cook de las observaciones de los modelos, se utiliza el comando “fitlm” de Matlab (MatLab®, 2011).

II.2.3.3 ANALISIS DE DISTRIBUCIÓN NORMAL DE RESIDUOS.

Este análisis se basa en comprobar la suposición de una distribución normal de las observaciones en el momento de construir el modelo. Esto se logra con la representación gráfica de la distribución de probabilidad normal (Figura 8), o el histograma de los residuos del modelo (Figura 9). Si se detectan observaciones que se alejan significativamente de la línea en la Figura 8, dichos valores se pueden etiquetar como atípicos (NIST/SEMATECH, 2012).

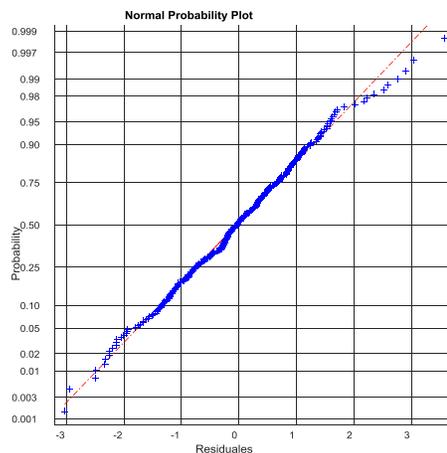


Figura 8. Distribución normal de los residuos del modelo (Probabilidad).

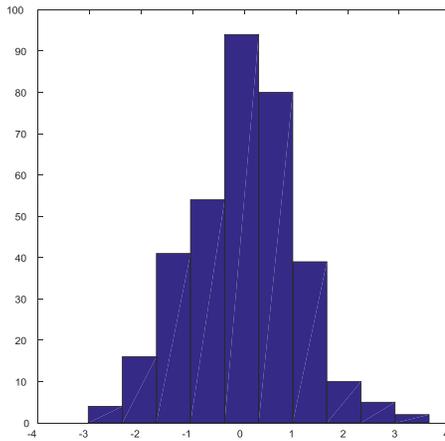


Figura 9. Distribución normal de los residuos del modelo (Histograma).

II.2.4 COEFICIENTE DE CORRELACION DE PEARSON.

Es una medida normalizada (en el intervalo $[0,1]$), que proporciona la capacidad del modelo en cuestión, de reproducir la respuesta real del sistema. Se denota como R^2 y se presenta en la ecuación (16). El coeficiente de correlación de Pearson es un patrón de comparación entre modelos, que tienen una misma finalidad, pero si estos modelos difieren en el número de predictores, es necesario aplicarle una transformación al coeficiente R^2 , que se conoce como R^2_{ajustado} (MatLab®, 2011) y que se presenta en la ecuación (17).

$$R^2 = 1 - \frac{SSR}{SST} \quad (16)$$

$$R^2_{\text{ajustado}} = 1 - \frac{(n-1)}{(n-m)} * \frac{SSR}{SST} \quad (17)$$

$$\bar{Y} = \frac{\sum_{i=1}^n Y}{n} \quad (18)$$

$$SST = \sum_{i=1}^n (Y - \bar{Y})^2 \quad (19)$$

$$SSR = \sum_{i=1}^{345} (Y - \hat{Y})^2 \quad (20)$$

El término \bar{Y} corresponde a la media de la respuesta real del sistema, n es el número de observaciones, m es el número de predictores en el modelo, SST es la suma del cuadrado de los errores, SSR es la suma del cuadrado de los residuos y \hat{Y} es la respuesta del modelo (ecuaciones (18) a (20)).

II.3 MODELOS DIFUSOS.

II.3.1 LOGICA DIFUSA.

La lógica difusa es una extensión de la lógica convencional, cuya diferencia es la introducción de cierta imprecisión o vaguedad en los conjuntos en estudio, con el fin de emular a la lógica humana. El concepto de lógica difusa fue presentado por Zadeh (Zadeh, 1965), quien define a los conjuntos difusos o “fuzzy sets” como aquellos conformados por las variables difusas o imprecisas en estudio. La idea fundamental de la lógica difusa es permitir que el grado de pertenencia de un valor contenido en un conjunto, pueda variar desde 0 hasta 1 en lugar de tomar sólo dichos valores extremos. La arquitectura de un modelo basado en lógica difusa es la siguiente:

- Variables de Entrada y Salida Difusas.
- Fuzzificación.
- Bloque de Inferencia Difusa.
- Defuzzificación.

II.3.1.1 VARIABLES DIFUSAS.

Las variables difusas están definidas o discretizadas en un conjunto universal U, que contiene a todos los posibles valores que puede tomar dicha variable en un

sistema. Es importante definir el conjunto universal para que los valores que tome la variable no sean erróneos (Ross, 2004). Estas variables pueden ser físicas, pueden ser estados de funcionamiento de un sistema, o también sistemas de comparación (Piegat, 2001).

Por ejemplo, para la medición del pH se define como conjunto universal U:

$$U = [0, 14]$$

La lógica difusa permite que las variables estén conformadas por conjuntos tales como muy ácido, medio ácido, poco ácido, neutro, poco básico, medio básico, etc. los cuales son interpretados de diferente forma, según lo establezca la persona que realiza el estudio.

II.3.1.2 CONJUNTOS DIFUSOS.

Los conjuntos difusos se utilizan para clasificar de manera cualitativa las variables que ya han sido discretizadas. Estos conjuntos se encuentran definidos por variables lingüísticas, que no tienen un límite específico dentro del conjunto universal U.

Por ejemplo, cada persona puede clasificar el sitio donde se encuentre en función de su sensación térmica como frío, tibio o caliente, y es aquí donde la imprecisión entra en juego. Una persona que vive en Noruega puede establecer como caliente un lugar con una temperatura de 20°C, mientras que para una persona que viva en Venezuela, caliente podría ser un sitio con una temperatura de 38°C. Esa es la razón por la cual se debe asignar un rango a los conjuntos difusos, que describan a la variable difusa, en función del conocimiento que se tenga sobre el fenómeno en estudio que representa dicha variable.

Una vez definidos los conjuntos difusos, se les debe asignar una función de pertenencia, la cual se describe a continuación.

II.3.1.3 FUNCIÓN DE PERTENENCIA.

Los conjuntos difusos están caracterizados por las funciones de pertenencia (FP), las cuales asignan un valor en el intervalo $[0,1]$ a la función que le corresponde el valor de la variable difusa en determinado momento. Así, 1 representa la pertenencia total y 0 ninguna pertenencia de la variable a determinado conjunto o función de pertenencia (Zadeh, 1965). Existen diferentes tipos de FP como se ve en la Figura 10. Las FP se denotan con el símbolo μ_A , en donde el sub-índice señala al conjunto difuso que representa (Piegat, 2001).

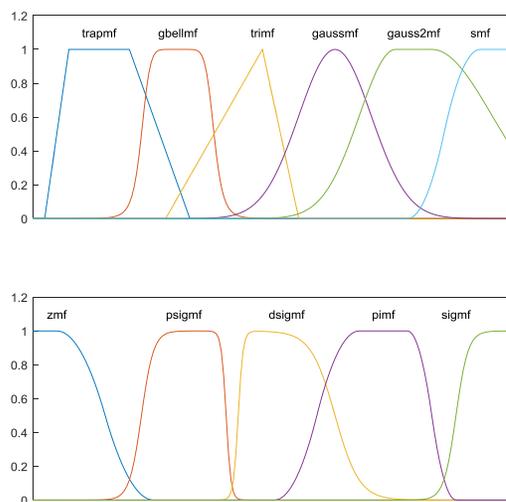


Figura 10. Funciones de pertenencia utilizadas en lógica difusa

- **FUNCIÓN DE PERTENENCIA TRIANGULAR (trimf).**

Son las más utilizadas, debido a su practicidad y simplicidad. Una de las mayores ventajas es que requieren poca información para ser generadas, pues solo dependen de 3 parámetros, como se muestra en la ecuación (21). Cabe mencionar, que estas funciones de pertenencia son de gran ayuda en los procesos con ausencia de grandes cantidades de datos (Piegat, 2001).

$$\mu(x; a, b, c) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & c \leq x \end{cases} \quad (21)$$

En donde a y c son los puntos de la base del triángulo y b es el punto donde se ubica el vértice superior (MatLab®, 2011).

- **FUNCION DE PERTENENCIA TRAPEZOIDAL (trapmf).**

Al igual que las FP triangulares, son muy utilizadas ya que requieren sólo 4 parámetros para ser generadas (ver ecuación (22)).

$$\mu(x; a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases} \quad (22)$$

En donde a y d son los puntos de la base y b y c son los hombros del trapecio (MatLab®, 2011).

- **FUNCION DE PERTENENCIA GAUSSIANA (gaussmf).**

Las FP gaussianas dependen de la media y desviación estándar del conjunto de datos contenidos en el conjunto difuso al cual pertenecen. Son un poco más complejas de usar, y su estructura se representa en la ecuación (23).

$$\mu(x; \sigma, c) = e^{-\frac{(x-c)^2}{2*\sigma^2}} \quad (23)$$

La elección del tipo de FP se realiza según los datos que se vayan a estudiar. Las más utilizadas son las de segmentos rectos (triangulares y trapezoidales). Durante el desarrollo de este trabajo de investigación, se utilizaran los tres tipos de FP. Para complementar las estructuras de las demás FP consultar Piegat (Piegat, 2001).

Ahora se da un ejemplo en donde se crea como variable difusa al pH, que está discretizada en el conjunto universal:

$$U = [0, 14]$$

Se definen 3 conjuntos difusos ácido, neutro y básico, y se le asigna a cada uno de ellos una FP. La imprecisión asociada a dicha variable se observa cuando los valores del pH se acercan a 7 ya que un valor de pH igual a 6.9, la muestra es relativamente neutra, casi neutra o muy poco ácida.

Se le asigna un rango a los conjuntos difusos, los cuales se muestran en la Tabla 2 y se representan en la Figura 11:

Tabla 2. Conjuntos difusos para el pH del agua

CONJUNTO DIFUSO	VALORES ASIGNADOS
Ácido	$0 < \text{pH} < 7$
Neutro	$6 < \text{pH} < 8$
Básico	$7 < \text{pH} < 14$

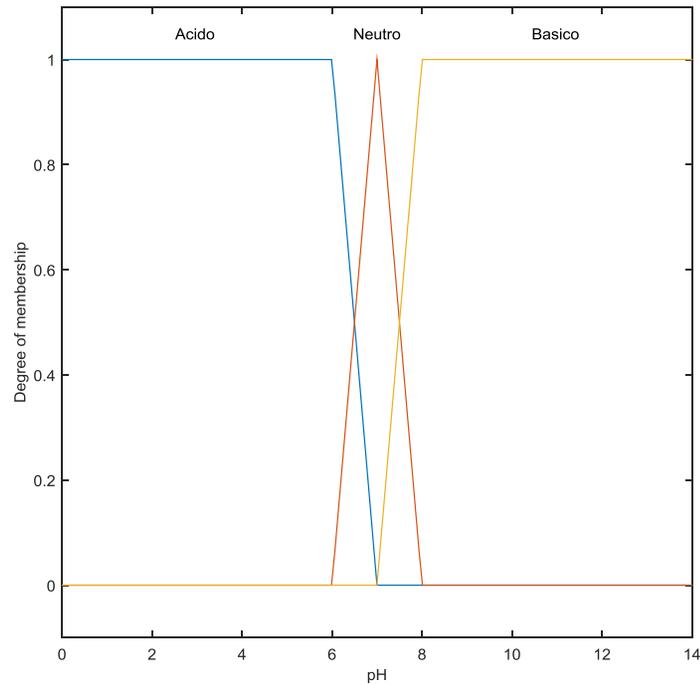


Figura 11. Funciones de pertenencia para la variable pH

Estas FP tienen las funciones algebraicas presentadas en las ecuaciones (24), (25) y (26).

$$\mu_{\text{Acido}}(x; 0,0,6,7) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x < 6 \\ 7-x & 6 \leq x < 7 \\ \frac{7-6}{7-6} & 7 \leq x \\ 0 & \end{cases} \quad (24)$$

$$\mu_{\text{Neutro}}(x; 6,7,8) = \begin{cases} 0, & x \leq 6 \\ \frac{x-6}{7-6}, & 6 \leq x \leq 7 \\ \frac{8-x}{8-7}, & 7 \leq x \leq 8 \\ 0, & 8 \leq x \end{cases} \quad (25)$$

$$\mu_{\text{Basico}}(x; 7,8,14,14) = \begin{cases} 0 & x < 7 \\ \frac{7-x}{7-6} & 7 \leq x < 8 \\ 1 & 8 \leq x < 14 \\ 0 & 14 \leq x \end{cases} \quad (26)$$

En la Figura 11, se puede observar que para valores cercanos a 7, existe una pertenencia tanto para el conjunto Neutro como para los demás conjuntos.

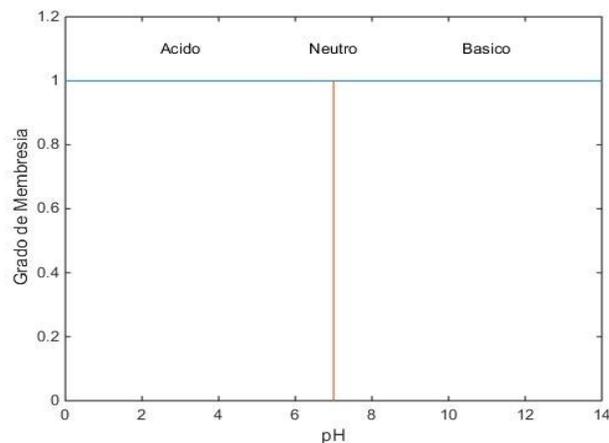


Figura 12. Funciones de pertenencia convencionales para la variable pH

El mismo ejemplo, en la lógica convencional, se presenta en la Figura 12, donde se obtiene el valor máximo de pertenencia justo en el valor de pH igual a 7.

II.3.1.4 FUZZIFICACIÓN.

Es el proceso mediante el cual se transforma un valor “Crisp” o real, en un valor difuso. Esto se hace sencillamente para observar que todas las cantidades que se consideraban determinantes, son realmente imprecisas o no determinísticas (Ross, 2004). Siguiendo el ejemplo anterior, al fuzzificar un pH igual a 6.9, se debe evaluar este valor sobre las funciones de pertenencia establecidas en cada conjunto de la variable.

Para el conjunto difuso ácido:

$$\mu_{\text{Acido}} = \frac{7 - 6.9}{7 - 6} \quad 6 \leq 6.9 < 7$$

$$\mu_{\text{Acido}} = 0.1$$

Para el conjunto difuso neutro:

$$\mu_{\text{Neutro}} = \frac{6.9 - 6}{7 - 6} \quad 6 \leq 6.9 \leq 7$$

$$\mu_{\text{Neutro}} = 0.9$$

Para el conjunto difuso básico:

$$\mu_{\text{Basico}} = 0 \quad 6 \leq 6.9 \leq 7$$

Una vez finalizada la etapa de fuzzificación, se pasa al bloque de inferencia difusa (BID), donde se reciben los resultados obtenidos del proceso de fuzzificación, y se calcula una respuesta difusa de salida (Piegat, 2001). Para llevar a cabo estos cálculos, el BID debe tener definidos los siguientes elementos:

- Reglas Base.
- Mecanismo de Inferencia.

II.3.1.5 REGLAS BASE.

Establecen la relación entre los conjuntos difusos que conforman a las variables de entrada y los conjuntos difusos que conforman a las variables de salida. Las reglas bases (RB) manejan la información según la relación que establezca el experto del proceso en estudio, y es la razón por la cual se debe “saber cómo” diseñarlas (Piegat, 2001).

Las RB utilizan expresiones del tipo si-entonces como se muestra en la ecuación (27), debido a que permiten inferir, derivar o concluir una solución conociendo un dato, antecedente o hipótesis (Ross, 2004).

$$\text{SI (entrada = valor) ENTONCES (salida = conclusión)} \quad (27)$$

Los sistemas donde hay múltiples variables de entrada y una sola variable de salida, se conocen como MISO (múltiples entradas y una sola salida, del inglés: Multiple Input-Single Output). En los sistemas MISO, las RB poseen en su estructura, operadores lógicos como AND – OR como se muestra en la ecuación (28), y es aquí donde ejecuta su función el mecanismo de inferencia.

$$\text{SI } (x_1 = A_{11}) \text{ AND } (x_2 = A_{21}) \text{ ENTONCES } (y = B_1) \quad (28)$$

El número de RB dependerá de la experiencia o del conocimiento que se tenga sobre el proceso en estudio. De forma alternativa se presenta en la Figura 13 uno de los métodos para determinar el número de RB necesarias para un modelo difuso (Piegat, 2001).

x_1	A_{11}	A_{12}	A_{13}	A_{14}
x_2	A_{21}	A_{22}	A_{23}	A_{24}
	B_1	B_2	B_3	B_4
	B_5	B_6	B_7	B_8
	B_9	B_{10}	B_{11}	B_{12}
	B_{13}	B_{14}	B_{15}	B_{16}

Figura 13. Estrategia para la determinación del número de RB (Piegat, 2001).

En la Figura 13, las variables A_{ij} representan a las FP de cada variable de entrada al sistema, y las variables B_k corresponden a las FP de la variable de salida.

Por lo tanto, para un sistema MISO, el número de RB será igual al producto del número de FP que se asignen a cada variable de entrada, cuando se utilizan solo operadores AND (ver ecuación (29)).

$$N^{\circ}_{RB} = N^{\circ}_{FP_{x1}} * N^{\circ}_{FP_{x2}} \quad (29)$$

Para el caso presentado en la Figura 13, el número de RB será entonces:

$$N_{RB}^{\circ} = 4 * 4 = 16$$

II.3.1.6 MECANISMO DE INFERENCIA.

Se encarga de seleccionar los valores necesarios para obtener la respuesta requerida (Zekai, 2010), según el tipo de operador lógico establecido en cada RB, ya sea el operador AND, OR, PROD o cualquier otro. Para realizar dicha selección, es necesario conocer los fundamentos de dichos operadores, que se describen a continuación.

- **INTERSECCIÓN (Operador AND).**

La intersección de dos conjuntos difusos, A y B, implica a la operación lógica AND, cuyo significado radica en seleccionar el valor más pequeño encontrado del grado de pertenencia entre dos conjuntos implicados en una RB. El operador AND es el más utilizado, debido a que le aporta una mayor suavidad a las superficies generadas por los modelos difusos (Piegat, 2001), junto con una mayor comprensión por la lógica humana. Este se define según la ecuación (30).

$$A \wedge B = \text{MIN}[\mu_A(x); \mu_B(x)] \quad (30)$$

- **UNIÓN (Operador OR).**

La unión de dos conjuntos difusos, A y B, implica a la operación lógica OR, cuyo significado radica en seleccionar el valor más grande encontrado del grado de pertenencia entre dos conjuntos difusos, para una misma variable (ver la ecuación (31)).

$$A \vee B = \text{MAX}[\mu_A(x); \mu_B(x)] \quad (31)$$

- **PRODUCTO (Operador PROD).**

Consiste en el producto de los grado de pertenencia de los conjuntos difusos implicados en una RB. Ver la ecuación (32).

$$\text{PROD}[\mu_A(x); \mu_B(x)] = \mu_A(x) * \mu_B(x) \quad (32)$$

II.3.1.7 DEFUZZIFICACIÓN.

Consiste en calcular el valor numérico real a partir de las conclusiones obtenidas en sistema de inferencia difuso para la(s) variable(s) de salida(s). Este procedimiento se lleva a cabo mediante el uso de métodos de defuzzificación, siendo el más utilizado el del Centroide (Ross, 2004).

- **MÉTODO DEL CENTROIDE.**

Este método asume que el valor “Crisp” o real que resulta de la conclusión del BID, es el centro de gravedad del área que lo contiene en la FP (Piegat, 2001). El centroide se calcula según la ecuación (33).

$$y^* = \frac{\int y * \mu_{salida} \cdot dy}{\int \mu_{salida} \cdot dy} \quad (33)$$

El método del centroide permite que todas las FP de salida que estén implicadas en el cálculo tomen parte en el proceso de defuzzificación. Además, garantiza una sensibilidad en el modelo difuso en el que otros métodos no lo hacen. Por otro lado, mientras más complejo sea el modelo difuso y más complejas sean las FP, habrá un alto costo en el cálculo.

Entendiendo la estructura de los modelos difusos, se explican continuación los dos tipos de modelos difusos más utilizados:

- Mamdani.
- Takagi-Sugeno.

II.3.2 MODELO DIFUSO TIPO MAMDANI.

Este modelo difuso fue desarrollado por el mismo autor (Mamdani, 1974), con la finalidad de establecer modelos que permitan generar curvas o superficies que

relacionen las variables de entrada con las variables de salida lo más aproximado a la realidad, utilizando FP, tanto para las entradas como para las salidas.

Los modelos tipo Mamdani son aquellos, cuyas RB definen un punto específico en el espacio, donde se representan las variables de entrada (Piegat, 2001) y su forma característica. Por ejemplo, para un sistema de dos entradas (x_1 y x_2) y una salida (y), se representa en la ecuación (34).

$$\text{SI } x_1 = A_1 \text{ AND } x_2 = B_1 \text{ ENTONCES } y = C_1 \quad (34)$$

Siendo A_1 , B_1 y C_1 conjuntos difusos definidos en las variables del modelo.

Estos modelos constan de un proceso de fuzzificación, inferencia difusa y defuzzificación por el método de centroide para obtener una respuesta, tal como se muestra en la Figura 14.

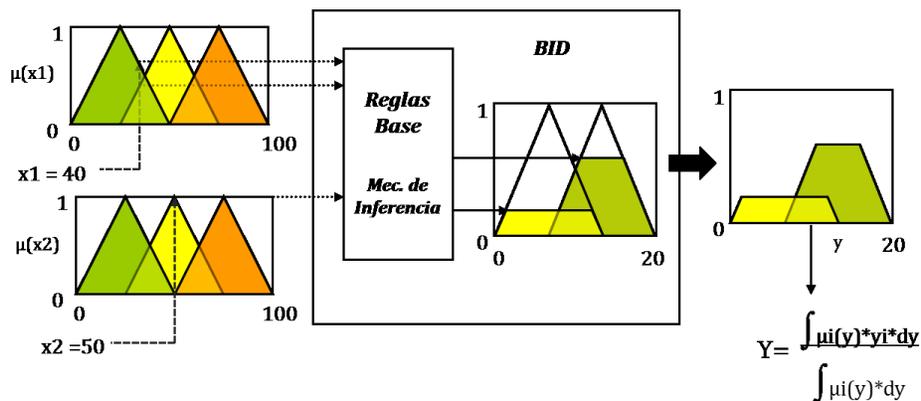


Figura 14. Arquitectura general de un modelo Difuso tipo Mamdani

II.3.3 MODELO DIFUSO TIPO TAKAGI-SUGENO

El modelo difuso de tipo Takagi-Sugeno (T-S) fue desarrollado en 1985 (Takagi & Sugeno, 1985). Se diferencia del de Mamdani, porque las FP de las variables de salida son ecuaciones algebraicas lineales o valores constantes.

Los modelos tipo T-S son aquellos cuyas RB definen una porción de espacio representado por una ecuación algebraica lineal. Estas porciones de espacio son definidas por quien está desarrollando el modelo.

La estructura de las RB, por ejemplo, la ecuación (35) muestra una regla base del sistema de inferencia difuso para un modelo del tipo T-S, basado en dos variables de entrada (x_1 y x_2) y una variable de salida (y):

$$\text{SI } x_1 = A1 \text{ AND } x_2 = B1 \text{ ENTONCES } y = b_0 + b_1 * x_1 + b_2 * x_2 \quad (35)$$

Siendo $A1$ y $B1$ conjuntos difusos definidos en las variables del modelo.

Estos modelos constan de un proceso de fuzzificación, inferencia difusa y una defuzzificación calculada en base al grado particular de activación de cada una de las conclusiones que arrojen las reglas base, tal como se muestra en la Figura 15.

Este cálculo se realiza utilizando las ecuaciones (36) o (37).

$$y = \frac{\sum \mu_{i\text{salida}} * y_i}{\sum \mu_{i\text{salida}}} \quad (36)$$

$$y = \frac{\sum \mu_{i\text{salida}} * (b_0 + b_1 * x_1 + b_2 * x_2)}{\sum \mu_{i\text{salida}}} \quad (37)$$

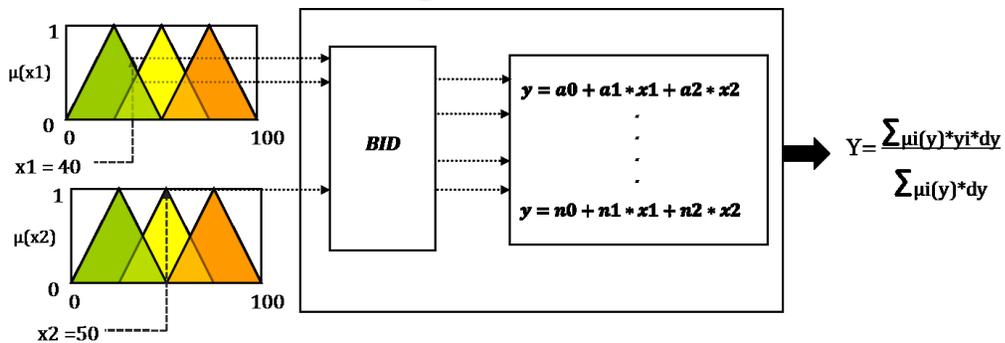


Figura 15. Arquitectura general de un modelo Difuso tipo T-S.

Cuando se tienen múltiples variables de entrada, resulta complejo desarrollar un modelo difuso de tipo T-S, en cuanto a la generación de las reglas base se refiere. Por esto, es necesario utilizar un método de simplificación del modelo como el método de clústeres, que permitan agrupar las observaciones y reconocer patrones, lo que automáticamente genera las reglas del modelo difuso.

Los métodos de clúster más utilizados para la generación de los modelos T-S son los siguientes:

- Grid Partitioning.
- Subtractive Cluster.
- Fuzzy C-Means Cluster.

En este trabajo de investigación, se utilizó el método de Grid Partitioning, el cual consiste en dividir la base de datos en sub-espacios rectangulares exactos mediante el uso de particiones en paralelo, lo que significa que cada uno de estos sub-espacios está definido por un hiperplano. El número de particiones depende del número de conjuntos difusos que se definan para cada variable de entrada. Por lo tanto, si se tienen 2 variables de entrada con 4 conjuntos difusos cada una, se obtienen 16 sub-espacios rectangulares exactos. Los hiperplanos encontrados para cada sub-espacio corresponden a las FP de la variable de salida del modelo. La desventaja de trabajar con este algoritmo de clústeres radica en que, a medida que se aumentan los conjuntos difusos de las variables de entrada, el número de reglas crece de manera desproporcionada, lo que aumenta el tiempo computacional en el momento de generar el modelo difuso (Suparta & Alhasa, 2016).

Para contrarrestar esta desventaja, se realiza una transformación de toda la base de datos, con el objetivo de disminuir las diferencias entre los órdenes de magnitud de las variables. Con esto, se le da un tratamiento equivalente a cada variable y, por lo tanto, disminuye el número de FP, lo que se traduce en un modelo más eficiente.

- **Transformación de los datos.**

Cuando los órdenes de magnitud entre las variables que conformaran un modelo difieren significativamente, se puede aplicar la transformación de datos o escalamiento de los mismos a un mismo rango, utilizando la ecuación (38).

$$X' = \frac{(Valor_{max\ deseado} - Valor_{min\ deseado})}{max(data) - min(data)} * (x - min(data)) + Valor_{min\ deseado} \quad (38)$$

En este trabajo, para realizar el método de Grid Partitioning, se utilizó el comando `genfis1` contenido en el toolbox `FuzzyLogicDesigner` (MatLab®, 2011).

II.3.4 SISTEMAS ADAPTATIVOS DE INFERENCIA NEURO-DIFUSA.

En este tipo de sistemas, las redes neuronales artificiales (RNA) están ligadas a la lógica difusa como método de optimización para determinar los mejores parámetros de los conjuntos difusos que conforman un modelo. El objetivo de este tipo de estrategia es tratar de imitar o emular el funcionamiento del cerebro humano, con la capacidad de aprender y procesar datos simultáneamente (Harris, 2000).

Las RNA pueden ser entrenadas en función de los datos de entrada y salida del sistema en estudio. La estructura básica de las RNA se muestra en la Figura 16.

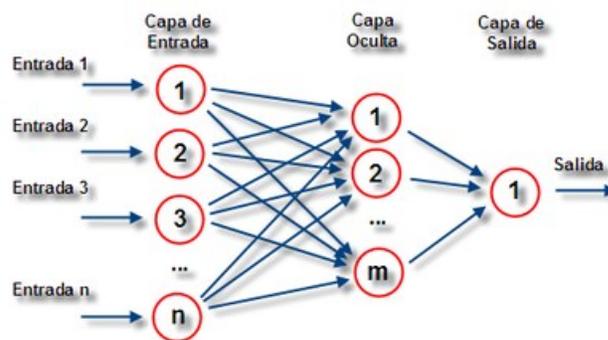


Figura 16. Estructura de una Red Neuronal Artificial.

Al mezclar una red neuronal con un modelo difuso del tipo Takagi-Sugeno, es posible encontrar los mejores parámetros de las funciones de membresía, que permitan realizar estimaciones de las variables de salida con mejores aproximaciones.

Los Sistemas Adaptativos de Inferencia Neuro-Difusa (del inglés ANFIS) fueron propuestos por Jyh-Shing y Jang (Jyh-Shing & Jang, 1991), quienes implementaron Redes Neuronales Artificiales con modelos difusos del tipo Takagi-Sugeno, con el fin de optimizar los parámetros de las funciones de pertenencia (FP).

Estos modelos buscan los mejores parámetros de las FP de las variables de salida, que disminuyan el error entre la respuesta del modelo y la respuesta real del sistema. Para esto, se aplica una red neuronal, que utilice alguno de los siguientes métodos de optimización:

- Backpropagation.

- Hybrid.

En este trabajo de investigación, se utilizó el método Hybrid. Para más detalles sobre este tema, referirse a (Suparta & Alhasa, 2016) y (T. Hagan, B. Demuth, Hudson Beale, & De Jesús, 2012).

II.4 ANTECEDENTES DE LA INVESTIGACIÓN

Los modelos estadísticos desarrollados para los procesos de potabilización no son precisos ni robustos, debido a la interacción altamente compleja de los parámetros involucrados. En 1985, Takagi y Sugeno (Takagi & Sugeno, 1985) propusieron un modelo difuso de dosificación de coagulantes para el proceso de potabilización del agua en Japón. Gracias al conocimiento de los operadores de la planta de potabilización, establecieron como parámetros importantes: temperatura, pH, alcalinidad y cloro residual. Tales autores utilizaron datos experimentales obtenidos entre junio y julio de 1985, para establecer los límites del modelo difuso. Seleccionaron como variables significativas para establecer las reglas difusas al pH, alcalinidad y temperatura. Los resultados obtenidos demuestran que la aplicación de la lógica difusa permite encontrar un modelo que se ajusta mejor a la realidad que los modelos empíricos convencionales.

Posteriormente, González (González, 2006) determinó los índices de calidad de las aguas superficiales basada en el uso de la lógica difusa. Utilizó dos modelos del tipo Takagi-Sugeno. Uno de ellos utiliza fórmulas de agregación para el cálculo de los Índices de calidad del agua. El otro modelo difuso fue desarrollado en base a variables fisicoquímicas y bacteriológicas del agua, como coliformes fecales, coliformes totales, pH, oxígeno disuelto, dureza, alcalinidad y turbidez.

Más tarde, Fernández, Carvajal y Colina (Fernández, Carvajal, & Colina, 2010) establecieron un modelo de Redes Neuronales para la estimación de los índices de calidad del agua, específicamente el índice para la materia orgánica (ICOMO), índice para la mineralización (ICOMI) y el índice de contaminación orgánica (OPI). Este modelo tuvo muy poca efectividad, y un alto gasto computacional. Es por eso que establecieron un modelo difuso tipo Mamdani, el cual se validó con un conjunto de datos hallados dentro de los rangos de la legislación colombiana para agua

potable y vertidos líquidos, demostrando las grandes potencialidades y ventajas que hacen de la lógica difusa una herramienta de base para el desarrollo de nuevos sistemas para la gestión del agua.

Más recientemente, Villareal y Caicedo (Villareal & Caicedo, 2013) presentaron cuatro modelos para la estimación de la dosificación de coagulante en el proceso de potabilización de agua, los cuales utilizan técnicas que incluyen redes neuronales artificiales, modelos tipo Takagi-Sugeno y Mamdani y estructuras de Sistemas de Inferencia Neuro Difusas Adaptativas (ANFIS). En dichos modelos, se incluye la experiencia de los operadores de planta, además de los datos experimentales, para establecer las reglas difusas. Los resultados obtenidos en los trabajos previos demuestran que la lógica difusa, las redes neuronales y las combinaciones de estas técnicas en los modelos neuro-difusos, resultan ser una excelente opción en los procesos de potabilización de agua.

II.5 APLICACIONES DE LÓGICA DIFUSA EN INGENIERÍA.

Además de los usos de esta técnica mencionados previamente, en 1974 Mamdani (Mamdani, 1974) desarrolló el primer controlador basado en lógica difusa de una planta dinámica.

En 2007 Gafanhao (Gafanhao, 2007) estableció un módulo de inferencia difusa que permite la predicción del punto de precipitación en los crudos venezolanos con discrepancias muy pequeñas con respecto a los resultados experimentales.

En ese mismo año, Jiménez (Jiménez, 2007) desarrolló un conjunto de modelos avanzados, (en los que se incluyen 45 de inferencia difusa del tipo Takagi-Sugeno) para la predicción a corto plazo de la producción eléctrica en parques eólicos, en función de la velocidad media horaria.

En 2008 Fallahpour, Fatehi, Araabi y Azizi (Fallahpour, Fatehi, Araabi, & Azizi, 2008) desarrollaron un controlador neuro-difuso para un proceso de producción de cemento, lo que permitió controlar variables como temperatura de operación, temperatura de pre-calentamiento, cantidad de oxígeno y contenido de CO₂.

En 2010 Tututi Ávila (Tututi Avila, 2010) estableció un sistema de control basado en lógica difusa, para un proceso de cristalización continuo no-isotérmico, que fue

capaz de predecir el tamaño medio del cristal y la temperatura del cristizador en estado estacionario, conociendo los modelos cinéticos de cristalización.

En 2013, Castaño, Hernández y Gallo (Castaño, Hernández, & Gallo, 2013) implementaron un controlador PI basado en lógica difusa, para el control y monitoreo de la temperatura de un horno de curado de prendas índigo (el índigo es un compuesto insoluble en agua que se usa para teñir tejidos de color azul).

www.bdigital.ula.ve

III. METODOLOGÍA

En esta sección, se muestra en detalle metodología utilizada para la obtención de los modelos lineal, no-lineal, Mamdani y ANFIS.

III.1 RECOPIACIÓN DE LOS DATOS.

En función de los objetivos planteados, fue necesario el modelado de cada una de las dosificaciones realizadas, tanto en una planta de potabilización de agua, como a nivel de laboratorio.

Para la aplicación de los modelos a nivel industrial, se recopiló la información de la planta de potabilización de agua ubicada en el Vigía, estado Mérida, donde se disponía de datos de pH, turbidez, color, alcalinidad, cloro residual y dosificación de sulfato de aluminio, policloruro de aluminio y cloro gaseoso. Para el procesamiento de los datos, se contó con una base de datos de 390 observaciones para la dosificación de sulfato de aluminio, 478 observaciones para la dosificación de policloruro de aluminio y 120 observaciones para la dosificación de cloro gaseoso en la planta de potabilización, correspondientes a los años 2005 al 2012 y algunos meses de 2016.

En el caso del laboratorio, se utilizó la metodología empleada por Mayorga (Mayorga, 2009), para simular la etapa de coagulación-floculación, en donde se tomaron como variables principales: pH, turbidez, temperatura y dosificación de sulfato de aluminio. Se tomaron muestras instantáneas en la fuente de agua en las inmediaciones de la Facultad de Ingeniería de la Universidad de Los Andes. Para la simular la etapa de coagulación-floculación, se agregó arcilla al agua, para aumentar su turbidez (ya que la fuente no presentaba grandes variaciones de este parámetro). Se aplicó la prueba de jarras, utilizando como coagulante al sulfato de aluminio, y se tomaron en total 35 mediciones.

III.2 PROCESAMIENTO DE LOS DATOS.

Los datos obtenidos se procesaron para eliminar cualquier punto vacío o errores de transcripción. Realizado esto, se dividió la base de datos en dos partes.

- Datos para modelar.
- Datos para validar.

El siguiente procedimiento fue el utilizado para la obtención de los modelos.

III.3 ESTRATEGIA DEL MODELADO ESTADÍSTICO.

III.3.1 MODELO LINEAL.

A continuación, se presenta el procedimiento utilizado para la construcción de los modelos lineales.

- Variables de Entrada y Salida.
- Ajuste por Mínimos Cuadrados.
- Análisis de Normalidad de los Datos.
- Análisis por distancia de Leverage y Cook.
- Ajuste por Mínimos Cuadrados excluyendo valores atípicos.
- Validación.

Se explicará cada una de las fases del procedimiento:

- Variables de Entrada y Salida.

En esta etapa se tomaron variables de entrada, que tuvieran relación con la variable de salida, ya sea por criterios propios o sugerencias de los expertos en el área.

- Ajuste por Mínimos Cuadrados.

Una vez seleccionadas las variables, se procedió a generar la matriz "X" y "Y" de las variables de entrada y salida respectivamente. Utilizando la ecuación (10) se realizó el ajuste por mínimos cuadrados, dando lugar a un modelo inicial. Con la disponibilidad del modelo inicial, se procedió a calcular los residuos para pasar a la siguiente etapa.

- Análisis de Normalidad de los Datos.

Se comprobó la distribución normal de los residuos graficando un histograma y la recta probabilidad normal. Además, se comprobó la presencia de posibles

observaciones atípicas, cuando los residuos de algunas observaciones se alejaban significativamente de la recta de distribución normal.

- Análisis por distancia de leverage y cook.

Se utilizaron los criterios de las distancias de leverage y Cook, las cuales fueron calculadas con el comando fitlm (MatLab®, 2011), y los resultados se graficaron en función del número de la observación, para detectar con mayor facilidad a las observaciones atípicas. Al finalizar este análisis, se seleccionaron los puntos atípicos, que fueron excluidos del modelo.

- Ajuste por Mínimos Cuadrados, excluyendo valores atípicos.

Se repitió el ajuste por mínimos cuadrados, excluyendo a los valores atípicos seleccionados en la fase anterior. Con esto, se obtuvo un modelo lineal final, al cual se le determinó el coeficiente de correlación de Pearson (R^2), para evaluar su desempeño con los datos del modelaje.

- Validación.

En esta etapa, se probó el modelo lineal final con los datos de validación, los cuales no han sido incluidos en el modelaje, y se evaluó el desempeño en función de los coeficientes de correlación de Pearson.

III.3.2 MODELO NO-LINEAL

Tiene un procedimiento similar al explicado para los modelos lineales, con algunas modificaciones que están establecidas en Ingels (M. Ingels, 1980). Consiste en los siguientes pasos:

- Representación gráfica de las variables de entrada con la variable de salida.
- Selección de las primeras variables.
- Ajuste por Mínimos Cuadrados.
- Adición de nuevas variables.
- Ajuste por Mínimos Cuadrados.
- Análisis de Normalidad de los Datos.
- Análisis por distancia de leverage y Cook.

- Ajuste por Mínimos Cuadrados, excluyendo valores atípicos.
- Validación.

Se explicará cada una de las fases del procedimiento, obviando los términos descritos en la sección III.3.1.

- Representación gráfica de las variables de entrada con la variable de salida.

En esta etapa, se graficaron todas las variables de entrada en función de la variable de salida, con la finalidad de buscar patrones de comportamiento similares a las funciones elementales mostradas en la Figura 5.

- Selección de las primeras variables.

Al determinar los patrones nombrados anteriormente, se seleccionaron dos variables de entrada cualesquiera, para comenzar la construcción del modelo no-lineal. Al realizar esta selección, se aplicó a cada variable la transformación de los datos de modelado según la función elemental a la cual se asemejaba su comportamiento.

- Adición de una nueva variable.

Para adicionar una nueva variable, se graficó la misma en función de los residuos del modelo anterior. En esta gráfica, se buscaron patrones de comportamiento, y se aplicó una transformación a los datos de la variable a incluir según la función elemental a la cual se asemejaba su comportamiento. Esta variable era incluida en la matriz “X” y se realizó nuevamente el ajuste por mínimos cuadrados. Este procedimiento se repitió hasta agregar todas las variables de entrada seleccionadas, y cuando se observó un comportamiento aleatorio de los residuales del modelo en cuestión.

III.4 ESTRATEGIA DEL MODELADO DIFUSO.

III.4.1 MODELO TIPO MAMDANI.

A continuación, se presenta el procedimiento utilizado para la construcción del modelo difuso tipo Mamdani.

- Variables de entrada y salida.
- Extracción del conocimiento y generación de reglas base.
- Modelo Mamdani inicial.
- Prueba del modelo inicial.
- Optimización del modelo Inicial.
- Validación.

Se explicarán cada una de las fases del procedimiento para la construcción de los modelos difusos de tipo mamdani, obviando los términos descritos en la sección III.3.1 y III.3.2.

- Extracción del conocimiento y generación de Reglas Base.

Para el desarrollo de los modelos difusos de tipo Mamdani, se realizó el proceso de extracción del conocimiento, en donde se tradujo el pensamiento del operador de la planta. La primera fase del proceso de extracción fue la creación de los conjuntos difusos que conformaron a las variables de entrada y salida, en donde el operador seccionó cada variable entre bajo, medio, alto y muy alto.

Luego, se utilizó un generador de reglas para que el operador estableciera las relaciones posibles entre los conjuntos de las variables de entrada y salida, que él mismo había creado en la fase anterior.

- Modelo Mamdani inicial.

Con la información contenida en el generador de reglas, se desarrollaron las RB con estructura SI-ENTONCES, que permitieron completar los modelos Mamdani iniciales.

- Prueba del modelo inicial.

Se probó el modelo inicial con los datos de modelación utilizada para los casos anteriores, y se utilizaron los coeficientes de correlación de Pearson para evaluar el desempeño del modelo Mamdani inicial.

- Optimización del modelo Inicial.

Al no proporcionar un buen desempeño en el momento de la prueba, se procedió a optimizar las funciones de pertenencia de las variables de entrada y salida por

ensayo y error, ya que no existen métodos de optimización para modelos del tipo Mamdani.

III.4.2 MODELO TIPO ANFIS.

A continuación, se presenta el procedimiento utilizado para la construcción del modelo difuso tipo ANFIS.

- Variables de entrada y salida.
- Transformación de los datos.
- Descubrimiento del conocimiento y entrenamiento del modelo.
- Validación del modelo.

Se explicarán cada una de las fases del procedimiento para la construcción de los modelos difusos tipo ANFIS, obviando los términos explicados detalladamente en las secciones III.3.1, III.3.2 y III.4.1.

- Transformación de los datos.

Se realizó la transformación de los datos de modelación, utilizando la ecuación (38). El valor máximo y mínimo se fijó en 1 y -1, respectivamente, para cada una de las variables.

$$X' = \frac{(Valor_{max\ deseado} - Valor_{min\ deseado})}{max(data) - min(data)} * (x - min(data)) + Valor_{min\ deseado}$$

- Descubrimiento del conocimiento y entrenamiento del modelo.

Una vez transformados los datos, se generó el modelo difuso tipo T-S con la opción de "Generate FIS" del toolbox Neuro-Fuzzy Designer (MatLab®, 2011) como se ve en la Figura 17, en donde se suministraron los parámetros de entrada: datos de modelación, número de conjuntos difusos de cada variable, tipo de FP para las variables de entrada y de salida y método de clúster a utilizar. Esta etapa permitió la generación las RB del modelo T-S. Luego, se procedió a entrenar al modelo con la opción "Train Now", en donde se suministraron los parámetros de entrada: error o tolerancia y sesiones de entrenamiento (epochs).

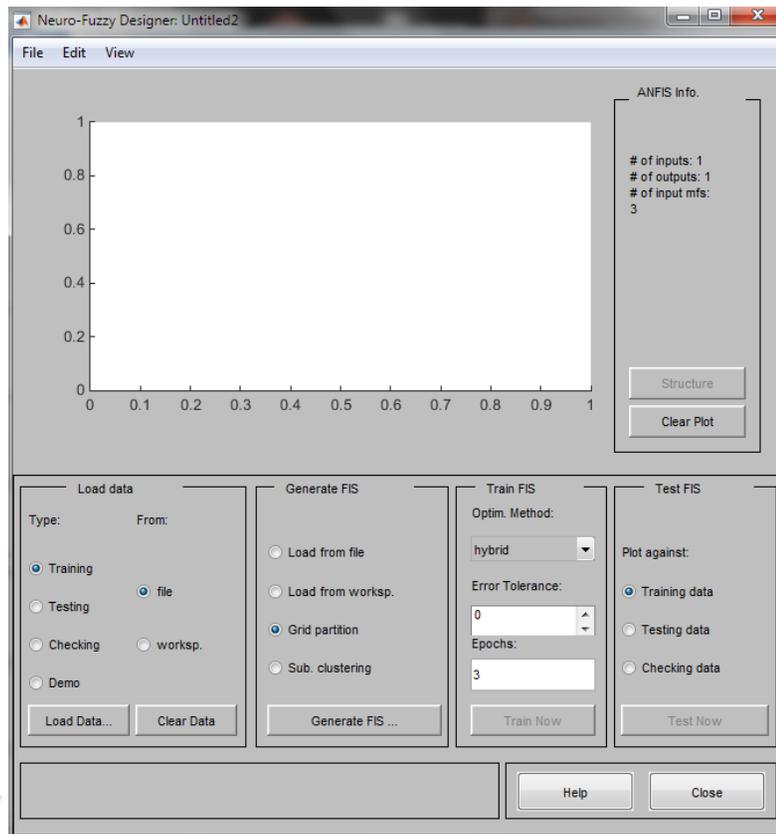


Figura 17. Neuro-Fuzzy Designer Toolbox (MatLab®, 2011).

Cuando finalizó la etapa de entrenamiento, se generaron las estructuras ANFIS como se ve en la Figura 18.

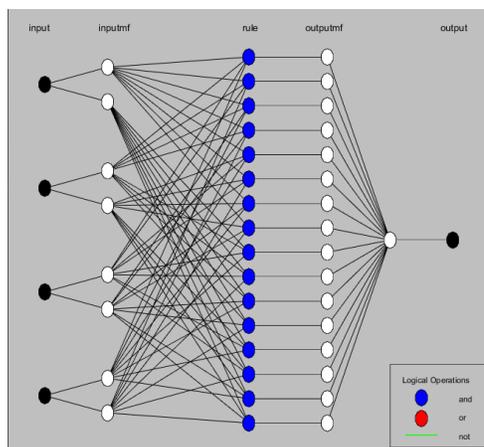


Figura 18. Neuro-Fuzzy Designer Toolbox “Red Neuronal” (MatLab®, 2011).

IV. RESULTADOS Y DISCUSIÓN

IV.1 DOSIFICACIÓN DE COAGULANTE EN LA PLANTA.

IV.1.1 SULFATO DE ALUMINIO.

IV.1.1.1 MODELO LINEAL.

Para el procesamiento, como se mencionó previamente, se disponía de una base de datos de 390 observaciones para la dosificación de sulfato de aluminio de la planta de potabilización de agua. La base de datos se dividió aleatoriamente, tal como se muestra en la Tabla 3.

Tabla 3. División de los datos para el modelo de dosificación de Sulfato de Aluminio

Datos	Numero de Observaciones
Modelar	345
Validar	45

Se siguió el procedimiento presentado en la sección III.3.1. Se tomaron como variables de entrada: pH, Turbidez (UNT), Color (UCV) y Alcalinidad (ppm). Como variable de salida, la Dosis de Sulfato de Aluminio (ppm). Se procedió a formar la matriz "X" y "Y", y utilizando la ecuación (10), se realizó el ajuste por mínimos cuadrados, dando como resultado la ecuación (39).

$$Y = \begin{bmatrix} 34.68 \\ 16.66 \\ \cdot \\ \cdot \\ \cdot \\ 38.4 \end{bmatrix}, X = \begin{bmatrix} 1 & 7.98 & 120 & 300 & 94 \\ 1 & 7.88 & 18 & 40 & 96 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 7.97 & 688 & 900 & 90 \end{bmatrix}$$

$$\beta = (X^T * X)^{-1} * X^T * Y$$

$$\beta = \begin{pmatrix} [1 & 7.98 & 120 & 300 & 94]^T \\ [1 & 7.88 & 18 & 40 & 96]^T \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ [1 & 7.97 & 688 & 900 & 90]^T \end{pmatrix}^{-1} * \begin{pmatrix} [1 & 7.98 & 120 & 300 & 94]^T \\ [1 & 7.88 & 18 & 40 & 96]^T \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ [1 & 7.97 & 688 & 900 & 90]^T \end{pmatrix} * \begin{pmatrix} [34.68] \\ [16.66] \\ \vdots \\ \vdots \\ [38.4] \end{pmatrix}$$

$$\beta = \begin{bmatrix} 96.092 \\ -5.4517 \\ -0.04084 \\ 0.043989 \\ -0.32393 \end{bmatrix}$$

$$\widehat{Sulf} = 96.092 - 5.4517 * pH - 0.04084 * Turbidez + 0.043989 * Color - 0.32393 * Alc \quad (39)$$

Se realizó el análisis de los residuos, para comprobar la distribución normal de los mismos, y la presencia de posibles observaciones atípicas. Los resultados se presentan en la Figura 19, en donde se observa que la gran mayoría de residuos siguen un comportamiento normalizado, pero se nota que las observaciones cuyos valores residuales están cercanos a -40 y 40 unidades se desvían significativamente de una distribución normal, por lo que se consideraron observaciones atípicas.

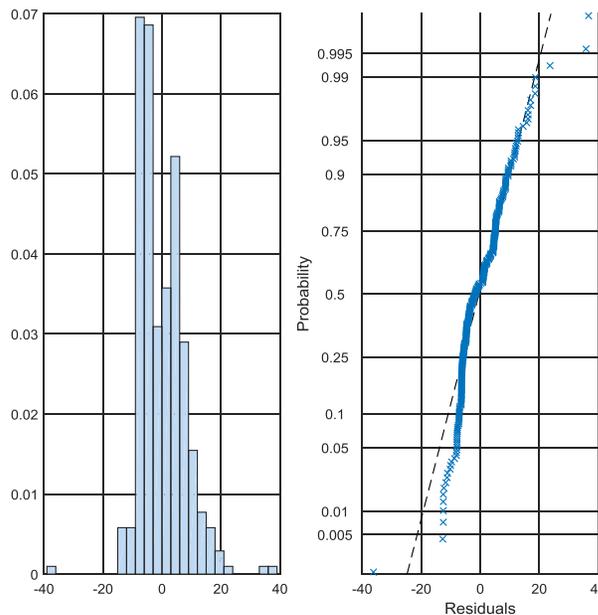


Figura 19. Distribución normal para los residuos del modelo lineal de Sulfato de Aluminio.

Posteriormente, se realizó el análisis mediante las distancias de leverage y Cook y los resultados se presentan en la Figura 20. Se observa que solo cuatro puntos se alejaron significativamente de los demás con respecto a las distancias de Cook, por lo que se consideran observaciones atípicas.

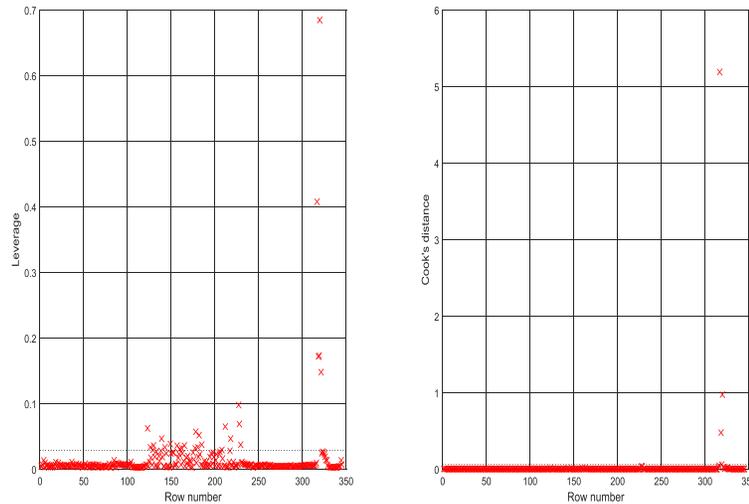


Figura 20. Análisis de leverage y Cook del modelo lineal de Sulfato de Aluminio.

El modelo que se obtuvo al excluir las observaciones atípicas se presenta en la ecuación (40). En este punto se da por finalizada la construcción del modelo lineal para la dosis de sulfato de aluminio en la planta, y se grafica el desempeño del modelo lineal con respecto a los datos de validación (ver Figura 21). Adicionalmente se calculan los coeficientes de correlación de Pearson, los cuales muestran cuantitativamente el desempeño que realiza el modelo al predecir las dosificaciones.

$$\widehat{\text{Sulf}} = 82.589 - 3.8059 * \text{pH} - 0.068617 * \text{Turbidez} + 0.064448 * \text{Color} - 0.33304 * \text{Alc} \quad (40)$$

$$R^2 = 0.548$$

$$R^2_{\text{ajustado}} = 0.543$$

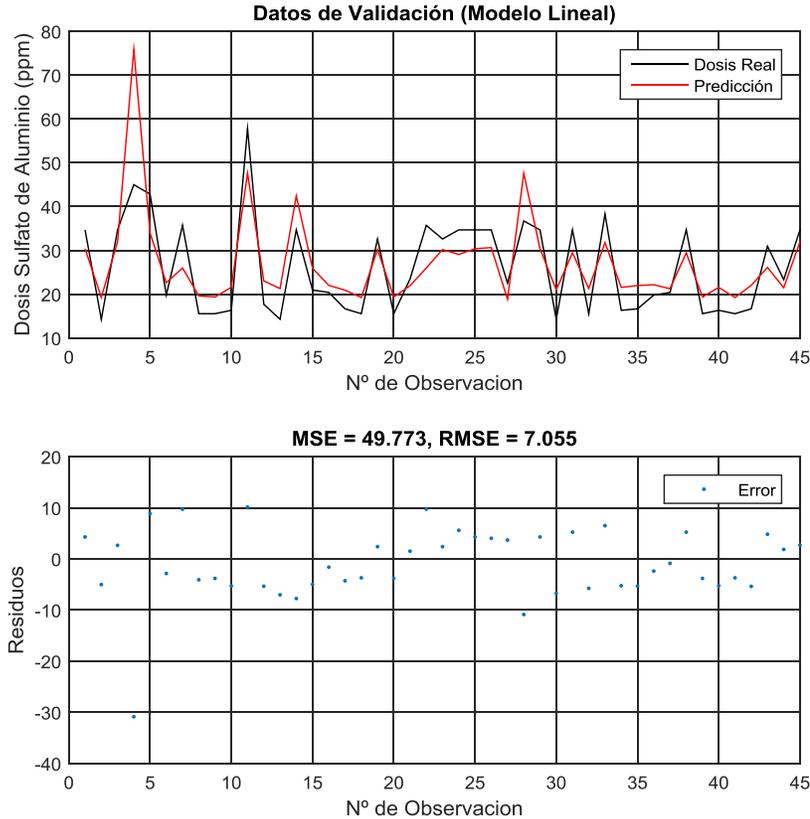


Figura 21. Desempeño del modelo lineal de Sulfato de Aluminio sobre los datos de validación.

Se puede observar que el proceso de dosificación de sulfato de aluminio no es satisfactoriamente descrito mediante modelos lineales, probablemente debido a interacciones no consideradas entre los parámetros incluidos en el modelo. Al observar la Figura 21, se puede comparar la línea de predicción con la línea de dosificación real, y es aquí donde el coeficiente de correlación de Pearson indica el nivel de desempeño para los datos de validación con 54.3% de precisión.

IV.1.1.2 MODELO NO-LINEAL EN PLANTA “SULFATO DE ALUMINIO”.

Se siguió el procedimiento presentado en la sección III.3.2 dando como resultado lo mostrado en la Figura 22, donde aparece cada variable de entrada en función de la dosis de sulfato de aluminio.

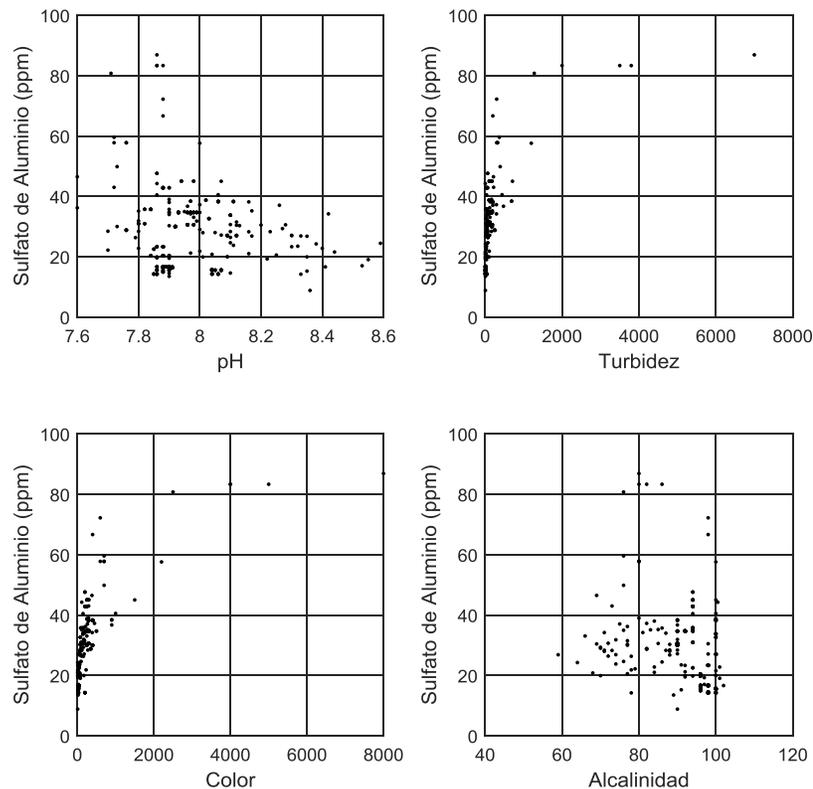


Figura 22. Representación de cada variable de entrada vs dosis de Sulfato de Aluminio.

En este caso, se busca un patrón que tenga relación con alguna de las funciones elementales mostradas en la Figura 5 (p. 17). El pH y la Alcalinidad presentan una tendencia cuadrática, mientras que la turbidez y el color muestran una tendencia logarítmica. Se aplica entonces el modelo no-lineal con las variables pH y turbidez (ecuaciones (41) y (42)) y la transformación correspondiente a los datos, según el patrón observado en la Figura 22.

$$\text{pH} = (\text{pH} - 8.1)^2 \quad (41)$$

$$\text{Turbidez} = \ln(\text{Turbidez}) \quad (42)$$

Con esta transformación de los datos, se crea la matriz "X", se aplica la ecuación (10) y se obtiene el modelo presentado en la ecuación (43).

$$\widehat{Sulf} = -2.2591 + 30.194 * (pH - 8.1)^2 + 7.6862 * \ln(Turb) \quad (43)$$

Para agregar el término correspondiente a la variable Color, se grafica la misma en función de los residuos de la ecuación (43), como se ve en la Figura 23.

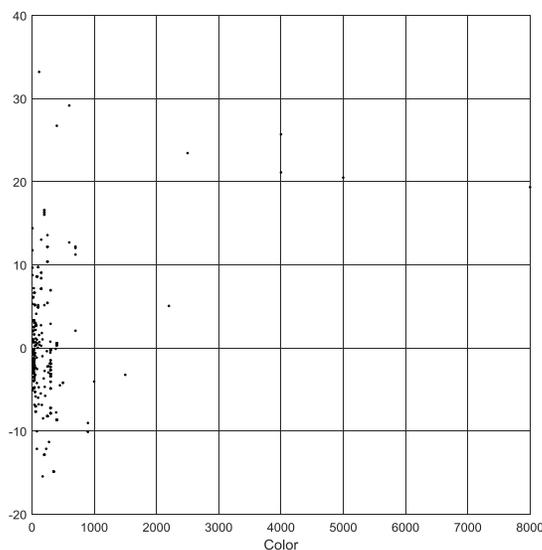


Figura 23. Representación del Color vs Residuales del modelo no-lineal (43)

En la Figura 23 se observa un patrón similar al de la función logarítmica, por lo tanto, se realiza la transformación correspondiente a la variable Color (ecuación (44)), se agrega a la matriz "X" y se realiza nuevamente la regresión por mínimos cuadrados, obteniendo la ecuación (45).

$$\text{Color} = \ln(\text{Color}) \quad (44)$$

$$\widehat{Sulf} = -15.005 + 29.194 * (pH - 8.1)^2 - 2.1342 * \ln(Turb) + 10.909 * \ln(\text{Color}) \quad (45)$$

Para agregar el término correspondiente a la variable Alcalinidad, ésta se grafica en función de los residuos de la ecuación (45), como se ve en la Figura 24. Aquí se observa un patrón cuadrático, por lo tanto, se realiza la transformación correspondiente a la variable Alcalinidad (ecuación (46)), se agrega a la matriz "X", y

se realiza nuevamente la regresión por mínimos cuadrado, obteniendo la ecuación (47).

$$\text{Alc} = (\text{Alc} - 85)^2 \quad (46)$$

$$\widehat{\text{Sulf}} = -18.56 + 36.67 * (\text{pH} - 8.1)^2 - 1.62 * \ln(\text{Turb}) + 10.8 * \ln(\text{Color}) + 0.015 * (\text{Alc} - 85)^2 \quad (47)$$

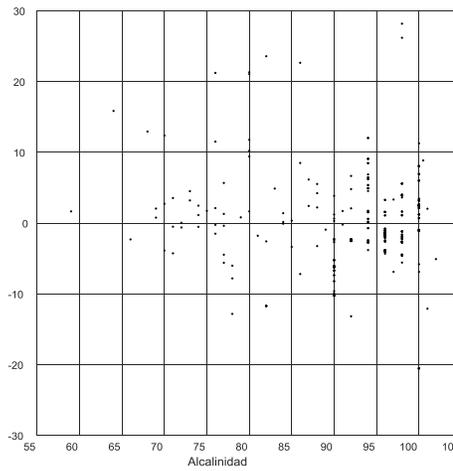


Figura 24. Representación de la Alcalinidad vs Residuales del modelo no-lineal (45)

Por último, se grafican todas las variables transformadas con los residuos de la ecuación (47) tal como se muestra en la Figura 25, con la finalidad de encontrar otros patrones y agregar nuevos términos.

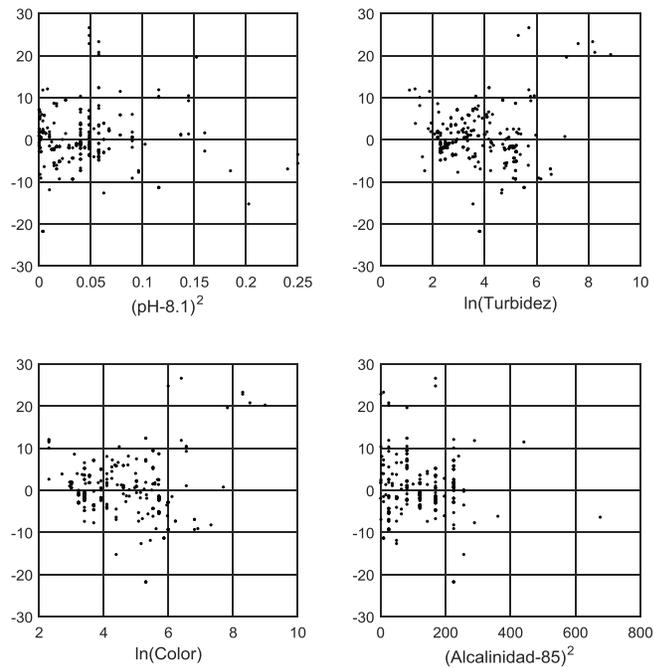


Figura 25. Representación de cada variable de entrada con los residuales del modelo (ecuación (47)).

En la Figura 25 no se observan patrones similares a funciones elementales, por lo tanto se considera terminada esta fase de construcción del modelo no-lineal, el cual queda descrito por la ecuación (47). Se aplica el análisis de normalidad de los datos, distancias de leverage y Cook para determinar las observaciones atípicas (Figura 26). Se nota que las observaciones con las cuales el modelo genera residuos menores a -20 y mayores a 20 unidades se alejan significativamente de una distribución normal, por lo tanto, se consideran puntos atípicos.

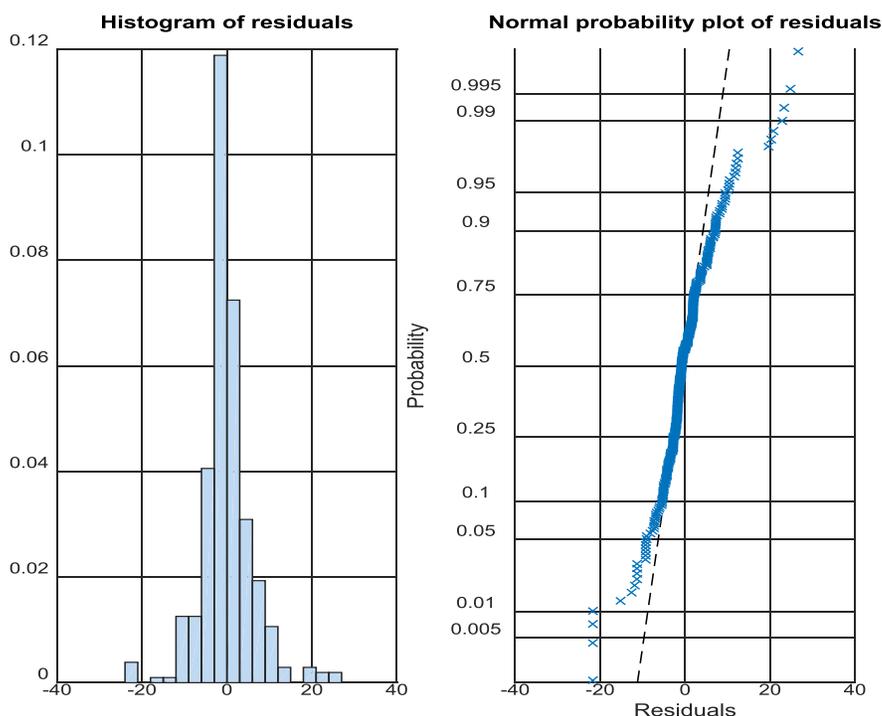


Figura 26. Distribución normal para los residuos del modelo no-lineal de Sulfato de Aluminio.

En la Figura 27 se observa una gran cantidad de puntos cuya distancia de Cook es bastante elevada con respecto a la media, por lo tanto se consideran puntos atípicos. El modelo no-lineal que se obtuvo al excluir las observaciones atípicas se presenta en la ecuación (48). En este punto, se da por finalizado el modelo no-lineal para la dosis de sulfato de aluminio en planta. En la Figura 28 se grafica el desempeño del modelo no-lineal con respecto a los datos de validación y se calculan los coeficientes de correlación de Pearson.

$$\widehat{\text{Sulf}} = -16.88 + 26.52 * (\text{pH} - 8.1)^2 - 4.95 * \ln(\text{Turb}) + 13.22 * \ln(\text{Color}) + 0.0143 * (\text{Alc} - 85)^2 \quad (48)$$

$$R^2 = 0.879$$

$$R^2_{\text{ajustado}} = 0.877$$

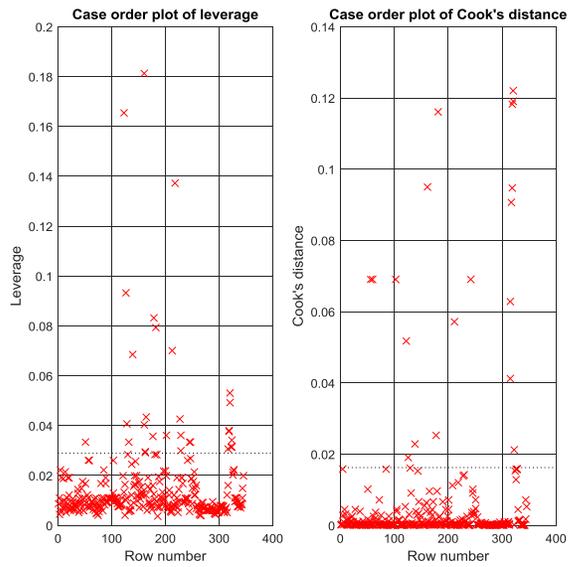


Figura 27. Análisis de Leverage y Cook del modelo no-lineal de Sulfato de Aluminio.

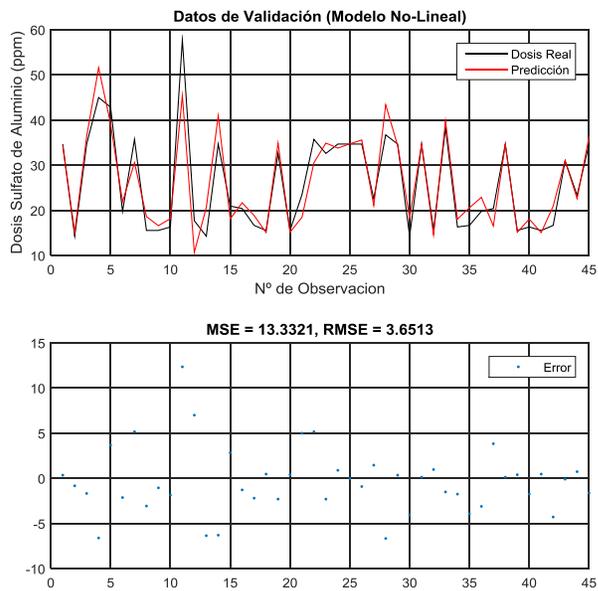


Figura 28. Desempeño del modelo no-lineal de Sulfato de Aluminio sobre los datos de validación.

Se puede comprobar que el proceso de dosificación de sulfato de aluminio es altamente no-lineal y que puede ser descrito con muy buen desempeño mediante la ecuación (48).

El modelo no-lineal tiene una mejor aproximación (87.7%) que el modelo lineal, pero a cambio, la ecuación es mucho más compleja.

IV.1.1.3 MODELO MAMDANI.

Se sigue el procedimiento presentado en la sección III.4.1. En este caso, el operador de la planta, define como variables de entrada la Turbidez y el Color, ya que es la forma más rápida para estimar un valor aproximado de la dosis de sulfato de aluminio. La variable de salida del modelo será la dosis de sulfato de aluminio. Se establecen los conjuntos difusos de cada una de las variables seleccionadas, tal como se ve en las Tabla 4, Tabla 5 y Tabla 6.

Tabla 4. Conjuntos difusos para la variable "Turbidez".

Conjuntos Difusos	Rango (UNT)
Baja	0 - 8
Media	6 - 50
Alta	40 - 300
Muy Alta	250 - 7000

Tabla 5. Conjuntos difusos para la variable "Color".

Variable Lingüística	Rango (UCV)
Bajo	0 - 20
Medio	12 - 100
Alto	80 - 600
Muy Alto	350 - 8000

Tabla 6. Conjuntos difusos para la variable “Dosis de sulfato de aluminio”.

Variable Lingüística	Rango (ppm)
Baja	0 - 15
Media	13 - 24
Alta	22 - 40
Muy Alta	35 -90

Conocidos los conjuntos de cada una de las variables, se toman las FP de tipo triangular, como se ve en la Figura 29.

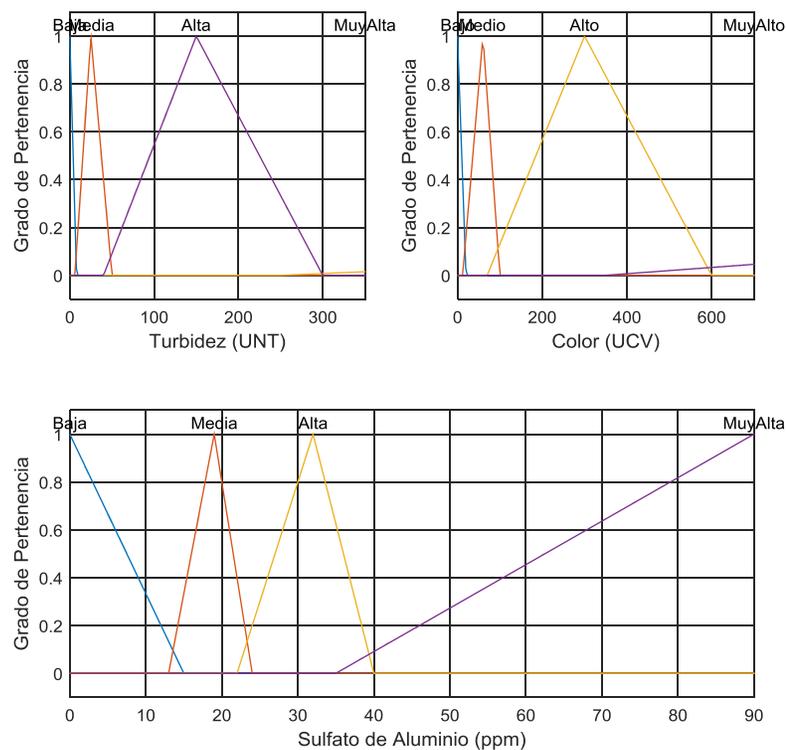


Figura 29. Funciones de pertenencia para el modelo Mamdani de dosificación de Sulfato de Aluminio.

Establecidas las FP de todas las variables, se procede a generar las reglas base del modelo inicial. En varias sesiones de trabajo, con la ayuda del experto de la planta de potabilización de agua de El Vigía, se desarrollan las reglas difusas, como se muestra en la Tabla 7.

Tabla 7. Reglas difusas generadas para el modelo de dosis de Sulfato de Aluminio.

GENERADOR DE REGLAS (SULFATO DE ALUMINIO)				
SULFATO DE ALUMINIO	COLOR			
TURBIDEZ	Bajo	Medio	Alto	Muy Alto
Baja	Baja	Media	Media	Alta
Media	Media	Media	Media	Alta
Alta	NA	NA	Alta	Alta
Muy Alta	NA	NA	Muy Alta	Muy Alta

En las casillas donde aparecen las siglas “NA (No Aplica)”, el experto argumentó que no era posible que se dieran dichas combinaciones o casos.

Al interpretar el generador de reglas, se obtiene como resultado las RB del modelo, que aparecen en la Tabla 8.

Tabla 8. Bloque de Reglas base para el modelo Mamdani de dosificación de Sulfato de Aluminio.

Nº	Estructura SI-ENTONCES
1	<i>SI Turbidez es Baja Y Color es Bajo ENTONCES Sulfato es Bajo</i>
2	<i>SI Turbidez es Baja Y Color es Medio ENTONCES Sulfato es Medio</i>
3	<i>SI Turbidez es Baja Y Color es Alto ENTONCES Sulfato es Medio</i>
4	<i>SI Turbidez es Baja Y Color es Muy Alto ENTONCES Sulfato es Alto</i>
5	<i>SI Turbidez es Media Y Color es Bajo ENTONCES Sulfato es Bajo</i>
6	<i>SI Turbidez es Media Y Color es Medio ENTONCES Sulfato es Medio</i>
7	<i>SI Turbidez es Media Y Color es Alto ENTONCES Sulfato es Medio</i>
8	<i>SI Turbidez es Media Y Color es Muy Alto ENTONCES Sulfato es Alto</i>
9	<i>SI Turbidez es Alta Y Color es Alto ENTONCES Sulfato es Alto</i>
10	<i>SI Turbidez es Alta Y Color es Muy Alto ENTONCES Sulfato es Alto</i>
11	<i>SI Turbidez es Muy Alta Y Color es Alto ENTONCES Sulfato es Muy Alto</i>
12	<i>SI Turbidez es Muy Alta Y Color es Muy Alto ENTONCES Sulfato es Muy Alto</i>

Con las reglas base y las variables de entrada y salida establecidas, se procede a construir el modelo difuso Mamdani para la dosis de sulfato de aluminio, con la ayuda del FuzzyLogicDesigner toolbox (MatLab®, 2011). Como este modelo no se construyó a partir de una base de datos, tuvo que ser probado con los datos de modelación, y se utilizó el coeficiente de correlación de Pearson para evaluar su desempeño, el cual se registra en la Tabla 9.

Tabla 9. Coeficientes de correlación de Pearson del modelo Mamdani inicial con datos de modelación.

R^2	$R^2_{ajustado}$
0.6	0.598

Posteriormente, se optimizan los parámetros de las FP del modelo Mamdani, tal como se muestra en la Figura 30.

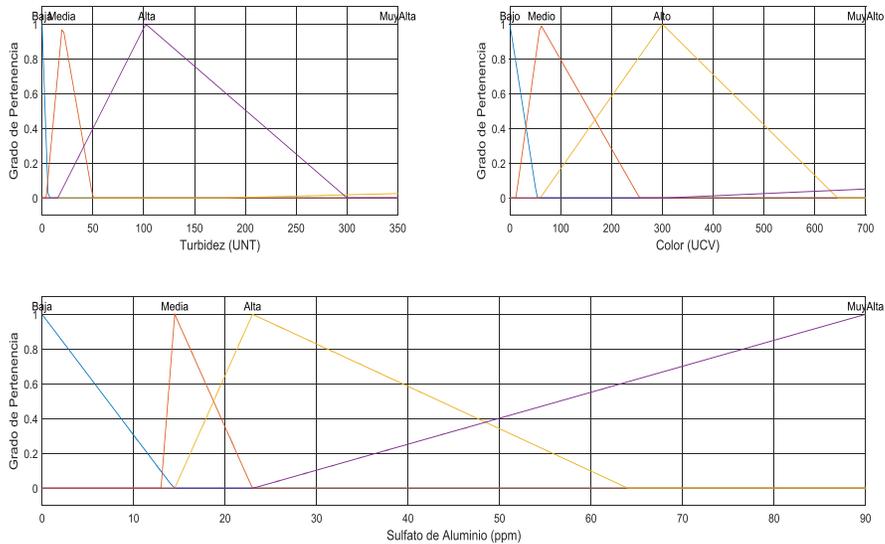


Figura 30. Funciones de pertenencia optimizadas para el modelo Mamdani de dosificación de Sulfato de Aluminio.

La optimización de las FP de las variables de entrada y salida se realiza mediante ensayo y error, hasta encontrar el mejor ajuste posible de los datos de modelación. Al obtener el máximo ajuste posible con los datos de modelación se procede a determinar el desempeño que tiene el modelo Mamdani con los datos de validación, y los resultados se representan en la Figura 31.

Al optimizar las FP del modelo se consigue una precisión del 81.4% con respecto a los datos de validación. Los resultados obtenidos con respecto a los datos de validación se consideran buenos, debido a que el modelo Mamdani solo utiliza dos variables de entrada. Adicionalmente puede decirse que, sin aplicar los análisis de residuos a los datos del modelo, se obtiene un comportamiento aleatorio, y que además, la línea de predicción se aproxima bastante a la dosis real de sulfato de aluminio.

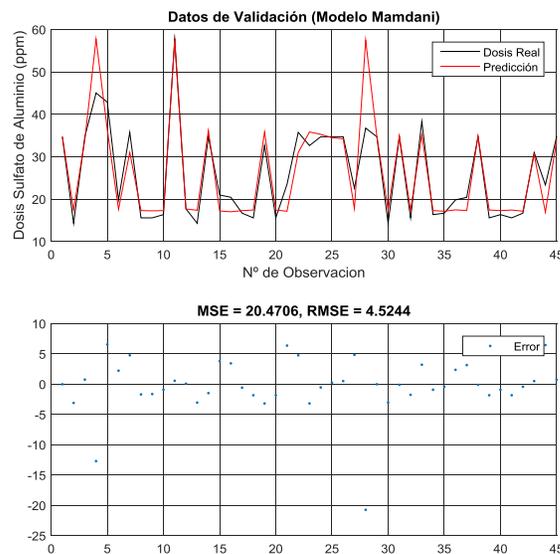


Figura 31. Desempeño del modelo Mamdani de Sulfato de Aluminio sobre los datos de validación.

En la Tabla 10 se registran los coeficientes de correlación de Pearson que se obtuvieron con los datos de modelación y los de validación.

Tabla 10. Nuevos coeficientes de correlación de Pearson modelo Mamdani.

Datos	R^2	$R^2_{ajustado}$
Validación	0.814	0.813

Con solo las variables Turbidez y Color se aproxima al desempeño realizado por el modelo no-lineal, el cual tiene una ecuación compleja. Se debe acotar que para los modelos difusos no se desprecia ninguna observación ya que no se consideran atípicas sino imprecisas.

IV.1.1.4 MODELO ANFIS EN PLANTA.

Se sigue el procedimiento presentado en la sección III.4.2. Se realiza la transformación de los datos de modelación utilizando la ecuación (38). El valor máximo y mínimo se fijó en 1 y -1 respectivamente para cada una de las variables.

$$X' = \frac{(Valor_{max\ deseado} - Valor_{min\ deseado})}{max(data) - min(data)} * (x - min(data)) + Valor_{min\ deseado}$$

Para demostrar que la transformación no afecta la relación entre las variables y que solo se trata de un escalado, se grafica el pH vs la Dosis de Sulfato de Aluminio antes de la transformación (Figura 32.a) y luego de haberla hecho (Figura 32.b).

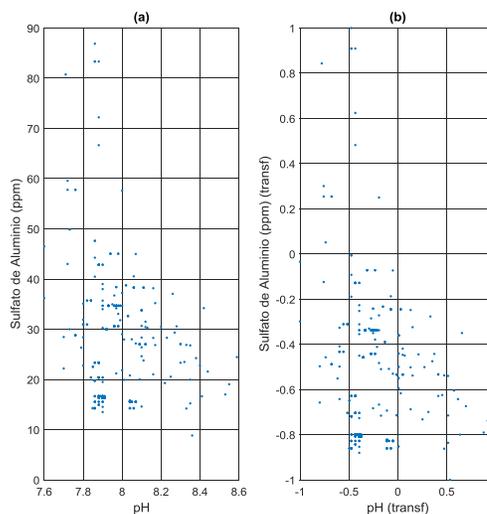


Figura 32. pH vs Sulfato de Aluminio: antes (a) y después (b) de la transformación de los datos.

Una vez transformados los datos se genera el modelo difuso tipo T-S con el toolbox FuzzyLogicDesigner (MatLab®, 2011) en el cual se elige el método de grid partitioning, en donde se le suministraron los parámetros de entrada presentados en

la Tabla 11. A cada variable se le asignan 2 conjuntos difusos, descritos por FP de tipo gaussiana, etiquetándolas como bajo y alto. Además, al tener dos conjuntos difusos para cada variable, se obtiene un total de 16 sub-espacios o hiperplanos, lo que significa que se tienen 16 reglas difusas para describir la dosificación de sulfato de aluminio.

Tabla 11. Parámetros de entrada para el Fuzzy Logic Designer Toolbox (MatLab®, 2011).

Variable de Entrada	Nº de Conjuntos Difusos
pH	2
Turbidez	2
Color	2
Alcalinidad	2
Tipo de FP para los conjuntos difusos	Ecuación Algebraica de FP
Gaussiana	$\mu(x; \sigma, c) = e^{-\frac{(x-c)^2}{2*\sigma^2}}$
Tipo de FP de Variable de Salida	Ecuación Algebraica de FP
Lineal	$y = b_0 + b_1 * pH + b_2 * Turbidez + b_3 * Color + b_4 * Alcalinidad$
Método de Clúster	
Grid Partitioning	

El modelo difuso que se genera tiene las FP de las variables de entrada que se muestran en la Figura 33.

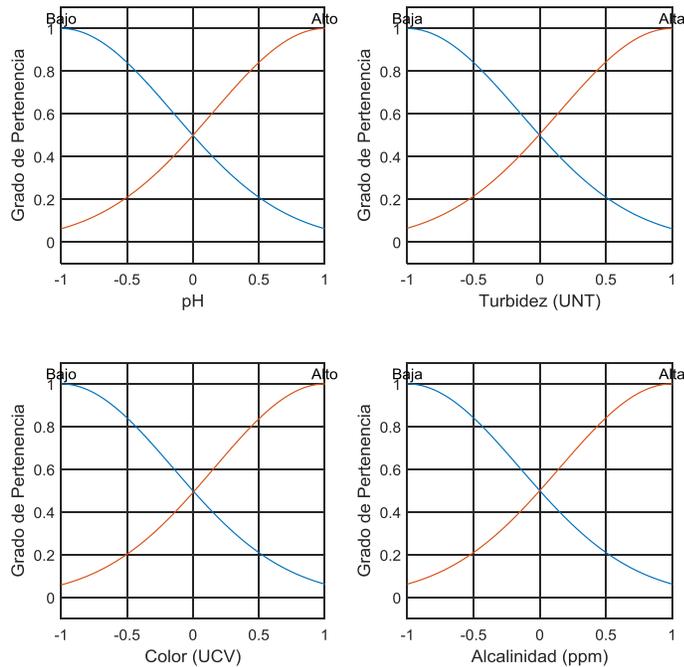


Figura 33. FP del modelo ANFIS de Sulfato de Aluminio.

Al modelo T-S generado, se le aplica una red neuronal, para encontrar los parámetros de las FP de la variable de salida que minimicen el valor de los residuos. Los parámetros de entrada al Fuzzy Logic Designer se presentan en la Tabla 12.

Tabla 12. Parámetros de entrada para el comando ANFIS (MatLab®, 2011).

Datos de entrada para entrenar:	Datos de modelación
Método de optimización:	Hybrid
Error o tolerancia	0.001
Numero de sesiones (epochs)	50

Los parámetros de las funciones de pertenencia optimizados se muestran en la Tabla 13, recordando que las FP de las variables de salida tienen la estructura siguiente:

$$FP \{ b_0 + b_1 * pH + b_2 * Turbidez + b_3 * Color + b_4 * Alcalinidad \}$$

Tabla 13. Parámetros de las FP de la variable Sulfato de Aluminio modelo ANFIS.

b_0	b_1	b_2	b_3	b_4
-2.67	67.24	-57.21	-0.45	9.22
-3.49	5.07	-4.94	0.78	-5.02
-63.93	-12.56	-20.23	21.15	-76.97
-70.74	-62.46	-20.17	30.09	56.07
122.28	15.33	80.13	77.23	7.47
80.19	30.23	93.37	-54.21	33.76
34.03	36.76	0.62	-72.03	30.60
29.48	13.96	-17.84	-73.17	98.71
-1.89	54.11	-51.95	1.49	6.30
-2.54	124.20	-85.32	-0.41	39.46
-107.97	-41.30	-8.37	43.13	-63.26
-44.11	-43.00	-8.11	20.13	-88.61
-32.62	-47.93	45.89	97.15	-63.06
-46.37	-33.00	59.20	-38.45	-96.46
35.98	-62.83	-38.49	-15.55	6.58
65.72	-77.30	-54.74	-19.30	2.45

Al finalizar la etapa de entrenamiento, se procede a evaluar el modelo ANFIS con los datos de validación. Los resultados se presentan en la Figura 34. Adicionalmente se calculan los coeficientes de correlación de Pearson.

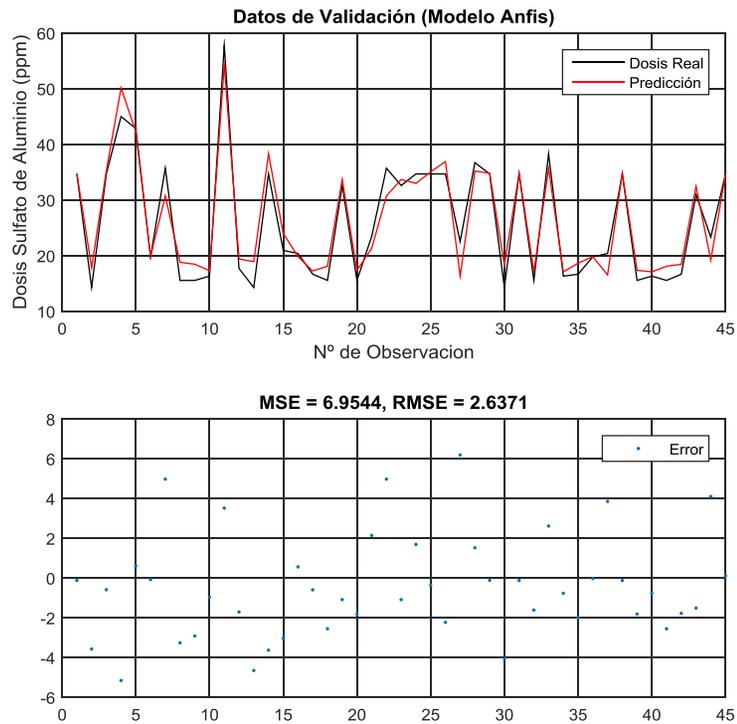


Figura 34. Desempeño del modelo ANFIS de Sulfato de Aluminio sobre los datos de validación.

Tabla 14. Coeficientes de correlación de Pearson modelo ANFIS.

Datos	R^2	$R^2_{ajustado}$
Validación	0.937	0.936

Se puede observar que el modelo ANFIS presenta el mejor desempeño de los cuatro utilizados ya que tiene el mejor ajuste de todos, arrojando un 93.7% de precisión, lo que comprueba que los modelos ANFIS tienen la capacidad de aprender de los datos de entrenamiento y de realizar estimaciones con un alto grado de precisión, haciéndolo un modelo más robusto. Al igual que el modelo Mamdani, la construcción de un modelo tipo ANFIS no tienen etapas de determinación de puntos atípicos, por lo tanto, se comprueba que los modelos difusos contienen

información valiosa de los datos que para modelos convencionales es desechada. En las Figura 35 y Figura 36 se comparan, respectivamente, los desempeños de los modelos estadísticos y difusos para la dosis de sulfato de aluminio, y en la Figura 37 se compararan los mejores modelos obtenidos.

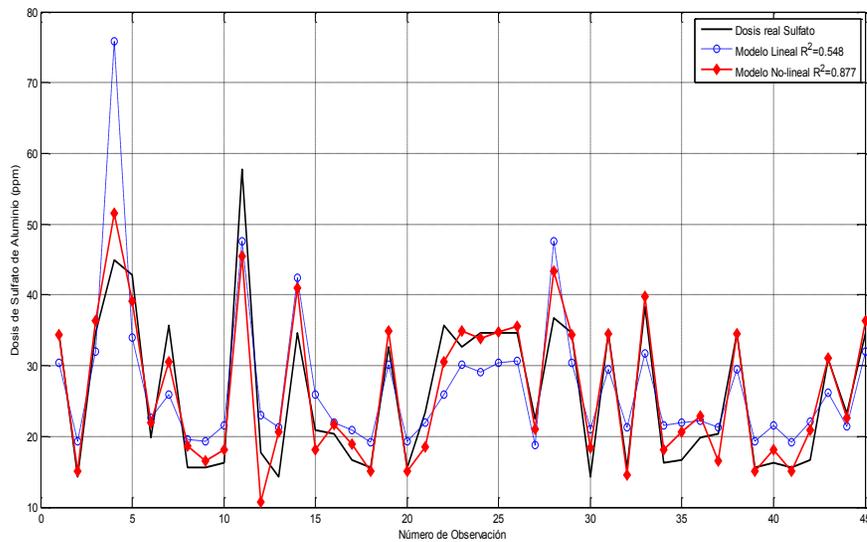


Figura 35. Comparación entre modelos estadísticos con datos de validación (Sulfato de Aluminio).

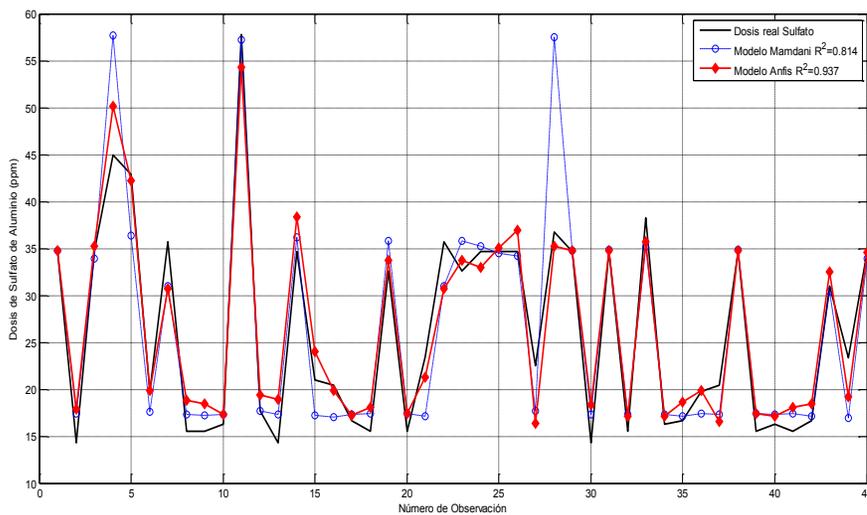


Figura 36. Comparación entre modelos difusos con datos de validación (Sulfato de Aluminio).

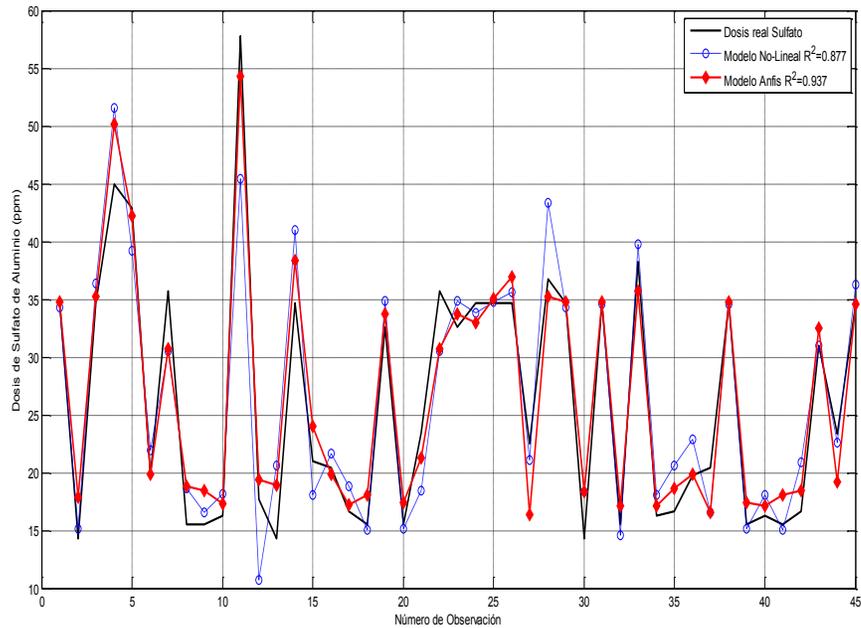


Figura 37. Comparación entre los mejores modelos (Sulfato de Aluminio).

www.bdigital.ula.ve

IV.1.2 POLICLORURO DE ALUMINIO, PAC.

En este caso, se consideran como variables de entrada para el modelo lineal y no-lineal: el pH, turbidez (UNT), color (UCV) y alcalinidad (ppm) y como variable de salida la dosificación de Policloruro de Aluminio (ppm). Siguiendo el procedimiento descrito previamente, se construyen los modelos estadísticos para la dosificación de PAC. En la Tabla 15, aparecen las ecuaciones correspondientes. La comparación entre modelos estadísticos y difusos con los datos de validación, aparecen en las figuras 38 y 39.

Tabla 15. Modelo lineal y no-lineal para la dosificación de PAC.

Modelo	Ecuación
Lineal	$\widehat{PAC} = 138.65 - 14.82 * pH + 0.016082 * Turbidez - 0.0042572 * Color + 0.21134 * Alc$
No-Lineal	$\widehat{PAC} = 137.49 - 19.166 * pH + 0.02196 * Turbidez + \frac{121.56}{Turbidez} + \frac{127.94}{Color} + 0.51793 * Alc$

Para el modelo difuso Mamdani, se toman como variables de entrada la turbidez y el color. Para el modelo ANFIS, se usan las mismas variables que en los modelos lineales y no-lineales. Las estructuras de los modelos difusos se presentan en la sección de ANEXOS VII.1

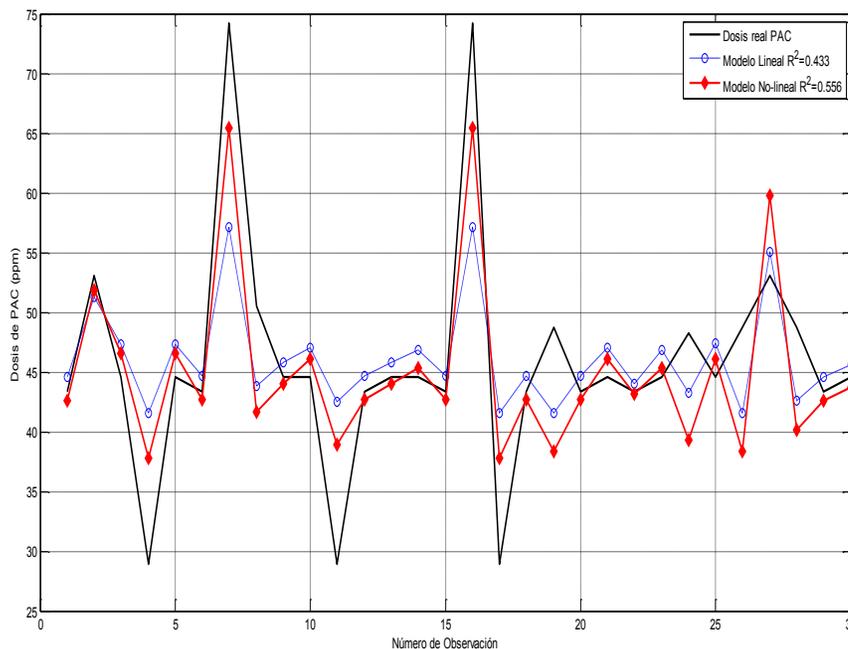


Figura 38. Comparación entre modelos estadísticos con datos de validación (PAC).

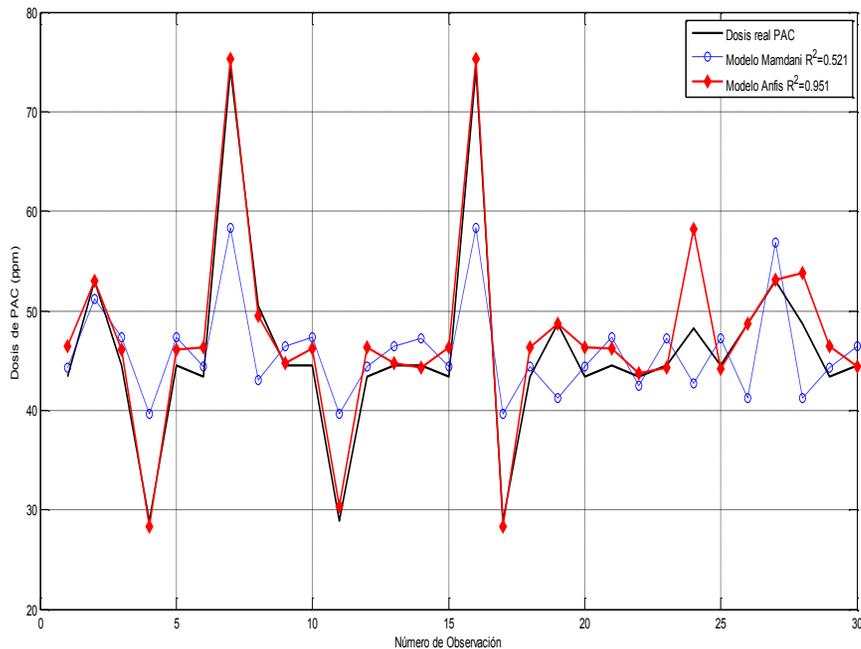


Figura 39. Comparación entre modelos difusos con datos de validación (PAC).

En la Tabla 16 se presentan los coeficientes de correlación de Pearson de cada modelo desarrollado. Se puede notar que el peor desempeño ocurre con el modelo lineal, con solo el 43.3%, de precisión con respecto a los datos de validación. El modelo Mamdani, con solo dos variables de entrada se aproxima al desempeño obtenido con el modelo no-lineal (al igual que en el caso de la dosificación de sulfato de aluminio). Sin embargo, estos tres modelos no logran ajustarse de buena manera a las dosificaciones reales realizadas en la planta, mientras que usando, el modelo ANFIS, se obtiene un desempeño muy satisfactorio, con una precisión del 95.1%.

Tabla 16. Comparación entre los modelos desarrollados con los datos de validación (PAC).

Modelo	R^2	$R^2_{ajustado}$
ANFIS	0.951	0.949
No-lineal	0.556	0.515
Mamdani	0.521	0.512
Lineal	0.433	0.392

En la Figura 40 se muestra la comparación entre el modelo no-lineal y el ANFIS, cuyos obtuvieron la mejor precisión.

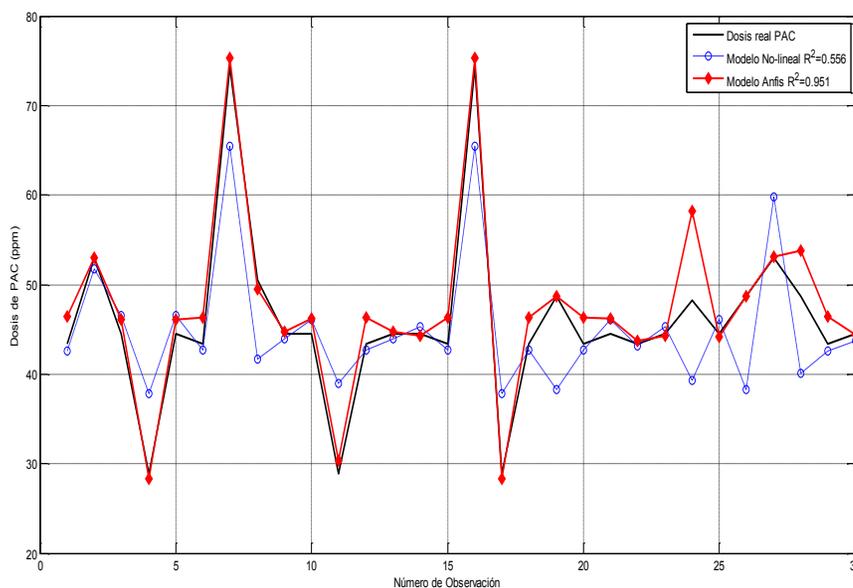


Figura 40. Comparación entre mejores modelos (PAC).

IV.2 ETAPA DE DESINFECCIÓN.

IV.2.1 CLORO GASEOSO.

Se toman como variables de entrada para el modelo lineal y no-lineal: el pH, alcalinidad (ppm) y cloro residual (ppm). Como variable de salida, la dosificación de cloro gaseoso (ppm). Siguiendo el procedimiento explicado antes, se determinan las ecuaciones para la dosificación de Cloro, utilizando los modelos estadísticos presentados en la Tabla 17.

Tabla 17. Modelo lineal y no-lineal para la dosificación de Cloro.

Modelo	Ecuación
Lineal	$\hat{Cl} = 4.15 - 0.15 * pH - 0.013 * Alcalinidad - 0.25 * Cl_{residual}$
No-Lineal	$\ln(\hat{Cl}) = 0.68 - 0.57 * (pH - 8)^2 + 0.00088 * (Alcalinidad - 90)^2 - 0.27 * Cl_{res}$

Para el modelo difuso Mamdani, se toman como variables de entrada el pH y el cloro residual. Para el modelo ANFIS, se usan las mismas variables que para los modelos lineal y no-lineal. Las estructuras de los modelos difusos se presentan en la sección de ANEXOS VII.1.

En la Tabla 18 se presentan los coeficientes de correlación obtenidos para cada modelo. Es notable que los modelos utilizados no fueron capaces de desempeñar un buen papel en cuanto a las predicciones de la dosificación de cloro se refiere, con una precisión máxima de 32,1%, correspondiente al modelo ANFIS.

Tabla 18. Comparación entre los modelos desarrollados con los datos de validación (Cloro).

Modelo	R^2	$R^2_{ajustado}$
ANFIS	0.321	0.26
No-lineal	0.206	0.092
Mamdani	0.023	-
Lineal	0.156	0.036

Esto se puede atribuir a que las variables de entrada consideradas (pH, Alcalinidad y Cloro residual) no captan de forma adecuada los cambios en la dosificación del cloro que se efectúa en la planta de potabilización. Otra variable adicional para la construcción de los modelos es el NMP (Número Más Probable de coliformes), la cual debería guardar estrecha relación con la dosificación de desinfectante, pero que no se pudo considerar, por la poca cantidad de información para este parámetro. La desventaja que presenta usar la variable NMP radica en el tiempo de espera por los resultados del ensayo, ya que se requiere de al menos 3 días para obtener un resultado confiable de la misma. Por lo tanto, si el objetivo es realizar una predicción de dosificación al afluente que llega en un momento específico a la planta, no se puede esperar tanto tiempo para determinar la dosis de cloro a aplicar, cuando ya el agua ha salido de la planta. Por lo tanto, no es posible determinar un modelo de dosificación de cloro usando NMP. Los resultados muestran que la dosificación de cloro pareciera ser independiente de las variables de entradas consideradas, y esto parece tener lógica puesto que el cloro dosificado se realiza con la finalidad de tener un efecto residual que garantice la eliminación de microorganismos patógenos en el efluente, independientemente de las variables de entrada antes mencionadas.

En la Figura 41 y Figura 42 se presentan los resultados de los modelos de dosificación de cloro a nivel de planta.

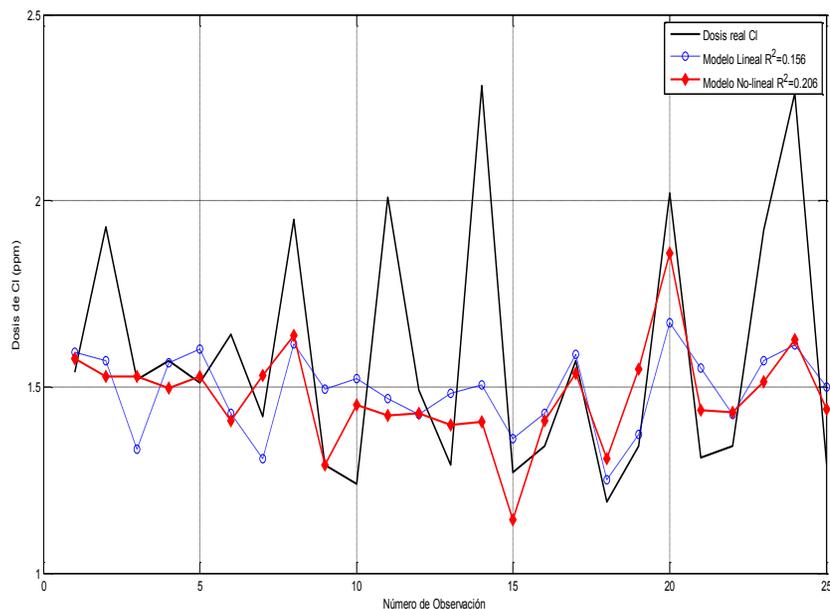


Figura 41. Comparación entre modelos estadísticos con datos de validación (Cloro).

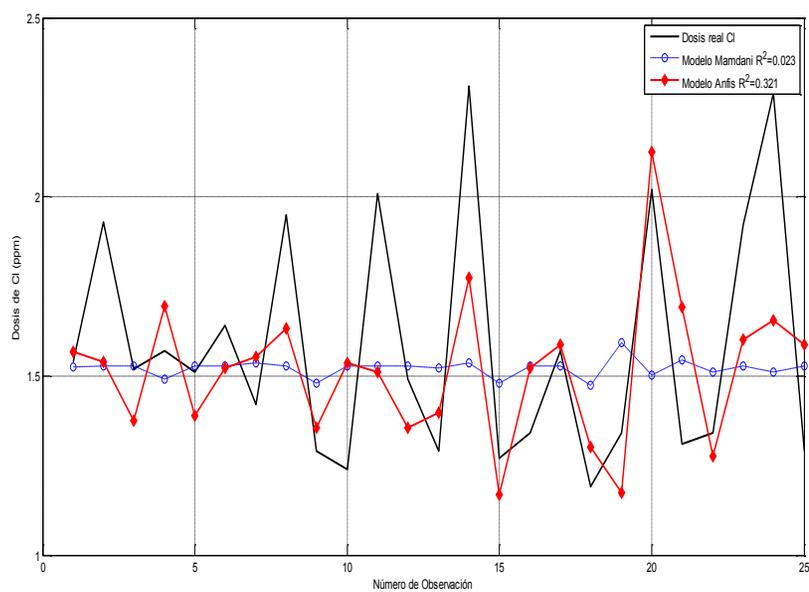


Figura 42. Comparación entre modelos difusos con datos de validación (Cloro).

IV.3 MODELOS OBTENIDOS EN EL LABORATORIO.

IV.3.1 SULFATO DE ALUMINIO.

A nivel de laboratorio, se tomaron como variables de entrada: pH, turbidez (UNT) y temperatura (°C) y como variable de salida, se utilizó la dosificación de sulfato de aluminio (ppm).

Los datos fueron recopilados siguiendo el procedimiento usado por Mayorga (Mayorga, 2009). Los datos se presentan en la Tabla 27 (ver sección ANEXOS VII.4).

Siguiendo el mismo procedimiento, se obtienen los modelos estadísticos presentados en la Tabla 19.

Tabla 19. Modelo lineal y no-lineal para la dosificación de Sulfato de aluminio en el Laboratorio.

Modelo	Ecuación
Lineal	$\widehat{Sulf} = -32.76 + 2.41 * pH + 0.047 * Turbidez + 0.91 * Temp$
No-Lineal	$\widehat{Sulf} = -11.47 + 0.204 * (pH - 8)^2 + 0.053 * Turbidez + 0.00004 * (Turb - 500)^2 + 0.53 * Temp$

Para el modelo Mamdani, se toman como variables de entrada la turbidez y el pH del agua. En el modelo ANFIS se usan las mismas variables del modelo lineal. Las estructuras de los modelos difusos se presentan en la sección de ANEXOS VII.1. Con los datos obtenidos en el laboratorio, se desarrollan modelos con muy buen desempeño para realizar predicciones, tal como se ve en la Tabla 20.

Los resultados obtenidos demuestran que, a nivel de laboratorio, no es necesario el desarrollo de un modelo difuso para describir el proceso de potabilización de agua, ya que con los modelos estadísticos se obtienen resultados aceptables. Lo interesante es que, en este caso, el modelo lineal también obtiene un desempeño alto, lo que significa que la cantidad de datos utilizados afecta directamente la

predicción, ya que sólo fueron utilizadas 9 observaciones (de 35 tomadas) para validar el modelo.

Tabla 20. Comparación entre los modelos desarrollados con los datos de validación.

Datos de Laboratorio

Modelo	R^2	$R^2_{ajustado}$
No-Lineal	0.976	0.971
Mamdani	0.966	0.965
ANFIS	0.959	0.955
Lineal	0.911	0.898

En la Figura 43 y Figura 44 se comparan, respectivamente, los desempeños de los modelos estadísticos y difusos para la dosis de sulfato de aluminio en el laboratorio, y en la Figura 45 se compararan los mejores modelos obtenidos.

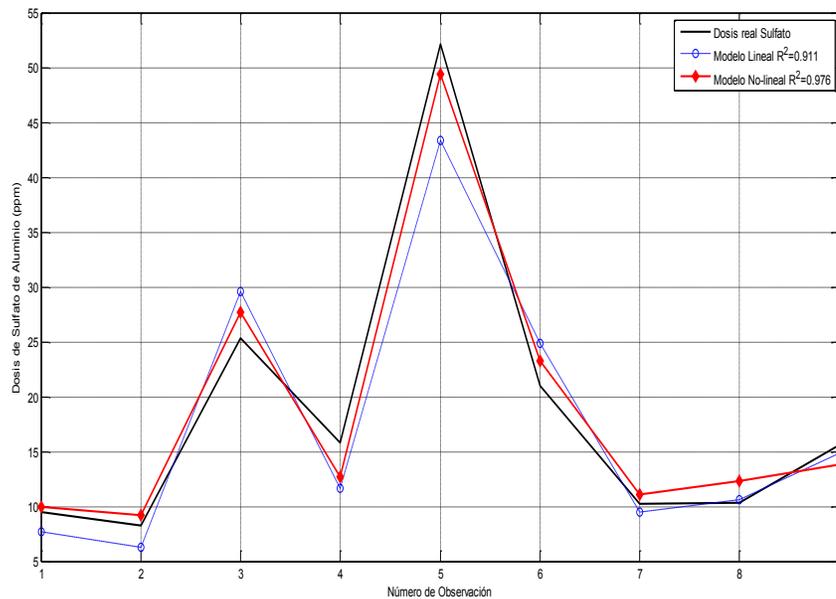


Figura 43. Comparación entre modelos estadísticos con datos de validación. Datos de Laboratorio

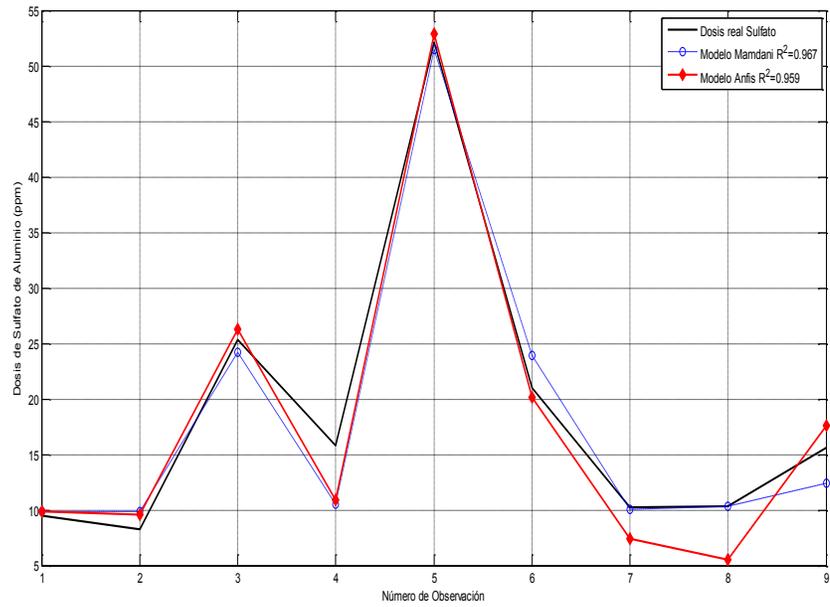


Figura 44. Comparación entre modelos difusos con datos de validación. Datos de Laboratorio

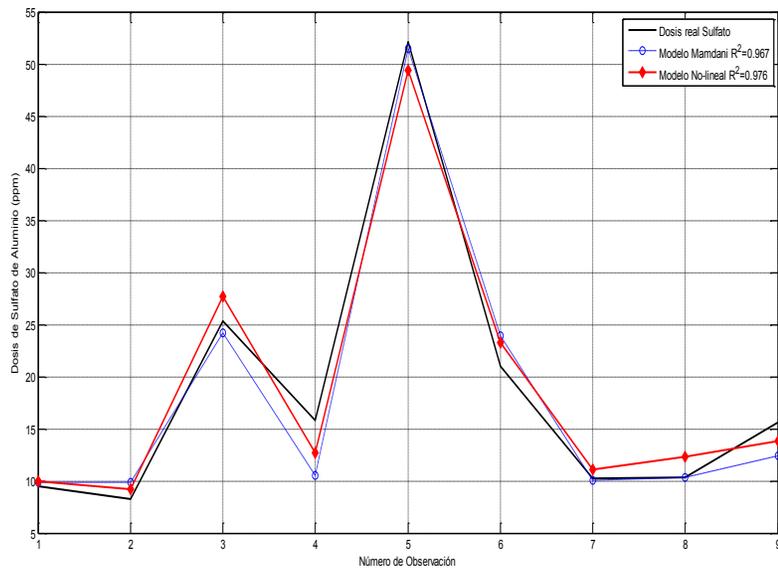


Figura 45. Comparación entre los mejores modelos. Datos de laboratorio (Sulfato de Aluminio).

V. CONCLUSIONES Y RECOMENDACIONES

Los modelos tipo ANFIS desarrollados para las etapas de dosificación de coagulantes a nivel de la planta de potabilización de agua obtuvieron los mejores desempeños debido a que fueron entrenados con toda la base de datos. Por otro lado, los modelos Mamdani, al ser más sencillos en su estructura, y al tener reglas base fundamentadas en el conocimiento que se tiene del proceso, tienen un desempeño más bajo. Esto indica que los modelos de tipo Mamdani podrían ser lo suficientemente buenos si en su estructura se adicionan más variables, que permitan describir el proceso de mejor forma, sabiendo que a medida que se agreguen variables aumentará el número de reglas base. De esta manera los modelos tipo Mamdani se basan en el entendimiento del proceso, mientras que los modelos ANFIS utilizan información “escondida” en la base de datos.

Los modelos tipo ANFIS tienen la ventaja de ser auto adaptativos. Esto significa que el desempeño inicial del modelo puede ser mejorado a medida que aumenta la cantidad de datos, proporcionando con ello mayor información al sistema de inferencia neuro-difuso, produciendo cada vez mejores resultados.

La desventaja de los modelos difusos se debe a su limitante de extrapolar más allá del universo descrito para cada variable. Por lo tanto, si hubiese un valor fuera de ese universo, el grado de pertenencia sería cero, lo que no tiene sentido a la hora de calcular una predicción. La solución a este problema se resume en la reestructuración de las funciones de membresía de las variables tanto de entrada como de salida.

Los modelos lineales desarrollados no proporcionan un buen desempeño, lo que indica que el proceso es altamente no lineal. Los modelos no-lineales desarrollados tienen buen desempeño en la etapa de dosificación de coagulante, pero la ecuación algebraica que los define es más compleja.

La etapa de dosificación de cloro no pudo ser modelada por ningún modelo utilizado en este trabajo de investigación, debido a que no se cuenta con variables que describan la dosis de cloro de forma adecuada. En la planta de potabilización de agua de El Vigía, se realiza la dosificación de cloro en función del cloro residual.

En otras palabras, se realiza un control manual por retroalimentación, midiendo el cloro residual en el efluente, y dosificando con cloro gaseoso a la entrada de la planta, a fin de garantizar la existencia de cloro residual y por ende, la ausencia de microorganismos patógenos en el efluente. Los resultados obtenidos permiten corroborar que una estrategia basada en modelos predictivos (convencionales y no convencionales) no es la forma adecuada para estimar la dosificación de cloro en este tipo de proceso en particular.

El modelo de laboratorio para la dosificación de sulfato de aluminio obtuvo muy buenos resultados para todos los modelos. Esto puede ser debido, por una parte a la baja cantidad de datos obtenida en el laboratorio con respecto a los obtenidos en planta, y por otro lado a que las variables de entrada (pH, turbidez y temperatura) presentan una alta correlación con la variable de salida (Sulfato de Aluminio).

www.bdigital.ula.ve

VI. REFERENCIAS

- A.N.C. (2007). *El Agua Potable Segura Es Esencial*. Recuperado el 8 de Julio de 2015, de Academia Nacional de Ciencias: <https://www.koshland-science-museum.org/water/html/es/Treatment/Coagulation-Flocculation.html>
- Black & Veatch, C. (2010). *White's Handbook of Chlorination and Alternative Disinfectants*. John Wiley & Sons.
- Campos Pinilla, C. (2003). Indicadores De Contaminacion Fecal En Aguas. En *Agua Potable Para Comunidades Rurales, Reuso y Tratamientos Avanzados De Aguas Residuales Domésticas*.
- Castaño, S. A., Hernández, D. S., & Gallo, J. H. (2013). Control y Monitoreo de Temperatura para un Horno de Curado de Prendas Indigo utilizando Logica Digusa. *Revista Politécnica*, 9(17), 69-81.
- Chatterjee, S., & Hadi, A. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*, 379-393.
- Cheremisinoff, N. P. (2002). *Handbook Of Water And Wastewater Treatment Technologies*. Butterworth-Heinemann.
- Dufour, A., Medema, G., Robertson, W., Waite, M., Hunter P., Kirby, R., y otros. (2003). *Assessing microbial safety of drinking water: Improving approaches and methods*. Ginebra (Suiza).
- Fallahpour, M., Fatehi, A., Araabi, B., & Azizi, M. (2008). *A Neuro-Fuzzy Controller for Rotary Cement Kilns*. Seoul: IFAC.
- Fernández, N., Carvajal, L., & Colina, E. (2010). Sistema Difuso Tipo Mamdani Para La Determinacion Generica De La Calidad Del Agua. *BISTUA: Revista de la Facultad de Ciencias Básicas*, 8(1).
- Gaceta Oficial de la Republica de Venezuela. (13 de Febrero de 1998). Normas Sanitarias de Calidad del Agua Potable. (36.395). Caracas, Distrito Capital, Venezuela.
- Gafanhao, M. (2007). *Aplicaciones de Lógica Difusa en la Predicción del Punto de Precipitación de los Asfaltenos*. Tesis de Grado, Universidad de Los Andes.
- González, I. (2006). *Establecimiento Y Determinación De Indices De Calidad Del Agua. Enfoque Basado En Lógica Difusa*. Universidad de Los Andes, Facultad de Ingeniería.
- Harris, J. (2000). *An Introduction to Fuzzy Logic Applications*. Kluwer Academic Publishers.
- Jiménez, L. (2007). *Modelos Avanzados Para La Predicción A Corto Plazo De La Producción Eléctrica En Parques Eólicos*. Tesis Doctoral, Universidad de La Rioja.
- Jyh-Shing, & Jang, R. (1991). Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm. 762 - 767.

- Lenntech. (2015). *LENNTECH*. Recuperado el 28 de Junio de 2015, de <http://www.lenntech.es/turbidez.htm>
- M. Ingels, R. (1980). How to use the computer to analyze test data. 145-156.
- Mamdani, E. (1974). Application Of Fuzzy Algorithms For Control Of Simple Dynamic Plant. *Proceedings IEEE*, 12(121), 1585 - 1588.
- MatLab®. (2011). Help Files for MatLab R2011a. *Matlab Function Reference: Fuzzy Logic Toolbox: Fuzzy Inference System Modeling*. The MathWorks, Inc.
- Mayorga, J. (2009). *Guía de Practicas del Laboratorio de Contaminación Ambiental*. Universidad De Los Andes, Facultad De Ingeniería.
- NIST/SEMATECH. (Abril de 2012). *e-Handbook of Statistical Methods*. Recuperado el 20 de Octubre de 2016, de <http://www.itl.nist.gov/div898/handbook/>
- OMS. (2008). Primer Apéndice a la Tercera Edición. En Organización Mundial de la Salud, *Guías Para La Calidad Del Agua Potable* (Vol. 1).
- OPS/CEPIS. (2002). Tratamiento De Agua. En Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente, *Operación y Mantenimiento De Plantas De Tratamiento De Agua Manual De Capacitacion De Operadores*.
- OSE. (2015). *Obras Sanitarias del Estado*. Recuperado el 2015 de Julio de 11, de <http://www.ose.com.uy>
- Parsons, S., & Jefferson, B. (2006). *Introduction To Potable Water Treatment Processes*. Blackwell Publishing Ltd.
- Perlman, H. (2015). *USGS Science for a Changing World*. Recuperado el 2015 de Junio de 30, de <http://water.usgs.gov/gotita/characteristics.html>
- Piegat, A. (2001). *Fuzzy Modeling And Control*. Zolnierska, Polonia: Springer-Verlag Berlin Heidelberg GmbH.
- Renatas, K., Ángel, A., & Perez, J. (2015). *Multicolinealidad y Observaciones Atípicas*. Recuperado el 20 de 11 de 2016, de http://www.uoc.edu/in3/emath/docs/T03_Multicolinealidad.pdf
- Ross, T. (2004). *Fuzzy Logic With Engineering Applications* (Segunda ed.). John Wiley & Sons, Ltd.
- Suparta, W., & Alhasa, K. (2016). *Modeling of Topospheric Delays Using ANFIS*. Springer.
- T. Hagan, M., B. Demuth, H., Hudson Beale, M., & De Jesús, O. (2012). *Neural Network Design*.
- Takagi, T., & Sugeno, M. (1985). Fuzzy Identification Of Systems And Its Applications To Modeling And Control. *IEEE Transactions On Systems, Man And Cybernetics*, 15(1), 116-132.
- Toprak, H. (2006). *Toprak Home page Wastewater Engineering Desing - Construction - Operation*. Recuperado el 25 de Junio de 2015, de <http://web.deu.edu.tr/atiksu/ana52/aryen2.html>

- Tututi Avila, S. (2010). *Control Del Proceso De Cristalización Continuo No-Isotérmico Empleando Lógica Difusa*. Tesis doctoral, Instituto Tecnológico de Celaya.
- Villareal, C. A., & Caicedo, E. (2013). Técnicas De Inteligencia Computacional Aplicadas A Modelos De Estimacion De Coagulante En El Proceso De Potabilizacion De Agua. *Revista Facultad de Ingeniería Universidad de Antioquia*, 1-15.
- Wang, L., Hung, Y.-T., & Shammass, N. (2005). *Physicochemical Treatment Processes Handbook of Environmental Engineering* (Vol. III). Totowa, New Jersey: Humana Press.
- Zadeh, L. (1965). Fuzzy Sets. *Information and Control*, 8, 338-353.
- Zekai, S. (2010). *Fuzzy Logic And Hydrological Modeling*. CRC Press Taylor & Francis Group.

www.bdigital.ula.ve

VII. ANEXOS

VII.1 ESTRUCTURA DE MODELOS DIFUSOS.

VII.1.1 MAMDANI.

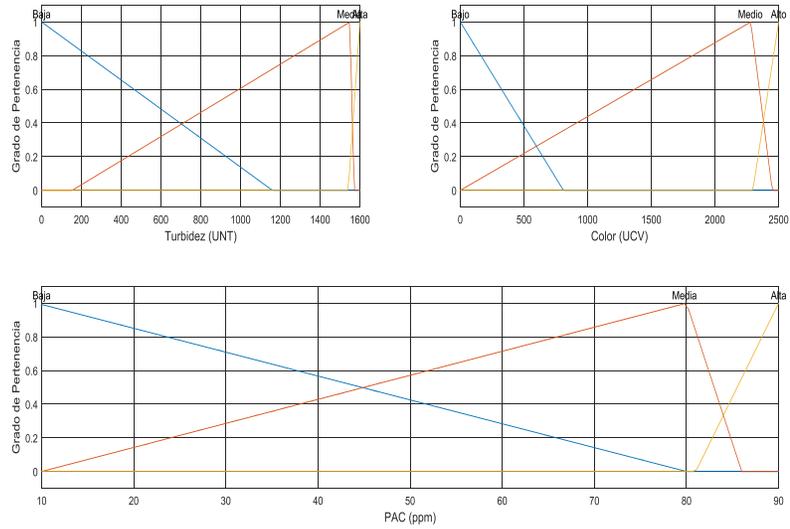


Figura 46. Funciones de pertenencia para el modelo Mamdani de dosificación de PAC.

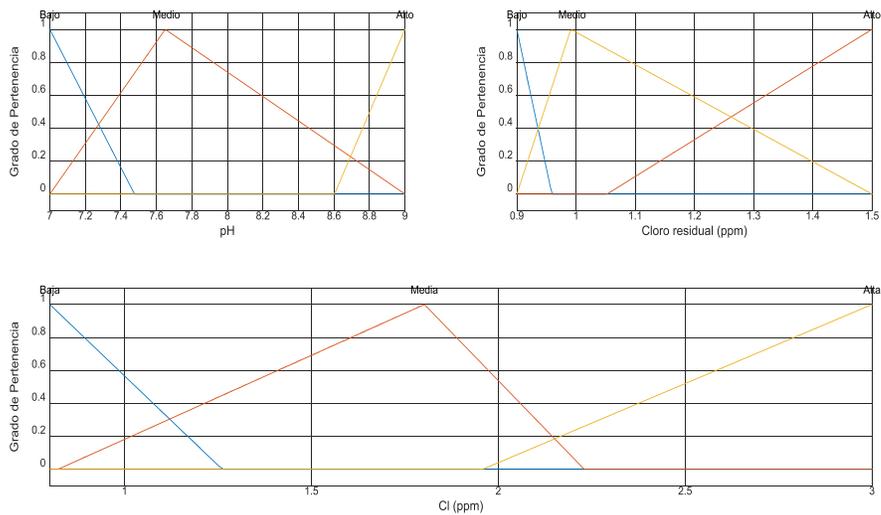


Figura 47. Funciones de pertenencia para el modelo mamdani de dosificación de Cloro.

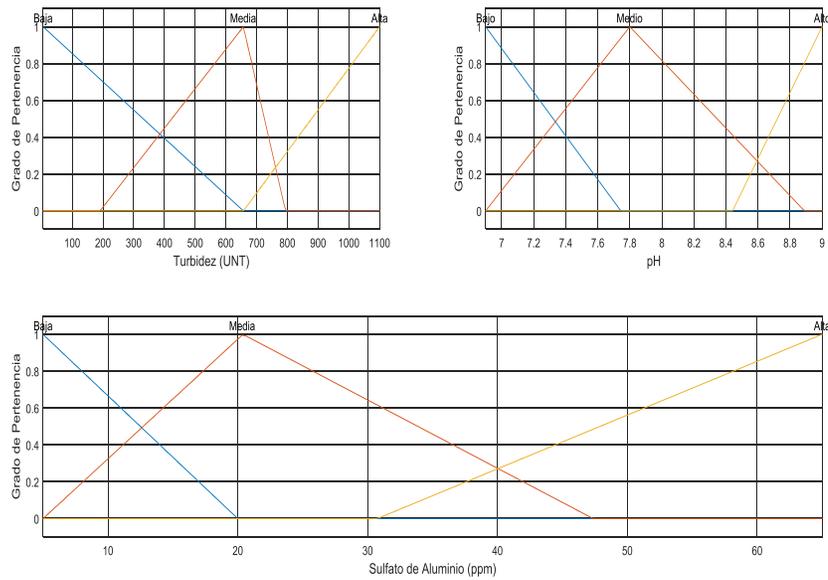


Figura 48. Funciones de pertenencia para el modelo Mamdani de dosificación de Sulfato de Aluminio (Laboratorio).

www.bdigital.ula.ve

VII.1.2 ANFIS.

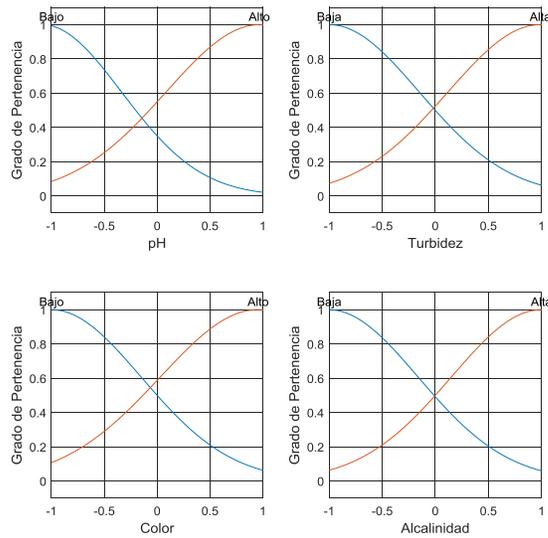


Figura 49. FP de las variables de entrada al modelo ANFIS de PAC.

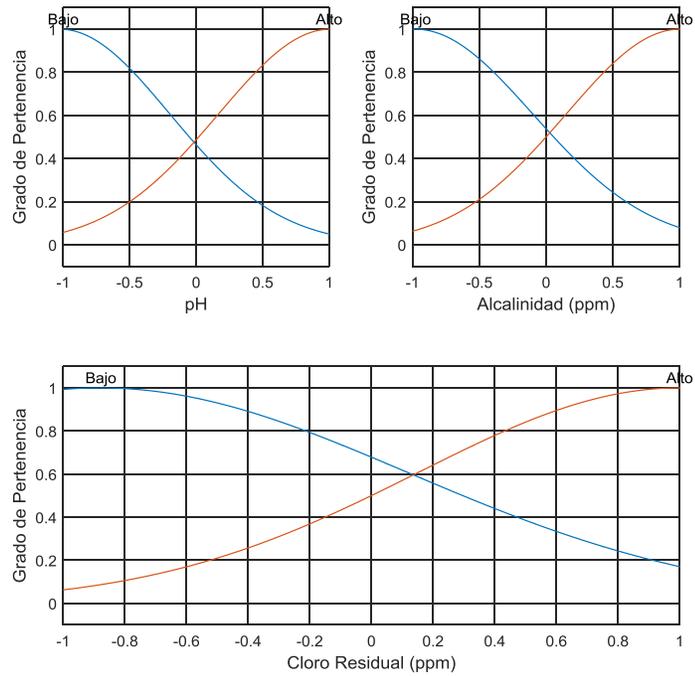


Figura 50. FP de las variables de entrada al modelo ANFIS de Cloro.

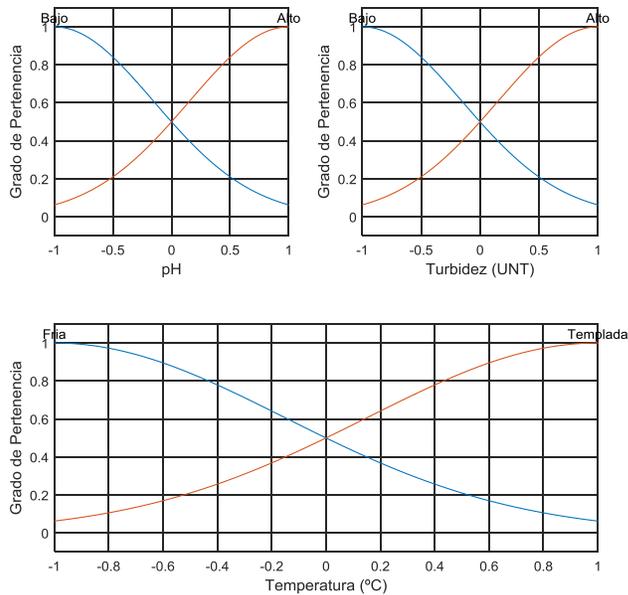


Figura 51. FP de las variables de entrada al modelo ANFIS de Sulfato de Aluminio (Laboratorio).

VII.2 GENERADOR DE REGLAS MODELOS MAMDANI.

Tabla 21. Generador de reglas difusas para el modelo de dosis de PAC en planta.

GENERADOR DE REGLAS (POLICLORURO DE ALUMINIO)			
PAC	COLOR		
TURBIDEZ	Bajo	Medio	Alto
Baja	Baja	Baja	Media
Media	Media	Media	Alta
Alta	NA	NA	Alta

Tabla 22. Generador de reglas difusas para el modelo de dosis de Cloro en planta.

GENERADOR DE REGLAS (CLORO)			
CLORO	CLORO RESIDUAL		
pH	Bajo	Medio	Alto
Bajo	Alta	Baja	Baja
Medio	Alta	Media	Baja
Alto	Alta	Media	Baja

Tabla 23. Generador de reglas difusas para el modelo de dosis de Sulfato de Aluminio en Laboratorio.

GENERADOR DE REGLAS (SULFATO DE ALUMINIO)			
SULFATO DE ALUMINIO	TURBIDEZ		
pH	Baja	Media	Alta

Bajo	Baja	Media	Alta
Medio	Baja	Media	Alta
Alto	Baja	Baja	Media

VII.3 TABLAS DE PARAMETROS DE FP DE SALIDA MODELOS ANFIS.

Tabla 24. Parámetros de las FP de la variable PAC en planta.

b_0	b_1	b_2	b_3	b_4
-20.76	-27.86	87.28	-13.36	31.36
16.04	40.88	-65.95	-11.46	-6.31
12.96	-82.40	112.70	24.18	-21.11
-117.66	72.87	-238.52	146.62	-65.58
42.99	27.05	54.71	50.10	-25.02
50.99	-39.68	58.40	-46.72	58.75
-28.22	12.08	-127.63	-149.07	8.36
32.35	51.22	70.17	43.12	18.68
-19.27	-75.70	46.18	6.39	-5.31
12.83	-31.61	45.02	10.06	6.84
90.31	-120.60	150.72	54.30	-39.45
-113.09	175.27	45.58	-59.98	85.15
-43.48	-35.67	-11.80	-77.35	-6.57
-3.33	-154.07	-67.42	-35.40	45.57
-21.07	124.29	-115.81	17.18	-2.10
127.67	28.93	-8.97	1.42	-82.98

Tabla 25. Parámetros de las FP de la variable Cloro en planta.

b_0	b_1	b_2	b_3
-23.18	8.09	20.03	1.81
0.25	6.58	11.17	-15.83
9.55	11.45	21.79	10.81
3.07	0.55	13.38	-18.48
-33.15	-10.98	-12.78	6.38
10.18	-8.51	-5.32	-2.82
37.82	-11.37	-27.03	-9.42
-20.78	-4.83	-12.54	22.30

Tabla 26. Parámetros de las FP de la variable Sulfato de Aluminio en Laboratorio.

b_0	b_1	b_2	b_3
0.44	0.60	-0.16	1.05
1.85	-1.44	1.28	-0.66
-1.67	0.60	-0.39	-0.92
0.63	2.59	1.50	0.61
0.53	2.98	-0.78	0.78
1.16	1.86	-0.33	0.16
2.08	2.65	1.16	-3.15
-0.84	1.36	-2.45	-0.77

VII.4 TABLA DE DATOS DEL LABORATORIO.

Tabla 27. Datos para los modelos de laboratorio de dosificación de Sulfato de Aluminio.

Muestra	pH	Turb (UNT)	Temp (°C)	Alumbre (ppm)	pH Final	Turb Final	Temp Final
1	8.1	171	20	12.5	7.5	1	20
2	7.8	75	20	9.5	6.9	4	20
3	8.1	92	19.5	9.2	7.3	2	19.5
4	7.4	189	19.5	15.8	6.7	1	19.5
5	7.2	167	20	10.3	6.4	0	20
6	8.5	210	21	15.3	7.8	3	21
7	8.2	93	21	10.2	7.1	0	21
8	7.2	163	21	12	6.3	1	21
9	7.6	967	20	58.4	6.9	4	20
10	7.4	602	20	29	7	2	20
11	7.7	454	19.5	21	7.1	1	19.5
12	7.9	470	19.5	21.6	6.9	1	19.5
13	7.7	229	19	16.2	6.8	4	19
14	7.8	558	20	26.8	7.3	3	20
15	7.4	853	20	52.1	6.8	5	20
16	7.5	251	21	17	7.1	2	21
17	7.6	40	20	7.4	6.9	0	20
18	8.1	304	19.5	18.7	7.4	1	19.5
19	8.2	501	19	23	7.1	1	19
20	8	223	19	16.5	7.3	4	19
21	8.1	275	21	19.1	7.6	5	21
22	7.5	381	21	20.8	6.8	4	21
23	7.9	831	20.5	48	6.9	5	20.5
24	7.8	120	20.5	10.9	7.1	3	20.5
25	8.1	20	20	7	7.2	2	20
26	8.4	870	19.5	51.4	7.6	1	19.5

27	8.2	1021	19.5	57.5	7.7	2	19.5
28	7.6	124	20	10.2	6.3	2	20
29	7.9	39	20	8.3	7	3	20
30	7.9	534	20	25.3	6.9	2	20
31	8	340	20	18.1	7	4	20
32	7.3	894	19.5	54.5	6.5	4	19.5
33	7.3	174	19.5	12.2	6.8	4	19.5
34	7.9	203	21	15.6	6.9	1	21
35	8	732	21	29.6	7.2	3	21

VII.5 FIGURAS DE LOS MODELOS OBTENIDOS EN PLANTA.

VII.5.1 MODELOS “SULFATO DE ALUMINIO”.

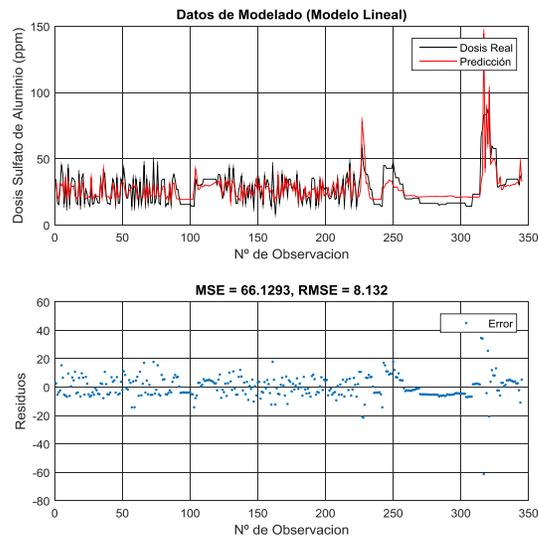


Figura 52. Desempeño del modelo lineal de Sulfato de Aluminio sobre los datos de modelado.

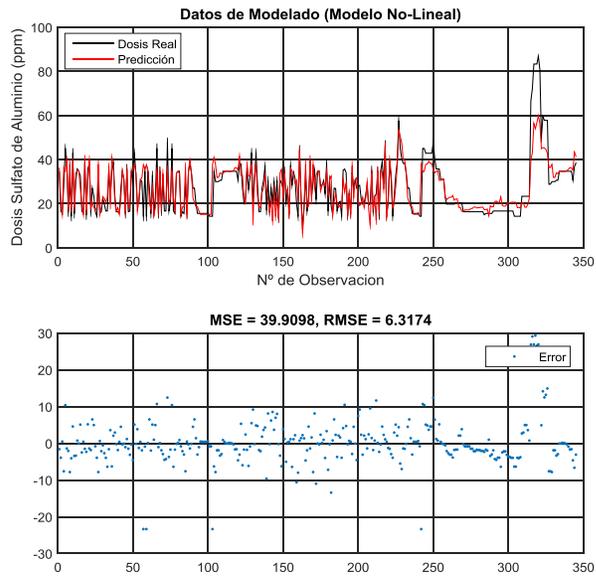


Figura 53. Desempeño del modelo no-lineal de Sulfato de Aluminio sobre los datos de modelado.

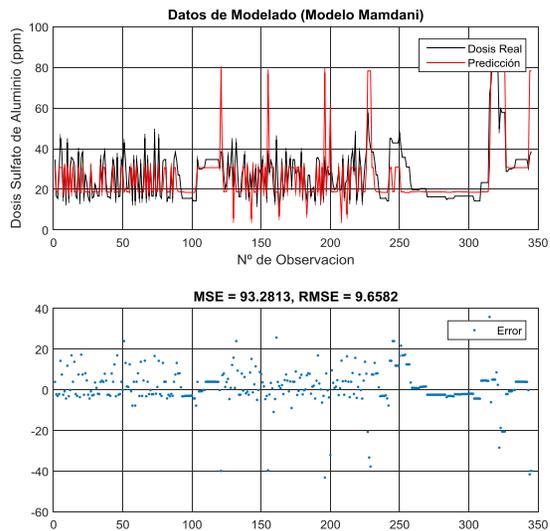


Figura 54. Desempeño del modelo Mamdani de Sulfato de Aluminio sobre los datos de modelado.

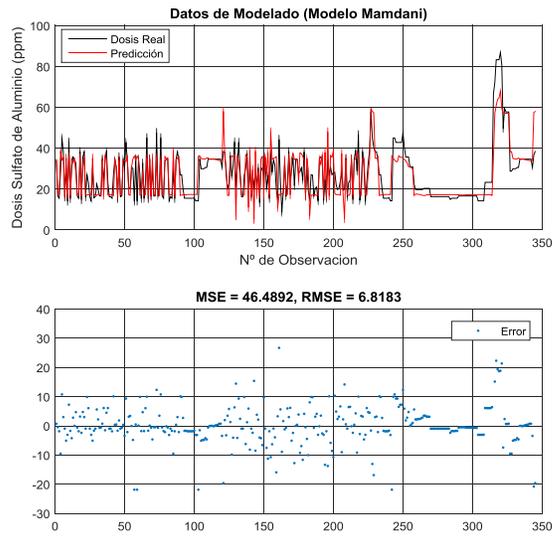


Figura 55. Nuevo desempeño del modelo Mamdani de Sulfato de Aluminio sobre los datos de modelado.

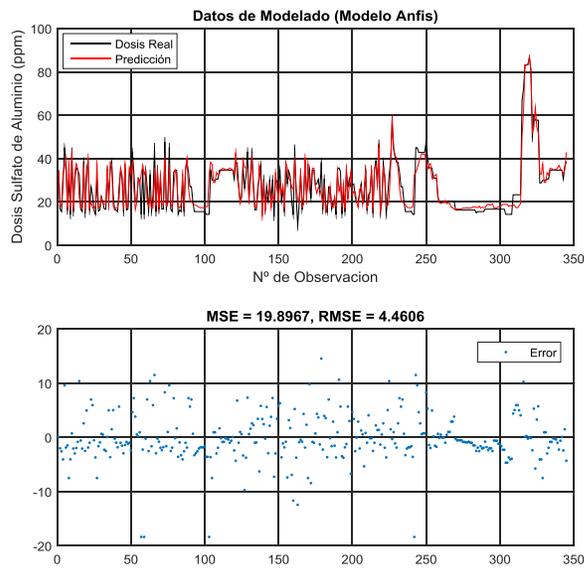


Figura 56. Desempeño del modelo ANFIS de Sulfato de Aluminio sobre los datos de modelado.

VII.5.2 MODELOS “PAC”

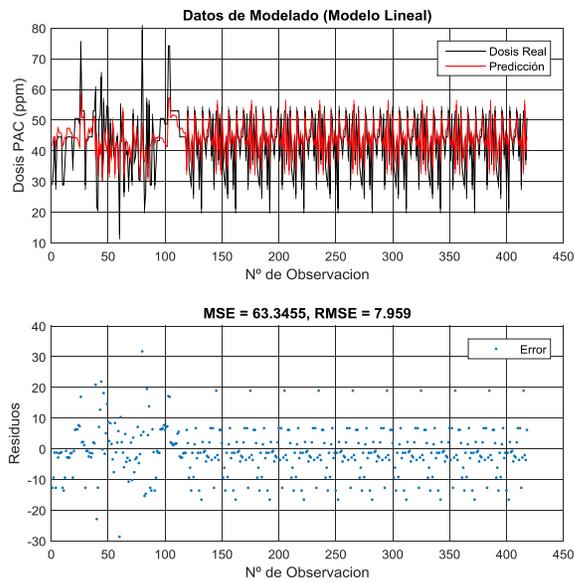


Figura 57. Desempeño del modelo lineal de PAC sobre los datos de modelado.

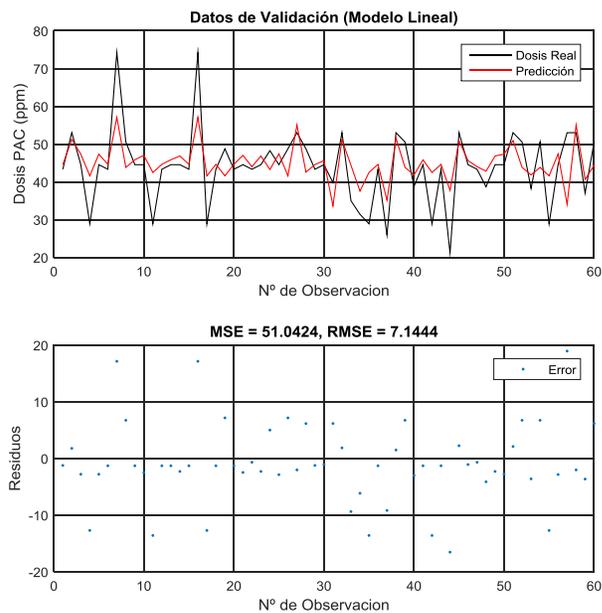


Figura 58. Desempeño del modelo lineal de PAC sobre los datos de validación.

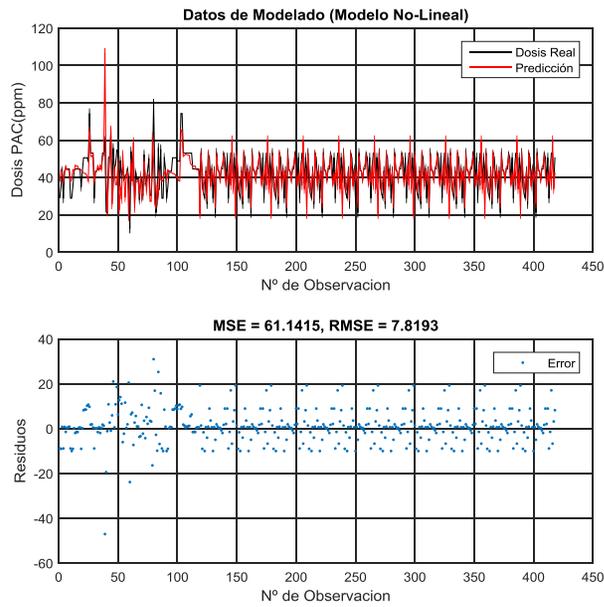


Figura 59. Desempeño del modelo no-lineal de PAC sobre los datos de modelado.

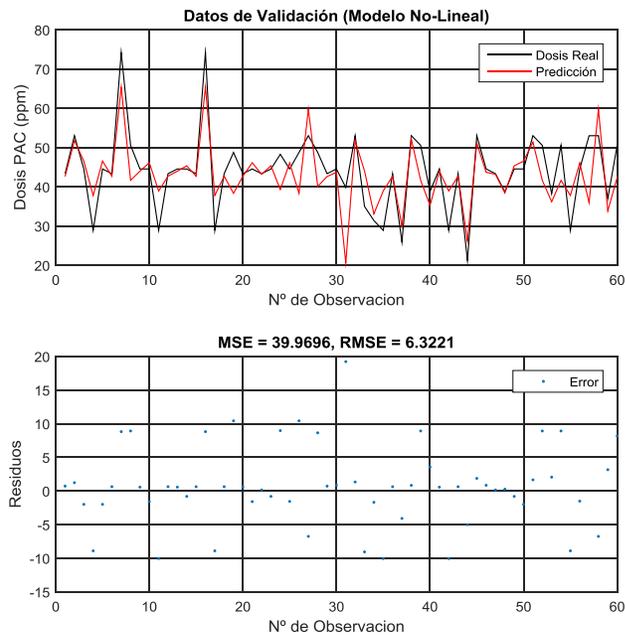


Figura 60. Desempeño del modelo no-lineal de PAC sobre los datos de validación.

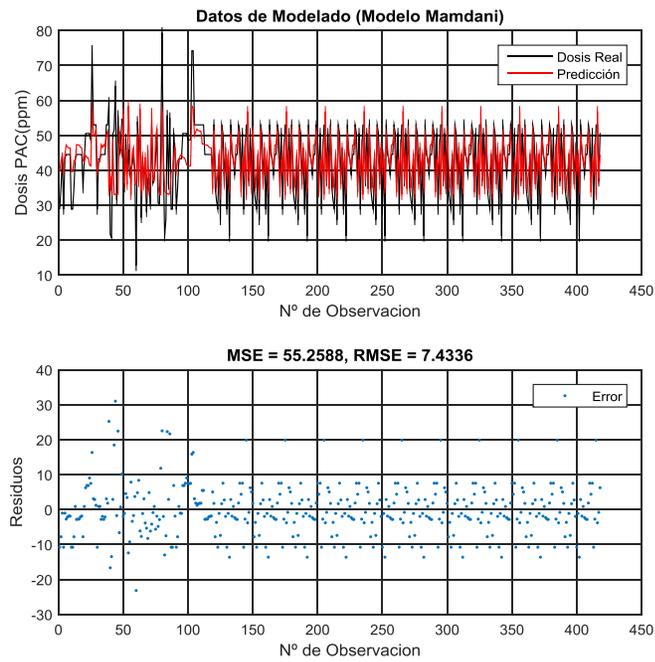


Figura 61. Desempeño del modelo Mamdani de PAC sobre los datos de modelado.

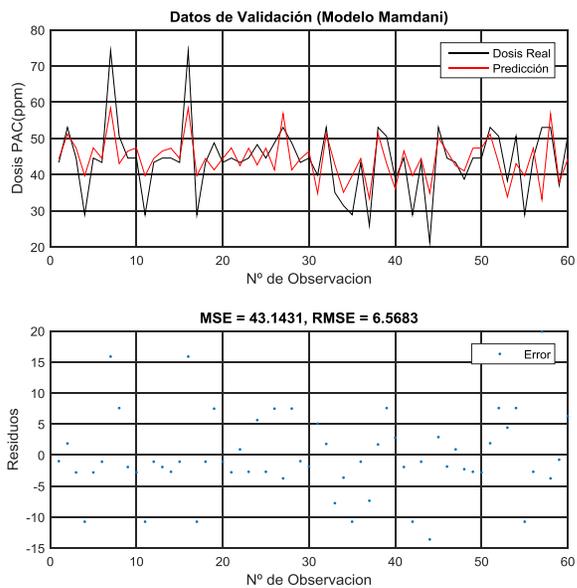


Figura 62. Desempeño del modelo Mamdani de PAC sobre los datos de validación.

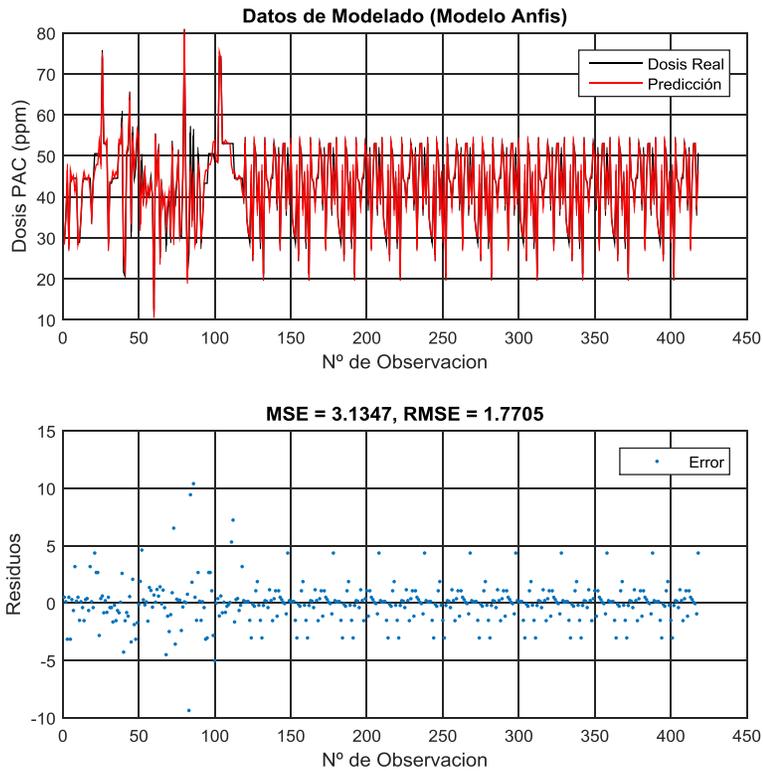


Figura 63. Desempeño del modelo ANFIS de PAC sobre los datos de modelado.

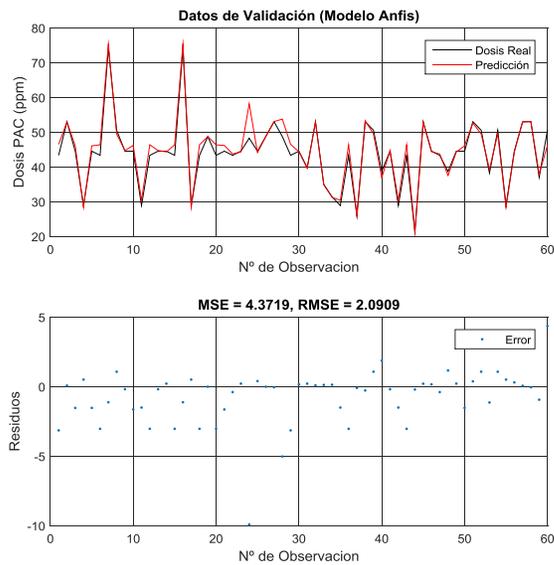


Figura 64. Desempeño del modelo ANFIS de PAC sobre los datos de validación.

VII.5.3 MODELOS “CLORO”

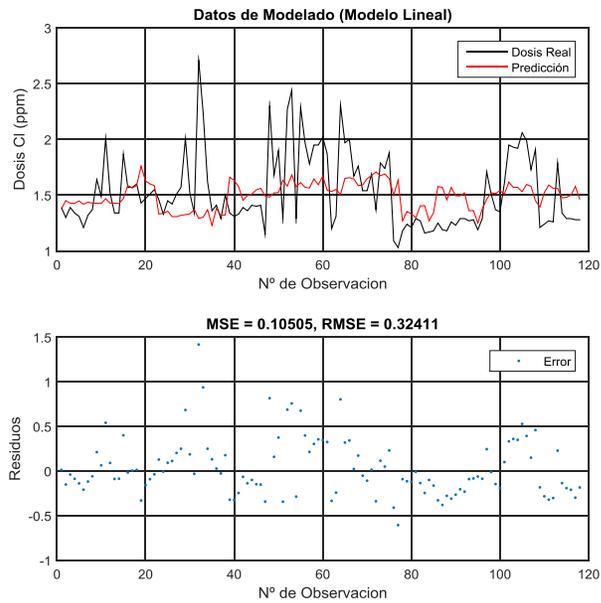


Figura 65. Desempeño del modelo lineal de Cloro sobre los datos de modelado.

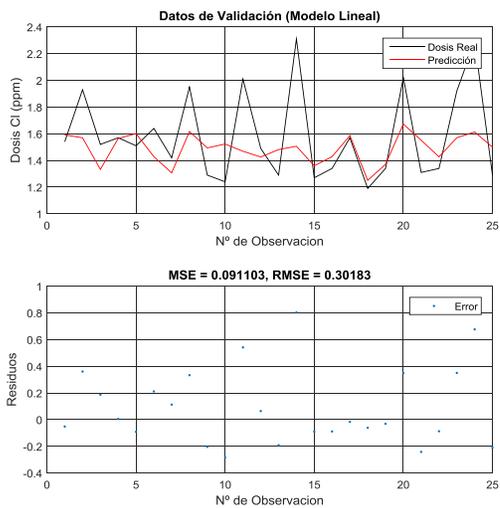


Figura 66. Desempeño del modelo lineal de Cloro sobre los datos de validación.

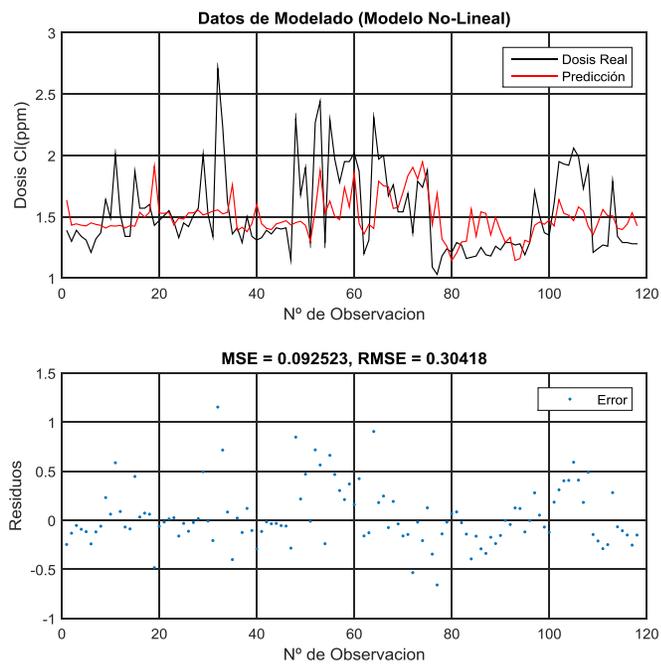


Figura 67. Desempeño del modelo no-lineal de Cloro sobre los datos de modelado.

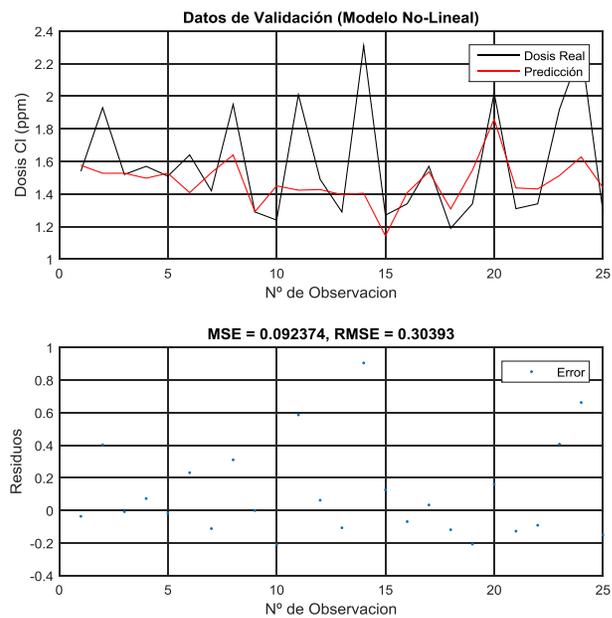


Figura 68. Desempeño del modelo no-lineal de Cloro sobre los datos de validación.

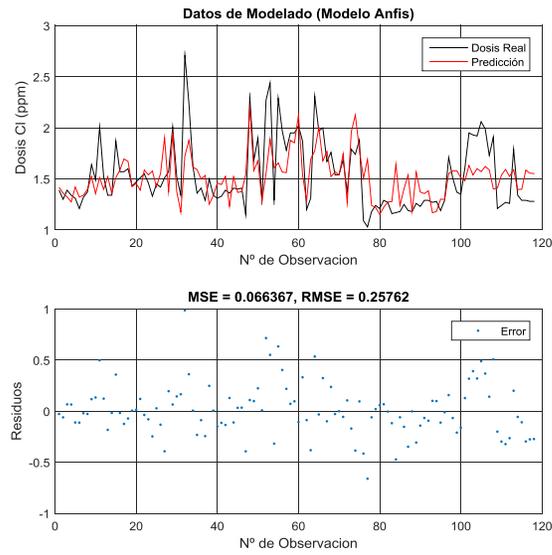


Figura 69. Desempeño del modelo ANFIS de Cloro sobre los datos de modelado.

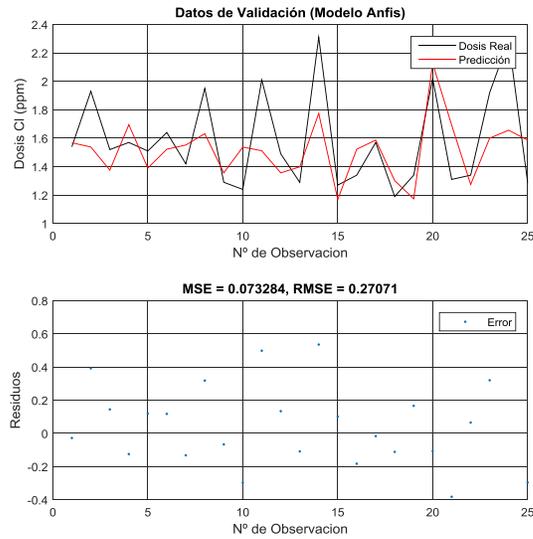


Figura 70. Desempeño del modelo ANFIS de Cloro sobre los datos de validación.