



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

**PREDICCIÓN DE CONTAMINANTES ATMOSFÉRICOS
MEDIANTE EL CLASIFICADOR GAMMA**

T E S I S

**QUE PARA OBTENER EL GRADO DE
MAESTRA EN CIENCIAS EN INGENIERÍA DE CÓMPUTO**

PRESENTA:

GUADALUPE DE LA LUZ SÁENZ MORALES

DIRECTORES DE TESIS:

**DR. CORNELIO YÁÑEZ MÁRQUEZ
DR. AMADEO JOSÉ ARGÜELLES CRUZ**



MÉXICO, D.F.

ENERO DE 2010



INSTITUTO POLITECNICO NACIONAL

SECRETARIA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 12:25 horas del día 17 del mes de noviembre de 2009 se reunieron los miembros de la Comisión Revisora de Tesis designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis de grado titulada:

“PREDICCIÓN DE CONTAMINANTES ATMOSFÉRICOS MEDIANTE EL CLASIFICADOR GAMMA”

SÁENZ
Apellido paterno

MORALES
materno

GUADALUPE DE LA LUZ
nombre(s)

Con registro:

A	0	8	0	3	1	1
---	---	---	---	---	---	---

aspirante al grado de: **MAESTRÍA EN CIENCIAS EN INGENIERÍA DE CÓMPUTO CON OPCIÓN EN SISTEMAS DIGITALES**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **SU APROBACIÓN DE LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Presidente

Secretario

Dr. Sergio Suárez Guerra

Dr. Oleksiy Pogrenbyak

**Primer vocal
(Director de tesis)**

**Segundo vocal
(Director de tesis)**

Dr. Cornelio Yáñez Márquez

Dr. Amadeo José Argüelles Cruz

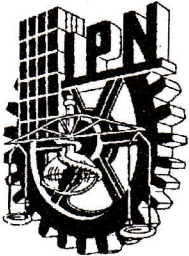
Tercer vocal

Dra. María Elena Acevedo Mosqueda

EL PRESIDENTE DEL COLEGIO



Dr. Jaime Álvarez Gallegos



CARTA CESIÓN DE DERECHOS

En la Ciudad de México el día 9 del mes Diciembre del año 2009, el (la) que suscribe Guadalupe de la Luz Sáenz Morales alumno (a) del Programa de Maestría en Ciencias en Ingeniería de Cómputo con opción en Sistemas Digitales con número de registro A080311, adscrito a Centro de Investigación en Computación, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de Dr. Cornelio Yáñez Márquez y Dr. Amadeo José Argüelles Cruz y cede los derechos del trabajo intitulado Predicción de Contaminantes Atmosféricos Mediante el Clasificador Gamma, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección gsaenza08@sagitario.cic.ipn.mx. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Guadalupe de la Luz Sáenz Morales.

Nombre y firma

Agradecimientos

Agradezco a Dios y a la Virgen de Guadalupe por darme la vida y la salud necesaria para culminar mis estudios.

Agradezco a mis papás, Juan José Sáenz González y Ma. de los Ángeles Morales Castillo, por todo su apoyo incondicional que me han brindado. Gracias a ellos he logrado finalizar esta etapa importante de mi vida. Gracias papá, porque con tu ejemplo he aprendido a terminar siempre todo lo que comienzo. Gracias mamá por tus consejos y porque siempre me alientas a seguir adelante, cambiando mi mentalidad para bien.

Gracias a mis hermanas, Ceci y Mari Jo por su apoyo, compañía y todos los momentos agradables que me brindaron, haciendo más grata mi estancia en México. Gracias Ceci porque con tu ejemplo he aprendido que nunca hay que conformarse con lo que hemos logrado, siempre hay que ver hacia el futuro. Gracias Mari Jo por tu optimismo que me ha hecho tomar la vida de un modo más alegre.

Agradezco al Dr. Cornelio Yáñez Márquez por todo el apoyo, comprensión y los consejos que me brindó durante la maestría. Porque con su ejemplo he aprendido que el trabajo en equipo es importante para avanzar en cualquier proyecto académico.

Agradezco también al Dr. Amadeo José Argüelles Cruz por su apoyo al haberme enseñado que la disciplina es necesaria para poder terminar mi trabajo de tesis.

Finalmente, agradezco a todos los compañeros del grupo Alfa-Beta por su amistad y todos los momentos agradables que pasé con ellos. Gracias Itzamá por la ayuda que me brindaste en el desarrollo de esta tesis. Gracias a Toño, Mario, Israel, Benjamín, Demetrio, Pancho, Edgar y los que me han faltado, por sus comentarios y opiniones que ayudaron a mejorar mi tesis.

Agradecimientos Institucionales

Agradezco al Instituto Politécnico Nacional y al Centro de Investigación en Computación (CIC). Por haberme brindado la oportunidad de estudiar la Maestría en Ciencias en Ingeniería de Cómputo.



Agradezco al grupo Alfa-Beta por haber tenido el honor de trabajar en él, aprendiendo en mi estancia muchas cosas didácticas, trabajo en equipo y por todo el apoyo que me brindaron en mi estancia de la maestría.



Finalmente, agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT), por el apoyo económico brindado durante mi estancia en el programa de maestría del CIC.

Número de registro de CONACyT : 257983

Resumen

En la presente tesis se aborda el tema de la predicción de contaminantes atmosféricos; más específicamente, de la ciudad de México y haciendo uso del clasificador Gamma. El clasificador Gamma es un modelo basado en las memorias asociativas Alfa-Beta, ambas creadas en el Centro de Investigación en Computación del IPN. Este clasificador entra en la categoría de enfoque asociativo de reconocimiento de patrones y ha demostrado su utilidad en diferentes aplicaciones.

Otra contribución de esta tesis es la propuesta de ciertas metodologías para realizar una predicción de contaminantes atmosféricos más precisa.

Se presentan varios experimentos de predicción de la concentración de 7 contaminantes de la Ciudad de México, así como un análisis de desempeño en la predicción de cada uno de ellos.

Con este trabajo de tesis se espera hacer una aportación importante en el campo de predicción de contaminantes atmosféricos. Un área que concierne no sólo a la Ciudad de México, sino a todo el mundo.

Abstract

In the current thesis the main topic is the prediction of environmental pollution; more specifically, in Mexico City and using the Gamma classifier. The Gamma classifier is a model based on the Alpha-Beta associative memories, both introduced at the Computing Research Center of the IPN. This classifier belongs to the associative approach pattern recognition and had proved its utility in different applications.

Other contributions of this work thesis are some methodologies proposed in order to do a better prediction.

Also, there are several experiments presented about concentration prediction's of seven kinds of pollutants in México City, as well as a performance analysis in prediction for each one.

With this work thesis we hope to do a very important contribution in the prediction of environmental pollution. A topic that is very important not only in Mexico City but also in all the world.

Índice general

1. Introducción	1
1.1. Antecedentes	1
1.2. Justificación	2
1.3. Objetivo	2
1.4. Contribuciones	3
1.5. Organización del Documento	3
2. Estado del Arte	4
2.1. Técnicas computacionales para la predicción...	5
2.2. Panorama nacional de la contaminación atmosférica	7
3. Materiales y Métodos	9
3.1. Los operadores α y β	9
3.2. El clasificador Gamma	10
3.2.1. El operador $v\beta$	10
3.2.2. Código binario Johnson-Möbius modificado	11
3.2.3. El operador Gamma de similitud	11
3.2.4. Algoritmo del clasificador Gamma	11
3.3. Sistema de Monitoreo Atmosférico (SIMAT)	13
3.3.1. Red Automática de Monitoreo Atmosférico (RAMA)	16
4. Modelo Propuesto	20
4.1. Conjunto fundamental, de prueba y clases	20
4.1.1. Conjunto fundamental	21
4.1.2. Conjunto de prueba	22
4.1.3. Clases	24
4.2. Predicción con el clasificador Gamma	24
4.3. Modelo propuesto	27

4.3.1. Ajuste de los datos obtenidos con el clasificador Gamma 27

5. Resultados y Discusión	32
5.1. Parámetros de prueba	32
5.1.1. Interpretación de las gráficas	35
5.2. Predicción del óxido de nitrógeno (NO _x)	35
5.3. Predicción del dióxido de azufre (SO ₂)	43
5.4. Predicción del monóxido de carbono (CO)	52
5.5. Predicción del ozono (O ₃)	61
5.6. Predicción del dióxido de nitrógeno (NO ₂)	63
5.7. Predicción de partículas menores a 10 micrómetros (PM ₁₀) . .	75
5.8. Predicción de partículas menores a 2.5 micrómetros (PM _{2.5}) .	85
5.9. Análisis de las predicciones realizadas...	87
5.10. Análisis de desempeño computacional	95
6. Conclusiones y Trabajo Futuro	97
6.1. Conclusiones	97
6.2. Trabajo futuro	98

Índice de cuadros

3.1.	Definición de los operadores Alfa y Beta.	10
3.2.	IMECA y sus implicaciones en la salud de la población.	17
3.3.	Índices de contaminación del NO _x , correspondiente al 1 de Enero del 2007 en las estaciones de Tacuba (TAC) y ENEP Acatlán (EAC).	19
4.1.	Ejemplo de organización del conjunto fundamental. (a) Datos originales. (b) Organización de los patrones.	22
4.2.	Organización del conjunto fundamental para la gráfica 4.1. (a) Datos originales. (b) Organización de los patrones.	23
4.3.	Ejemplo de organización del conjunto de prueba. (a) Datos originales. (b) Organización de los patrones.	24
4.4.	Ejemplo de organización de las clases. (a) Datos originales. (b) Organización de las clases.	25
5.1.	Relación de patrones por cada mes del año 2008. n es la dimensión de prueba.	33
5.2.	Predicción del NO _x para el año 2008.	36
5.3.	Predicción del NO _x para el año 2009, entrenando el clasificador con el año 2007.	39
5.4.	Predicción del NO _x para el año 2009, entrenando el clasificador con el año 2008.	42
5.5.	Porcentajes de mejora en la predicción del NO _x para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008	42
5.6.	Predicción de SO ₂ para el año 2008.	46
5.7.	Predicción de SO ₂ para el año 2008.	47

5.8. Predicción del SO ₂ para el año 2009, entrenando el clasificador con el año 2007.	48
5.9. Predicción del SO ₂ para el año 2009, entrenando el clasificador con el año 2008.	48
5.10. Porcentajes de mejora en la predicción del SO ₂ para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008	51
5.11. Predicción de CO para el año 2008.	56
5.12. Predicción de CO para el año 2008.	57
5.13. Predicción del CO para Enero y Marzo del 2009, entrenando el clasificador con el año 2007.	58
5.14. Predicción del CO para Enero y Marzo del 2009, entrenando el clasificador con el año 2008.	58
5.15. Porcentajes de mejora en la predicción del CO para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008	60
5.16. Predicción de O ₃ para el año 2008.	65
5.17. Predicción de O ₃ para el año 2008.	66
5.18. Predicción del O ₃ para el año 2009, entrenando el clasificador con el año 2007.	67
5.19. Predicción del O ₃ para el año 2009, entrenando el clasificador con el año 2008.	67
5.20. Porcentajes de mejora en la predicción del O ₃ para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008	69
5.21. Predicción de NO ₂ para el año 2008.	70
5.22. Predicción de NO ₂ para el año 2008.	71
5.23. Predicción del NO ₂ para el año 2009, entrenando el clasificador con el año 2007.	74
5.24. Predicción del NO ₂ para el año 2009, entrenando el clasificador con el año 2008.	74
5.25. Porcentajes de mejora en la predicción del NO ₂ para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008	76
5.26. Predicción de PM ₁₀ para el año 2008.	77
5.27. Predicción de PM ₁₀ para el año 2008.	78
5.28. Predicción de PM ₁₀ para el año 2009, entrenando el clasificador con el año 2007.	81

5.29. Predicción de PM_{10} para el año 2009, entrenando el clasificador con el año 2008.	81
5.30. Porcentajes de mejora en la predicción de PM_{10} para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008	84
5.31. Predicción de $PM_{2.5}$ para el año 2008.	89
5.32. Predicción de $PM_{2.5}$ para el año 2008.	90
5.33. Predicción de $PM_{2.5}$ para el año 2009, entrenando el clasificador con el año 2007.	91
5.34. Predicción de $PM_{2.5}$ para el año 2009, entrenando el clasificador con el año 2008.	91
5.35. Porcentajes de mejora en la predicción de $PM_{2.5}$ para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008	92
5.36. Comparación de las predicciones realizadas con otros métodos computacionales. <i>ND</i> , No Definido. <i>ppm</i> , partes por millón. $\mu g/m^3$, microgramos sobre metro cúbico. μm , micrómetros. . .	94
5.37. Tiempos de procesamiento del clasificador Gamma	96

Índice de figuras

3.1.	Diagrama de bloques del algoritmo del clasificador Gamma. . .	14
3.2.	Diagrama de bloques de la fase de aprendizaje del algoritmo del clasificador Gamma.	15
3.3.	Diagrama de bloques de la fase de clasificación del algoritmo del clasificador Gamma.	15
4.1.	Ejemplo ilustrativo.	23
4.2.	Predicción con el clasificador Gamma del contaminante monóxido de carbono (CO), correspondiente al mes de Marzo del 2008.	26
4.3.	Ejemplo ilustrativo. (a) Gráfica de los valores de los puntos tabulados en (b)	28
4.4.	Cálculo de pendientes de los datos de la figura 4.3 (b) . (a) Pendientes de obtenidas. (b) Valores de las pendientes calculadas.	29
5.1.	Predicción del contaminante óxido de nitrógeno (NOx), correspondiente al año 2008.	37
5.2.	Predicción del contaminante óxido de nitrógeno (NOx), correspondiente al mes de Marzo del 2008.	38
5.3.	Predicción del contaminante óxido de nitrógeno (NOx), correspondiente al mes de Enero del 2009, entrenando el clasificador con el año 2007 y con el año 2008.	40
5.4.	Predicción del contaminante óxido de nitrógeno (NOx), correspondiente al mes de Marzo del 2009, entrenando el clasificador con el año 2007 y con el año.	41
5.5.	Predicción del contaminante dióxido de azufre (SO ₂), correspondiente al año 2008.	44

5.6. Predicción del contaminante dióxido de azufre (SO ₂), correspondiente al mes de Enero del 2008.	45
5.7. Predicción del contaminante dióxido de azufre (SO ₂), correspondiente al mes de Enero del 2009, entrenando el clasificador con el año 2007 y con el año 2008.	49
5.8. Predicción del contaminante dióxido de azufre (SO ₂), correspondiente al mes de Marzo del 2009, entrenando el clasificador con el año 2007 y con el año 2008.	50
5.9. Predicción del contaminante monóxido de carbono (CO), correspondiente al año 2008, entrenando el clasificador con el año 2007.	52
5.10. Predicción del contaminante monóxido de carbono (CO), correspondiente al mes de Marzo del 2008.	53
5.11. Predicción del contaminante monóxido de carbono (CO), correspondiente al mes de Enero del 2009, entrenando el clasificador con el año 2007.	55
5.12. Predicción del contaminante monóxido de carbono (CO), correspondiente al mes de Marzo del 2009, entrenando el clasificador con el año 2008.	59
5.13. Predicción del contaminante ozono (O ₃), correspondiente al año 2008, entrenando el clasificador con el año 2007.	61
5.14. Predicción del contaminante ozono (O ₃), correspondiente al mes de Enero del 2008.	62
5.15. Predicción del contaminante ozono (O ₃), correspondiente al mes de Enero del 2009, entrenando el clasificador con los años 2007 y 2008.	64
5.16. Predicción del contaminante ozono (O ₃), correspondiente al mes de Marzo del 2009, entrenando el clasificador con los años 2007 y 2008.	68
5.17. Predicción del contaminante dióxido de nitrógeno (NO ₂), correspondiente al año del 2008.	72
5.18. Predicción del contaminante dióxido de nitrógeno (NO ₂), correspondiente al mes de Enero del 2008.	73
5.19. Predicción del contaminante dióxido de nitrógeno (NO ₂), correspondiente al mes de Enero del 2009.	75
5.20. Predicción del contaminante dióxido de nitrógeno (NO ₂), correspondiente mes de Marzo del 2009.	76
5.21. Predicción de partículas menores a diez micrómetros (PM ₁₀), correspondiente al año 2008.	79

5.22. Predicción de partículas menores a diez micrómetros (PM_{10}), correspondiente a Enero del 2008.	80
5.23. Predicción del contaminante PM_{10} , correspondiente al mes de Enero del 2009.	82
5.24. Predicción del contaminante (PM_{10}), correspondiente mes de Marzo del 2009.	83
5.25. Predicción de partículas menores a 2.5 micrómetros ($PM_{2.5}$), correspondiente al año 2008.	85
5.26. Predicción del contaminante $PM_{2.5}$, correspondiente al mes de Marzo del 2008.	86
5.27. Predicción del contaminante $PM_{2.5}$, correspondiente al mes de Enero del 2009.	87
5.28. Predicción del contaminante ($PM_{2.5}$), correspondiente mes de Marzo del 2009.	88

Capítulo 1

Introducción

En esta tesis se desarrolla una propuesta para la realización de un sistema de predicción de contaminantes atmosféricos mediante el uso del clasificador de patrones Gamma y las memorias asociativas Alfa-Beta.

1.1. Antecedentes

La contaminación atmosférica se puede definir como la presencia de elementos químicos en la atmósfera, que en cierto nivel de concentración pueden dañar a las plantas, los animales las personas y las construcciones. Esto trae como consecuencia a los humanos enfermedades, principalmente, de tipo respiratorias que pueden derivar problemas mayores que llegan a ser de tipo mortal [7].

La contaminación atmosférica es un problema que atañe no sólo a la ciudad de México, sino a todo el mundo. Diversos estudios se han realizado para atacar este problema y se han desarrollado técnicas computacionales con el fin de predecir este tipo de comportamiento en la contaminación atmosférica y tomar medidas preventivas al respecto.

Algunas de estas técnicas desarrolladas incluyen las redes neuronales [5], [23], [15], [8], [4], [12], [22], [14]; regresión lineal [31]; modelo de computación aritmética de tiempo de retardo cerebral (TD-CMAC), que es un derivado de las redes neuronales [20]; redes neuronales con lógica difusa [19]; lógica difusa

[11], [13], [18]; series de tiempo [10] y enfoque probabilístico [24], [32]. Todas ellas han demostrado ser competitivas dentro del rubro de la predicción de contaminación atmosférica.

Actualmente, una nueva técnica de predicción se ha estado desarrollando en el Centro de Investigación en Computación del Instituto Politécnico Nacional. Esta técnica es basada en el Clasificador Gamma [30], el cual se deriva de las memorias asociativas Alfa-Beta [16], y que ha demostrado ser bastante competitivo con las técnicas computacionales antes mencionadas [28].

1.2. Justificación

La contaminación atmosférica es uno de los principales problemas que sufre la ciudad de México; debido a ella la población ha desarrollado principalmente enfermedades de tipo respiratorias. El conocer de antemano cómo estarán los niveles de contaminación atmosférica en cierto día del año, permite al gobierno alertar a la población para que tome medidas preventivas al respecto. A nivel mundial y también nacional se han desarrollado modelos computacionales para llevar a cabo la predicción de contaminantes atmosféricos. Uno de estos modelos es el clasificador Gamma, basado en las memorias asociativas Alfa-Beta y que ha demostrado buenos resultados en este rubro. Debido a ello, se pretende utilizar este clasificador para mejorar la predicción de contaminantes atmosféricos de la ciudad de México.

1.3. Objetivo

El objetivo de esta tesis es desarrollar un sistema de predicción del indicador de la calidad del aire IMECA de la ciudad de México, mediante la aplicación del Clasificador Gamma y los modelos asociativos Alfa-Beta.

1.4. Contribuciones

Las contribuciones de este trabajo de tesis son las siguientes:

- Un sistema predictor de la calidad del aire IMECA de la ciudad de México.
- Una nueva técnica de predicción basada en el uso del clasificador Gamma y las memorias asociativas Alfa-Beta.

1.5. Organización del Documento

Esta tesis consta de 6 capítulos y una bibliografía. Una breve descripción de cada capítulo se da a continuación:

- *Capítulo 1:* Se da una breve introducción a la tesis, así como sus antecedentes, objetivo, justificación y contribuciones.
- *Capítulo 2:* Estado del arte. En este capítulo se describen detalladamente los antecedentes que llevaron a la realización de esta tesis.
- *Capítulo 3:* Se presentan los materiales y métodos que se utilizaron para desarrollar la tesis.
- *Capítulo 4:* En este capítulo se presenta el modelo propuesto para la predicción de contaminantes.
- *Capítulo 5:* Los resultados experimentales y la discusión de ellos se muestran en este capítulo.
- *Capítulo 6:* Se establecen las conclusiones del trabajo realizado así como sus perspectivas o trabajo futuro a realizar.

En la bibliografía se muestran las referencias mencionadas en la tesis. A fin de que el lector pueda consultarlas libremente en caso de querer conocer más acerca de ellas.

Capítulo 2

Estado del Arte

La atmósfera terrestre es una delgada capa de gases que está compuesta por los siguientes elementos y compuestos [6]:

- 78 % de Nitrógeno (N_2).
- 21 % de Oxígeno (O_2).
- 0.9 % de Argón (Ar).
- 0.06 % de otros gases raros, y
- 0.04 % de Bióxido de Carbono (CO_2).

La correcta correlación de estos elementos proporciona al planeta Tierra la protección de agentes dañinos como los rayos ultravioleta. También brinda al planeta cierta temperatura para hacer posible la existencia de vida en ella. Al momento de alterarse este equilibrio de elementos se presentan graves problemas como el aumento de temperatura en la Tierra. Esto se puede deber a la contaminación atmosférica.

La contaminación atmosférica se puede definir como la presencia en exceso de elementos y compuestos químicos en la atmósfera, que son distintos a los que componen la atmósfera terrestre. La presencia de estos elementos y compuestos trae como consecuencia a la población enfermedades, principalmente, de tipo respiratorias que pueden derivar problemas mayores que llegan a ser de tipo mortal [7].

2.1. Técnicas computacionales para la predicción de contaminantes atmosféricos

La contaminación atmosférica es un problema que atañe no sólo a la ciudad de México, sino a todo el mundo. Diversos estudios se han realizado para atacar este problema y se han desarrollado técnicas computacionales con el fin de predecir este tipo de comportamiento en la contaminación atmosférica y tomar medidas preventivas al respecto.

Dentro de estas técnicas computacionales podemos mencionar las siguientes:

- *Regresión Lineal.*- Método matemático que involucra variables dependientes con independientes y una variable aleatoria. En el trabajo realizado en [31] se desarrolla un modelo de predicción de contaminación atmosférica basado en regresión lineal en el cual los resultados obtenidos no son muy precisos, esto debido a la falta de variables a manejar dentro del modelo.
- *Redes Neuronales.*- Basados en el modelo de la neurona del cerebro humano. Diversas investigaciones se han realizado utilizando esta técnica para la predicción de contaminantes atmosféricos. Algunos agregan un preprocesamiento, antes de utilizar las redes neuronales con el fin de obtener mejores resultados [15], [4]. Otros, juegan con el modelado de las redes neuronales obteniendo también buenos resultados [14], [22], [12], [8], [5] y [23].
- *TD-CMAC.*- Sus siglas significan: Time Delay Cerebellar Model Arithmetic Computer. Es un modelo derivado de las redes neuronales que tiene la ventaja de necesitar menos memoria para aprender que las redes neuronales. En [20] se utiliza esta técnica para predecir contaminantes atmosféricos obteniendo resultados similares a los modelos desarrollados con redes neuronales.
- *Lógica Difusa.*- Es una técnica computacional que sirve para la modelación de sistemas inciertos con el fin de tomar decisiones en ausencia de información completa y precisa [3]. La aplicación de esta técnica

para la predicción de contaminantes atmosféricos se puede observar en [11] y [18], el primero utiliza un preprocesamiento con regresión lineal y el segundo un preprocesamiento con algoritmos genéticos. Esto con el fin de obtener una mejor predicción.

- *Redes Neuronales con Lógica Difusa.*- Es una combinación de las redes neuronales con lógica difusa. Esto da como resultado un mejor desempeño de la aplicación en problemas como la predicción de contaminantes atmosféricos. Un ejemplo de estos trabajos se encuentran en [19] y [13].
- *Series de Tiempo.*- Son secuencias, discretas o continuas, de datos cuantitativos asignados para momentos específicos en el tiempo y estudiados con respecto a las estadísticas de su distribución en el tiempo [27]. Este método utilizado en predicción no resulta muy eficiente, como es mostrado en [10], donde realiza un predictor de contaminación de Ozono (O_3) y lo compara con un modelo de red neuronal, resultando mejor la red neuronal para la realización de la predicción.
- *Enfoque Probabilístico.*- Como su nombre lo dice, utiliza métodos basados en la probabilidad y estadística. En [24] y [32] se utiliza este enfoque para la predicción de contaminantes atmosféricos. Más específicamente en [32] se realiza un modelo de árboles probabilísticos para la predicción de Ozono (O_3), resultando éste superior a los otros modelos basados en la probabilidad y la estadística.
- *Clasificador Gamma.*- Modelo basado en las memorias asociativas Alfa-Beta [16]. Es un clasificador de alto desempeño [30] que ha demostrado robustez en el campo de predicción de contaminantes atmosféricos, en especial para el óxido de nitrógeno (NO_x) [29] y el dióxido de azufre (SO_2) [28], donde mostraba errores de predicción en los cambios drásticos de niveles altos a bajos en comparación con los datos reales. Una explicación más profunda del clasificador Gamma se dará en el capítulo 3.

En la tarea de desarrollar un predictor de contaminantes atmosféricos preciso, se han observado grandes esfuerzos realizados a nivel mundial, como los modelos descritos anteriormente.

2.2. Panorama nacional de la contaminación atmosférica

El gobierno de México posee un organismo dedicado al cuidado del medio ambiente. Este es mejor conocido como la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). Una de sus principales actividades es el vigilar el cumplimiento de protocolos para la protección del medio ambiente. Entre ellos, aquellos que controlen la emisión de contaminantes atmosféricos que traen como consecuencia el incremento de gases de efecto invernadero. Como es sabido, estos gases contribuyen en el incremento de temperatura del planeta ocasionando con ello cambios climáticos que afectan el entorno de supervivencia de todo ser vivo. Algunos antecedentes que proporciona la SEMARNAT respecto al problema de contaminación ambiental son descritos a continuación [6].

En 1997 México firma el Protocolo de Kioto [17]. Este protocolo establece principalmente que los países participantes adquieren el compromiso de reducir sus emisiones de dióxido de carbono (CO_2) en un 5.2% en comparación con los niveles que se tenían en 1990, todo esto en el período 2008-2012. Posteriormente, en el año 2000 el Senado de la República ratificó el protocolo de Kioto. En el año 2000 México se posicionó a nivel mundial en el lugar 93 con respecto a las emisiones de CO_2 , 6.4 toneladas emitidas por habitante, un poco abajo del promedio mundial que fue de 6.55 toneladas. Con estos antecedentes, en el año 2005 se crea la Comisión Intersecretarial de Cambio Climático (CICC) cuyo propósito es coordinar la formulación de políticas de acción climática [6].

Por otro lado, la Secretaría de Salud es un organismo gubernamental dedicado entre sus principales funciones a la prevención y control de enfermedades en la población del país. En su Programa Nacional de Salud 2007-2012 [7] afirma que la cuarta causa de mortalidad en mujeres y la quinta de hombres en el país es debida a las enfermedades de tipo EPOC (Enfermedad Pulmonar Obstructiva Crónica). La EPOC es una enfermedad que se caracteriza por la obstrucción de las vías respiratorias que se puede manifestar mediante bronquitis crónica o enfisema. Esto puede derivar en problemas de salud mayores como cáncer de pulmón o enfermedades del corazón. Uno de los factores para desarrollar estas enfermedades es la contaminación atmosférica.

Las partículas contaminantes suspendidas en el aire se relacionan con enfermedades como el cáncer pulmonar y del corazón. Mientras que el Ozono (O_3) es relacionado con afecciones agudas del tracto respiratorio. Pese a que las concentraciones de Ozono han ido descendiendo, en la Ciudad de México se han registrado más de 200 días al año en que se rebasan estas. Mientras que en Guadalajara estos índices se rebasan en más de 50 días al año [7].

A nivel local, el gobierno del Distrito Federal posee un sistema de monitoreo atmosférico, mejor conocido como SIMAT. Esta red se encarga de capturar los niveles de contaminación del Valle de México y el Distrito Federal para determinar si existe o no contingencia ambiental en el área que se esté monitoreando [2]. Se verá más a detalle este sistema en el capítulo 3 de esta tesis.

Algunas medidas que ha tomado el gobierno del Distrito Federal para la reducción de contaminantes atmosféricos son: la implementación de un programa de circulación vehicular, la prohibición de la quema de cualquier residuo sólido al aire libre, la suspensión de actividades públicas que obstruyan el tránsito vehicular como el bacheo o pintado de banquetas, entre otras [21]. De igual manera, se implementó un programa para mejorar la calidad del aire en la zona del Valle de México. Este consiste en un plan de 8 años para reducir las emisiones de partículas contaminantes. El programa es mejor conocido como PROAIRE [1].

En el área de investigación científica nacional se han realizado algunas contribuciones para la predicción de contaminantes atmosféricos. Más específicamente, en la predicción del ozono de la ciudad de México. Algunos de ellos han sido de enfoque de razonamiento fuzzy inductivo (FIR) [18] y [25]. Otros utilizan el enfoque probabilístico [24]. Todos ellos han contribuido en el esfuerzo de aportar una solución al problema de la contaminación atmosférica de la Ciudad de México.

Hasta la fecha, lo anteriormente mencionado es parte de lo que se ha estado haciendo en el país para mejorar la calidad del aire y disminuir la contaminación atmosférica.

Capítulo 3

Materiales y Métodos

En este capítulo se presentan en diferentes secciones los materiales y métodos necesarios para el desarrollo de la tesis.

En la primera sección se dan a conocer los operadores α y β , fundamentales para la implementación del clasificador Gamma. En la segunda sección se describe el clasificador Gamma y finalmente en la tercera sección se menciona en forma más detallada la base de datos contenida en la Red Automática de Monitoreo Atmosférico (RAMA), tomada del Sistema de Monitoreo Atmosférico (SIMAT) del D.F.

3.1. Los operadores α y β

Dentro del campo de reconocimiento de patrones se encuentra un área de investigación llamada memorias asociativas; una de estas clases de memorias son las memorias asociativas *alfa-beta*, que contienen dos operadores, como su nombre lo dice, *alfa* y *beta*.

Los operadores α y β creados en el Centro de Investigación en Computación [16], están basados en operaciones binarias entre dos conjuntos A y B definidos de la siguiente manera:

$$A = \{0, 1\} \quad \text{y} \quad B = \{00, 01, 10\}$$

El operador α es utilizado en la fase de aprendizaje y el operador β en la fase de recuperación. El cuadro 3.1 muestra la representación de estos dos

operadores dados los conjuntos A y B arriba mencionados. Una descripción más detallada de estos operadores se puede encontrar en [16].

$\alpha : A \times A \rightarrow B$			$\beta : B \times A \rightarrow A$		
x	y	$\alpha(x, y)$	x	y	$\beta(x, y)$
0	0	01	00	0	0
0	1	00	00	1	0
1	0	10	01	0	0
1	1	01	01	1	1
			10	0	1
			10	1	1

Cuadro 3.1: Definición de los operadores Alfa y Beta.

3.2. El clasificador Gamma

El clasificador Gamma es un clasificador de patrones de alto desempeño basado en las memorias asociativas *alfa-beta* y cuyo principal operador es el operador de similitud γ . El clasificador Gamma trabaja en conjunto con los operadores α y β , el operador $v\beta$ y con patrones codificados con el método de Johnson-Möbius modificado. En las siguientes subsecciones se describirán brevemente estas herramientas útiles para trabajar con el clasificador Gamma y finalmente se describirá el algoritmo del clasificador. Información más detallada acerca de este clasificador y sus componentes se puede encontrar en [30].

3.2.1. El operador $v\beta$

Dados el conjunto $A = \{0, 1\}$, un número $n \in \mathbb{Z}^+$ y $x \in A^n$ un vector binario de dimensión n , con la i -ésima componente representada por x_i . El operador $v\beta(x)$ se define de la siguiente manera:

$$v\beta(x) = \sum_{i=1}^n \beta(x_i, x_i)$$

3.2.2. Código binario Johnson-Möbius modificado

El código de Johnson-Möbius modificado convierte un conjunto de números enteros en sus representaciones binarias mediante los siguientes pasos:

1. Sea un conjunto de números reales $\{r_1, r_2, \dots, r_i, \dots, r_n\}$ donde n es un número entero positivo fijo.
2. Restar el mínimo encontrado en el conjunto a cada uno de sus componentes, con el fin de quedar solamente con enteros positivos en el conjunto.
3. Escalar los números del conjunto transformado a cierto número fijo de decimales.
4. Concatenar $e_m - e_j$ ceros por la derecha con e_j unos, donde e_m es el número mayor en el conjunto y e_j es el número a codificar.

3.2.3. El operador Gamma de similitud

Sean el conjunto $A = \{0, 1\}$, dos números $n, m \in \mathbb{Z}^+$, $n \leq m$, $x \in A^n$ y $y \in A^m$ dos vectores binarios, n -dimensional y m -dimensional, respectivamente, con la i -ésima componente representada por x_i y y_i respectivamente; y θ un número entero no negativo. El operador gamma de similitud $\gamma_g(x, y, \theta)$ se calcula así:

$$\gamma_g = \begin{cases} 1 & \text{si } m - v\beta[\alpha(x, y) \bmod 2] \leq \theta \\ 0 & \text{en otro caso} \end{cases}$$

3.2.4. Algoritmo del clasificador Gamma

Sean los números $k, m, n, p \in \mathbb{Z}^+$, $\{x^\mu | \mu = 1, 2, \dots, p\}$ el conjunto fundamental de patrones de cardinalidad p , donde $\forall \mu x^\mu \in \mathbb{R}^n$ y $y \in \mathbb{R}^n$ es un vector real n -dimensional a ser clasificado. Se asume que el conjunto fundamental está particionado en m clases diferentes, donde cada clase tiene cardinalidad k_i , $i = 1, 2, \dots, m$, por lo que

$$\sum_{i=1}^m k_i = p$$

Para clasificar el patrón y se realiza lo siguiente:

1. Codificar cada componente de cada patrón del conjunto fundamental con el código Johnson-Möbius modificado, obteniéndose un valor

$$e_m = \bigvee_{i=1}^p x_j^i$$

por cada componente. Así, la componente x_j^i se transforma en un vector binario de dimensión $e_m(j)$.

2. Calcular el parámetro de paro

$$\rho = \bigwedge_{j=1}^n e_m(j)$$

3. Codificar cada componente del patrón a clasificar con el código Johnson-Möbius modificado, utilizando las mismas condiciones que se utilizaron para codificar las componentes de los patrones fundamentales. En caso de que alguna componente del patrón a clasificar sea mayor al e_m correspondiente ($y_\xi > e_m(\xi)$), igualar esa componente a $e_m(\xi)$ y guardar su valor anterior en la variable $nmax_\xi$.
4. Realizar una transformación de índices en los patrones del conjunto fundamental, de manera que el índice único que tenía un patrón originalmente en el conjunto fundamental, por ejemplo x^μ , se convierta en dos índices: uno para la clase (por ejemplo la clase i) y otro para el orden que le corresponde a ese patrón dentro de la clase (por ejemplo ω). Bajo estas condiciones ejemplificadas, la notación para el patrón x^μ será ahora, con la transformación, $x^{i\omega}$. Lo anterior se realiza para todos los patrones del conjunto fundamental.
5. Inicializar θ a 0.
6. Realizar la operación $\gamma_g(x_j^{i\omega}, y_j, \theta)$ para cada clase y para cada componente de cada uno de los patrones fundamentales que corresponden a esa clase, y del patrón a clasificar, considerándose $nmax_\xi$ como la dimensión del patrón binario y_ξ .
7. Calcular la suma ponderada c_i de los resultados obtenidos en el paso 6, para cada clase $i = 1, 2, \dots, m$:

$$c_i = \frac{\sum_{\omega=1}^{k_i} \sum_{j=1}^n \gamma_g(x_j^{i\omega}, y_j, \theta)}{k_i}$$

8. Si existe más de un máximo entre las sumas ponderadas por clase, incrementar θ en 1 y repetir los pasos 6 y 7 hasta que exista un máximo único, o se cumpla con la condición de paro $\theta \geq \rho$.
9. Si existe un máximo único, asignar al patrón a clasificar la clase correspondiente a ese máximo:

$$C_y = C_j \text{ tal que } \bigvee_{i=1}^m c_i = c_j$$

10. En caso contrario: si λ es el índice más pequeño de clase que corresponde a uno de los máximos, asignar al patrón a clasificar la clase C_λ .

El algoritmo descrito se muestra en la figura 3.1. El algoritmo del clasificador Gamma se puede dividir en dos fases, una de aprendizaje y la otra de recuperación. En la fase de aprendizaje se codifican los patrones del conjunto fundamental y del conjunto de prueba con el código Johnson-Möbius modificado, se calcula el parámetro de paro ρ , se transforman los índices de los patrones fundamentales, se inicializa θ en cero y se realiza la operación $\alpha(x_j^{i\omega}, y_j) \bmod 2$. Este algoritmo se ilustra en la figura 3.2. La fase de recuperación o de clasificación consta de calcular la diferencia entre la dimensión n de los patrones y el operador $v\beta$ de $\alpha(x_j^{i\omega}, y) \bmod 2$, para después calcular el operador Gamma de similitud y continuar con los pasos 7 a 10. Esta fase se ilustra en el diagrama de la figura 3.3.

3.3. Sistema de Monitoreo Atmosférico (SIMAT)

La Secretaría del Medio Ambiente del Distrito Federal cuenta con un sistema de monitoreo atmosférico (SIMAT) [2], que se encarga de medir la calidad del aire de la ciudad de México. Gracias a esto se pueden tomar medidas preventivas para la población y con ello evitar la proliferación de enfermedades respiratorias principalmente.

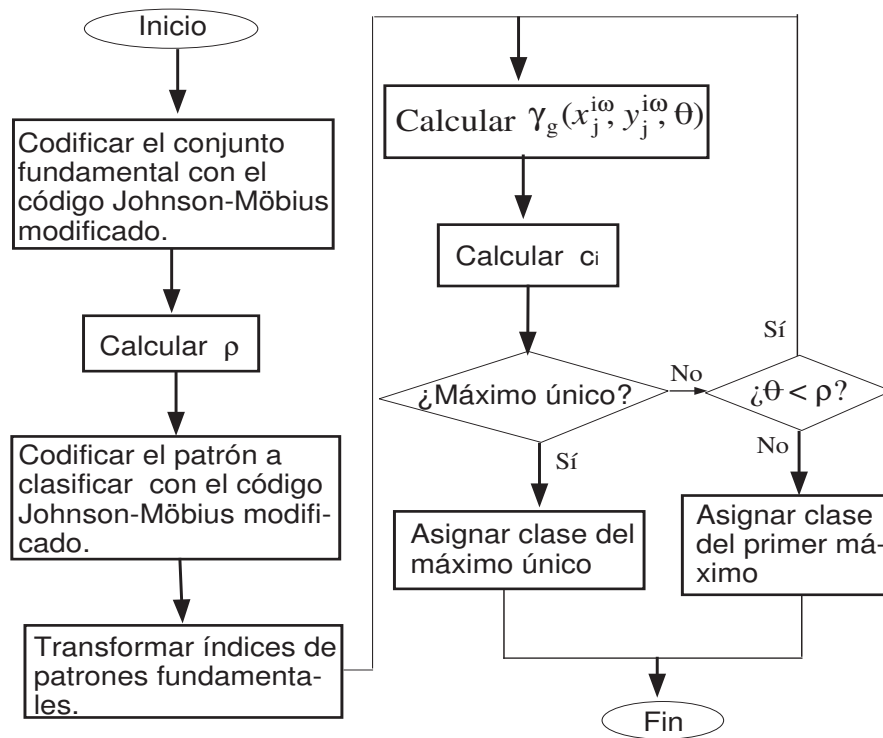


Figura 3.1: Diagrama de bloques del algoritmo del clasificador Gamma.

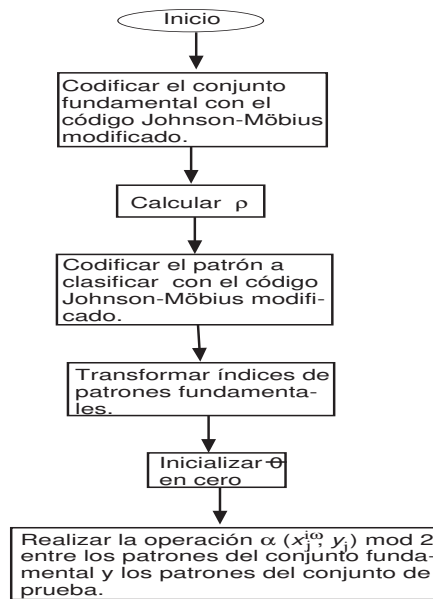


Figura 3.2: Diagrama de bloques de la fase de aprendizaje del algoritmo del clasificador Gamma.

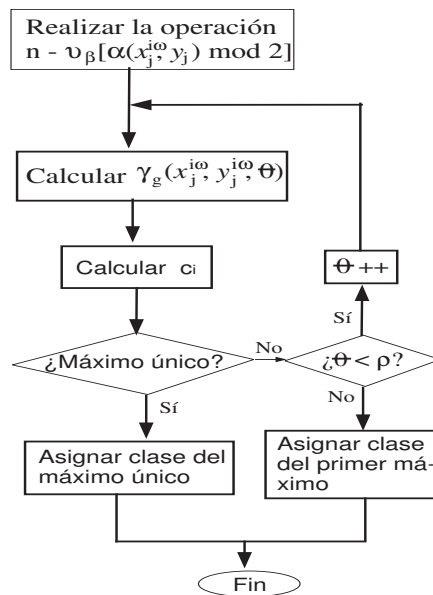


Figura 3.3: Diagrama de bloques de la fase de clasificación del algoritmo del clasificador Gamma.

Una de las tareas principales del SIMAT es el difundir los niveles IMECA existentes en la ciudad de México. El IMECA (Índice Metropolitano de la Calidad del Aire) mide a cada hora cinco diferentes tipos de contaminantes:

- Ozono (O_3)
- Partículas menores a diez micrómetros (PM_{10})
- Dióxido de azufre (SO_2)
- Dióxido de nitrógeno (NO_2)
- Monóxido de carbono (CO)

Tomando en cuenta los contaminantes mencionados y de acuerdo con el cuadro 3.2, dentro de la página oficial del SIMAT [2], a cada hora se publica la calidad del aire existente en diversas zonas de la ciudad de México.

Actualmente el SIMAT está conformado por la Red Automática de Monitoreo Atmosférico (RAMA), la Red Manual de Monitoreo Atmosférico (REDMA), la Red de Depósito Atmosférico (REDDA) y la Red de Meteorología y Radiación Solar (REDMET). Así mismo, cuenta con un laboratorio móvil de monitoreo atmosférico, un laboratorio de estándares, un laboratorio de análisis ambiental y un centro de información de la calidad del aire (CICA).

En la siguiente sección se describirá detalladamente la RAMA, ya que es el objeto de estudio para el desarrollo de esta tesis.

3.3.1. Red Automática de Monitoreo Atmosférico (RAMA)

La Red Automática de Monitoreo Atmosférico (RAMA) se encarga de medir en forma permanente y continua los siguientes tipos de contaminantes [2]:

- Ozono (O_3)
- Partículas menores a diez micrómetros (PM_{10})

IMECA	Color	Calidad del aire	Recomendaciones y riesgos
0-50	verde	Buena	Adecuada para llevar a cabo actividades al aire libre.
51-100	amarillo	Regular	Posibles molestias en niños, adultos mayores y personas con enfermedades respiratorias o cardiovasculares.
101-150	naranja	Mala	Evite las actividades al aire libre. Posibles efectos adversos a la salud, en particular niños, adultos mayores y personas con enfermedades respiratorias o cardiovasculares.
151-200	rojo	Muy mala	Evite salir de casa y mantenga las ventanas cerradas. Efectos adversos a la salud de la población en general. Se agravan los síntomas en niños, adultos mayores y personas con enfermedades respiratorias o cardiovasculares.
>200	morado	Extremadamente mala	Proteja su salud, no salga de casa o permanezca en lugares cerrados. Efectos graves a la salud de la población en general. Se pueden presentar complicaciones en niños, adultos mayores y personas con enfermedades respiratorias o cardiovasculares.

Cuadro 3.2: IMECA y sus implicaciones en la salud de la población.

- Dióxido de azufre (SO_2)
- Dióxido de nitrógeno (NO_2)
- Monóxido de carbono (CO)
- Partículas menores a 2.5 micrómetros ($PM_{2.5}$)
- Óxido de nitrógeno (NO_x)

Estas mediciones son importantes en la evaluación y difusión del índice metropolitano de la calidad del aire (IMECA) y la rapidez con que se envían y reciben permite alertar a la población en caso de presentarse alguna contingencia ambiental.

La RAMA está conformada por 34 estaciones de monitoreo ubicados en diferentes puntos del Valle de México. 21 estaciones se encuentran en el Distrito Federal y 13 estaciones en el Estado de México. Los criterios tomados para la colocación estratégica de estas estaciones se basan en la densidad de población, la distribución de fuentes de emisión, la meteorología y la topografía.

Cada uno de los contaminantes evaluados en cada estación se guardan dentro de una base de datos, existente desde el año de 1986. Los datos se muestran organizados por año, fecha, hora, estación y el valor de la concentración. Algunas concentraciones se muestran en partes por millón (ppm) y otras en microgramos por metro cúbico ($\mu g/m^3$). Un ejemplo de la organización de los datos se muestra en el cuadro 3.3. En éste aparecen los valores obtenidos del contaminante óxido de nitrógeno (NO_x) en las estaciones de Tacuba (TAC), ubicada en la delegación Miguel Hidalgo del D.F. y ENEP Acatlán (EAC), ubicada en el municipio de Naucalpan, Edo. de México, durante el primer día de Enero del 2007. Los datos están dados en partes por millón (ppm). Cabe aclarar que estos valores son solamente índices de concentración, es decir, no se toman en cuenta factores como viento, humedad, entre otros.

Otra característica que se puede observar de la RAMA es su historial técnico, donde se presentan en forma anual las estaciones que han sido dadas de alta y baja, así como los tipos de contaminantes que miden.

Fecha	Hora	TAC	EAC
01/01/2007	1	0.045	0.033
01/01/2007	2	0.034	0.028
01/01/2007	3	0.028	0.024
01/01/2007	4	0.023	0.016
01/01/2007	5	0.016	0.018
01/01/2007	6	0.017	0.015
01/01/2007	7	0.016	0.012
01/01/2007	8	0.017	0.015
01/01/2007	9	0.017	0.015
01/01/2007	10	0.021	0.016
01/01/2007	11	0.024	0.019
01/01/2007	12	0.024	0.026
01/01/2007	13	0.020	0.029
01/01/2007	14	0.030	0.029
01/01/2007	15	0.022	0.033
01/01/2007	16	0.016	0.030
01/01/2007	18	0.016	0.022
01/01/2007	19	0.018	0.025
01/01/2007	20	0.027	0.023
01/01/2007	21	0.036	0.032
01/01/2007	22	0.042	0.031
01/01/2007	23	0.036	0.044
01/01/2007	24	0.031	0.055

Cuadro 3.3: Índices de contaminación del NO_x, correspondiente al 1 de Enero del 2007 en las estaciones de Tacuba (TAC) y ENEP Acatlán (EAC).

En este capítulo se han presentado de manera detallada las herramientas que serán útiles para el desarrollo de esta tesis. En el siguiente capítulo se mostrará la aplicación de estos conceptos así como algunas propuestas para el desarrollo de este trabajo.

Capítulo 4

Modelo Propuesto

En este capítulo se dará a conocer la metodología para llevar a cabo la predicción de contaminantes atmosféricos. Las secciones 4.1 y 4.2 son metodologías que la autora ya había realizado [28] y [29]. En la primera sección hablaremos del arreglo de los datos de entrenamiento y prueba. En la segunda sección se realizará una inspección más cercana al problema a resolver en la predicción. Finalmente, en la tercera sección se planteará el algoritmo para realizar la predicción de contaminantes atmosféricos.

Como se mencionó en el párrafo anterior, existe una metodología para realizar la predicción de contaminantes atmosféricos basada en el arreglo de los datos en patrones contenidos en un conjunto fundamental, uno de prueba y otro de clases para posteriormente, implementar el clasificador Gamma y realizar la predicción. En las siguientes dos secciones hablaremos de esta metodología.

4.1. Conjunto fundamental, de prueba y clases

En el capítulo 3 se mencionó la organización de la red automática de monitoreo atmosférico (RAMA), más específicamente, la base de datos de la RAMA. Esta base de datos es de gran importancia en nuestro estudio para llevar a cabo la predicción de contaminantes atmosféricos. Como se recordará, los valores monitoreados de cada contaminante se encuentran respaldados en

archivos electrónicos de forma anual, donde en cada año se observan los valores medidos en cada estación a cada hora durante los 365 días del año. Estos valores servirán para la organización de los datos necesarios para llevar a cabo la predicción.

Los datos se tomarán de una sola estación y se organizarán dentro de k patrones representados de la siguiente forma

$$\mathbf{x}^k = \begin{pmatrix} x_1^k \\ x_2^k \\ x_3^k \\ \vdots \\ x_n^k \end{pmatrix}$$

donde

x^k Es el k -ésimo patrón.

x_n^k Es el n -ésimo valor del patrón x^k

Con esta explicación podemos definir el conjunto fundamental y el conjunto de prueba.

4.1.1. Conjunto fundamental

El conjunto fundamental es el conjunto de patrones de dimensión n , que contendrán los valores medidos de un contaminante durante cierto periodo. Por dimensión se entiende como el total de datos que contendrá cada patrón. Por ejemplo, para un conjunto de datos mostrados en el cuadro 4.1 (a). Si se desea que la dimensión del conjunto fundamental sea de 3, los patrones contendrán los valores asociados de tal manera que el primer patrón contendrá los valores de los índices 0 a 2, el segundo patrón contendrá los valores de los índices 1 a 3, el tercer patrón contendrá los valores de los índices 2 a 4 y así sucesivamente hasta llegar al último patrón que contendrá los valores de los índices x_{n-2} a x_n . Como se muestra en el cuadro 4.1 (b).

Para ilustrar lo anterior, consideremos un ejemplo numérico. Sea un conjunto de puntos dados graficados como en la figura 4.1. Si se desean organizar

índice	valor		
0	x_0	$\mathbf{x}^1 = \begin{pmatrix} x_0^1 \\ x_1^1 \\ x_2^1 \end{pmatrix}$	$\mathbf{x}^2 = \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \end{pmatrix}$
1	x_1		
2	x_2		
3	x_3		
4	x_4		
\vdots	\vdots		
$k-2$	x_{n-2}	$\mathbf{x}^3 = \begin{pmatrix} x_2^3 \\ x_3^3 \\ x_4^3 \end{pmatrix} \cdots$	$\mathbf{x}^k = \begin{pmatrix} x_{n-2}^k \\ x_{n-1}^k \\ x_n^k \end{pmatrix}$
$k-1$	x_{n-1}		
k	x_n		

(a)
(b)

Cuadro 4.1: Ejemplo de organización del conjunto fundamental. (a) Datos originales. (b) Organización de los patrones.

patrones cuya dimensión sea $n = 3$, entonces los patrones estarán organizados como en el cuadro 4.2. El total de patrones creados fue de 3, al momento de entrenar el clasificador y realizar las pruebas el primer patrón creado, x^1 , se mantendrá igual ya que éste no cuenta con algún valor anterior para compararse. Por lo tanto, el total de patrones del conjunto fundamental se reduce a 2. Expresado de otra manera:

$$Totaldepatrones = m - n$$

donde m es el número de elementos del conjunto fundamental, en este ejemplo, $m = 5$.

4.1.2. Conjunto de prueba

El conjunto de prueba es el conjunto de patrones de dimensión igual al del conjunto fundamental y que contendrán los valores medidos del mismo contaminante del conjunto fundamental en un periodo distinto a este. Los patrones se arreglan de igual manera que los del conjunto fundamental, cada patrón del conjunto de prueba se denotará por \tilde{x} . Tomaremos el mismo ejemplo del conjunto fundamental para organizar los patrones del conjunto

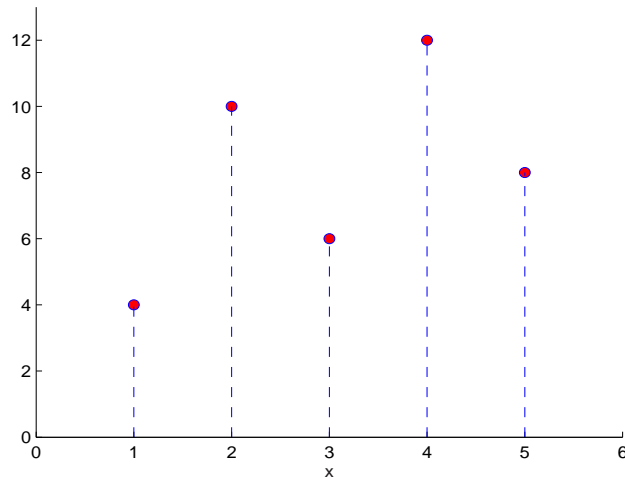


Figura 4.1: Ejemplo ilustrativo.

x	valor
1	4
2	10
3	6
4	12
5	8

(a)

$$\mathbf{x}^1 = \begin{pmatrix} 4 \\ 10 \\ 6 \end{pmatrix} \quad \mathbf{x}^2 = \begin{pmatrix} 10 \\ 6 \\ 12 \end{pmatrix}$$

(b)

$$\mathbf{x}^3 = \begin{pmatrix} 6 \\ 12 \\ 8 \end{pmatrix}$$

Cuadro 4.2: Organización del conjunto fundamental para la gráfica 4.1. (a) Datos originales. (b) Organización de los patrones.

de prueba con una dimensión $n = 3$, los datos se toman como se muestra en el cuadro 4.4.

índice	valor		
0	\tilde{x}_0	$\tilde{\mathbf{x}}^1 = \begin{pmatrix} \tilde{x}_0^1 \\ \tilde{x}_1^1 \\ \tilde{x}_2^1 \end{pmatrix}$	$\tilde{\mathbf{x}}^2 = \begin{pmatrix} \tilde{x}_1^2 \\ \tilde{x}_2^2 \\ \tilde{x}_3^2 \end{pmatrix}$
1	\tilde{x}_1		
2	\tilde{x}_2		
3	\tilde{x}_3		
4	\tilde{x}_4		
\vdots	\vdots		
$k - 2$	\tilde{x}_{n-2}	$\tilde{\mathbf{x}}^3 = \begin{pmatrix} \tilde{x}_2^3 \\ \tilde{x}_3^3 \\ \tilde{x}_4^3 \end{pmatrix} \cdots$	$\tilde{\mathbf{x}}^k = \begin{pmatrix} \tilde{x}_{n-2}^k \\ \tilde{x}_{n-1}^k \\ \tilde{x}_n^k \end{pmatrix}$
$k - 1$	\tilde{x}_{n-1}		
k	\tilde{x}_n		

(a)
(b)

Cuadro 4.3: Ejemplo de organización del conjunto de prueba. (a) Datos originales. (b) Organización de los patrones.

4.1.3. Clases

Las clases, denotadas por c_i , son conjuntos que contienen cierto número de patrones del conjunto fundamental en ellos. En el caso de la aplicación del clasificador Gamma para la predicción, las clases estarán ordenadas en forma de vector unidimensional. Es decir, si el conjunto fundamental es de dimensión $n = 3$, la primera clase estará conformada por el dato x_{n+1} , la segunda clase será el valor del dato x_{n+2} y así sucesivamente. Tomando el ejemplo del cuadro 4.1 (a), la primera clase será el valor correspondiente a x_3 , la segunda clase será el valor correspondiente a x_4 y así sucesivamente. Esto se muestra en el cuadro 4.4.

4.2. Predicción con el clasificador Gamma

Una vez organizados los patrones, se procede a aplicar el clasificador Gamma en ellos para realizar la predicción. El clasificador se entrena con el

índice	valor		
0	x_0		
1	x_1	$\mathbf{c}_1 = (x_3)$	$\mathbf{c}_2 = (x_4)$
2	x_2		
3	x_3		
4	x_4		
\vdots	\vdots		
$i - 2$	x_{n-2}	$\mathbf{c}_3 = (x_5) \cdots$	$\mathbf{c}_i = (x_n)$
$i - 1$	x_{n-1}		
i	x_n		

(a)
(b)

Cuadro 4.4: Ejemplo de organización de las clases. (a) Datos originales. (b) Organización de las clases.

conjunto fundamental; una vez que aprendió empieza a clasificar los patrones del conjunto del prueba. El resultado arrojado por el clasificador será el valor de la clase seleccionada. Un ejemplo de la aplicación de esta técnica de predicción se muestra en la figura 4.2.

La figura 4.2 muestra el resultado de la predicción realizada por el clasificador Gamma sobre el contaminante monóxido de carbono (CO), correspondiente al mes de Marzo del 2008, en la estación de Tacuba. El conjunto fundamental eran los datos correspondientes al año 2007, mientras que los datos del conjunto prueba fueron los correspondientes al mes de Marzo del 2008. Ambos conjuntos tenían una dimensión $n = 10$. Aunque la gráfica superior muestre una predicción más precisa, si observamos la gráfica inferior, es una inspección más cercana de una región de la gráfica original. Esta inspección demuestra que los datos predichos no son lo suficientemente precisos, incluso se pueden observar los puntos donde ocurren máximos y mínimos que aumentan más el error en la predicción. Estos tipos de errores son los más importantes de corregir. En la siguiente sección se planteará un modelo para mejorar la predicción de contaminantes.

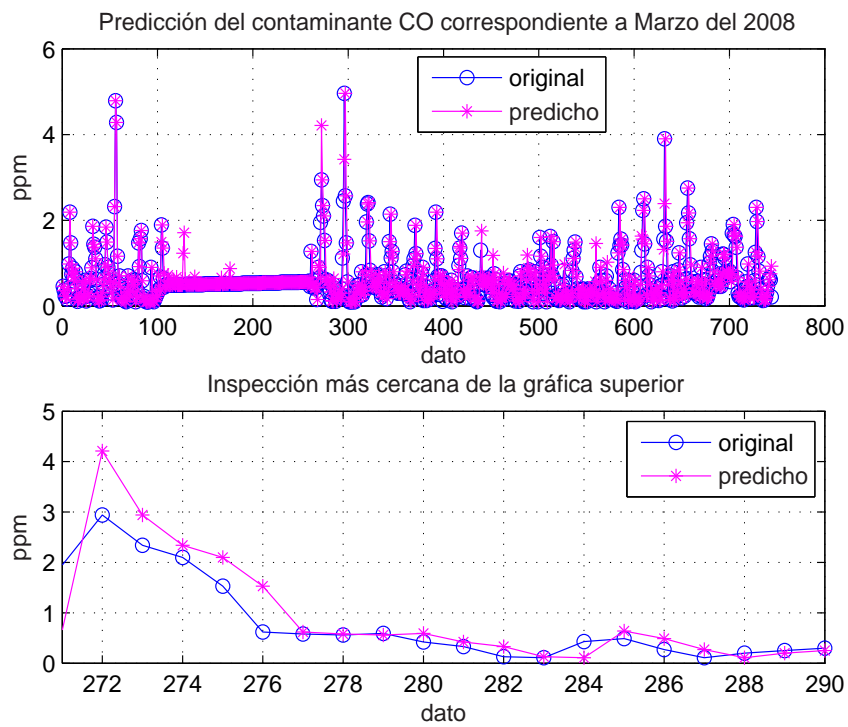


Figura 4.2: Predicción con el clasificador Gamma del contaminante monóxido de carbono (CO), correspondiente al mes de Marzo del 2008.

4.3. Modelo propuesto

Como se pudo observar en la gráfica superior de la figura 4.2, el resultado de la predicción con el clasificador Gamma muestra un desempeño aceptable; no obstante, una inspección más cercana de ese resultado, mostrada en la gráfica inferior, indica que la predicción muestra ciertos errores sobre todo en los puntos donde ocurren máximos o mínimos. Es por ello, que nos enfocaremos en esos puntos. Con el fin de minimizar estos errores, se propone un procesamiento de los resultados arrojados por el clasificador. De igual manera, se propone experimentar con diferentes dimensiones en los patrones del conjunto fundamental y de prueba, para obtener con ello un mejor resultado en la predicción con el clasificador. Así, el algoritmo propuesto se describe en los siguientes pasos:

1. Organizar los patrones de los conjuntos fundamental y de prueba, así como las clases. Esta organización se describió en la sección 4.1.
2. Aplicar el clasificador Gamma a los patrones creados en el paso 1.
3. Los resultados obtenidos del clasificador se procesan para realizar un ajuste de la siguiente manera:
 - Se detectan los cambios de signo en los datos obtenidos del clasificador.
 - Se realiza un procesamiento matemático en los datos donde se detectaron los cambios de signo, con el fin de ajustar ese valor a un valor más cercano al real.

Los pasos 1 y 2 ya fueron explicados en las secciones 4.1 y 4.2, ahora se describirá el paso 3 que es el ajuste de los datos obtenidos con el clasificador.

4.3.1. Ajuste de los datos obtenidos con el clasificador Gamma

El primer paso para realizar el ajuste de los datos obtenidos con el clasificador Gamma es el detectar cambios de signo en ellos. Por ejemplo, consideremos los puntos de la gráfica 4.3(a) cuyos valores tabulados se muestran

en (b). Se calculan: la pendiente m_1 de la recta que pasa por los puntos (a, b) y (c, d) , y la pendiente m_2 que pasa por los puntos (c, d) y (e, f) , para luego comparar los signos de ambas pendientes; si los signos son iguales, se procede a calcular la pendiente m_3 de la recta que pasa por los puntos (e, f) y (g, h) , y a continuación se comparan los signos de las pendientes m_2 y m_3 . Si los signos son iguales, se calcula la siguiente pendiente y se hace de nuevo la comparación. Así se continúa con este proceso, hasta encontrar tres puntos (x_i, y_i) , (x_{i+1}, y_{i+1}) , (x_{i+2}, y_{i+2}) , de manera que los signos de las dos pendientes sean diferentes: ahí hay un cambio de signo. La pendiente se calcula mediante la fórmula 4.1.

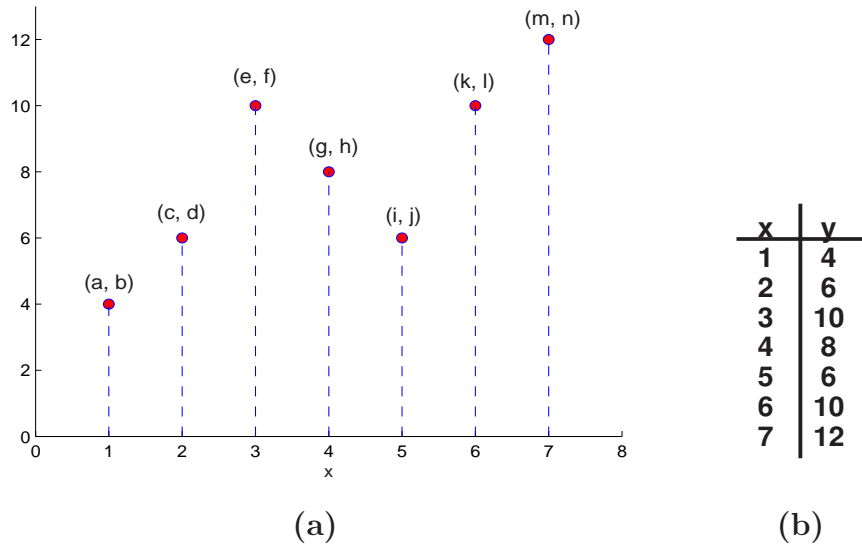


Figura 4.3: Ejemplo ilustrativo. (a) Gráfica de los valores de los puntos tabulados en (b).

$$m = \frac{y_2 - y_1}{x_2 - x_1} \tag{4.1}$$

donde m es la pendiente de la recta definida por los puntos (x_2, y_2) y (x_1, y_1) .

El intervalo del eje de las ordenadas es siempre el mismo, en este caso, la unidad; por lo que la ecuación 4.1 se puede reducir a la fórmula 4.2

$$m = y_2 - y_1 \quad (4.2)$$

Cuando la pendiente sube su valor es positivo, cuando baja es negativo. Esto se ilustra en la figura 4.4 (a), mientras que (b) representa los valores tabulados de las pendientes. Si el resultado de una pendiente calculada es de signo contrario al de la pendiente anterior, se guarda el dato para realizar el ajuste necesario. En este ejemplo, los valores a guardar son los correspondientes a los intervalos $x = 3$ y $x = 5$.

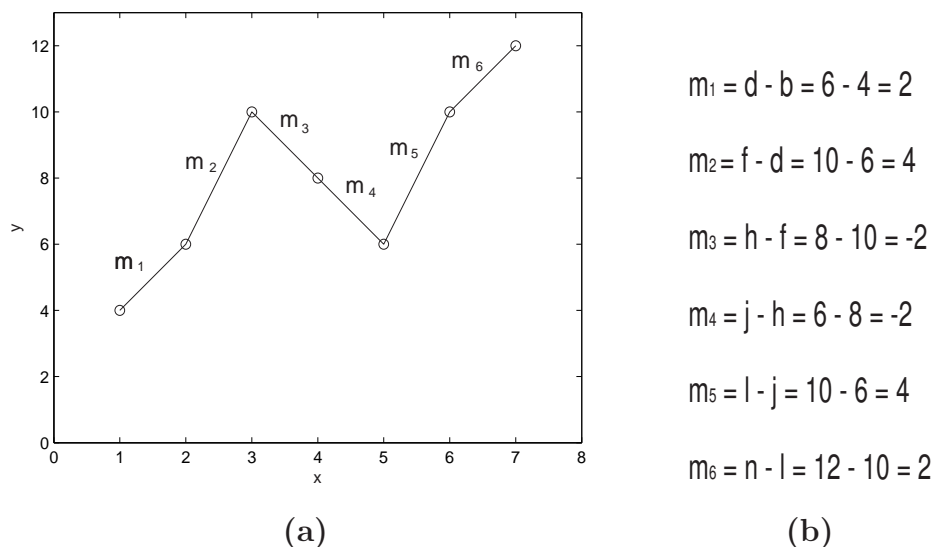


Figura 4.4: Cálculo de pendientes de los datos de la figura 4.3 (b). (a) Pendientes de obtenidas. (b) Valores de las pendientes calculadas.

Existen diversas técnicas para ajustar puntos en una recta, como las diferencias finitas, la interpolación de Newton, entre otras; algunas de ellas funcionan solamente para funciones de primer orden o para aproximar los puntos a una ecuación específica. Como se puede observar en la predicción del monóxido de carbono (CO) de la figura 4.2, el comportamiento del contaminante no se aproxima a alguna función de orden n . Se pueden emplear

métodos heurísticos como los algoritmos genéticos ó la lógica difusa, pero computacionalmente hablando, son técnicas que requieren de mucha prueba y error para encontrar una solución óptima. La solución ideal para ajustar estos puntos es encontrar algún algoritmo matemático que trabaje con datos que no necesariamente se ajusten a alguna función específica pero que sí ofrezcan una aproximación al conjunto de datos de donde se desea encontrar el valor del punto. Es por ello, que para ajustar los datos donde ocurrieron cambios de signo se propone aplicar dos técnicas matemáticas conocidas: la interpolación de Lagrange y la extrapolación de Richardson. Estas dos técnicas tratan de encontrar el valor de un punto dentro de un conjunto de datos tabulados y son de las más precisas en cuanto al tema de interpolación numérica, además de ser fáciles de programar.

La interpolación de Lagrange es un método matemático que sirve para determinar el valor de un punto en un conjunto de datos que pueden estar representados en intervalos iguales o distintos. Una definición más formal es la siguiente: *Dados un conjunto de coordenadas $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$ que definen una función tabular y y es el valor de la función para un valor x dado. La interpolación de Lagrange se define mediante la fórmula:*

$$y = \frac{(x - x_2)(x - x_3)\dots(x - x_n)}{(x_1 - x_2)(x_1 - x_3)\dots(x_1 - x_n)}y_1 + \frac{(x - x_1)(x - x_3)\dots(x - x_n)}{(x_2 - x_1)(x_2 - x_3)\dots(x_2 - x_n)}y_2 + \frac{(x - x_1)(x - x_2)\dots(x - x_n)}{(x_3 - x_1)(x_3 - x_2)\dots(x_3 - x_n)}y_3 + \dots + \frac{(x - x_1)(x - x_2)\dots(x - x_n)}{(x_n - x_1)(x_n - x_2)\dots(x_n - x_{n-1})}y_n$$

La extrapolación de Richardson es un método matemático empleado para calcular valores en funciones, integrales, derivadas, entre otros. La fórmula de Richardson se emplea comúnmente en el cálculo de integrales. Dado un conjunto de valores equidistantes, se requiere conocer el valor de la integral para un intervalo Q dado. La fórmula de Richardson se describe de la siguiente manera:

$$Q = \frac{\left(\frac{h_2}{h_1}\right)^n Q_1 - Q_2}{\left(\frac{h_2}{h_1}\right)^n - 1}$$

donde

Q es la integral a evaluar.

Q_1 y Q_2 son los valores de las integrales calculadas al dividir en sub-intervalos a Q .

h_1 y h_2 son los intervalos en que se dividieron Q_1 y Q_2 .

n es el orden de error de las integrales calculadas.

Se propone que las integrales Q_1 y Q_2 sean calculadas mediante la fórmula de Simpson 1/3, expresada de la siguiente manera:

$$Q_1 = \frac{h_1}{3} [y_0 + y_n + 2 \sum \text{ordenadas de orden par} + 4 \sum \text{ordenadas de orden impar}]$$

Los resultados al aplicar esta metodología se observarán en el siguiente capítulo.

Capítulo 5

Resultados y Discusión

En el presente capítulo veremos los resultados obtenidos al aplicar la metodología propuesta en el capítulo 4. En la primera sección daremos a conocer los parámetros determinados para la realización de las pruebas. Mientras que en las siguientes secciones mostraremos los resultados obtenidos en la predicción de diversos contaminantes tomados de la base de datos de la red automática de monitoreo atmosférico (RAMA). Finalmente, se presentará un análisis de desempeño computacional de la metodología propuesta en el capítulo 4.

5.1. Parámetros de prueba

Para la realización de la predicción de contaminantes atmosféricos, se determinaron ciertos parámetros que describiremos a continuación:

1. Los contaminantes analizados fueron de la misma estación. Se tomó la estación de Tacuba como referencia para los contaminantes NO_x, SO₂, O₃, CO y NO₂. Para los contaminantes PM₁₀ y PM_{2.5} se tomó la estación de Merced debido a que la estación de Tacuba no monitorea estos contaminantes.
2. El conjunto fundamental para cada contaminante correspondió a los datos medidos en el año 2007. Es decir, obtuvimos un total de 8760- n patrones, donde n es la dimensión. La conformación del conjunto fundamental fue explicado en la sección 4.1.1.

3. El conjunto de prueba de cada contaminante correspondió a un mes del año 2008. La relación del total de patrones empleados por cada mes se muestra en el cuadro 5.1. Un mes de 30 días contiene 720 datos, mientras que un mes de 31 días 744 datos, Febrero fue un mes de 29 días, por lo que se tienen 696 datos. Cada dato corresponde al valor obtenido del contaminante en una hora, de esta manera, los primeros 24 datos corresponden al primer día del mes, del dato 25 al 48 son los valores correspondientes al segundo día del mismo mes y así sucesivamente. Por ejemplo, si se prueban los patrones del conjunto de prueba con una dimensión $n = 10$, para un mes de 31 días tendríamos 734 patrones, para un mes de 30 días tendríamos 710 patrones y para el mes de Febrero tendríamos 686 patrones.

Mes	Patrones
Enero	744- n
Febrero	696- n
Marzo	744- n
Abril	720- n
Mayo	744- n
Junio	720- n
Julio	744- n
Agosto	744- n
Septiembre	720- n
Octubre	744- n
Noviembre	720- n
Diciembre	744- n

Cuadro 5.1: Relación de patrones por cada mes del año 2008. n es la dimensión de prueba.

4. Por cada contaminante se realizaron pruebas para determinar la dimensión de los patrones donde se obtuvo el menor error de predicción. El criterio para tomar la variación de las dimensiones en los patrones se basó en la observación de la variación en los errores obtenidos en la predicción de uno de los contaminantes. Inicialmente se realizaron pruebas variando las dimensiones en los patrones de uno en uno hasta llegar a $n = 10$; al no presentar mucha variación en los resultados, se decidió empezar a variar la dimensión en incrementos de 10, hasta

llegar a $n = 100$; finalmente, se decidió hacer una última prueba con $n = 150$.

5. Para medir el desempeño de la predicción se utilizaron dos parámetros: El *Root Mean Square Error* (*RMSE*), que es el error típico de la estima del valor predicho sobre el valor original y es determinado por la fórmula 5.1. El segundo parámetro es el *BIAS*, mostrado en la ecuación 5.2, que nos da un índice de cuánto se está subestimando ó sobreestimando el resultado, es decir, si existe un *BIAS* positivo, quiere decir que la predicción obtenida está sobreestimada, mientras que si éste es negativo se considera que la predicción obtenida está subestimada; un *BIAS* cercano a cero indica que la predicción se acercó lo más posible al valor real.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (5.1)$$

$$BIAS = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (5.2)$$

donde

P_i Es el i -ésimo valor predicho.

O_i Es el i -ésimo valor original.

n Es el total de patrones del conjunto de prueba.

6. Para realizar el ajuste de los resultados obtenidos con el clasificador Gamma, utilizando interpolación de Lagrange, se tomaron los dos datos anteriores y los dos datos siguientes al dato donde se detectó el cambio de signo. Esto se decidió porque experimentalmente dio mejor resultado.
7. En el ajuste por extrapolación de Richardson se tomaron los datos de igual manera que en la interpolación de Lagrange. También se realizó otra forma de ajuste, tomando ahora los dos datos anteriores al dato donde se encontró el cambio de signo.
8. Para aplicar la fórmula de Richardson se tomó $n = 2$, debido a que es el orden de error del método de Simpson $1/3$.

9. Los intervalos h_1 y h_2 para la extrapolación de Richardson fueron de 1 y 2, respectivamente.

5.1.1. Interpretación de las gráficas

Con el fin de demostrar los resultados obtenidos de las predicciones, se realizaron sus respectivas gráficas. Para mejor entendimiento de éstas, se explicará brevemente los términos utilizados en ellas.

El significado de los términos utilizados en el eje de las ordenadas en las gráficas de las figuras, correspondientes a las predicciones de los meses de Enero y Marzo del año 2009, es el siguiente: *Pred. Gamma* es la predicción realizada con el clasificador Gamma; *Aj. Lagrange* es el ajuste de los datos arrojados por el clasificador utilizando la extrapolación de Lagrange; *Aj. Richardson1* se refiere a la extrapolación de Richardson aplicada a los datos dados por el clasificador, utilizando los dos valores anteriores y los dos siguientes al punto donde se detectó el cambio de signo; *Aj. Richardson2* es la extrapolación de Richardson aplicada en los dos valores anteriores al punto donde se detectó el cambio de signo, sobre los resultados obtenidos del clasificador.

En las gráficas de predicción de un mes del año 2008, el eje de las ordenadas, denominada *dato*, se refiere a una hora dentro del mismo mes, ya que los valores se capturan cada hora. Es decir, el primer día de un mes corresponde a los datos del 1 al 24, el segundo día, a los datos del 25 al 48, el tercer día, los datos del 49 al 72, y así sucesivamente.

Los resultados obtenidos en la predicción de contaminantes atmosféricos con los parámetros especificados se muestran en las siguientes secciones.

5.2. Predicción del óxido de nitrógeno (NOx)

Para el óxido de nitrógeno se realizaron pruebas en cada mes del año 2008. Los resultados se muestran en el cuadro 5.2. Donde se muestra la mejor dimensión en los patrones del conjunto fundamental y el de prueba y los

RMSE y BIAS calculados para cada método de ajuste. Los valores en negritas son los mejores obtenidos en la predicción de cada mes. Se puede observar que de los meses de Abril a Octubre el menor error se obtuvo al aplicar la interpolación de Lagrange, mientras que para los meses restantes el menor error se obtuvo al aplicar la extrapolación de Richardson, tomando los dos valores anteriores y los dos siguientes al dato donde se detectó el cambio de signo. Este cuadro se representa en la gráfica superior de la figura 5.1.

La gráfica inferior de la figura 5.1 muestra el RMSE calculado solamente en los puntos donde se detectaron cambios de signo. Aquí se puede observar que en general las técnicas de ajuste por interpolación de Lagrange y extrapolación de Richardson disminuyen los errores obtenidos por el clasificador Gamma.

Mes	Di- men- sión	Predicción		Ajuste		Ajuste	
		Gamma		Lagrange		Richardson1	
		RMSE (milésimas)	BIAS (milésimas)	RMSE (milésimas)	BIAS (milésimas)	RMSE (milésimas)	BIAS (milésimas)
Enero	150	35.80861	.92876	34.51973	-.56855	34.3083	.040323
Febrero	150	30.37697	3.14244	28.882	1.655	28.771	2.421
Marzo	70	25.75904	.75672	25.17276	.11425	24.66223	.79301
Abril	100	18.57066	8.52778	18.911	6.54	18.998	7.601
Mayo	150	44.65528	34.48853	41.257	31.628	42.075	32.711
Junio	150	63.9543	54.56667	60.855	51.869	61.57	52.66
Julio	150	63.75949	54.85618	60.759	52.105	61.657	53.116
Agosto	150	24.09234	13.62634	22.623	12.013	22.935	12.953
Septiembre	150	87.19345	76.275	85.028	74.379	85.648	75.021
Octubre	50	26.99378	16.74731	25.162	14.688	25.227	15.644
Noviembre	150	30.66807	.48889	30.149	-.4	29.772	.308
Diciembre	150	36.21952	1.43413	35.251	.312	34.789	1.038

Cuadro 5.2: Predicción del NOx para el año 2008.

En la figura 5.2 se muestra la gráfica de la predicción del mes de Marzo con la dimensión $n = 70$ en sus patrones, mostrada en el cuadro 5.2. La gráfica inferior izquierda muestra una fracción de la gráfica superior, que es la predicción hecha con el clasificador Gamma. Mientras que la gráfica inferior

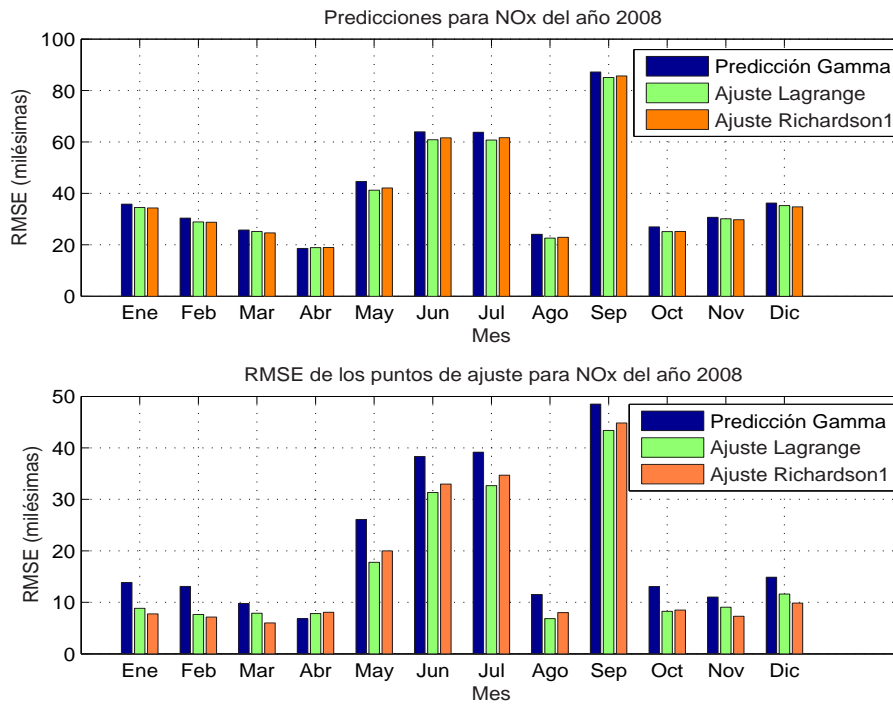


Figura 5.1: Predicción del contaminante óxido de nitrógeno (NOx), correspondiente al año 2008.

derecha muestra el ajuste realizado con la fórmula de Richardson, utilizando los dos datos anteriores y los dos siguientes al dato donde se encontró el cambio de signo. Se puede observar en esta última gráfica cómo el ajuste funcionó para uno de los puntos, mientras que para otro lo acercó más al valor real.

Una vez obtenidas las dimensiones ideales de cada mes, se procedió a calcular un promedio de dimensión, resultando en 130. Se decidió probar la dimensión promedio junto con la encontrada para los meses de enero y marzo en la predicción del año 2009, entrenando al clasificador Gamma con el año 2007. El resultado se muestra en el cuadro 5.3. Donde se puede observar que los mejores resultados fueron con la dimensión promedio, de 130, y con el ajuste utilizando la fórmula de Richardson con los dos datos anteriores y los dos siguientes al dato donde se detectó el cambio de signo.

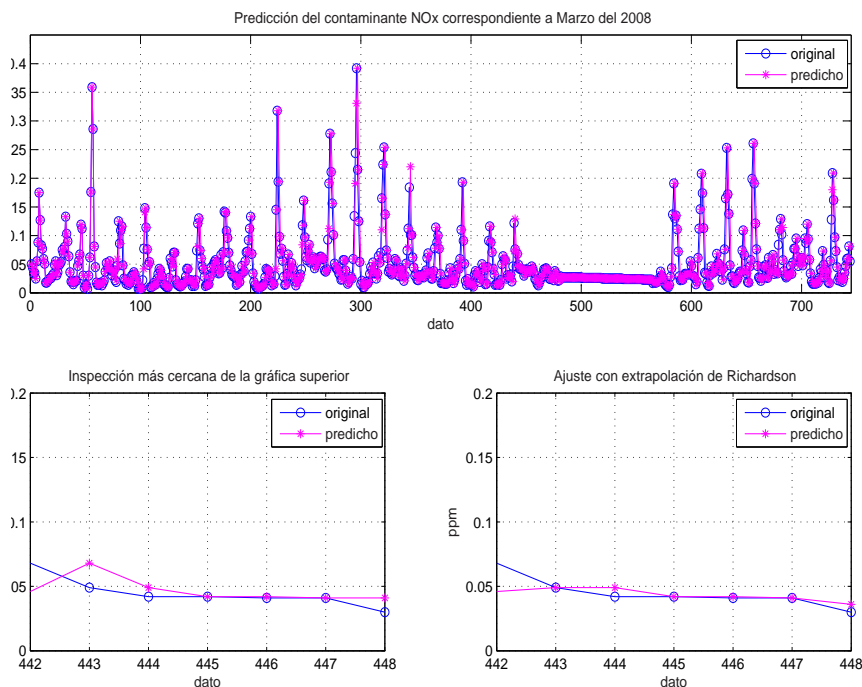


Figura 5.2: Predicción del contaminante óxido de nitrógeno (NOx), correspondiente al mes de Marzo del 2008.

Otro experimento también fue realizado, la predicción del NOx utilizando ahora el año 2008 como conjunto de entrenamiento para el clasificador y probando en los meses de Enero y Marzo del 2009 con la dimensión promedio y la dimensión ideal calculadas para el año 2008. Los resultados se muestran en el cuadro 5.4. Se puede observar que el mejor resultado obtenido para Enero fue con una dimensión de 150 y realizando el ajuste con la fórmula de Richardson utilizando los dos valores anteriores y los dos valores siguientes al dato donde ocurrió el cambio de signo. Para el mes de Marzo el mejor resultado obtenido fue con la dimensión promedio encontrada, 130, y utilizando la interpolación de Lagrange como técnica de ajuste. En los cuadros 5.3 y 5.4 se marcan con negritas los mejores valores obtenidos en las predicciones realizadas.

La representación gráfica de los cuadros 5.3 y 5.4 se puede observar en las figuras 5.3 y 5.4. La gráfica inferior en ambas figuras muestra el RMSE calculado únicamente en los puntos donde se realizaron los ajustes. Para el mes de Enero del 2009, entrenando con el año 2007, se observa en la gráfica inferior que el menor error obtenido fue al aplicar el ajuste con la fórmula de Richardson1 y con una dimensión de 130. En el mes de Marzo del 2009, entrenando con el año 2007, el menor error obtenido fue con Richardson1 y con una dimensión de 130, esto se muestra en la gráfica 5.4. La gráficas inferiores de las figuras 5.3 y 5.4 muestran el RMSE de los puntos donde se realizaron los ajustes observándose que el mínimo error obtenido fue al aplicar la técnica de ajuste de Richardson1 con una dimensión de 150. En la gráfica inferior de la figura 5.4 el RMSE de los puntos de ajuste, entrenando el clasificador con el año 2008, presentó un mínimo error con la técnica de ajuste de Lagrange con una dimensión de 130.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	130	Predicción Gamma	34.28598	.40726
		Ajuste Lagrange	32.34379	-.73118
		Ajuste Richardson1	32.092	-.13172
		Ajuste Richardson2	34.58931	-1.02823
Enero	150	Predicción Gamma	34.91249	2.5326
		Ajuste Lagrange	33.03454	1.4664
		Ajuste Richardson1	32.7464	1.99597
		Ajuste Richardson2	34.97088	1.0457
Marzo	130	Predicción Gamma	28.40978	13.28763
		Ajuste Lagrange	27.33553	11.3975
		Ajuste Richardson1	27.104388	10.263441
		Ajuste Richardson2	28.62785	7.88978
Marzo	70	Predicción Gamma	31.07059	14.92204
		Ajuste Lagrange	28.0505	11.85753
		Ajuste Richardson1	28.11172	12.97581
		Ajuste Richardson2	30.1286	11.21371

Cuadro 5.3: Predicción del NOx para el año 2009, entrenando el clasificador con el año 2007.

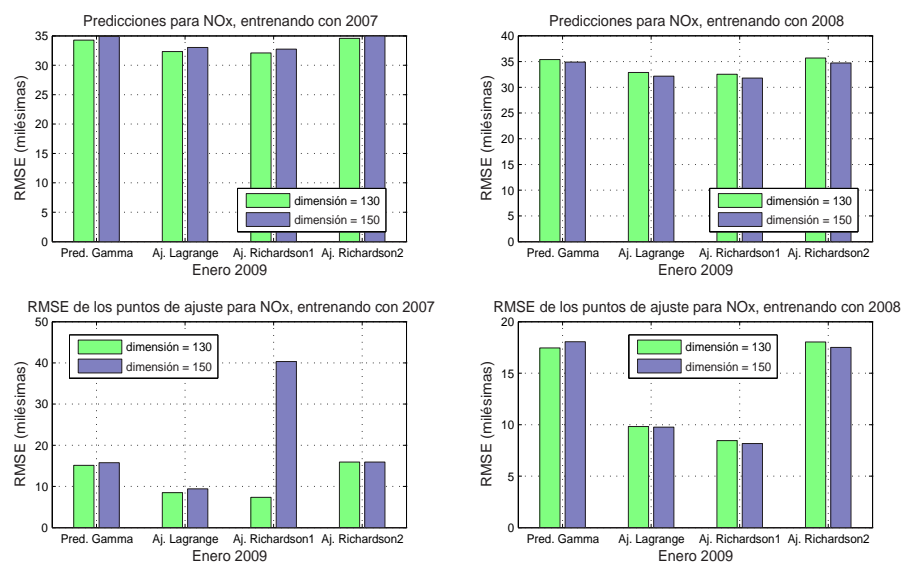


Figura 5.3: Predicción del contaminante óxido de nitrógeno (NOx), correspondiente al mes de Enero del 2009, entrenando el clasificador con el año 2007 y con el año 2008.

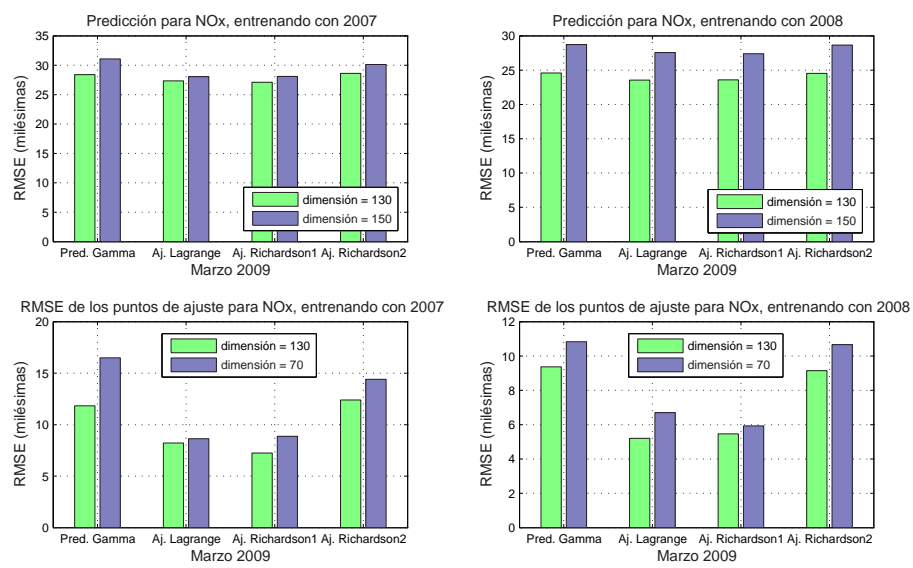


Figura 5.4: Predicción del contaminante óxido de nitrógeno (NOx), correspondiente al mes de Marzo del 2009, entrenando el clasificador con el año 2007 y con el año.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	130	Predicción Gamma	35.40672	35.0941
		Ajuste Lagrange	32.8898	1.42473
		Ajuste Richardson1	32.53982	2.2621
		Ajuste Richardson2	35.69899	1.23522
Enero	150	Predicción Gamma	34.9065	4.014478
		Ajuste Lagrange	32.1603	1.95968
		Ajuste Richardson1	31.78914	2.76747
		Ajuste Richardson2	34.73571	1.71505
Marzo	130	Predicción Gamma	24.5891	6.53495
		Ajuste Lagrange	23.54848	5.54167
		Ajuste Richardson1	23.58766	6.20833
		Ajuste Richardson2	24.5264	49.0457
Marzo	70	Predicción Gamma	28.73049	8.47312
		Ajuste Lagrange	27.5629	6.79973
		Ajuste Richardson1	27.39999	7.60618
		Ajuste Richardson2	28.66571	5.78763

Cuadro 5.4: Predicción del NOx para el año 2009, entrenando el clasificador con el año 2008.

Ajuste Mes	Análisis del mes completo			Análisis de los puntos de ajuste		
	Ajuste Lagrange	Ajuste Richardson1	Ajuste Richardson2	Ajuste Lagrange	Ajuste Richardson1	Ajuste Richardson2
	Enero 2008	3.6 %	4.19 %		36.13 %	44.03 %
Marzo 2008	2.28 %	4.26 %		19.13 %	38.69 %	
Enero 2009	5.66 %	6.4 %	-.88 %	43.92 %	51.34 %	-5.27 %
Marzo 2009	3.78 %	4.59 %	-.77 %	30.59 %	38.79 %	-4.77 %
Enero* 2009	7.87 %	8.93 %	.49 %	45.92 %	54.76 %	3.09 %
Marzo* 2009	4.23 %	4.07 %	.25 %	44.44 %	41.66 %	2.38 %

Cuadro 5.5: Porcentajes de mejora en la predicción del NOx para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008

Para la mayoría de las pruebas realizadas con este contaminante se observa que el ajuste, utilizando la fórmula de Richardson con los dos valores anteriores al dato donde ocurre el cambio de signo, dio un resultado ligeramente superior al de la predicción con el clasificador Gamma. Un análisis de porcentajes de mejora en las técnicas de ajuste propuestas se muestra en el cuadro 5.5, estos porcentajes fueron calculados con respecto al valor obtenido con el clasificador Gamma en la mejor dimensión. Los mejores porcentajes para cada mes en las diferentes dimensiones utilizadas están marcados en negritas.

5.3. Predicción del dióxido de azufre (SO₂)

Para el dióxido de azufre (SO₂) se realizaron las predicciones del año 2008. Encontrándose también la dimensión de los patrones donde la predicción tuvo menor error. Estos resultados se muestran en los cuadros 5.6 y 5.7, donde los mínimos errores obtenidos en la predicción se marcan en negritas. Se puede observar que, en general, la mejor técnica de ajuste fue la de la extrapolación de Richardson aplicada sobre los dos valores anteriores al punto donde se detectó el cambio de signo. Estos cuadros se representan de manera gráfica en la figura 5.5. La parte superior de esta gráfica muestra el RMSE obtenido en cada uno de los meses. Mientras que la gráfica inferior representa el RMSE obtenido solamente en los puntos donde se detectaron cambios de signo; esto con la finalidad de demostrar cómo disminuye el error al aplicar las técnicas de ajuste propuestas.

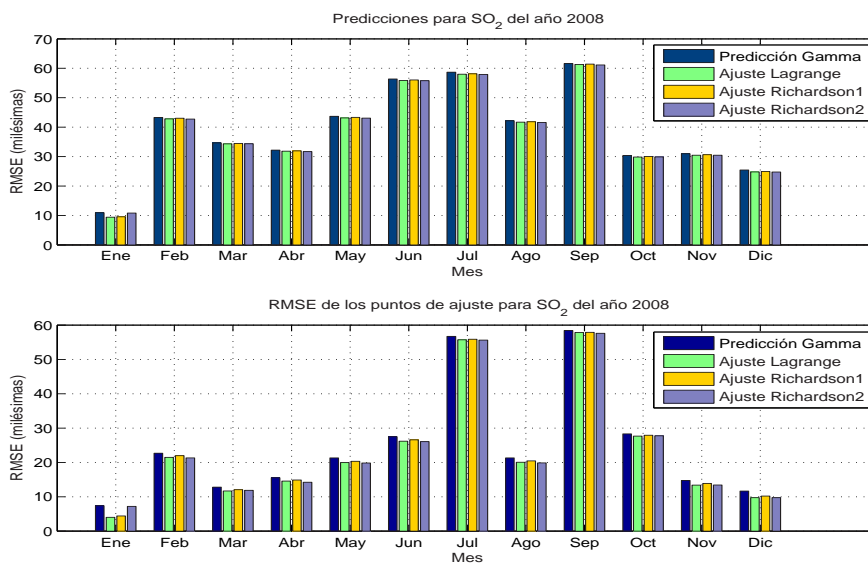


Figura 5.5: Predicción del contaminante dióxido de azufre (SO₂), correspondiente al año 2008.

La gráfica de la predicción del SO₂ correspondiente al mes de Enero del 2008 utilizando el clasificador Gamma se muestra en la figura 5.6. La gráfica inferior izquierda muestra una región de la gráfica superior. Mientras que la

gráfica inferior derecha muestra la misma región ajustada con la interpolación de Lagrange. Es clara la mejoría en el ajuste utilizando esta técnica.

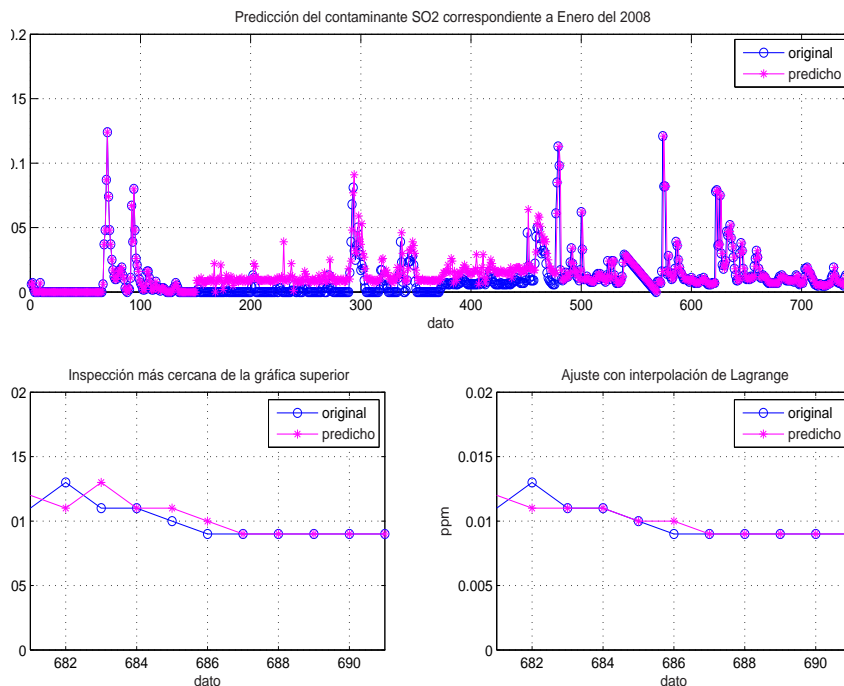


Figura 5.6: Predicción del contaminante dióxido de azufre (SO₂), correspondiente al mes de Enero del 2008.

La predicción de los meses de Enero y Marzo del año 2009, entrenando el clasificador con los datos correspondientes al año 2007, es mostrada en el cuadro 5.8. Esta predicción se realizó con las dimensiones encontradas en el año 2008. Se puede observar que para ambos meses la mejor técnica de ajuste fue la interpolación de Lagrange, estos valores aparecen marcados en negritas.

De igual manera, se realizaron las predicciones de los meses de Enero y Marzo del 2009, esta vez entrenando el clasificador con los datos pertenecientes al año 2008. Los resultados se muestran en el cuadro 5.9. Para ambos meses, la mejor técnica de ajuste resultó ser la extrapolación de Richardson aplicada a los dos valores anteriores y los dos siguientes al punto donde se

detectó el cambio de signo. Dichos valores aparecen marcados en negritas.

Los cuadros 5.8 y 5.9 se muestran de forma gráfica en las figuras 5.7 y 5.8. Las gráficas inferiores de ambas figuras reflejan el RMSE obtenido de los puntos donde se detectaron los cambios de signo, para posteriormente aplicar las técnicas de ajuste propuestas. Se puede concluir, para ambas gráficas, que el mínimo error obtenido en la predicción ocurre cuando el clasificador es entrenado con los datos del año 2008.

Un análisis de porcentajes de mejora en las técnicas de ajuste propuestas se muestra en el cuadro 5.10, estos porcentajes fueron calculados con respecto al valor obtenido con el clasificador Gamma en la mejor dimensión. Los porcentajes más altos se marcaron en negritas.

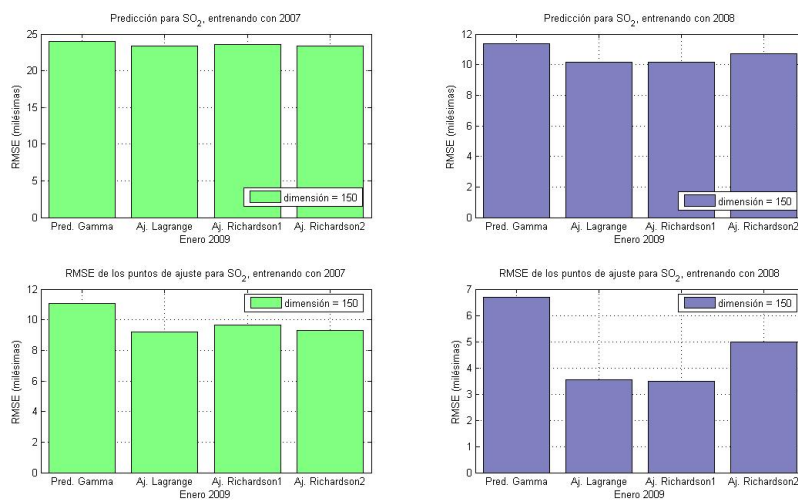


Figura 5.7: Predicción del contaminante dióxido de azufre (SO₂), correspondiente al mes de Enero del 2009, entrenando el clasificador con el año 2007 y con el año 2008.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	10.92004	4.02688
		Ajuste Lagrange	9.40366	3.80242
		Ajuste Richardson1	9.54672	3.92473
		Ajuste Richardson2	10.81106	3.54839
Febrero	150	Predicción Gamma	43.30693	37.83477
		Ajuste Lagrange	42.82632	37.40086
		Ajuste Richardson1	43.03262	37.63075
		Ajuste Richardson2	42.73169	37.18678
Marzo	80	Predicción Gamma	34.67136	31.96909
		Ajuste Lagrange	34.31811	31.68414
		Ajuste Richardson1	34.45755	31.83871
		Ajuste Richardson2	34.36833	31.60349
Abril	150	Predicción Gamma	32.19737	27.89861
		Ajuste Lagrange	31.84049	27.61111
		Ajuste Richardson1	31.95735	27.76389
		Ajuste Richardson2	31.70421	27.34583
Mayo	150	Predicción Gamma	43.67664	38.70565
		Ajuste Lagrange	43.16001	38.27823
		Ajuste Richardson1	43.32014	38.44086
		Ajuste Richardson2	43.07985	38.16398
Junio	150	Predicción Gamma	56.3608	50.05417
		Ajuste Lagrange	55.87127	49.62917
		Ajuste Richardson1	56.0349	49.77639
		Ajuste Richardson2	55.79275	49.50833

Cuadro 5.6: Predicción de SO₂ para el año 2008.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Julio	150	Predicción Gamma	58.663	52.21237
		Ajuste Lagrange	57.97936	51.63306
		Ajuste Richardson1	58.14389	51.79301
		Ajuste Richardson2	57.90932	51.54032
Agosto	150	Predicción Gamma	42.18662	37.40995
		Ajuste Lagrange	41.69097	36.96505
		Ajuste Richardson1	41.85277	37.11962
		Ajuste Richardson2	41.58954	36.81317
Septiembre	150	Predicción Gamma	61.60393	54.71667
		Ajuste Lagrange	61.31123	54.44167
		Ajuste Richardson1	61.41325	54.5375
		Ajuste Richardson2	61.13892	54.23194
Octubre	90	Predicción Gamma	30.36521	27.51344
		Ajuste Lagrange	29.84026	25.55914
		Ajuste Richardson1	30.04466	25.81586
		Ajuste Richardson2	29.89943	25.35215
Noviembre	150	Predicción Gamma	30.97629	26.825
		Ajuste Lagrange	30.47832	26.47778
		Ajuste Richardson1	30.66406	26.66806
		Ajuste Richardson2	30.47055	26.32083
Diciembre	150	Predicción Gamma	25.435693	20.9879
		Ajuste Lagrange	24.7829	20.53629
		Ajuste Richardson1	24.94367	20.76478
		Ajuste Richardson2	24.75897	20.25403

Cuadro 5.7: Predicción de SO₂ para el año 2008.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	24.03864	20.08199
		Ajuste Lagrange	23.40274	19.68817
		Ajuste Richardson1	23.56982	19.90188
		Ajuste Richardson2	23.41873	19.49597
Marzo	80	Predicción Gamma	32.69818	30.23522
		Ajuste Lagrange	32.07793	29.72446
		Ajuste Richardson1	32.33642	30.00403
		Ajuste Richardson2	32.1487	29.71371

Cuadro 5.8: Predicción del SO₂ para el año 2009, entrenando el clasificador con el año 2007.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	11.374254	2.88172
		Ajuste Lagrange	10.183729	2.576613
		Ajuste Richardson1	10.166689	2.727151
		Ajuste Richardson2	10.705114	2.38172
Marzo	80	Predicción Gamma	8.326825	1.852151
		Ajuste Lagrange	8.127429	1.716398
		Ajuste Richardson1	8.082239	1.787634
		Ajuste Richardson2	8.091546	1.66129

Cuadro 5.9: Predicción del SO₂ para el año 2009, entrenando el clasificador con el año 2008.

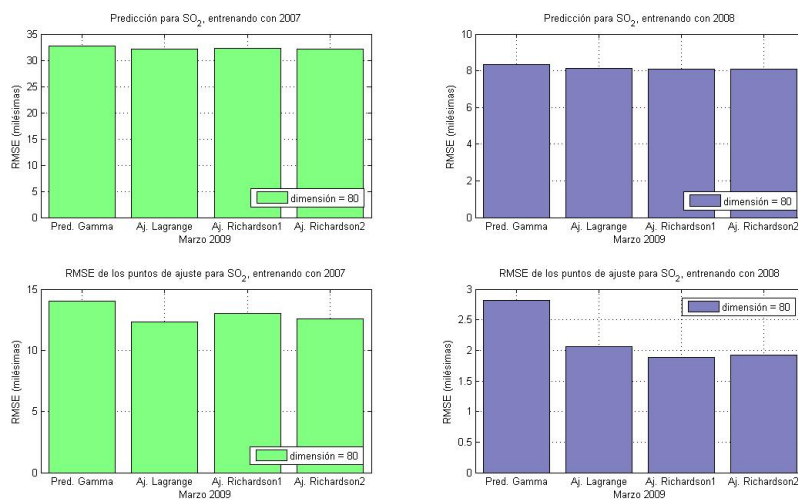


Figura 5.8: Predicción del contaminante dióxido de azufre (SO₂), correspondiente al mes de Marzo del 2009, entrenando el clasificador con el año 2007 y con el año 2008.

Ajuste Mes	Análisis del mes completo			Análisis de los puntos de ajuste		
	Ajuste Lagrange	Ajuste Richardson1	Ajuste Richardson2	Ajuste Lagrange	Ajuste Richardson1	Ajuste Richardson2
Enero 2008	13.89 %	12.58 %	.99 %	45.8 %	40.45 %	2.62 %
Marzo 2008	1.02 %	.62 %	.87 %	8.72 %	5.77 %	7.16 %
Enero 2009	2.65 %	1.95 %	2.58 %	16.89 %	12.73 %	15.95 %
Marzo 2009	1.9 %	1.1 %	1.7 %	12.2 %	7.34 %	10.33 %
Enero* 2009	10.47 %	10.62 %	5.88 %	47.01 %	47.70 %	25.4 %
Marzo* 2009	2.39 %	2.94 %	2.83 %	26.87 %	33.03 %	31.63 %

Cuadro 5.10: Porcentajes de mejora en la predicción del SO₂ para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008

5.4. Predicción del monóxido de carbono (CO)

Las predicciones realizadas para el monóxido de carbono (CO) fueron las correspondientes a cada uno los meses del año 2008. Estos resultados se muestran en los cuadros 5.11 y 5.12, donde los mejores se marcan en negritas. De igual manera se encontraron las mejores dimensiones de los patrones en cada mes. En general, la mejor técnica de ajuste estuvo repartida casi equitativamente en las 2 técnicas de extrapolación de Richradson propuestas, solamente en un mes, Mayo, resultó la interpolación de Lagrange como mejor técnica de ajuste.

Una representación gráfica de estos cuadros es mostrada en la figura 5.9. En la gráfica inferior de la figura se muestra el RMSE calculado sobre los puntos donde se detectaron cambios de signo, observándose en general un mínimo error con la técnica de ajuste de Richardson1.

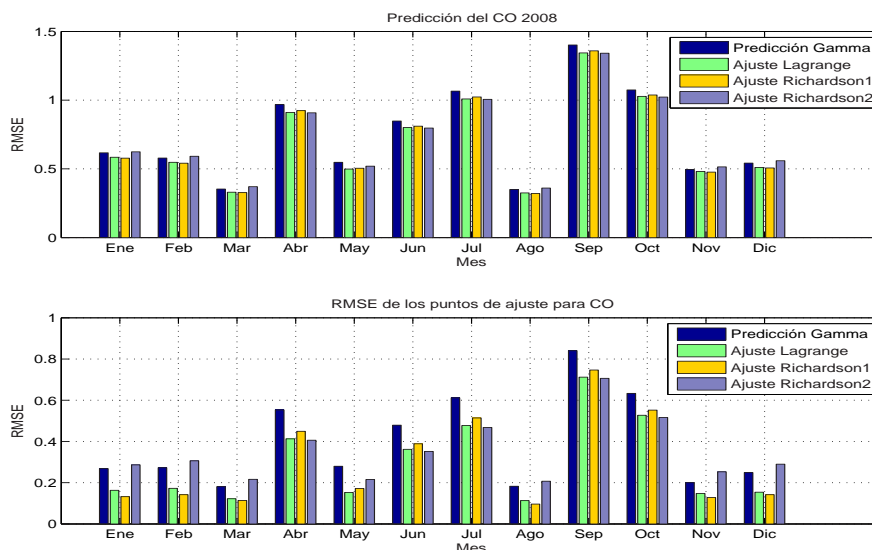


Figura 5.9: Predicción del contaminante monóxido de carbono (CO), correspondiente al año 2008, entrenando el clasificador con el año 2007.

En la figura 5.10 se aprecia la predicción realizada para el mes de Marzo del 2008. La gráfica inferior izquierda muestra una parte de la gráfica su-

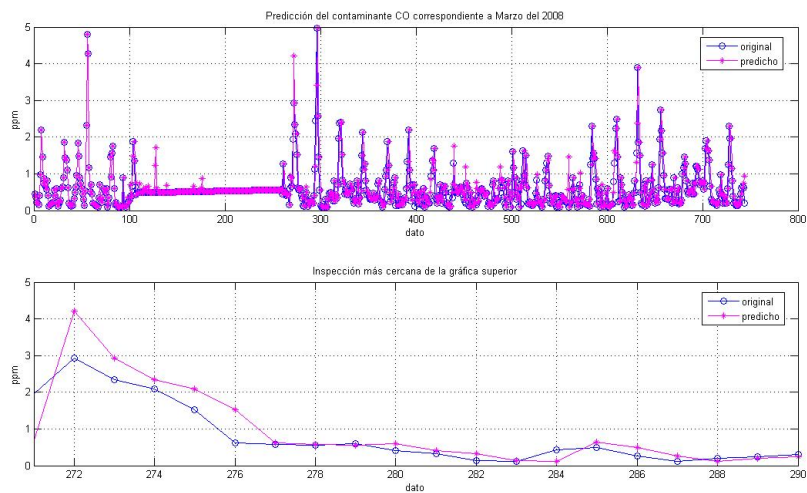


Figura 5.10: Predicción del contaminante monóxido de carbono (CO), correspondiente al mes de Marzo del 2008.

perior, que es la predicción realizada con el clasificador Gamma. En la parte inferior derecha se observa la misma región de la gráfica izquierda, con el ajuste realizado utilizando la fórmula de Richardson calculado con los dos valores anteriores y los dos siguientes al dato donde se detectó el cambio de signo. Aunque no se pudo realizar un ajuste más exacto, se puede observar que los puntos ajustados minimizaron sus errores con respecto a los valores originales.

El cuadro 5.13 muestra las predicciones realizadas para los meses de Enero y Marzo del 2009, entrenando el clasificador con el año 2007 y utilizando las dimensiones encontradas en el año 2008. Para el mes de Enero el mejor resultado obtenido fue utilizando la fórmula de Richardson, calculado con los dos valores anteriores y los dos siguientes al dato donde se detectó un cambio de signo. En el mes de Marzo, el mejor resultado obtenido fue utilizando la interpolación de Lagrange como técnica de ajuste.

Otra prueba fue realizada con la predicción de los meses de Enero y Marzo del 2009, entrenando el clasificador con los datos del año 2008 y utilizando las mismas dimensiones encontradas en el 2008 para cada uno de estos meses. Los resultados se muestran en el cuadro 5.14, donde los mejores se marcan con negritas.

Las figuras 5.11 y 5.12 muestran de forma gráfica los datos representados en los cuadros 5.13 y 5.14. En la parte inferior de ambas gráficas se puede observar el RMSE calculado de los puntos donde se realizaron los ajustes. Se observa que para el mes de Enero 2009, mostrado en ambos cuadros, el mínimo error obtenido fue con el ajuste utilizando Richardson1. Para la predicción del mes de Marzo del 2009, la fórmula de Lagrange resultó mejor técnica de ajuste cuando el clasificador se entrenó con el año 2007; no obstante, cuando el clasificador se entrenó con el año 2008, la mejor técnica de ajuste fue utilizando la fórmula de Richardson1.

Un análisis de porcentajes de mejora en las técnicas de ajuste propuestas se muestra en el cuadro 5.15, estos porcentajes fueron calculados con respecto al valor obtenido con el clasificador Gamma en la mejor dimensión. Los mejores porcentajes se encuentran marcados en negritas.

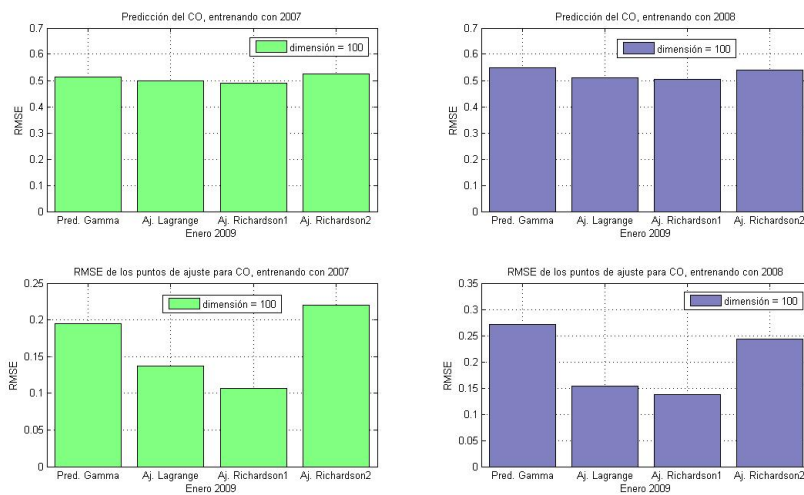


Figura 5.11: Predicción del contaminante monóxido de carbono (CO), correspondiente al mes de Enero del 2009, entrenando el clasificador con el año 2007.

Mes	Dimensión	Técnica	RMSE	BIAS
Enero	100	Predicción Gamma	.61708014	.0387644
		Ajuste Lagrange	.58448868	-.00030914
		Ajuste Richardson1	.57767901	.01858871
		Ajuste Richardson2	.62412476	-.00139785
Febrero	150	Predicción Gamma	.57887041	.1124569
		Ajuste Lagrange	.54793162	.07365661
		Ajuste Richardson1	.54091786	.09468966
		Ajuste Richardson2	.59188999	.05921264
Marzo	90	Predicción Gamma	.3532591	.02846505
		Ajuste Lagrange	.33043674	.00475806
		Ajuste Richardson1	.32762751	.0143018
		Ajuste Richardson2	.37034131	-.00642473
Abril	150	Predicción Gamma	.96750215	.758375
		Ajuste Lagrange	.91035574	.7088625
		Ajuste Richardson1	.92378574	.72692917
		Ajuste Richardson2	.90795166	.69829444
Mayo	50	Predicción Gamma	.54749982	.29272849
		Ajuste Lagrange	.49845416	.24222849
		Ajuste Richardson1	.50430696	.26.034812
		Ajuste Richardson2	.51976373	.23388306
Junio	150	Predicción Gamma	.84785882	.66470833
		Ajuste Lagrange	.80067654	.62037917
		Ajuste Richardson1	.81052417	.63485694
		Ajuste Richardson2	.79710385	.60310972

Cuadro 5.11: Predicción de CO para el año 2008.

Mes	Dimensión	Técnica	RMSE	BIAS
Julio	150	Predicción Gamma	1.06517913	.88103495
		Ajuste Lagrange	1.00874877	.82809812
		Ajuste Richardson1	1.02320439	.84817876
		Ajuste Richardson2	1.00537455	.81581586
Agosto	150	Predicción Gamma	.34948407	.02635753
		Ajuste Lagrange	.32514175	.00225134
		Ajuste Richardson1	.32070131	.01219489
		Ajuste Richardson2	.36024295	-.0084167
Septiembre	150	Predicción Gamma	1.4008783	1.21055556
		Ajuste Lagrange	1.34344752	1.160225
		Ajuste Richardson1	1.35827239	1.17654444
		Ajuste Richardson2	1.34113988	1.15364167
Octubre	150	Predicción Gamma	1.07384678	.91369624
		Ajuste Lagrange	1.02701919	.8726129
		Ajuste Richardson1	1.03740047	.88562769
		Ajuste Richardson2	1.02251355	.8589422
Noviembre	150	Predicción Gamma	.49539448	.021375
		Ajuste Lagrange	.48064069	.00243472
		Ajuste Richardson1	.47623617	.013675
		Ajuste Richardson2	.51421106	-.01021528
Diciembre	100	Predicción Gamma	.54154871	.07577957
		Ajuste Lagrange	.51001449	.04220565
		Ajuste Richardson1	.50691658	.05743414
		Ajuste Richardson2	.55881963	.02231586

Cuadro 5.12: Predicción de CO para el año 2008.

Mes	Dimen- sión	Técnica	RMSE	BIAS
Enero	100	Predicción Gamma	.51421775	.04553763
		Ajuste Lagrange	.4978839	.03182796
		Ajuste Richardson1	.49090505	.04267473
		Ajuste Richardson2	.5242692	.01853495
Marzo	90	Predicción Gamma	1.14319139	1.00037634
		Ajuste Lagrange	1.10607061	.96575269
		Ajuste Richardson1	1.12046559	.98297043
		Ajuste Richardson2	1.10997463	.95926075

Cuadro 5.13: Predicción del CO para Enero y Marzo del 2009, entrenando el clasificador con el año 2007.

Mes	Dimen- sión	Técnica	RMSE	BIAS
Enero	100	Predicción Gamma	.549696397	.123198925
		Ajuste Lagrange	0.508887457	.073733871
		Ajuste Richardson1	.504972376	.092302419
		Ajuste Richardson2	.538240804	.071759409
Marzo	90	Predicción Gamma	.465005347	.169744624
		Ajuste Lagrange	.440665157	.131657258
		Ajuste Richardson1	.440454891	.150513441
		Ajuste Richardson2	.458814928	.128248656

Cuadro 5.14: Predicción del CO para Enero y Marzo del 2009, entrenando el clasificador con el año 2008.

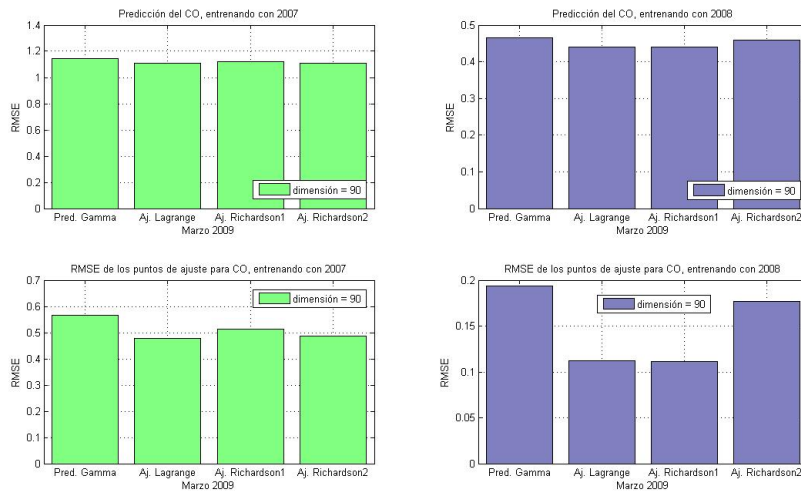


Figura 5.12: Predicción del contaminante monóxido de carbono (CO), correspondiente al mes de Marzo del 2009, entrenando el clasificador con el año 2008.

Ajuste Mes	Análisis del mes completo			Análisis de los puntos de ajuste		
	Ajuste	Ajuste	Ajuste	Ajuste	Ajuste	Ajuste
	Lagrange	Richardson1	Richardson2	Lagrange	Richardson1	Richardson2
Enero 2008	5.28 %	6.39 %	-1.14 %	39.11 %	50.65 %	-6.79 %
Marzo 2008	6.46 %	7.25 %	-4.84 %	32.33 %	37.21 %	-19.52 %
Enero 2009	3.18 %	4.53 %	-1.95 %	29.55 %	45.16 %	-12.57 %
Marzo 2009	3.25 %	1.99 %	2.91 %	16.02 %	9.6 %	14.18 %
Enero* 2009	7.42 %	8.14 %	2.08 %	43.42 %	49.24 %	10.34 %
Marzo* 2009	5.23 %	5.28 %	1.33 %	42.19 %	42.84 %	9.06 %

Cuadro 5.15: Porcentajes de mejora en la predicción del CO para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008

5.5. Predicción del ozono (O_3)

Para el contaminante ozono (O_3) se realizó la predicción correspondiente al año 2008. Los cuadros 5.16 y 5.17 muestran los resultados obtenidos, resaltando en negritas los mejores valores obtenidos con las técnicas de ajuste propuestas. Se puede observar que en general, la mejor técnica de ajuste resultó ser la interpolación de Lagrange. La figura 5.13 muestra de forma gráfica los datos representados en los cuadros 5.16 y 5.17. En la parte inferior de la gráfica se puede observar el RMSE calculado de los puntos donde se realizaron los ajustes. Con esto se resalta la mejora realizada con las técnicas de ajuste propuestas.

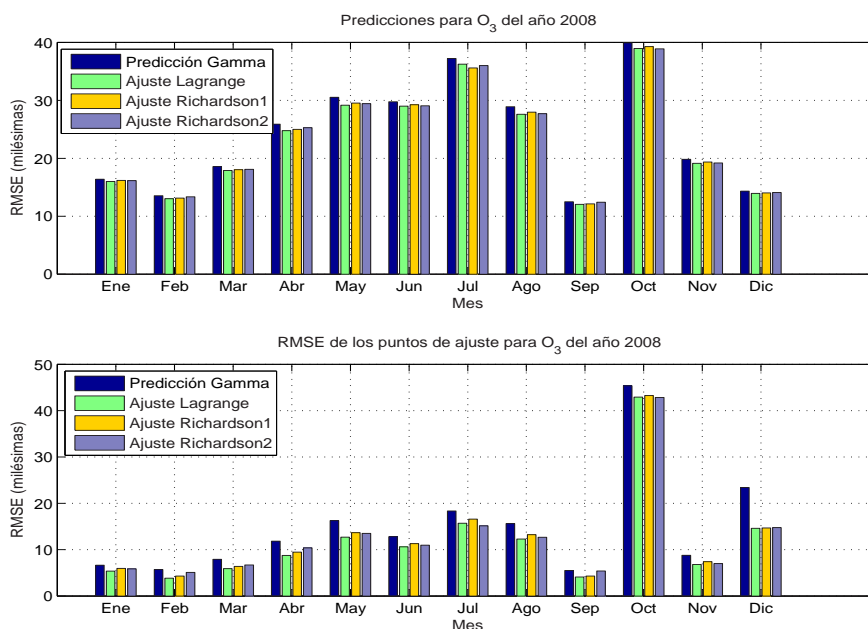


Figura 5.13: Predicción del contaminante ozono (O_3), correspondiente al año 2008, entrenando el clasificador con el año 2007.

La gráfica de la predicción realizada para el mes de Enero del 2008 se puede observar en la figura 5.14. La gráfica superior muestra la predicción realizada con el clasificador Gamma, una observación particular de este contaminante es que su comportamiento es muy variable por lo que se presen-

taron varios errores en la predicción. A pesar de esto, se puede observar en la gráfica inferior derecha cómo se logra ajustar el punto correspondiente al dato 205, mismo que se muestra en la gráfica inferior izquierda, que es una inspección más cercana de la gráfica superior.

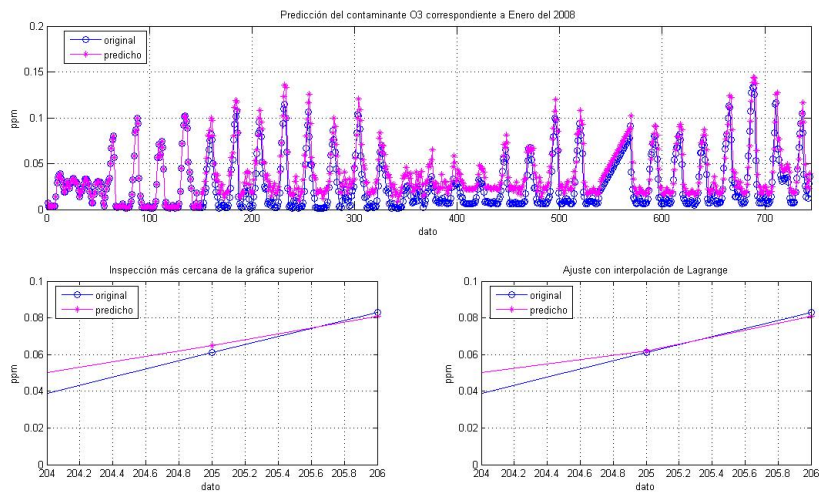


Figura 5.14: Predicción del contaminante ozono (O_3), correspondiente al mes de Enero del 2008.

Se realizaron predicciones para los meses de Enero y Marzo del 2009, entrenando el clasificador con el año 2007 y utilizando las dimensiones encontradas en el año 2008. Los resultados se muestran en el cuadro 5.18. Se

puede observar que, para ambos meses, el mejor resultado obtenido en la predicción fue utilizando la interpolación de Lagrange como técnica de ajuste. Los mejores resultados se muestran en negritas dentro de dicho cuadro.

Por otro lado, se realizaron las predicciones de los meses de Enero y Marzo del 2009, esta vez entrenando el clasificador con los valores correspondientes al año 2008. Los resultados se muestran en el cuadro 5.19. Para el mes de Enero, la mejor técnica de ajuste resultó ser la extrapolación de Richardson aplicada a los 2 valores anteriores y los dos siguientes al dato donde se detectó el cambio de signo. Mientras que para el mes de Marzo, la mejor técnica de ajuste fue la interpolación de Lagrange. Estos valores se muestran en negritas dentro del cuadro.

La representación gráfica de los cuadros 5.18 y 5.19 se muestra en las figuras 5.15 y 5.16. En la gráfica inferior de ambas figuras se observa que el mínimo error obtenido sobre los puntos donde se realizaron los ajustes fue aplicando la interpolación de Lagrange.

Un análisis de porcentajes de mejora en las técnicas de ajuste propuestas se muestra en el cuadro 5.20, estos porcentajes fueron calculados con respecto al valor obtenido con el clasificador Gamma en la mejor dimensión. Los mejores porcentajes se marcaron en negritas a fin de facilitar la observación de estos.

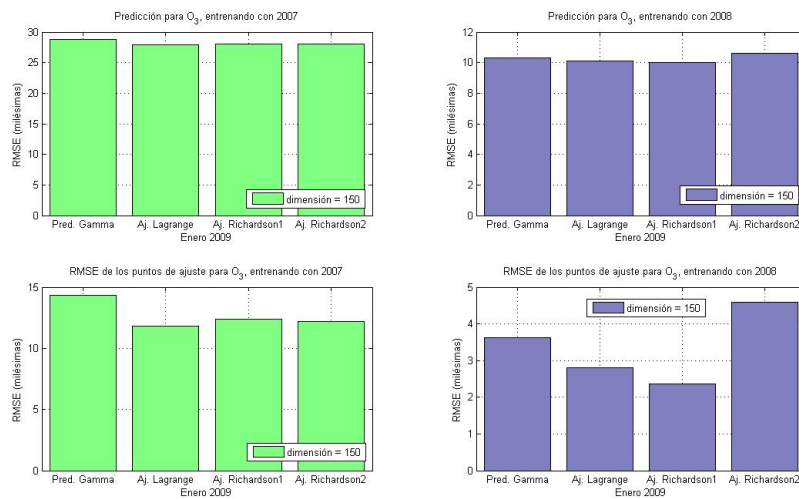


Figura 5.15: Predicción del contaminante ozono (O_3), correspondiente al mes de Enero del 2009, entrenando el clasificador con los años 2007 y 2008.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	16.38909	11.66667
		Ajuste Lagrange	16.00831	11.2043
		Ajuste Richardson1	16.18625	11.47312
		Ajuste Richardson2	16.14043	11.13441
Febrero	150	Predicción Gamma	13.54425	8.4727
		Ajuste Lagrange	13.02413	7.79454
		Ajuste Richardson1	13.13447	8.03161
		Ajuste Richardson2	13.35555	7.63793
Marzo	100	Predicción Gamma	18.55586	12.80645
		Ajuste Lagrange	17.89515	12.21237
		Ajuste Richardson1	18.03551	12.48118
		Ajuste Richardson2	18.10476	12.1371
Abril	70	Predicción Gamma	25.89208	20.86111
		Ajuste Lagrange	24.76142	19.72222
		Ajuste Richardson1	24.99908	20.11528
		Ajuste Richardson2	25.28175	19.69167
Mayo	150	Predicción Gamma	30.53609	25.12769
		Ajuste Lagrange	29.15734	23.90323
		Ajuste Richardson1	29.53725	24.33065
		Ajuste Richardson2	29.42395	23.73656
Junio	100	Predicción Gamma	29.75243	25.20139
		Ajuste Lagrange	28.99919	24.46389
		Ajuste Richardson1	29.25349	29.76667
		Ajuste Richardson2	29.07292	24.29028

Cuadro 5.16: Predicción de O₃ para el año 2008.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Julio	150	Predicción Gamma	37.22602	31.98387
		Ajuste Lagrange	36.24804	31.03898
		Ajuste Richardson1	36.58776	31.43548
		Ajuste Richardson2	36.00967	30.56452
Agosto	150	Predicción Gamma	28.90934	24.07258
		Ajuste Lagrange	27.60235	22.83065
		Ajuste Richardson1	27.95888	23.27419
		Ajuste Richardson2	27.7082	22.69624
Septiembre	150	Predicción Gamma	12.50217	7.36528
		Ajuste Lagrange	12.0643	6.72222
		Ajuste Richardson1	12.13684	7.00278
		Ajuste Richardson2	12.43081	6.39167
Octubre	150	Predicción Gamma	39.93141	34.67608
		Ajuste Lagrange	38.96083	33.71505
		Ajuste Richardson1	39.27522	34.07796
		Ajuste Richardson2	38.88894	33.47581
Noviembre	90	Predicción Gamma	19.82906	15.41389
		Ajuste Lagrange	19.13791	14.70417
		Ajuste Richardson1	19.3505	15
		Ajuste Richardson2	19.18868	14.66111
Diciembre	80	Predicción Gamma	14.32444	7.1707
		Ajuste Lagrange	13.95085	6.58333
		Ajuste Richardson1	14.02796	6.86694
		Ajuste Richardson2	14.08905	6.59274

Cuadro 5.17: Predicción de O₃ para el año 2008.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	28.8287	24.29032
		Ajuste Lagrange	27.8941	23.40323
		Ajuste Richardson1	28.08888	23.65995
		Ajuste Richardson2	28.01063	23.29973
Marzo	100	Predicción Gamma	16.57822	11.09543
		Ajuste Lagrange	15.93594	10.26882
		Ajuste Richardson1	16.05087	10.60887
		Ajuste Richardson2	16.3163	10.14382

Cuadro 5.18: Predicción del O₃ para el año 2009, entrenando el clasificador con el año 2007.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	10.300068	3.180108
		Ajuste Lagrange	10.094114	2.598118
		Ajuste Richardson1	10.004905	2.826613
		Ajuste Richardson2	10.604194	2.33871
Marzo	100	Predicción Gamma	12.229108	4.919355
		Ajuste Lagrange	11.584658	3.959677
		Ajuste Richardson1	11.587094	4.317204
		Ajuste Richardson2	12.199011	3.936828

Cuadro 5.19: Predicción del O₃ para el año 2009, entrenando el clasificador con el año 2008.

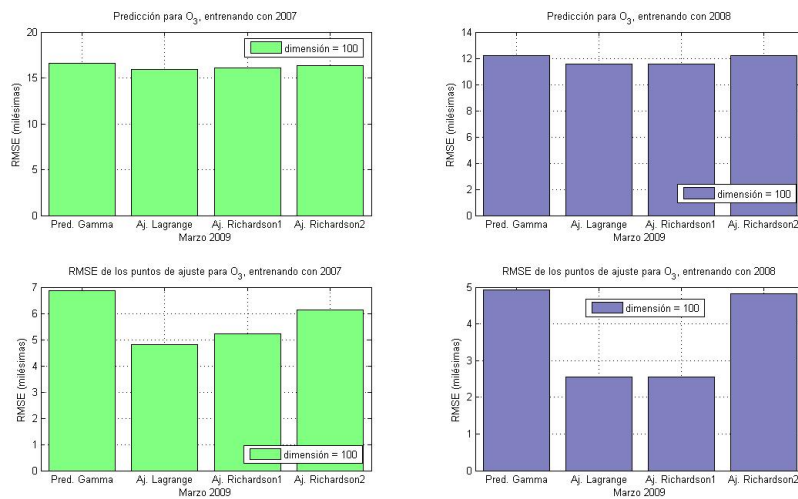


Figura 5.16: Predicción del contaminante ozono (O_3), correspondiente al mes de Marzo del 2009, entrenando el clasificador con los años 2007 y 2008.

Ajuste Mes	Análisis del mes completo			Análisis de los puntos de ajuste		
	Ajuste	Ajuste	Ajuste	Ajuste	Ajuste	Ajuste
	Lagrange	Richardson1	Richardson2	Lagrange	Richardson1	Richardson2
Enero 2008	2.32 %	1.24 %	1.52 %	19.28 %	10.33 %	11.9 %
Marzo 2008	3.56 %	2.8 %	2.43 %	25.33 %	19.44 %	15.46 %
Enero 2009	3.24 %	2.57 %	2.84 %	17.68 %	13.91 %	14.83 %
Marzo 2009	3.87 %	3.18 %	1.58 %	29.97 %	23.96 %	11.08 %
Enero* 2009	1.99 %	2.87 %	-2.95 %	22.55 %	34.62 %	-26.53 %
Marzo* 2009	5.27 %	5.25 %	.25 %	48.02 %	48.13 %	1.97 %

Cuadro 5.20: Porcentajes de mejora en la predicción del O₃ para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008

5.6. Predicción del dióxido de nitrógeno (NO₂)

Para el dióxido de nitrógeno (NO₂) se realizó la predicción de todo el año 2008, realizando la predicción en forma mensual. De igual manera, se buscó la mejor dimensión que minimizara el error en la predicción, para después aplicar las técnicas de ajuste propuestas. Los resultados se muestran en los cuadros 5.21 y 5.22.

Mes	Dimen- sión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	17.22987	9.72849
		Ajuste Lagrange	16.16406	8.42473
		Ajuste Richardson1	16.13418	8.849462
		Ajuste Richardson2	16.49075	7.86425
Febrero	150	Predicción Gamma	21.29399	15.53448
		Ajuste Lagrange	20.56508	14.4296
		Ajuste Richardson1	20.56508	14.91954
		Ajuste Richardson2	20.28543	13.75718
Marzo	60	Predicción Gamma	23.9055	18.93817
		Ajuste Lagrange	22.271288	17.81855
		Ajuste Richardson1	22.99679	18.255376
		Ajuste Richardson2	22.54175	17.29167
Abril	150	Predicción Gamma	34.6917	29.06667
		Ajuste Lagrange	28.11544	14.50278
		Ajuste Richardson1	33.64172	28.245833
		Ajuste Richardson2	33.03996	27.27778
Mayo	150	Predicción Gamma	33.99049	28.61425
		Ajuste Lagrange	31.96548	26.81855
		Ajuste Richardson1	32.44174	27.401882
		Ajuste Richardson2	31.98435	26.53091
Junio	150	Predicción Gamma	45.11421	39.50278
		Ajuste Lagrange	43.62918	38.14722
		Ajuste Richardson1	44.03594	38.580556
		Ajuste Richardson2	43.32582	37.78472

Cuadro 5.21: Predicción de NO₂ para el año 2008.

La gráfica de los cuadros 5.21 y 5.22 se muestra en la parte superior de la figura 5.17. En ella se puede observar que la mejor técnica de ajuste fue en general la interpolación de Lagrange. La gráfica inferior muestra un análisis de los puntos donde se realizaron los ajustes. En ésta, se puede observar que el RMSE en general disminuye, comparado con el RMSE de la gráfica superior. Con ello, se puede apreciar la mejora al aplicar las técnicas de ajuste propuestas.

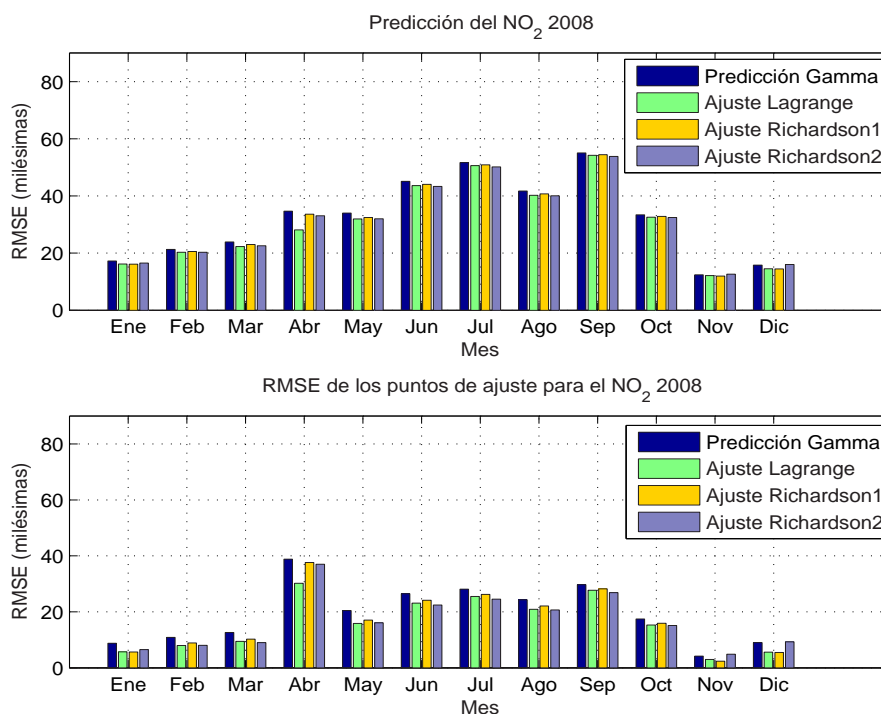


Figura 5.17: Predicción del contaminante dióxido de nitrógeno (NO₂), correspondiente al año del 2008.

La gráfica de la predicción para el mes de Enero del 2008 se aprecia en la figura 5.18. Para este mes en particular, la mejor técnica de ajuste resultó ser la extrapolación de Richardson aplicada en los 2 valores anteriores y siguientes al dato donde se detectó el cambio de signo, como se puede apreciar en la gráfica inferior derecha. En esta gráfica aunque no se logra un ajuste perfecto, los valores se aproximan en general a los valores originales.

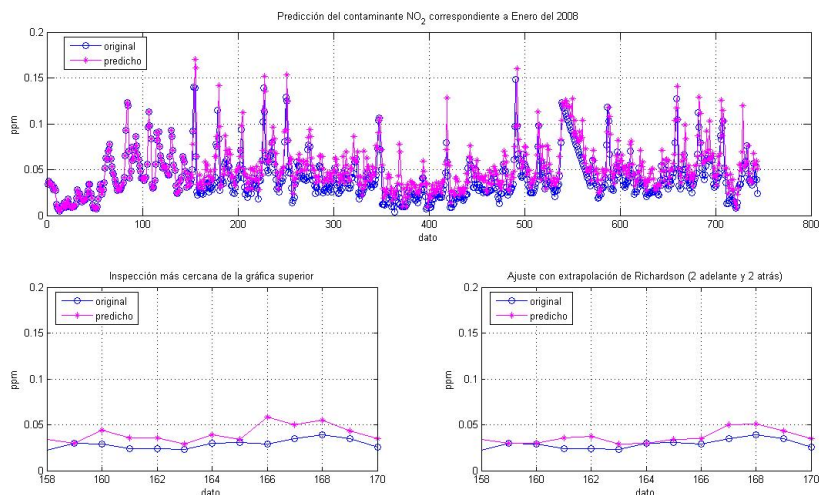


Figura 5.18: Predicción del contaminante dióxido de nitrógeno (NO₂), correspondiente al mes de Enero del 2008.

Se realizaron también predicciones del año 2009 para los meses de Enero y Marzo; esta vez en dos modalidades: entrenando el clasificador con el año 2007 y entrenando el clasificador con el año 2008. Los resultados de estas predicciones se muestran en los cuadros 5.23 y 5.24, donde los mejores valores se marcan en negrillas con el fin de facilitar su observación. En ellos se puede observar que la predicción del mes de Enero no varió mucho, en cambio, para el mes de Marzo hubo una mejora considerable en la predicción obtenida con

el entrenamiento del año 2008.

Las gráficas de los cuadros 5.23 y 5.24 se muestran en las figuras 5.19 y 5.20. En la parte superior de estas figuras se muestra el RMSE de la predicción obtenida en el mes completo junto con las técnicas de ajuste propuestas; mientras que la parte inferior muestra solamente un análisis del RMSE obtenido sobre los puntos donde se realizaron los ajustes con las técnicas propuestas.

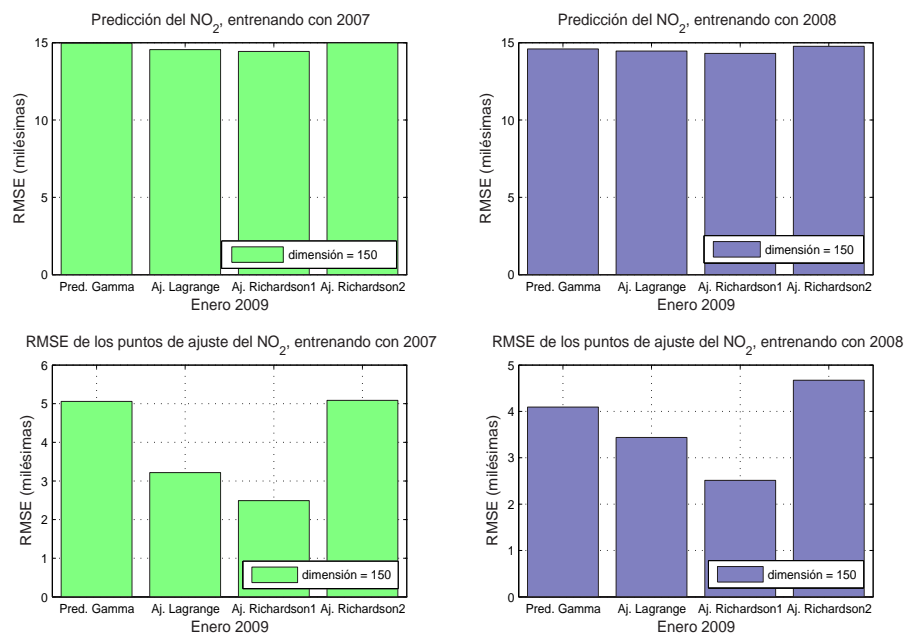


Figura 5.19: Predicción del contaminante dióxido de nitrógeno (NO_2), correspondiente al mes de Enero del 2009.

Un análisis de porcentajes de mejora en las técnicas de ajuste propuestas se muestra en el cuadro 5.25, estos porcentajes fueron calculados con respecto al valor obtenido con el clasificador Gamma en la mejor dimensión. Los mejores porcentajes se marcaron en negritas a fin de facilitar la observación de estos.

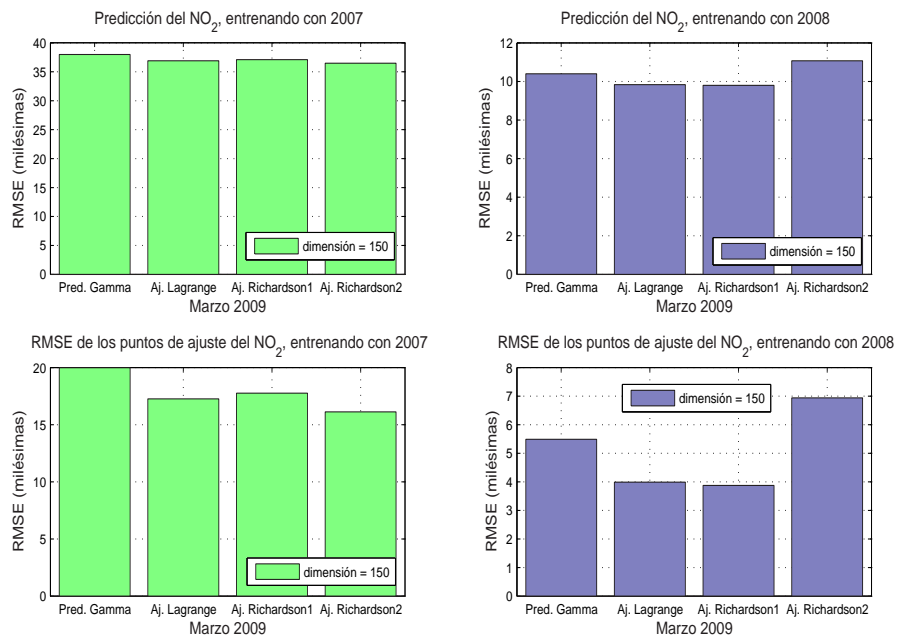


Figura 5.20: Predicción del contaminante dióxido de nitrógeno (NO₂), correspondiente mes de Marzo del 2009.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Julio	150	Predicción Gamma	51.6654	45.65995
		Ajuste Lagrange	50.5892	44.6922
		Ajuste Richardson1	50.90449	45.017473
		Ajuste Richardson2	50.15455	44.18011
Agosto	150	Predicción Gamma	41.70085	36.26747
		Ajuste Lagrange	40.21406	34.891133
		Ajuste Richardson1	40.72907	35.467742
		Ajuste Richardson2	40.0735	34.64382
Septiembre	150	Predicción Gamma	55.03134	48.59028
		Ajuste Lagrange	54.21115	47.85417
		Ajuste Richardson1	54.42556	48.066667
		Ajuste Richardson2	53.85099	47.39306
Octubre	150	Predicción Gamma	33.41642	28.90726
		Ajuste Lagrange	32.56934	28.18414
		Ajuste Richardson1	32.82527	28.466398
		Ajuste Richardson2	32.47863	27.78629
Noviembre	100	Predicción Gamma	12.40699	1.95833
		Ajuste Lagrange	12.10733	1.28194
		Ajuste Richardson1	11.98263	1.647222
		Ajuste Richardson2	12.63351	1.05
Diciembre	150	Predicción Gamma	15.82023	6.26075
		Ajuste Lagrange	14.51677	4.56989
		Ajuste Richardson1	14.47676	5.125
		Ajuste Richardson2	16.88873	3.97043

Cuadro 5.22: Predicción de NO₂ para el año 2008.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	14.9616	2.20968
		Ajuste Lagrange	14.54937	1.84005
		Ajuste Richardson1	14.43194	2.1922
		Ajuste Richardson2	14.98794	17.5672
Marzo	150	Predicción Gamma	37.98985	30.87097
		Ajuste Lagrange	36.91009	29.77957
		Ajuste Richardson1	37.10081	30.13441
		Ajuste Richardson2	36.48236	29.18817

Cuadro 5.23: Predicción del NO₂ para el año 2009, entrenando el clasificador con el año 2007.

Mes	Dimensión	Técnica	RMSE (milésimas)	BIAS (milésimas)
Enero	150	Predicción Gamma	14.59779	.70565
		Ajuste Lagrange	14.46232	.48387
		Ajuste Richardson1	14.30984	.76344
		Ajuste Richardson2	14.76441	.25941
Marzo	150	Predicción Gamma	10.39845	3.18952
		Ajuste Lagrange	9.837729	2.133065
		Ajuste Richardson1	9.799811	2.436828
		Ajuste Richardson2	11.070765	1.475806

Cuadro 5.24: Predicción del NO₂ para el año 2009, entrenando el clasificador con el año 2008.

Ajuste Mes	Análisis del mes completo			Análisis de los puntos de ajuste		
	Ajuste Lagrange	Ajuste Richardson1	Ajuste Richardson2	Ajuste Lagrange	Ajuste Richardson1	Ajuste Richardson2
Enero 2008	6.19 %	6.36 %	4.29 %	35.15 %	35.46 %	26.2 %
Marzo 2008	6.84 %	3.8 %	5.7 %	25.07 %	18.6 %	28.9 %
Enero 2009	2.76 %	3.54 %	-1.18 %	36.39 %	50.74 %	-1.56 %
Marzo 2009	2.84 %	2.34 %	3.97 %	13.62 %	11.13 %	19.27 %
Enero* 2009	.93 %	1.97 %	-1.14 %	15.99 %	38.61 %	-14.18 %
Marzo* 2009	5.39 %	5.76 %	-6.46 %	27.32 %	29.4 %	-26.44 %

Cuadro 5.25: Porcentajes de mejora en la predicción del NO₂ para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008

5.7. Predicción de partículas menores a 10 micrómetros (PM₁₀)

Se realizaron varias predicciones con respecto al contaminante PM₁₀. La primera fue la predicción del año 2008, que se muestra en los cuadros 5.26 y 5.27. Los valores de este contaminante están dados en microgramos sobre metro cúbico ($\mu g/m^3$). La representación gráfica de estos cuadros se muestra en la figura 5.21. En ella se puede observar que en la mayor parte de los meses la mejor técnica de ajuste fue la de la fórmula de Lagrange. La gráfica inferior de la gráfica muestra el RMSE calculado sobre los puntos donde se detectaron cambios de signo y se aplicaron las técnicas de ajuste propuestas.

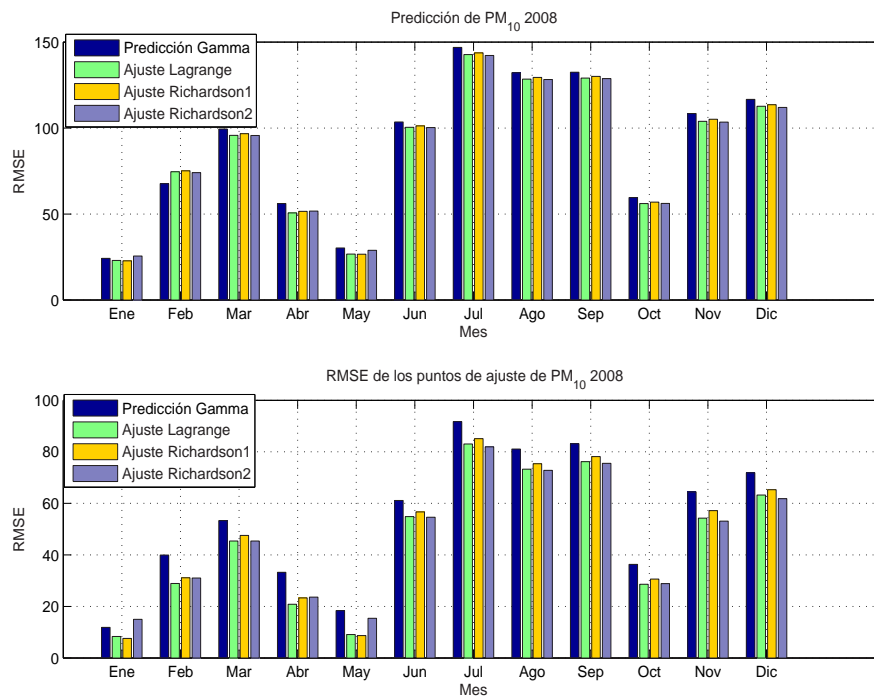


Figura 5.21: Predicción de partículas menores a diez micrómetros (PM₁₀), correspondiente al año 2008.

La predicción para el mes de Enero del 2008 se muestra en la gráfica su-

perior de la figura 5.23. La gráfica inferior izquierda muestra una inspección cercana de una región de la gráfica superior. Mientras que la gráfica inferior derecha muestra la misma región ajustada con la fórmula de Richardson aplicada a los 2 valores anteriores y siguientes de donde se detectó el cambio de signo.

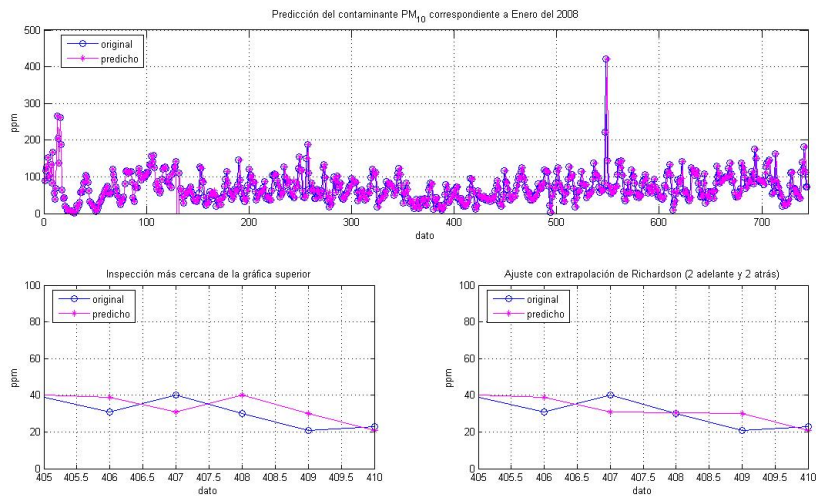


Figura 5.22: Predicción de partículas menores a diez micrómetros (PM_{10}), correspondiente a Enero del 2008.

Otras pruebas realizadas fueron las predicciones de los meses de Enero y Marzo del 2009, utilizando las dimensiones encontradas en el 2008 para cada

mes. Estos resultados se muestran en los cuadros 5.28 y 5.29. Los mejores resultados se marcan en negritas, para facilitar la observación de éstos. Se puede observar que existe mayor error en la predicción cuando el conjunto fundamental está conformado por los datos del 2007.

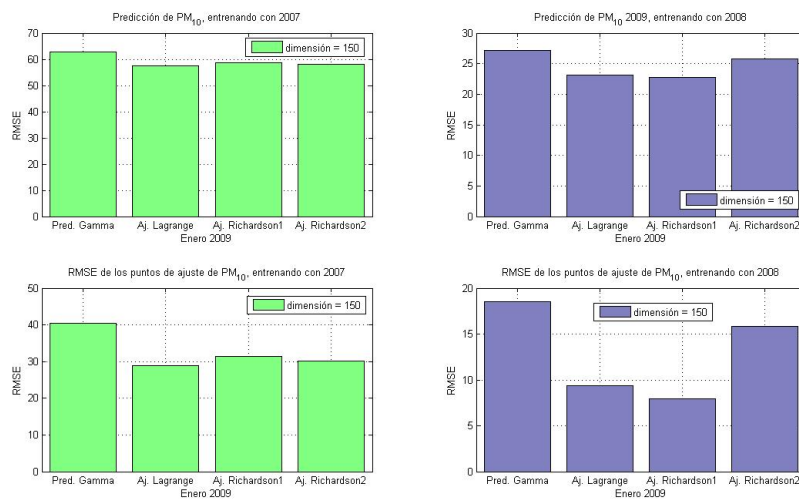


Figura 5.23: Predicción del contaminante PM_{10} , correspondiente al mes de Enero del 2009.

La representación gráfica de los cuadros 5.28 y 5.29 se muestran en las figuras 5.23 y 5.24. En ambos meses se observa menor error de predicción cuando se entrena el clasificador con los datos del año 2008. Con esto se pue-

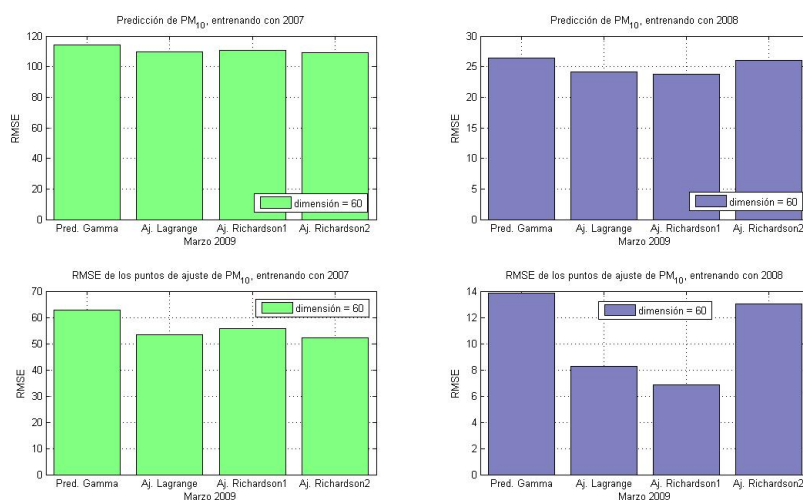


Figura 5.24: Predicción del contaminante (PM_{10}), correspondiente mes de Marzo del 2009.

de decir que el comportamiento del contaminante fue parecido para ambos meses en los años 2008 y 2009.

Un análisis de porcentajes de mejora en las técnicas de ajuste propuestas se muestra en el cuadro 5.30, estos porcentajes fueron calculados con respecto al valor obtenido con el clasificador Gamma en la mejor dimensión. Los mejores porcentajes se marcaron en negritas a fin de facilitar la observación de estos.

Mes	Dimensión	Técnica	RMSE	BIAS
Enero	150	Predicción Gamma	24.2407836	.13172043
		Ajuste Lagrange	23.0275335	-1.27643145
		Ajuste Richardson1	22.8194223	-.72506586
		Ajuste Richardson2	25.5873252	-2.06869624
Febrero	150	Predicción Gamma	67.6461062	53.7341954
		Ajuste Lagrange	74.6532415	49.7710718
		Ajuste Richardson1	75.1495771	50.8512141
		Ajuste Richardson2	74.1016743	48.8513721
Marzo	60	Predicción Gamma	99.4500604	91.7526882
		Ajuste Lagrange	95.7818844	88.7121398
		Ajuste Richardson1	96.7266597	89.684672
		Ajuste Richardson2	95.7739085	87.7047513
Abril	100	Predicción Gamma	56.0734389	42.858333
		Ajuste Lagrange	50.6792243	38.5159736
		Ajuste Richardson1	51.58277	40.0382014
		Ajuste Richardson2	51.7314247	37.8509222
Mayo	150	Predicción Gamma	30.2770363	9.09946237
		Ajuste Lagrange	26.7128286	5.57907124
		Ajuste Richardson1	26.6002091	6.7460457
		Ajuste Richardson2	28.930754	9.2502823
Junio	90	Predicción Gamma	103.558391	95.1013889
		Ajuste Lagrange	100.443532	92.4768542
		Ajuste Richardson1	101.315394	93.3567917
		Ajuste Richardson2	100.323684	91.9618722

Cuadro 5.26: Predicción de PM₁₀ para el año 2008.

Mes	Dimensión	Técnica	RMSE	BIAS
Julio	150	Predicción Gamma	146.932657	130.288978
		Ajuste Lagrange	142.769828	126.621418
		Ajuste Richardson1	143.734422	127.550934
		Ajuste Richardson2	142.260252	125.894257
Agosto	150	Predicción Gamma	132.173634	116.931452
		Ajuste Lagrange	128.49461	113.719085
		Ajuste Richardson1	129.458658	114.688621
		Ajuste Richardson2	128.286567	113.352593
Septiembre	150	Predicción Gamma	132.479228	117.068056
		Ajuste Lagrange	129.129238	114.03495
		Ajuste Richardson1	130.068119	114.943907
		Ajuste Richardson2	128.802895	113.624839
Octubre	150	Predicción Gamma	59.5630031	50.2271505
		Ajuste Lagrange	56.1029944	47.5109825
		Ajuste Richardson1	56.9412438	48.3968804
		Ajuste Richardson2	56.2020572	46.7685121
Noviembre	150	Predicción Gamma	108.441721	93.6736111
		Ajuste Lagrange	103.93991	89.6687514
		Ajuste Richardson1	105.78518	91.0576486
		Ajuste Richardson2	103.456633	88.6067889
Diciembre	150	Predicción Gamma	116.726072	102.11828
		Ajuste Lagrange	112.678804	98.7007191
		Ajuste Richardson1	103.622414	99.7001237
		Ajuste Richardson2	112.061979	97.617078

Cuadro 5.27: Predicción de PM₁₀ para el año 2008.

Capítulo 5 5.7. Predicción de partículas menores a 10 micrómetros (PM_{10})

Mes	Dimen- sión	Técnica	RMSE	BIAS
Enero	150	Predicción Gamma	62.826609	51.266129
		Ajuste Lagrange	57.5710093	47.0103078
		Ajuste Richardson1	58.5436171	48.2904745
		Ajuste Richardson2	58.0804805	46.2152016
Marzo	60	Predicción Gamma	114.06755	106.592359
		Ajuste Lagrange	109.592359	102.523296
		Ajuste Richardson1	110.594788	103.675036
		Ajuste Richardson2	109.001358	101.484614

Cuadro 5.28: Predicción de PM_{10} para el año 2009, entrenando el clasificador con el año 2007.

Mes	Dimen- sión	Técnica	RMSE	BIAS
Enero	150	Predicción Gamma	27.1576689	5.5336012
		Ajuste Lagrange	23.0989049	2.49260753
		Ajuste Richardson1	22.6899568	3.56160618
		Ajuste Richardson2	25.7943354	1.7928629
Marzo	60	Predicción Gamma	26.3916429	6.69623656
		Ajuste Lagrange	24.1332108	3.29815995
		Ajuste Richardson1	23.7242909	4.41674059
		Ajuste Richardson2	25.9895802	2.41934946

Cuadro 5.29: Predicción de PM_{10} para el año 2009, entrenando el clasificador con el año 2008.

Capítulo 5 5.7. Predicción de partículas menores a 10 micrómetros (PM₁₀)

Ajuste Mes	Análisis del mes completo			Análisis de los puntos de ajuste		
	Ajuste Lagrange	Ajuste Richardson1	Ajuste Richardson2	Ajuste Lagrange	Ajuste Richardson1	Ajuste Richardson2
Enero 2008	23 %	5.86 %	-5.55 %	29.89 %	36.04 %	-26.12 %
Marzo 2008	3.69 %	2.74 %	3.7 %	14.8 %	10.81 %	14.84 %
Enero 2009	8.37 %	6.82 %	7.55 %	28.29 %	22.48 %	25.2 %
Marzo 2009	3.92 %	3.04 %	4.44 %	14.88 %	11.38 %	16.99 %
Enero* 2009	14.95 %	16.45 %	5.02 %	49.58 %	56.9 %	14.2 %
Marzo* 2009	8.56 %	10.11 %	1.52 %	40.5 %	50.69 %	6.15 %

Cuadro 5.30: Porcentajes de mejora en la predicción de PM₁₀ para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008

5.8. Predicción de partículas menores a 2.5 micrómetros ($PM_{2.5}$)

La predicción de partículas menores a 2.5 micrómetros se realizó para el año 2008 en cada uno de los meses. Los resultados se muestran en los cuadros 5.31 y 5.32, donde los mejores valores aparecen en negritas. La representación gráfica de esta predicción se muestra en la figura 5.25. En ella se puede apreciar que en la mayoría de los meses la mejor técnica de ajuste fue al utilizar la fórmula de Lagrange. La gráfica superior muestra el RMSE obtenido en cada uno de los meses. Mientras que la gráfica inferior de la figura muestra el RMSE obtenido sobre los puntos donde se realizaron los ajustes.

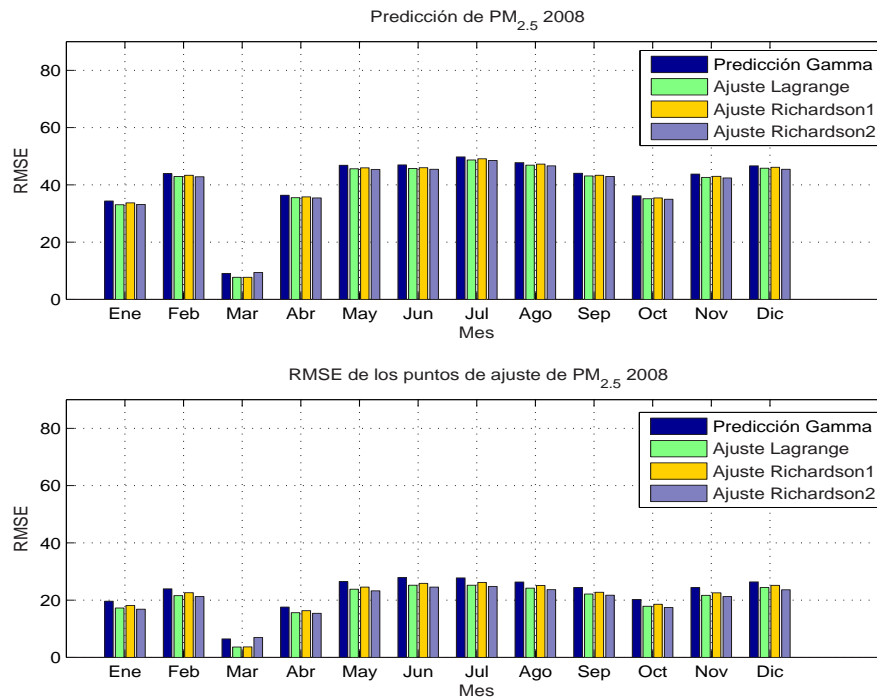


Figura 5.25: Predicción de partículas menores a 2.5 micrómetros ($PM_{2.5}$), correspondiente al año 2008.

La predicción para el mes de marzo del 2008 se muestra en la gráfica superior de la figura 5.26. En la gráfica inferior izquierda se muestra una

Capítulo 55.8. Predicción de partículas menores a 2.5 micrómetros ($PM_{2.5}$)

parte de la gráfica superior, mientras que la gráfica inferior derecha muestra la misma región ajustada con la mejor técnica, en este caso, la interpolación de Lagrange.

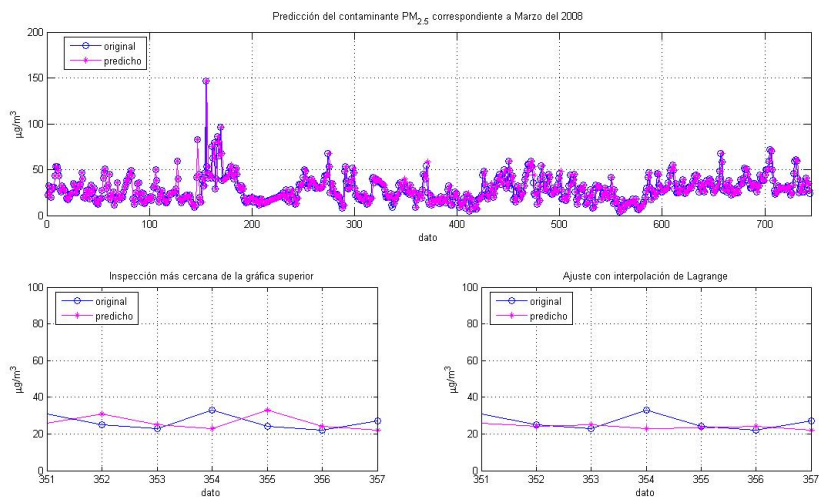


Figura 5.26: Predicción del contaminante $PM_{2.5}$, correspondiente al mes de Marzo del 2008.

Otra prueba realizada con este contaminante fue la predicción de los meses de Enero y Marzo del 2009. Entrenando el clasificador con el año 2007 y después con el año 2008. Los resultados se muestran en los cuadros 5.33 y 5.34. En donde se puede observar que el menor error de predicción para

Capítulo 55.8. Predicción de partículas menores a 2.5 micrómetros ($PM_{2.5}$)

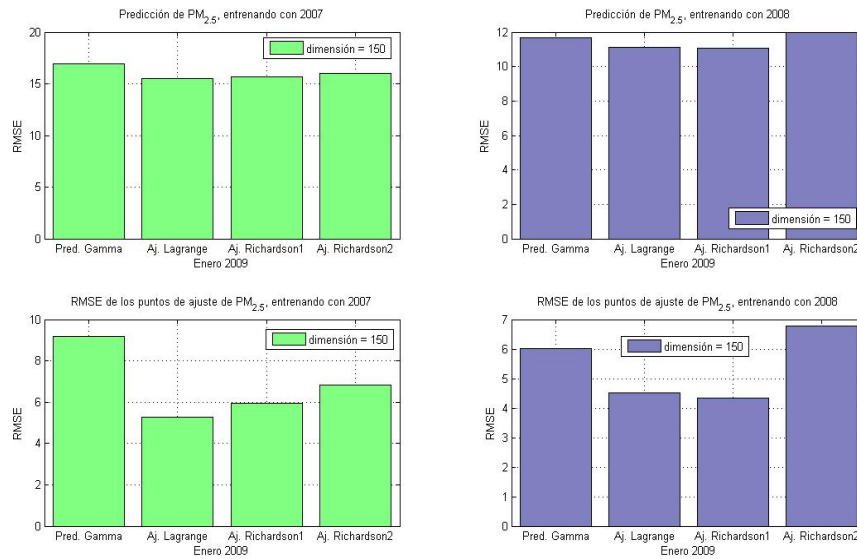


Figura 5.27: Predicción del contaminante $PM_{2.5}$, correspondiente al mes de Enero del 2009.

ambos meses se obtiene al entrenar el clasificador con patrones que estén conformados con datos de un año anterior. La representación gráfica de estos cuadros se muestra en las figuras 5.27 y 5.28.

Un análisis de porcentajes de mejora en las técnicas de ajuste propuestas se muestra en el cuadro 5.35, estos porcentajes fueron calculados con respecto al valor obtenido con el clasificador Gamma en la mejor dimensión. Los mejores porcentajes se marcaron en negritas a fin de facilitar la observación de estos.

Capítulo 55.8. Predicción de partículas menores a 2.5 micrómetros ($PM_{2.5}$)

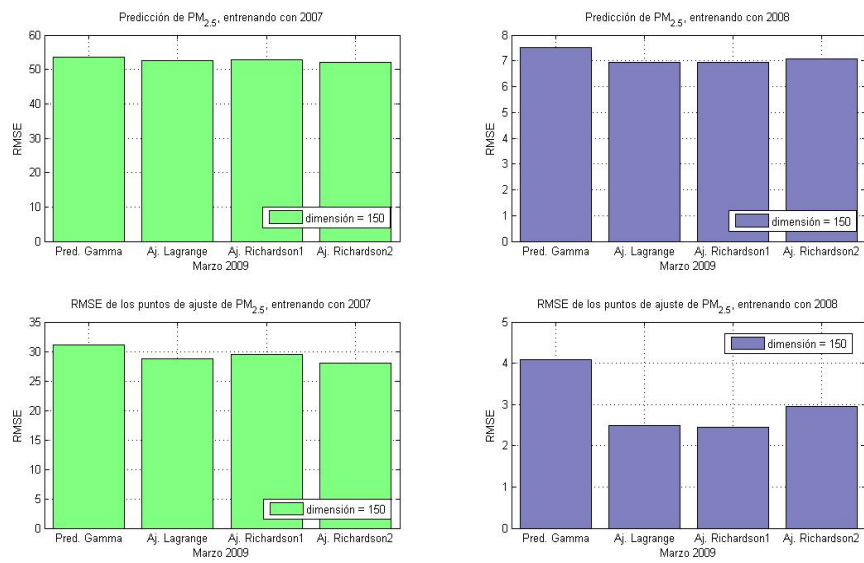


Figura 5.28: Predicción del contaminante ($PM_{2.5}$), correspondiente mes de Marzo del 2009.

Mes	Dimensión	Técnica	RMSE	BIAS
Enero	150	Predicción Gamma	34.348799	29.891129
		Ajuste Lagrange	33.0669721	29
		Ajuste Richardson1	33.6748555	29
		Ajuste Richardson2	33.1366919	29
Febrero	150	Predicción Gamma	43.9106329	38.1752874
		Ajuste Lagrange	42.9227117	37.2634052
		Ajuste Richardson1	43.3457807	37.7129511
		Ajuste Richardson2	42.7978454	36.9706264
Marzo	150	Predicción Gamma	9.05805647	.08333
		Ajuste Lagrange	7.678719985	-.42069489
		Ajuste Richardson1	7.69528927	-.194
		Ajuste Richardson2	9.35768445	-.68333
Abril	50	Predicción Gamma	36.3522313	33.2597222
		Ajuste Lagrange	35.520401	33
		Ajuste Richardson1	35.8048798	33
		Ajuste Richardson2	35.4150891	32
Mayo	150	Predicción Gamma	46.8050464	41.0860215
		Ajuste Lagrange	45.6204777	40.0443481
		Ajuste Richardson1	45.94627	40.3794879
		Ajuste Richardson2	45.3928547	39.7083199
Junio	150	Predicción Gamma	46.9349875	41.270833
		Ajuste Lagrange	45.72092919	40.2236097
		Ajuste Richardson1	45.9973904	40.5041611
		Ajuste Richardson2	45.4291488	39.8428917

Cuadro 5.31: Predicción de PM_{2.5} para el año 2008.

Mes	Dimensión	Técnica	RMSE	BIAS
Julio	150	Predicción Gamma	49.7658901	44.0282258
		Ajuste Lagrange	48.66490094	43.0132164
		Ajuste Richardson1	49.0718921	43.4274933
		Ajuste Richardson2	48.4841243	42.7634355
Agosto	150	Predicción Gamma	47.7571421	42.0268817
		Ajuste Lagrange	46.84752554	41.191082
		Ajuste Richardson1	47.2217772	41.5877406
		Ajuste Richardson2	46.6280469	40.878125
Septiembre	150	Predicción Gamma	44.0424669	38.5444
		Ajuste Lagrange	43.08508535	37.730325
		Ajuste Richardson1	43.3377798	38.0250861
		Ajuste Richardson2	42.903566	37.5322444
Octubre	150	Predicción Gamma	36.139589	31.3924731
		Ajuste Lagrange	35.12088762	30.4778212
		Ajuste Richardson1	35.4196779	30.8093602
		Ajuste Richardson2	34.9517548	30.1123011
Noviembre	150	Predicción Gamma	43.7222293	38.0972222
		Ajuste Lagrange	42.56553364	37.0844875
		Ajuste Richardson1	42.9304599	37.4681361
		Ajuste Richardson2	42.4033063	36.7198931
Diciembre	150	Predicción Gamma	46.6484652	41.0416667
		Ajuste Lagrange	45.78293733	40.2681465
		Ajuste Richardson1	46.1075305	40.6014812
		Ajuste Richardson2	45.4378824	39.8233226

Cuadro 5.32: Predicción de PM_{2.5} para el año 2008.

Capítulo 55.8. Predicción de partículas menores a 2.5 micrómetros ($PM_{2.5}$)

Mes	Dimen- sión	Técnica	RMSE	BIAS
Enero	150	Predicción Gamma	16.9089911	11
		Ajuste Lagrange	15.505938	10
		Ajuste Richardson1	15.7090768	10
		Ajuste Richardson2	15.9859947	9
Marzo	150	Predicción Gamma	53.5621059	47.5174731
		Ajuste Lagrange	52.4728927	46.5058266
		Ajuste Richardson1	52.7799188	46.8125766
		Ajuste Richardson2	52.1743978	46.1614234

Cuadro 5.33: Predicción de $PM_{2.5}$ para el año 2009, entrenando el clasificador con el año 2007.

Mes	Dimen- sión	Técnica	RMSE	BIAS
Enero	150	Predicción Gamma	11.6548711	1.30376344
		Ajuste Lagrange	11.0983348	.41375538
		Ajuste Richardson1	11.0456451	.7258871
		Ajuste Richardson2	11.9912688	.0221774
Marzo	150	Predicción Gamma	7.5200449	4.22580645
		Ajuste Lagrange	6.9426264	3.50851613
		Ajuste Richardson1	6.92892823	3.75252688
		Ajuste Richardson2	7.08303758	3.19802419

Cuadro 5.34: Predicción de $PM_{2.5}$ para el año 2009, entrenando el clasificador con el año 2008.

Capítulo 55.8. Predicción de partículas menores a 2.5 micrómetros (PM_{2.5})

Ajuste Mes	Análisis del mes completo			Análisis de los puntos de ajuste		
	Ajuste	Ajuste	Ajuste	Ajuste	Ajuste	Ajuste
	Lagrange	Richardson1	Richardson2	Lagrange	Richardson1	Richardson2
Enero 2008	3.73 %	1.96	3.53 %	12.21 %	7.76 %	14.33 %
Marzo 2008	15.23 %	15.04 %	3.3 %	44.55 %	43.8 %	-7.57 %
Enero 2009	8.3 %	7.09 %	5.46 %	42.81 %	35.26 %	25.65 %
Marzo 2009	2.03 %	1.46 %	2.59 %	7.76 %	5.53 %	9.96 %
Enero* 2009	4.78 %	5.22 %	-2.89 %	25.08 %	27.82 %	-12.94 %
Marzo* 2009	7.68 %	7.86 %	5.81 %	38.94 %	40.27 %	27.87 %

Cuadro 5.35: Porcentajes de mejora en la predicción de PM_{2.5} para los años 2008 y 2009, entrenando el clasificador con el año 2007. * Entrenando el clasificador con el año 2008

5.9. Análisis de las predicciones realizadas sobre los contaminantes atmosféricos

En las secciones 5.2 a 5.8, se presentaron los resultados de las predicciones realizadas con el clasificador Gamma y las técnicas de ajuste aplicadas que mejoraron la predicción realizada con dicho clasificador. El cuadro 5.36 muestra el desempeño del clasificador Gamma en conjunto con la técnica de ajuste que mejor resultado dio para la predicción del año 2008, comparado con otras técnicas de predicción realizadas en diferentes partes del mundo.

Como se puede observar en el cuadro 5.36, se presenta otra medida de evaluación de desempeño de los algoritmos, que no fue determinada como tal para la presente tesis pero que, sin embargo, otros trabajos si lo utilizaron; por ende, se decidió también realizar el cálculo correspondiente con el fin de tener un valor comparable. Dicha medida se llama Error Medio Cuadrático (MSE) y es el promedio del error elevado al cuadrado entre el valor predicho y el valor original. El MSE se determina mediante la fórmula 5.3.

$$MSE = \frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2 \quad (5.3)$$

donde P_i es el valor predicho, O_i es el valor original y n es el número de muestras evaluadas.

Otra observación en los datos mostrados en el cuadro 5.36, es la unidad en que se miden los contaminantes. La mayoría de los artículos no especificaban las unidades, esto hace un poco desapareja la comparación entre estos algoritmos y la metodología propuesta en la presente tesis. Sin embargo, en la predicción del contaminante ozono (O_3), existe una comparación directa entre el algoritmo de Razonamiento Fuzzy Inductivo (FIR) y el clasificador Gamma con la técnica de ajuste de Interpolación de Lagrange. En esta comparación se puede observar que la predicción realizada con el clasificador en conjunto con la técnica de ajuste propuesta fue mejor.

De acuerdo con las pruebas presentadas en el presente trabajo, se puede determinar que en general, la interpolación de Lagrange y la extrapolación de Richardson, ambas aplicadas en los dos valores anteriores y los dos siguientes a los puntos donde ocurrieron cambios de signo, son las mejores opciones

Algoritmo	Contaminante	Unidad	Resultado RMSE	Resultado MSE
TD-CMAC [20]	CO	ND	1.589	ND
TW-CMAC [20]	CO	ND	1.721	ND
Gamma con Ajuste Richardson1	CO	ppm	0.32070131	0.012194892
TD-CMAC [20]	SO ₂	ND	5.973	ND
TW-CMAC [20]	SO ₂	ND	11.129	ND
Online SVM [26]	SO ₂	$\mu\text{g}/\text{m}^3$	12.96, 10.9	ND
Gamma con Ajuste Lagrange	SO ₂	ppm	0.009403657	0.0000884288
TD-CMAC [20]	NO ₂	ND	6.87	ND
TW-CMAC [20]	NO ₂	ND	12.337	ND
Gamma con Ajuste Richardson1	NO ₂	ppm	0.01198263	0.00014358
FIR [18]	O ₃	ppm	0.0176	ND
DPNN [13]	O ₃	ND	13.0403	ND
lm [14]	O ₃	ND	20.3	ND
gam [14]	O ₃	ND	14.1	ND
loess [14]	O ₃	ND	13.2	ND
ppr1 [14]	O ₃	ND	15.7	ND
ppr3 [14]	O ₃	ND	11.8	ND
nn311 [14]	O ₃	ND	16.3	ND
nn331x [14]	O ₃	ND	13.2	ND
EFP [25]	O ₃	ppm	ND	25.4041
Comp. Link [25]	O ₃	ppm	ND	25.0474
Ward Link [25]	O ₃	ppm	ND	28.4126
K-means [25]	O ₃	ppm	ND	31.1698
Red Neuronal [9]	O ₃	$\mu\text{g}/\text{m}^3$	9.43	ND
Red Neuronal [8]	O ₃	ppb	15	ND
Gamma con Ajuste Lagrange	O ₃	ppm	0.0120643	0.0001455473
Redes Neuronales con Lógica Difusa [19]	PM ₁₀	μm	ND	0.87
Gamma con Ajuste Richardson1	PM ₁₀	$\mu\text{g}/\text{m}^3$	22.8194223	520.726
Gamma con Ajuste Lagrange	PM _{2.5}	$\mu\text{g}/\text{m}^3$	7.678719985	58.9627406
Gamma con Ajuste Lagrange	NO _x	ppm	0.022623	0.0005118

Cuadro 5.36: Comparación de las predicciones realizadas con otros métodos computacionales. *ND*, No Definido. *ppm*, partes por millón. $\mu\text{g}/\text{m}^3$, microgramos sobre metro cúbico. μm , micrómetros.

para obtener una predicción con menos error.

En las predicciones realizadas para los meses de Enero y Marzo del año 2009, en general, para todos los contaminantes analizados, la predicción que menos error dio fue al determinar el conjunto fundamental con los datos del año 2008. Con ello se puede deducir que el comportamiento de los contaminantes para estos meses, varía anualmente, es decir, los datos obtenidos 2 años atrás no son los mismos a los obtenidos actualmente.

En las predicciones realizadas para algunos contaminantes del mes de Marzo del 2009, se observa un menor error al entrenar el clasificador con los datos del año 2007. Con esto se puede deducir que el comportamiento del contaminante para el año 2009 fue parecido al del año 2007. Esto se puede deber a factores externos, por ejemplo, que haya existido o no período vacacional en esa fecha y por consiguiente los niveles de contaminación hayan variado.

5.10. Análisis de desempeño computacional

El desempeño computacional es un factor importante en la predicción de contaminantes atmosféricos. Hay que ver la importancia de si se desea mayor rapidez o mayor precisión en la obtención de resultados. En nuestro caso, la precisión es más importante que la rapidez.

Se realizó un estudio de tiempos de implementación del clasificador Gamma. Las pruebas fueron realizadas en una máquina con procesador Intel Celerino de 2 GB de RAM, utilizando el programa Borland C++ 5.02. Los resultados se muestran en el cuadro 5.37. Donde se puede observar que el mayor tiempo de procesamiento se obtiene cuando los patrones del conjunto fundamental y de prueba tienen una dimensión de 150. A pesar de ello, no es muy tardado el procedimiento.

De igual manera, se midieron los tiempos de procesamiento de los algoritmos de ajuste de los resultados arrojados por el clasificador. Obteniéndose para la interpolación de Lagrange un tiempo de procesamiento de 4 segundos. Mientras que el algoritmo de la extrapolación de Richardson tuvo un tiempo

	744 Patrones	720 Patrones
Dimensión	Tiempo (seg.)	Tiempo (seg.)
10	8.8	34.6
20	32.8	69.4
30	50.5	95.9
40	77.8	143.6
50	100.6	178.2
60	124.5	215
70	148	250.9
80	170	288
90	191	316.7
100	215	350.8
150	323.3	495

Cuadro 5.37: Tiempos de procesamiento del clasificador Gamma

de procesamiento de 5.3 segundos. Ambos algoritmos fueron implementados en Borland C++ 5.02. En general, se puede decir que la predicción realizada con las características descritas es rápida y eficiente.

Capítulo 6

Conclusiones y Trabajo Futuro

En este capítulo se darán las conclusiones generales de la presente tesis desarrollada. De igual manera, se presentarán algunas propuestas de trabajo futuro relacionadas con este trabajo.

6.1. Conclusiones

El modelo propuesto para mejorar la predicción de contaminantes atmosféricos resultó eficaz, ya que se pudo lograr la disminución de error en la predicción.

Los contaminantes analizados demostraron variabilidad en su comportamiento mensual, por lo que la predicción y búsqueda de dimensión ideal en los patrones de los conjuntos fundamental y de prueba se realizaron de forma mensual.

Las técnicas de ajuste implementadas como la interpolación de Lagrange y la extrapolación de Richardson, aplicada en los dos datos anteriores y los dos datos siguientes al dato donde se detectó un cambio de signo, demostraron mejoras en las predicciones obtenidas con el clasificador Gamma.

Por otro lado, la extrapolación de Richardson calculada con los dos valores anteriores al dato donde se realizó el cambio de signo, mostró un desempeño inferior al de las otras técnicas de ajuste propuestas. Resultando algunas ve-

ces con mayor error que con el resultado obtenido con el clasificador Gamma.

En los experimentos realizados se observó que para obtener una mejor precisión en la predicción de contaminantes atmosféricos, la dimensión de los patrones del conjunto fundamental y de prueba debe estar entre 50 y 150.

El modelo de predicción propuesto tiene la restricción de que los datos con que se realizan las predicciones son solamente temporales y no se toman en cuenta otros factores como: viento, cambio de temperatura, entre otros.

Estos factores pueden afectar la predicción de los contaminantes atmosféricos, por ejemplo, realizar la predicción en un día que resultó festivo y que al año siguiente no lo fue. A pesar de ello, los resultados fueron buenos.

Los resultados obtenidos de las predicciones realizadas mostraron competitividad con otras técnicas computacionales realizadas en otras partes del mundo y a nivel nacional.

6.2. Trabajo futuro

Realizar predicciones con diferentes metodologías, por ejemplo, que los conjuntos fundamentales correspondan a un día de la semana, a una semana entera o una estación del año entera, todas ellas de varios años y probarlo en un año diferente.

Se propone realizar predicciones tomando en cuenta varias estaciones de monitoreo atmosférico, con el fin de ver la relación entre ellas para realizar la predicción.

Un trabajo futuro muy importante es el de realizar un fundamento teórico para calcular la dimensión ideal de cada contaminante en sus patrones del conjunto fundamental y de prueba.

Bibliografía

- [1] *Programa Para Mejorar La Calidad Del Aire De La Zona Metropolitana Del Valle De México 2002-2012*. Disponible en www.sma.df.gob.mx.
- [2] *Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT)*. Disponible en www.sma.df.gob.mx/simat/pnimeca.htm.
- [3] George Bojadziev and Maria Bojadziev. *Advances In Fuzzy Systems-Applications and Theory. Fuzzy Sets, Fuzzy Logic, Applications*. volumen 5, edit. World Scientific, USA, 1995.
- [4] M. Boznar. *Pattern selection strategies for a neural network-based short term air pollution prediction model*. Intelligent Information Systems, 1997. IIS '97. Proceedings, págs. 340-344.
- [5] Surajit Chattopadhyay. *Prediction of mean monthly total ozone time series - application of radial basis function network*. International Journal of Remote Sensing, volumen 28, págs. 4037-4046, 2007.
- [6] Comisión Intersecretarial de Cambio Climático. *Programa Especial de Cambio Climático 2008-2012*. Disponible en www.semarnat.gob.mx.
- [7] Secretaría de Salud. *Programa Nacional de Salud 2007-2012 Por un México sano: construyendo alianzas para una mejor salud*. ISBN 978-970-721-414-9, México, D.F., 2007.
- [8] Alain-Louis Dutot, Joseph Rynkiewicz, Frédy E. Steiner, and Julien Rude. *A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions*. Environmental Modelling and Software, volumen 22, No. 9, págs. 1261-1269, 2007.

-
- [9] Salazar-Ruiz E. *Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence-based models in Mexicali, Baja California (Mexico) and Calexico, California (US)*. Environmental Modelling and Software, vol. 23 (8) 1056-1069 (2008).
- [10] R. Foxall, I. Krcmar, G. Cawley, S. Dorling, and D.P. Mandic. *Nonlinear modelling of air pollution time series*. Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on, volumen 6, págs. 3505-3508.
- [11] S. Fujita and H. Tamura. *Development of a Short-term Prediction Model for Predicting Photochemical Oxidants in a Local Area*. SICE-ICASE, 2006. International Joint Conference, págs. 2184-2189.
- [12] M.M. Kamal, R. Jailani, and R.L.A. Shauri. *Prediction of Ambient Air Quality Based on Neural Network Technique*. Research and Development, 2006. SCORed 2006. 4th Student Conference on, págs. 115-119.
- [13] Sungshin Kim, Jaeyong Kim, Chong-Bum Lee, and Min-Young Kim. *Fuzzy decision support system to the prediction of ozone concentrations*. Industrial Electronics, 2001. Proceedings. ISIE 2001. IEEE International Symposium on, volumen 1, págs. 602-607.
- [14] S. S. Lee. *Predicting atmospheric ozone using neural networks as compared to some statistical methods*. Northcon 95. IEEE Technical Applications Conference and Workshops Northcon95, pág. 101.
- [15] P. Mlakar and M. Boznar. *Perceptron neural network-based model predicts air pollution*. Intelligent Information Systems, 1997. IIS '97. Proceedings, págs. 345-349.
- [16] Cornelio Yáñez Márquez. *Memorias Asociativas basadas en Relaciones de Orden y Operadores Binarios*. Tesis de doctorado. Centro de Investigación en Computación, IPN, México, D.F., Febrero del 2002.
- [17] United Nations. *Kyoto Protocol to The United Nations Framework Convention on Climate Change*. 1997.

- [18] Angela Nebot and Violeta Mugica. *Finding qualitative patterns in ozone behavior*. Artificial Intelligence, 2006. MICAI '06. Fifth Mexican International Conference on, págs. 91-100.
- [19] Michael Negnevitsky and Galina Kelareva. *Air Quality Prediction Using a Neuro-Fuzzy System*. FUZZ-IEEE, págs. 481-484, 2001.
- [20] A.M. Rahmani, M. Teshnehlab, M. Abbaspour, and S. Setayeshi. *Short-term prediction of air pollution using td-cmac neural network model*. World Automation Congress, 2004. Proceedings, volumen 17, págs. 357-362.
- [21] Órgano del Gobierno del Distrito Federal. Gaceta Oficial del Distrito Federal. Disponible en www.sma.df.gob.mx, Junio del 2008.
- [22] J. Skrzypski, E. Jach-Szakiel, and W. Kaminski. *Neural Models for Prediction of Maximum Daily Particulate Matter PM10 Concentration in the Air in Big Cities as Ecological Safety Management Tools*. Systems Engineering, 2008. ICSENG '08. 19th International Conference on, págs. 141-146.
- [23] S. I. V. Sousa, F. G. Martins, M. C. M. Alvim-Ferraz, and M. C. Pereira. *Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations*. Environmental Modelling and Software, volumen 22, págs. 97-103, 2007.
- [24] Luis Enrique Sucar, Joaquín Pérez-Brito, J. Carlos Ruiz-Suárez, and Eduardo F. Morales. *Learning Structure from Data and Its Application to Ozone Prediction*. Appl. Intell, volumen 7, No. 4, págs. 327-338, 1997.
- [25] Fernando Vázquez Torres, Pilar Gómez Miranda, and Alfonso Leobardo Zarco Istiga. *Aplicación de algoritmos de clustering desarrollados en el entorno FIR a la predicción de la concentración de ozono*. Artículo derivado del proyecto de investigación titulado “Desarrollo de nuevas técnicas informáticas, para la creación de modelos para identificar sistemas complejos” Clave: 20040871.
- [26] Wang W., Men C., and Lu W. *Online prediction model based on support vector machine*. Neurocomputing, vol.71 (4-6) 550-558 (2008).

-
- [27] Norbert Wiener. *Time Series*. edit. M.I.T, Cambridge, Massachussets, USA, 1964.
- [28] Cornelio Yáñez-Márquez, Itzamá López-Yáñez, and Guadalupe de la Luz Sáenz Morales. *Analysis and Prediction of Air Quality Data with the Gamma Classifier*. CIARP, págs. 651-658, 2008.
- [29] Cornelio Yáñez-Márquez, Itzamá López-Yáñez, and Guadalupe de la Luz Sáenz Morales. *Application of the Gamma Classifier to Environmental Data Prediction*. Proc. Electronics, Robotics and Automotive Mechanics Conference CERMA 2008, IEEE Computer Society. ISBN: 978-0-7695-3320 (2008) 80-84.
- [30] Itzamá López Yáñez. *Clasificador Automático de Alto Desempeño*. Tesis de maestría. Centro de Investigación en Computación, IPN, México, D.F., Junio del 2007.
- [31] Lee Won Young, Choi Nam Ho, Shim Kyu Il, Han Sang Ok, and Kim Jong Suk. *The reliability of pollution prediction with regression analysis and the possibility of dispersion and receptor models*. Systems Engineering, 2008. ICSENG '08. 19th International Conference on, volumen 3, págs. 1035-1038.
- [32] Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, and Xiangshang Li. *Forecasting Skewed Biased Stochastic Ozone Days: Analyses and Solutions*. ICDM, págs. 753-764, 2006.