UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE INGENERÍA

INSTITUTO DE INGENIERÍA

"Estudio del riesgo relativo de muerte en la población de Mexicali por efecto de los principales contaminantes del aire"

TESIS

que presenta para obtener el grado de MAESTRO EN CIENCIAS EN ELÉCTRICA

ALICIA PAOLA MORENO FLORES

DIRECTOR DE TESIS:
Dr. MARCO ANTONIO REYNA CARRANZA

Mexicali, B. C. Noviembre del 2011

RESUMEN de la Tesis de Alicia Paola Moreno, presentada como requisito parcial para la obtención del grado de MAESTRO EN CIENCIAS EN ELÉCTRICA. Mexicali, Baja California, México. Noviembre de 2011.

Estudio del riesgo relativo de muerte en la población de Mexicali por efecto de los principales contaminantes del aire

Resumen aprobado por:

_____

Dr. Marco Antonio Reyna Carranza
Director de tesis

CUERPO DEL RESUMEN

La contaminación aérea de Mexicali tiene efectos perjudiciales para la salud de sus habitantes. Cuantificar el riesgo que agrega la polución a los casos de muerte natural en esta ciudad es un dato importante de aportar para mejorar las políticas de salud ambiental de la región. Este trabajo propone cuantificar el riesgo relativo (RR) de muerte que presenta la población mexicalense ante la exposición al aire contaminado. Para ello se consideraron en el estudio los principales contaminantes del aire de la región: PM_{10}, $PM_{2.5}$, CO, NO_2, O_3 y SO_2. Las series de tiempo que presentaron faltantes fueron imputadas utilizando cuatro métodos, de los cuales tres fueron desarrollados en esta investigación. Una vez identificada la mejor imputación de cada serie, se procedió al cálculo del RR de muerte mediante la construcción de modelos regresivos de Poisson. El estudio consideró el análisis de un periodo de cinco años (2003-2007) y por temporadas (frío y calor). Se encontró que algunos de los contaminantes que no exceden la norma producen riesgo relativo de muerte en la población como fue el caso del SO_2. También se observó que los contaminantes CO, $PM_{2.5}$, SO_2, y PM_{10}, incrementan en mayor medida el RR de muerte durante la temporada invernal.

ABSTRACT of the thesis presented by Alicia Paola Moreno Flores, in order to obtain the MASTER of SCIENCE DEGREE in ELECTRICAL SCIENCE. Mexicali, Baja California. Mexico. November, 2011.

Estudio del riesgo relativo de muerte en la población de Mexicali por efecto de los principales contaminantes del aire

Relative Risk of Death by Effect of the Main Air Pollutants in Mexicali, Baja California, Mexico

Approved by:

_____

Dr. Marco Antonio Reyna Carranza
Thesis Advisor

ABSTRACT BODY

Mexicali's air pollution has adverse effects on the health of its inhabitants. It is important to quantify the relative risk of death that the air pollution provokes in the population because this permits to improve environmental health policies of the region, and therefore the quality of life of the people improves. The study uses the most important air pollutants of the locality (PM10, PM2.5, CO, NO2, O3, and SO2) for estimating the relative risk of death. The missing values of the time-series are imputed using four methods; three of the methods are developed in this research. Once the best imputations are identified on each time-series, the RR of death is computed by means of the construction of Poisson regression models. In the study are considered a period of five years (2003-2007) and two seasons (winter and summer). The results show that some of the pollutants induce relative risk of death even though the air pollution levels do not exceed the Mexican standards; such is the case of the SO2. Also, pollutants such CO, PM2.5, SO2, and PM10 increase the relative risk of death more in winter than in summer.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por apoyarme y confiar en mí durante el transcurso de la investigación, sin su ayuda esto no hubiese sido posible.

Agradezco de forma muy especial al Dr. Marco Antonio Reyna Carranza por su valiosa ayuda ya que son demasiadas cosas las que tengo reconocerle, principalmente están: la asesoría apoyo y comprensión en los terribles momentos de confusión. Pero la más importante sin duda es su dirección durante este trabajo.

A mis Sinodales:
Dr. Efraín Carlos Nieblas Ortiz.
Dr. Miguel Enrique Bravo Zanoguera.
MC. Carmen Gorety Soria Rodriguez.
Por otorgarme un poco de su valioso tiempo, además de sus aportaciones, las cuales definitivamente fueron un parte aguas en este proyecto.

A todos los organizadores del Pan American Health Care Exchanges (PAHCE) y del 1er Congreso Regional de Bioingeniería, en especial a sus presidentes por abrirme un espacio para difundir el proyecto.

A mis Padres:
Héctor y Berthita, por apoyarme, quererme, darme la mano, mostrarme el camino, en fin tantas y tantas cosas. Todo esto no solo en el transcurso de este proyecto, sino durante el caminar de mi existencia.

A cada una de las personas y asociación que de alguna manera ayudaron en este proceso como: Emmanuel, Minerva, Luis Ángel, Francisco Júnior, Luis, Juanita, Roberto, Andrea, Víctor, TOASTMASTER, Reyes, León, Lupita, Juan, Miriam y Keyla.

Definitivamente un gran agradecimiento a Dios Nuestro Señor, pues él unió todo el rompecabezas para que las cosas se dieran de la mejor manera.

ÍNDICE **Página**

Introducción **1**

Capítulo 1 ..... **4**
Atmósfera y Contaminación **4**
1.1 Composición química de la atmósfera 4
1.2 Contaminación del aire 8
1.3 Contaminantes primarios y secundarios 8
1.4 Contaminantes criterio 9
1.5 Normas para medir la contaminación del aire 10

Capítulo 2 ..... **12**
Efectos de la contaminación del aire en la salud humana **12**
2.1 Historia 12
2.2 Materia Fraccionada Respirable (PST, PM_{10}, $PM_{2.5}$) 15
2.2.1 Constitución química 15
2.2.2 Afectación a la salud 16
2.2.3 Norma Mexicana NOM-025-SSA1-1993 Material Particulado 17
2.3 Gases (CO, O_3, NO_2, SO_2) 18
2.3.1 Constitución química 18
2.3.2 Afectación a la salud 20
2.3.3 Normas Mexicanas:
NOM-021-SSA1-1993 Monóxido de Carbono
NOM-020-SSA1-1993 Ozono
NOM-023-SSA1-1993 Dióxido de Nitrógeno
NOM-022-SSA1-1993 Dióxido de Azufre. 22

Capítulo 3 ..... **24**
Imputación de Series de Tiempo **24**
3.1 Imputación de datos en las series de tiempo 24
3.2 Imputación de factor de grupo 26
3.2.1 Factor de Serie (GG) 26
3.2.2 Factor de Ciclo 28
3.3 Imputación de promediado 31
3.3.1 Promediado por unidad (UA) 31
3.3.2 Imputación de promediado por exclusión (EA) 33
3.4 Automatic ARMA (AA) 36
3.4.1 Metodología utilizada por ARMA 36
3.4.2 AA por bloques (BAA) 38
3.4.3 AA por acumulado (AAA) 40
3.5 Amelia 42
3.6 Comparación 46
3.6.1 Comparación por Ventiles. 47
3.6.2 Comparación por Rangos Interventílicos. 48

Capítulo 4 51
Asociación de los principales contaminantes del aire con el riesgo relativo de muerte en la población de Mexicali 51
4.1 Introducción 51
4.2 Materiales 52
4.2.1 Lugar de estudio 52
4.2.1 Variables de estudio 52
4.2.1.1 Series de tiempo clínicas 53
4.2.1.2 Series de tiempo de la contaminación 54
4.2.1.3 Series de tiempo meteorológicas. 57
4.2.1.4 Variables Categóricas. 57
4.2.1.5 Colinealidad entre variables independientes 58
4.3 Método. 59
4.3.1 Crear Modelo Básico o Basal. 60
4.3.2 Método de selección de variables. 63
4.3.3 Elaborar Modelo con variables explicativas. 64
4.3.4 Pasos para la toma de decisión de los Métodos de imputación a utilizar. 66
4.4 Resultados. 68
4.4.1 Modelo Basal 68
4.4.1 Modelos con ajuste por contaminante. 72
4.5 Discusiones 77
4.6 Conclusión. 80

Refrencia Bibliográfica 81

APÉNDICE 90
APÉNDICE A. TABLA QUE INDICA LOS VALORES DE CORRELACIÓN DE LAS SERIES IMPUTADAS COMPARADAS CONTRA LA SERIE ORIGINAL 91
APÉNDICE B. ESTADÍSTICA DECRIPTIVA DE LAS VARIABLES CLÍNICAS, METEOROLÓGICAS Y CONTAMINANTES 92
APÉNDICE C. DIAGRAMA DE CAJA Y EJES DE LOS CONTAMINANTE 93
APÉNDICE D. CORRELACION ENTRE VARIABLES INDEPEDIENTES 94

Índice de figuras

Figura **Página**

1.1. Capas de la atmósfera terrestre. **6**
1.2. Porcentaje de gases en la atmósfera. **7**
3.1. Diagrama de flujo para el proceso de imputación. **25**
3.2. Cómo establecer la imputación por el método GG **27**
3.3. Representación de la imputación por YG **28**
3.4. Cómo se ordena para el cálculo del patrón UA **31**
3.5. Esquema de cómo organizar los datos para el método EA. **34**
3.6. Orden de los datos para establecer los modelos predictivos de los faltantes con el tratamiento BAA. **38**
3.7. Esquema de cómo organizar los datos para el método AAA. **40**
3.8. Matriz Y, empleada por Amelia II. **43**
3.9. Rutina del Programa Amelia **44**
3.10. Gráfica de distribución de Ventiles **47**
3.11. Orden de ventiles para crear los rangos interventílicos. **49**
4.1. Distribución de Estaciones de Monitoreo en Mexicali BC. **56**
4.2. Correlaciones entre la variable respuesta muertes y variables independientes del modelo básico, a) serie de tiempo de temperatura, b) humedad relativa y c) IRAs. **69**
4.3. Comportamiento diario mostrado por la variable respuesta muertes y valores pronosticados por el modelo basal. **71**
4.4. Riesgo relativo por temporadas de la ciudad de Mexicali BC **76**
C.1. Diagramas de caja y ejes de los contaminantes. **93**
D.1. Matriz de correlaciones entre variables del modelo basal. **94**
D.2. Matriz de correlaciones entre las variables del modelo basal y los contaminantes (PM10, CO, PM2.5, NO2, O3, SO2) **95**

Índice de tablas

Tabla **Página**

3.1. Ejemplo de imputación de grupo con YG en el contaminante $PM_{2.5}$ **29**
3.2. Ejemplo de imputación de grupo con GG en el contaminante PM_{10} **30**
3.3. Organización de los datos para el cálculo del patrón en el método promedios por unidad **32**
3.4. Ejemplo de imputación utilizando el patrón en el método promediado por unidad..... **33**
3.5. Ejemplo de cómo crear un patrón en el tratamiento promedio por exclusión (EA) **35**
3.6. Ejemplo de selección de datos por BAA..... **39**
3.7. Ejemplo de selección de datos por Acumulado AA **40**
3.8. Serie separada por ventiles..... **48**
3.9. Rangos Interventílicos..... **50**
4.1. Contaminantes monitoreados por estación **56**
4.2. Pruebas de normalidad de los residuales del modelo basal **71**
4.3. RR de muerte por efecto del contaminante PM_{10} **72**
4.4. RR de muerte por efecto del CO **73**
4.5. RR de muerte por efecto del $PM_{2.5}$ **73**
4.6. RR de muerte por efecto del NO_2 ..... **74**
4.7. RR de muerte por efecto del O_3 ..... **74**
4.8. RR de muerte por efecto del SO_2..... **75**
4.9. Coeficientes de regresión mostrados por periodos..... (frío, calor y cinco años) con los métodos que mostraron tener la mejor imputación **75**
4.10. Valores del RR de muerte mostrados para los tres periodos del estudio, con métodos de imputación sobresalientes **76**
A.1. Resultados de las comparaciones de las series de tiempo imputadas con los diferentes métodos, comparados con la serie original **91**
B.1. Estadística descriptiva de las variables Clínicas, Meteorológicas y Contaminantes, por el periodo de cinco años, además de las temporadas de invierno y calor. **92**

Introducción

La atmósfera terrestre es la capa gaseosa que rodea a la tierra, consta de una antigüedad aproximada de 4600 millones de años. Al principio de su formación, gran parte de las sustancias que la conformaban se perdían o fugaban al espacio, después de sufrir varios cambios en su composición e irse agregando nuevas sustancias a ella (como lo fue en su tiempo el oxígeno), esta masa gaseosa dio lugar a varias capas con diferentes funciones y composiciones hasta llegar a su constitución actual [1, 2].

Para su estudio la atmósfera, se ha dividido en distintas capas: desde la que se encuentra situada al exterior del planeta hasta la que tiene contacto directo con la superficie terrestre. Se mencionan a continuación [2, 3].

- Termósfera.
- Mesósfera.
- Estratósfera.
- Tropósfera.

La atmósfera se encuentra primordialmente constituida por dos gases los cuales son conocidos como [1, 2]:

- Nitrógeno.
- Oxígeno.

Dentro de la última capa atmosférica se presentan algunas sustancias que producen complicaciones en la vida vegetal, animal y humana; a tal grado, que pueden llegar a generar muertes prematuras, además de ser los responsables de los cambios climáticos que afectan al planeta [4].

La introducción de estas sustancias perjudiciales se debe a fuentes naturales y antropogénicas, siendo éstas, las que contribuyen mayoritariamente; a estos agentes se les denomina contaminantes y se sabe que seis de éstos son los que más afectan a la población [5]:

- PM_{10} (Partículas con un diámetro menor a 10 micrómetros).
- $PM_{2.5}$ (Partículas con un diámetro menor a 2.5 micrómetros).
- Monóxido de Carbono.
- Ozono.
- Dióxido de Nitrógeno.
- Dióxido de Azufre.

Las PM_{10} y $PM_{2.5}$ son partículas que pueden estar constituidas por sólidos o líquidos que se encuentran flotando a la deriva en el ambiente, estos contaminantes resultan ser nocivos para el ser humano ya que por su tamaño suele ser imposible filtrarlas en las vías respiratorias, por lo cual se conducen a través del sistema respiratorio e incluso llegan al sistema circulatorio, produciendo complicaciones y daños a cada sistema [6].

El monóxido de carbono es un gas conformado por Carbono y Oxígeno. Esta combinación es venenosa para el ser humano, a la vez que resulta ser peligrosa, porque al no tener olor ni sabor, produce debilidad y confusión [7-9].

Tanto el Dióxido de Nitrógeno como el Ozono al encontrarse en presencia del ser humano, provocan problemas, que principalmente se manifiestan en el sistema respiratorio [1, 10].

Los contaminantes han sido estudiados durante varios años, encontrándose que las características del propio medio provocan que se incremente o aminore su efecto, debido a la interacción del mismo contaminante con la naturaleza, por lo cual es necesario realizar un estudio que los contemple y de esta forma se pueda señalar, cuáles son los contaminantes que mayor influencia tienen en determinada región, en el riesgo relativo de muerte de la población expuesta, y que a su vez, esto pueda servir como base para poder establecer medidas ecológicas que favorezcan a los habitantes de la región y su medio ambiente.

Capítulo 1

Atmósfera y Contaminación

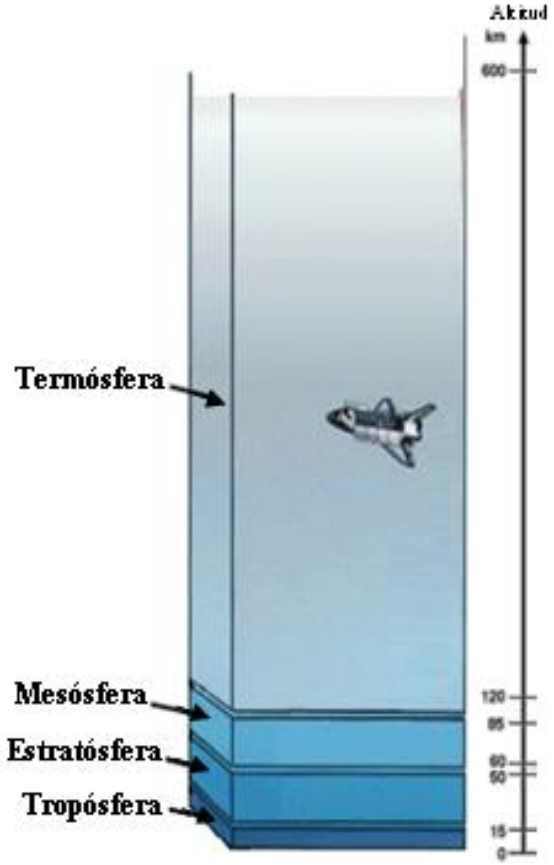
1.1 Composición química de la atmósfera

La tierra está envuelta en una manta compuesta por una mezcla gaseosa, a dicha composición se le conoce como aire, a toda esta extensión se le ha nombrado atmósfera. Las propiedades que presenta dicha mezcla interactúan entre sí para cumplir con diferentes funciones, como son: ayudar en la regulación y bloqueo de los rayos solares que inciden sobre todo el globo terráqueo, provocar las reducción de los cuerpos celestes que ingresan a ella, así como contribuir al control de la temperatura dentro de la misma. Se sabe que la atmósfera tiene una altura aproximada de 600 kilómetros (km), por tal razón se llegó al acuerdo de identificar los espacios dentro de la misma, para lo cual se propuso dividirla en capas según el gradiente de temperatura que se presenta a diferentes alturas [1, 11].

Es de esta manera como se dividieron las capas dentro del globo terráqueo, pero para poder identificarlas entre sí, se dieron a la tarea de diferenciarlas con nombres, como puede apreciarse en la figura 1.1. A la capa externa de la tierra se le conoce como Termósfera. Está situada entre los 120 kilómetros a los 600 kilómetros de la superficie

terrestre, es la capa que presenta temperatura más elevadas, esto se debe a la incidencia directa de los rayos solares sobre la misma, al mismo tiempo en ella se forma el fenómeno denominado aurora boreal. Debajo de esta capa se sitúa la Mesósfera que se encuentra ubicada por encima de los 60 kilómetros y por debajo de los 120 km. La estratósfera es la siguiente capa, en la cual se encuentra posicionada la famosa capa de ozono que protege a la tierra de los rayos solares directos. Debido a que esta capa no está en interacción con la humanidad, no se considera una sustancia nociva para ella; dicha capa se encuentra sobre los 15 km y bajo los 40 km; finalmente, se encuentra la capa que es la que interactúa mayormente con la humanidad; además de encontrarse en contacto directo con la superficie terrestre, se le ha denominado Tropósfera, la altura en que se ubica es a 15 km. Un dato interesante es que contiene la mayor cantidad de masa total de los gases, esto quiere decir que en esta capa se encuentran el 80% de la masa total de gases existente en la atmósfera, por consiguiente contaminar esta capa repercute afectando las concentraciones de casi la totalidad de la masa gaseosa total terrestre. Esta es la capa donde se producen la mayoría de los fenómenos climáticos [12, 13].

La atmósfera en condiciones óptimas se encuentra constituida por determinada composición de gases los cuales se aprecian en la Figura 1.2. En dicha estructura destacan algunos elementos como el Nitrógeno, este gas es el que mayormente la conforma, ya que constituye el 78 porciento (%) de los gases presentes en la misma, el Oxígeno figura con un porcentaje de 20.9, el Argón está presente en un 0.93%, de la misma manera se encuentran dispersos otros gases como lo es el Dióxido de Nitrógeno, vapor de agua, Ozono, Azufre, los cuales se encuentran en menores cantidades [6, 14].



Fuente: NASA.

Figura 1.1. Capas de la atmósfera terrestre.

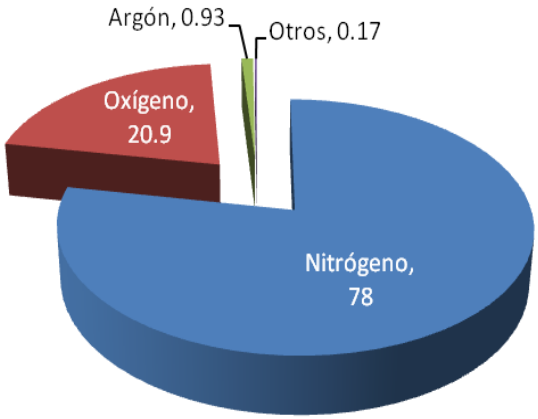


Figura 1.2. Porcentaje de gases en la atmósfera.

Se sabe que de toda la masa de gases distribuida en la atmósfera, el 80% de la concentración total se encuentra dentro de la Tropósfera, lo cual indica que en los primeros 15 km se encuentra la mayor parte de los gases de la misma, si a esto se le suma la cuestión de que en esta capa se detectan ciertas sustancias que originan o producen alteraciones y complicaciones que modifican las concentraciones de los elementos; esto conllevará a transformar las cantidades que mantienen el balance de los gases de la mayor parte de la concentración terrestre, teniendo como consecuencia un desequilibrio ecológico, el cual se traduce en perjuicios tanto a la vida animal, vegetal como al medio ambiente. Quien primordialmente sufre estos daños es la población expuesta. Los principales síntomas que muestra la ciudadanía en general, derivados de estos agentes son [6, 15]:

- Aumentos en los ataques de Asma.
- Incremento en la admisión hospitalaria y salas de emergencias.
- Muertes prematuras.

Se ha detectado que al incrementarse la concentración o prolongarse la exposición, se agrava la situación de la persona afectada, pudiendo desembocar en muertes prematuras; por otra parte, se tiene que estas sustancias son las responsables de los conocidos cambios climáticos que azotan a la humanidad entera [6].

1.2 Contaminación del aire

Históricamente la relación entre el hombre y el medio ambiente ha provocado varias series de cambios que a su vez han propiciado consecuencias ambientales como la contaminación hídrica, suelo y aire. Esta última se define como la introducción y difusión de cualquier sustancia ya sea sólida, líquida o gas en la atmósfera que provoca cambios tanto en la funciones de eliminación y reducción naturales propias de cada sitio; por lo anterior, se sabe que se produce en la atmósfera, cierta adición de sustancias que originalmente no se encontraban en ella o no estaban en esas cantidades, esto a su vez conlleva a provocar alteraciones en la composición de la atmosfera transformando sus condiciones naturales, provocando la pérdida del equilibrio ecológico, seguida por una serie de modificaciones y cambios como lo son [14, 16]:

- Los daños a la salud humana.
- Afectaciones a la fauna y flora.
- Produce perjuicios al ecosistema.

Una vez identificados los deterioros ocasionados por las sustancias es momento de saber ¿Cómo se les denomina a éstas? A estos elementos se les conoce como contaminantes.

1.3 Contaminantes primarios y secundarios

Al trabajar con contaminantes es necesario identificar el tipo de sustancia que se desea manejar, para lo cual existe el procedimiento de caracterización, mismo que establece que hay que conocer cuál fue la manera cómo dicha sustancia se introdujo a la atmósfera, ya que esa es la base para definir la clasificación del contaminante mediante esta información se sabe si el contaminante es de índole primaria o secundaria.

Los contaminantes primarios, son aquellas sustancias que se caracterizan por ser irradiadas directamente a la atmósfera. Tanto las fuentes antropogénicas como las naturales son las encargadas de verter la mayor cantidad del contaminante en la capa terrestre

conocida como Tropósfera, como ejemplo de contaminantes esparcidos por este tipo de clasificación se encuentran [2, 17]:

- Dióxido de Azufre (SO_2).
- Dióxido de Nitrógeno (NO_2).
- Monóxido de Carbono (CO).

Los contaminantes secundarios no son emitidos directamente a la Tropósfera, sino que se producen en la misma, derivados de diversos procesos que se forman en la atmósfera, entre los cuales destacan las reacciones químicas, hidrólisis y oxidación, un ejemplo claro de esta última se encuentra en la formación que particularmente constituye al Ozono (O_3) [2, 17].

1.4 Contaminantes criterio

Los contaminantes del ambiente se han catalogado como contaminantes criterio y no criterio. Los contaminantes criterio son seis y se han reconocido como nocivos para la salud y el bienestar de los seres humanos. El nombre de contaminantes criterio es debido a que estas sustancias constituyeron la base para constituir los estándares de la calidad del aire en los Estados Unidos Americanos (EUA), dichos estándares se plantearon con el propósito de establecer niveles tolerables que protejan la salud humana, el medio ambiente y el bienestar de la población potencialmente expuesta [18]. Debido a la globalización se puede decir que hoy por hoy el término "contaminantes criterio" ha sido adoptado en muchos países y los contaminantes que conforman la clasificación son [19, 20]:

- SO_2.
- NO_2.
- Material Partículado (PM).
- CO.
- O_3.
- Plomo (Pb).

Los efectos de los contaminantes del aire sobre la salud de los seres vivos son muy diversos, por esa razón en todo el mudo se han llevado a cabo distintos estudios con el

objetivo de encontrar esa pequeña línea que existe entre la exposición al contaminante y la afectación que este produce, por lo cual resulta necesario mostrar los principales contaminantes que se encuentran presentes en el aire.

1.5 Normas para medir la contaminación del aire

La humanidad requiere de ciertos elementos para su prevalencia, dentro de las cuales se encuentra el aire. En promedio, una persona inhala 10 a 20 m^3 de aire por día; sin embargo, para que este recurso satisfaga las necesidades propias de cada uno de los organismos debe encontrarse libre de contaminantes, o por lo menos, que las sustancias dañinas presentes se encuentren en niveles que no pongan en riesgo la salud de la población expuesta, es por eso que varios países y organizaciones han realizado investigaciones con el fin de detectar los valores marginales para cada uno de los contaminantes que afectan a la salud humana. Encontrar valores que no representen peligro alguno a la población expuesta, suele ser complejo, por lo que muchos de los organismos investigadores de estos fenómenos, alrededor de los años 1980s, optaron por identificar niveles aceptables de riesgo para el ser humano [21].

Los valores o niveles marginales deben estarse renovando, no deben de quedar estáticos, ya que por diversos efectos pueden producirse cambios atmosféricos que no se presentaron cuando se realizaron las pruebas anteriores para establecer estos datos; es por eso que han continuado en constante investigación por muchos organismos alrededor del mundo, gracias a ello en la actualidad nos podemos dar cuenta de que un valor marginal de cierto contaminante puede no ser el nivel estandarizado más adecuado para todo el país, que lo más congruente sería tener valores para determinadas regiones contemplando sus situaciones geográficas y climatológicas. Además también se ha visto la necesidad de que el ajuste de estos límites no sólo sean dictados por las afectaciones que produce a la salud humana si no que también contemple a la fauna, flora y elementos que conforman la naturaleza del planeta [22, 23].

En el proceso de búsqueda de estos niveles, se percataron de que la distribución de respuesta de la población expuesta a cierta sustancia, la mayoría de las veces no se

presentaba igual, lo que producía un conflicto al momento de indicar un riesgo aceptable, al investigar la raíz de la problemática, identificaron que unas personas resultan tener una mayor sensibilidad que otras; por consiguiente, necesitaron caracterizar estos comportamientos e identificar cuál es la causa que los originan [22, 23].

Capítulo 2

Efectos de la contaminación del aire en la salud humana

2.1 Historia

La problemática de las afectaciones a la salud derivadas de una atmósfera contaminada, no es nueva, hay documentos que muestran estos acontecimientos y fueron redactados hace varios años. Una carta escrita en el año 73 dC por Plinio el Joven y dirigida al historiador Tácito da prueba de la antigüedad del tema. En el documento se expresa que las cenizas y rocas encendidas arrojadas por el volcán Vesubio provocaron la sepultura de la ciudad de Pompeya y Herculano. Varias personas perdieron la vida a consecuencia de la contaminación persistente por 19 horas. Entre las víctimas de la catástrofe se encontraba Plinio el viejo quien fuese la razón de la escritura de la carta [24, 25]. El hecho descrito en la carta fue resultado de la naturaleza; no obstante, en la actualidad muchos de los contaminantes son conducidos a la atmósfera por acciones humanas. Este tipo de fuente es mejor conocida como antropogénica. Como ejemplo de ello se tienen la catástrofe ocurridas en Londres, en 1952, la cual cobró la vida de 4000 personas. Este suceso fue el principal detonador de los estudios de contaminación del aire [26, 27].

Al ver la gravedad del problema de salud los países empezaron a buscar soluciones, es así como en 1970 en los Estados Unidos de América, se da la orden de crear la Agencia de Protección al Ambiente (Environmental Protection Agency por sus siglas en Inglés EPA) con el fin de limpiar la tierra, agua y aire del país [18]. Esta agencia en 1990 modificó los estándares nacionales de la calidad del aire ambiental a los que nombró National Ambient Air Quality Standards (NAAQS), con el propósito de establecer niveles apropiados de concentración de contaminantes en el aire para la población, basados en la determinación de los riesgos que estos representan para la salud humana. Así, dichas normas establecen las concentraciones aceptables durante diferentes periodos de exposición, ya que bajo algunas circunstancias se reconoce que una concentración menor pero durante un periodo más prolongado representa un riesgo para la salud de la población [23].

En México, en el año de 1972 se dio el primer paso al establecer la Subsecretaría para el mejoramiento del medio ambiente, dentro de la Secretaría de Salubridad y Asistencia. Al pasar de los años, la Subsecretaría se fue reformulando convirtiéndose en Secretaría. Pasó por distintos nombres, se desprendieron algunas otras Secretarías y en el año 2000 se instituyó la Secretaría del Medio Ambiente y Recursos Naturales (SEMARNAT), que es como actualmente se conoce. Esta secretaría es el órgano responsable de evaluar la evidencia de los impactos en la salud por la contaminación del aire, y de establecer los límites permisibles de los contaminantes en la atmósfera en México [28].

En la actualidad, estas agencias y secretarías han impulsado algunos trabajos, investigaciones, incluso guías con el propósito de mantener en constante monitoreo los contaminantes existentes en la atmósfera, sus transformaciones y las consecuencias que estos producen en la población humana, así como en la fauna, flora y medio ambiente del sitio expuesto.

Gran cantidad de los estudios realizados sobre la contaminación del aire, han sido con el propósito de contar con la información necesaria, para valorar de una forma

aproximada, cual es la concentración que representa un menor riesgo en la población. A su vez, emplear la misma como base en los estándares. Si se deseara que el estándar indique con una absoluta seguridad, a partir de que concentración se empieza a notar algún efecto dañino en la población y además sustentar esta información con una alta probabilidad, se necesitaría realizar un estudio que contemple algunos elementos que por su naturaleza resultan ser complicados de obtener, como lo son [29]:

- La concentración-respuesta entre los individuos en relación con todas las fuentes de exposición.
- Los tipos de efectos tóxicos sucintados en los contaminantes y sus combinaciones.
- La existencia o no del umbral para cada específico efecto tóxico de cada sustancia.
- La significancia de las iteraciones entre las poluciones.
- Variación en sensibilidad
- Niveles de exposición entre la población humana.

Lo que vuelve a este estudio una tarea muy complicada y en algunos casos casi imposible es contar con la información necesaria. Por ello se opta por realizar estudios ecológicos que involucran series de tiempo.

Debido a lo anterior, existe la necesidad de monitoreo constante de los estándares con el propósito de mantener actualizados los mismos con los cambios que surgen día a día en todo el mundo, pudiendo de esta forma brindar una mejor calidad de vida tanto a la población humana como al medio ambiente [29].

La cantidad de concentración de un contaminante puede intoxicar a la población de manera crónica o aguda. Donde un efecto crónico producido por bajas concentraciones del contaminante durante un lapso prolongado. Mientras que el efecto agudo se presenta cuando se tienen altas concentración en lapsos cortos y es el tipo de efecto que se pretende encontrar en esta investigación [30].

2.2 Materia Fraccionada Respirable (PST, PM_{10}, $PM_{2.5}$)

2.2.1 Constitución química

El material particulado es conocido a su vez como partículas contaminantes, está compuesto de una mezcla compleja de materiales suspendidos en el aire, que pueden variar significativamente en tamaño, forma y composición, dependiendo fundamentalmente de su origen, como ejemplo se tiene en la atmósfera, substancias y partículas muy distintas como de amonio, agua, polvo, polen, hollín (carbón), metales (plomo, cadmio), asbesto, sales, pequeñas gotas de ácido (nitratos, sulfatos), dioxinas, pesticidas, entre otras [31].

Las partículas pueden ser vertidas a la atmósfera mediante procesos naturales o antropogénicos. En el primer caso suelen ser el resultado de la polinización de las plantas e incendios forestales, primordialmente. En cuanto a las fuentes derivadas de la mano del hombre se sabe que cubren un amplio espectro: desde la quema de combustibles hasta la fumigación de los campos agrícolas. Las partículas se categorizan en dos tipos dependiendo de la manera en la que fueron liberadas: si las partículas fueron emitidas directamente se les nombra partículas primarias; pero si el elemento del que están conformadas las partículas sufrió alguna reacción con el medio ambiente después de ser liberadas, se le denomina partículas secundarias. Las partículas secundarias comúnmente reaccionan con ciertos gases presentes en la atmósfera entre los que primordialmente se encuentran [6, 31]:

- Óxidos de nitrógeno.
- Óxidos de azufre.
- Amoniaco.
- Compuestos orgánicos.

Las partículas suspendidas en la atmósfera que tienen composición tanto líquida como sólida han sido catalogadas en tres conjuntos dependiendo del diámetro que presenta cada partícula. Se tiene que las partículas con diámetro aerodinámico igual o menor a 100 micrómetros (µm) se han nombrado como Partículas Suspendidas Totales (PST). Una de las características más predominantes de las partículas que cuentan con un diámetro próximo a los 20 µm es que estas permanecen unas cuantas horas suspendidas en el

ambiente; en esta categoría engloba a otros dos, uno es nombrado Partículas con diámetro menor a diez µm (PM_{10}) como su nombre hace referencia son todas aquellas partículas que tienen un diámetro igual o menor a diez µm. El último grupo se encuentra contenido en los dos anteriores debido a las dimensiones que maneja en la partícula, las cuales son diámetro menor o igual a dos punto cinco µm y es llamado $PM_{2.5}$ [32, 33].

2.2.2 Afectación a la salud

Los conductos del sistema respiratorio son utilizados por las partículas para ingresar al organismo, durante cada inhalación se recolecta una mezcla de aire con partículas, la mezcla es propagada por todo el sistema, las partículas que poseen diámetros mayores a 6 µm se depositan en las vías respiratorias altas (nariz, senos paranasales, faringe, laringe, tráquea); mientras las pequeñas se introducen aún más llegando a las vías respiratorias bajas (Pulmón, bronquios primarios y secundarios, bronquiolos, alvéolo). El daño que las partículas provocan en la salud de las personas, se ve incrementado conforme a la profundad que logren la partícula antes de alojarse. Es conveniente señalar que si los individuos practican la respiración bucal, esto perjudica aún más la salud ya que las partículas puedan viajar a una mayor profundidad dentro del sistema respiratorio [34, 35].

Los pulmones cuentan con un complejo sistema de filtros naturales, dichos filtros funcionan mediante la creación de mucosidades, las cuales capturan cualquier partícula que se encuentre en el sitio, a su vez también se cuenta con las vellosidades, ellas tienen la función ir recorriendo las partículas o sustancias que fueron capturadas hasta lograr expulsarlas del pulmón mediante una tos o estornudo; una vez fuera del pulmón las partículas pueden salir al exterior del organismo o dirigirse al sistema digestivo. Las partículas más pequeñas logran esquivar estos filtros e internarse dentro de los pulmones hasta alcanzar un acercamiento directo con los alvéolos, teniendo como consecuencia que el organismo prolongue el lapso de tiempo requerido para expulsarlos, dicho lapso puede prolongarse unas cuantas semanas o años [32, 34, 35].

Si el individuo se mantiene en una constante exposición al aire contaminado, le origina una afectación en la limpieza de sus pulmones además de disminuir las defensas

naturales de los mismos, teniendo como consecuencia el arribo de una cantidad mayor de partículas a lugares más profundos con una mayor facilidad. Este fenómeno desata en el sujeto, enfermedades respiratorias, afecciones al corazón y al sistema circulatorio. Dependiendo de la composición de las partículas, éstas pueden pasar de los pulmones al torrente sanguíneo por medio de los alvéolos; una vez en el torrente sanguíneo provoca alteraciones de los niveles en los componentes de la sangre, Además de generar riesgos o daños en los órganos donde se alojan. Resulta evidente que el daño originado por las partículas sobre una persona está en función del tamaño de las partículas, por lo cual se entiende que entre menor sea el tamaño de la partícula mayor será su repercusión a la salud [34, 35].

Las partículas pueden provocar diferentes reacciones a los individuos que se encuentran en contacto directo con ellas, si la exposición se caracterizó por ser moderada (no intervino una densa cantidad de partículas) las repercusiones pueden expresarse como una simple irritación en los ojos, molestias en vías respiratorias o congestión nasal. Si la exposición aumenta se presentarán infecciones, sinusitis, alergias, resfríos, tos, ronquera entre otras; al elevarse la exposición provoca una disminución de las capacidades pulmonares del individuo lo que desemboca en bronquitis, neumonía, asma, cáncer pulmonar, debilita al sujeto a tal grado que se produce la muerte prematura de la persona [34].

2.2.3 Norma Mexicana NOM-025-SSA1-1993 Material Particulado

La modificación de la norma NOM-025-SSA1-1993 Criterios para evaluar la calidad del aire ambiente; con respecto a material particulado, fue elaborada por organizaciones consientes y preocupadas por la salud de los ciudadanos mexicanos como la Secretaría de Salud, Comisión Federal para la Protección contra Riesgos Sanitarios, Instituto Nacional de Enfermedades Respiratorias, Instituto Nacional de Cancerología, Instituto Nacional de Salud Pública, Centro Nacional de Salud Ambiental, Secretaría de Medio Ambiente y Recursos Naturales entre algunos mas organismos tanto públicos como privados; se unieron en año de 2002 y establecen los nuevos lineamientos que deben de ser respetados dentro de todo el territorio que comprende al país. Los cuales indican:

Para las Partículas PM_{10}:

- Debe de manejar valores de 120 microgramos sobre metro cúbico ($\mu g \cdot m^{-3}$) en promedio con base de 24 horas. Si excede la cantidad se estará violando la norma.

- El promedio de todo el año se estableció en 50 $\mu g \cdot m^{-3}$.

Para las Partículas $PM_{2.5}$:

- Se implementan promedios diarios con un máximo de 65 $\mu g \cdot m^{-3}$.
- El promedio anual queda establecido con un máximo de 15 $\mu g \cdot m^{-3}$.

2.3 Gases (CO, O_3, NO_2, SO_2)

2.3.1 Constitución química

El monóxido de carbono (CO) es un gas considerado altamente venenoso y peligroso para el ser humano, sus característica son: inodora, incolora y toxica; debido a estas cualidades resulta ser de especial cuidado, ya que no se puede ver, oler o sentir por el ser humano. Éste contaminante es el resultado de la combustión del aceite, carbón mineral, madera, carbón vegetal, propano, gas natural, entre otros, que al quemarse no lo hace completamente, por consiguiente se genera el CO y la molécula de este contaminante se constituye por un átomo de carbono ligado a otro átomo de oxígeno [7, 8].

El Ozono (O_3) es un gas y en un principio se creía que era el "olor de la electricidad" se percibe como un olor muy particular. Pero no fue hasta 1840 cuando recibe su nombre a mano del científico Alemán Christian Schoenbein. La molécula del ozono está conformada por tres átomos de oxígeno, un átomo más que el gas requerido por los humanos para vivir (O_2). Si el O_3 se presenta fuera de la Tropósfera es benéfico tanto para el planeta como para la vida humana, pero si el O_3 se manifiesta dentro de ella, se considera un contaminante y se le denomina ozono superficial; la razón de considerarse de esta manera, obedece a lo nocivo que resulta ser para la salud de la población y el medio ambiente (Fauna y Flora) expuestos. Para su constitución necesita de la reacción de óxidos

de nitrógeno y compuestos orgánicos volátiles (COVs) junto con una elevada radiación solar, por tratarse de una reacción química y no ser directamente emitido a la atmósfera se le cataloga como contaminante secundario. Debido a esto se tiene generación de ozono en zonas industriales y en grandes ciudades, este fenómeno no es sorprendente ya que los elementos necesarios para la formación del contaminante se encuentran en un alto volumen en estas áreas, aunado a esto tenemos las altas temperaturas de los meses más calurosos del año [32, 36].

Unas de las principales características del contaminante Dióxido de Nitrógeno ($NO_{2)}$ es predominantemente corrosivo, altamente oxidante, reactivo, soluble en agua, a su vez tiene una apariencia física café-rojiza, estructuralmente su molécula se constituye por átomo de nitrógeno y dos átomos de oxígeno, a este contaminante es común encontrarlo formando la conocida neblina o smog fotoquímico [37, 38]. El NO_2 es un contamínate secundario y la forma en la que comúnmente se produce es mediante, una oxidación del compuesto oxido de nitrógeno (NO). Convencionalmente este compuesto puede llegar a la atmósfera mediante la generación del NO, originada en los cilindros de combustión interna, más la combinación del oxígeno como se aprecia en la reacción química presentada en la Ecuación 1 [39].

$$2NO(g) + O_2(g) \rightarrow 2NO_2$$ **Ecuación 2.1**.

Los óxidos de nitrógeno pueden ser generados por fuentes antropogénicas como son la quema de combustibles fósiles originadas tanto por fuentes fijas como móviles, también puede producirse por la fabricación de ácido nítrico, el uso de explosivos, el consumo de gas L.P. y el proceso de soldadura. Por otro lado, la madre naturaleza es capaz de generar óxidos de nitrógeno mediante la actividad bacteriana, volcánica, y descargas eléctricas originadas en la atmósfera. Si se hace la comparación de que fuente contribuye con la mayor producción del contaminante, sin lugar a dudas la fuente originada por el ser humano es la que destaca [39, 40].

Dentro del grupo de los ácidos sulfúricos se encuentra el contaminante Dióxido de Azufre, mismo que se reconoce por ser sumamente reactivo, además de no tener color y poseer un aroma penetrante. Su estructura molecular está dada por la combinación de un átomo de azufre con dos de oxígeno. Existen varias reacciones químicas que pueden dar origen a la composición del SO_2: la más sencilla de ellas se presenta mediante la combinación de dos átomos de oxígeno interactuando con uno de azufre tal y como se muestra en la Ecuación 2 [2, 41]:

$$S + O_2 \rightarrow SO_2 \qquad \textbf{Ecuación 2.2}$$

El contaminante SO_2 se considera primario, ya que es irradiado directamente como residuo a la atmósfera por ciertas actividades humanas como lo son el proceso de refinería, la combustión tanto del petróleo como del carbón, fabricación de acero y papel, entre otras, Una vez liberado el contaminante parte se deposita en el suelo y la mitad de él se convierte en la base para crear los iones de sulfato (SO_4^{2-}) [2, 41].

2.3.2 Afectación a la salud

El CO es considerado uno de los contaminantes más peligrosos, debido a sus particulares características suele retardar o anular la capacidad de respuesta de la persona expuesta, originándole confusión y debilidad e impidiendo o afectando el raciocinio de la persona para retirarse del lugar o pedir ayuda. El CO utiliza al sistema respiratorio para ingresar al organismo humano, una vez que se encuentra en la membrana alveolo-capilar, la emplea como conducto para entrar al sistema circulatorio, provocando en él un bloqueo en la absorción del oxígeno dificultando la formación de hemoglobina y favoreciendo la elaboración de carboxihemoglobina en 200 a 300 veces más que la hemoglobina. En un organismo la exposición a este contaminante origina la desestabilización del sistema nervioso, afectando a su vez algunos otros órganos vitales, lo anterior obedece a la escasez de oxígeno a nivel celular, además de incapacitar a las células para absorberlo [9].

Las primeras reacciones provocadas por el O_3 al ingresar al sistema respiratorio son principalmente las siguientes:

- Tos.
- Ardor de garganta.
- Incomodidad en el pecho.

De continuar con la exposición y/o incrementarse el volumen de la misma, el individuo afectado entra a una segunda fase, donde se le manifiestan ciertos desórdenes como:

- Se dificulta la respiración profunda.
- Se reducen las funciones pulmonares.

Las personas que se encuentran en esta fase comúnmente presentan "respiración superficial", este tipo de respiración suele ser evidente en individuos expuestos a O_3 que practican actividades al aire libre como hacer ejercicio o trabajar. Dicha respiración se caracteriza por ser rápida y superficial, más de lo habitual. Una vez que el contaminante se encuentre en las vías respiratorias bajas, tenderá a disolverse en la capa delgada del fluido epitelial a lo largo de las vías que conducen a los pulmones, provocando daños a las células, si se disminuye o erradicarse la exposición, las células se regeneran y las células viejas se desprenden como si se tratase de piel quemada, pero de continuar la exposición, el deterioro celular continuará avanzando [10, 42].

Tratándose del grupo vulnerable de asmáticos el O_3 les producirá aumento en la sensibilidad a los agentes desencadenantes de las alergias, lo que conlleva ha disminuciones de función pulmonar más severas, produciendo ataques que requerirán de auxilio tanto de medicamento como de médicos para poder controlarlos. La Bronquitis y el Enfisema pulmonar, al igual que algunas otras enfermedades crónicas situadas en los pulmones, pueden versen agravados por el O_3, al mismo tiempo que el sistema inmunológico del organismo humano ya que reduce la habilidad de detectar las bacterias infecciosas en el

sistema respiratorio. Aunado a eso, constantes exposiciones a O_3 producen avances en el deterioro natural de las funciones del pulmón [42].

El contaminante NO_2 produce irritación en ojos, nariz, laringe y faringe, si se expone constantemente puede conducir a bronquitis aguda o crónica. En pacientes asmáticos desencadena una disminución en la función pulmonar que puede culminar en una muerte por obstrucción crónica pulmonar, además de incrementar el riesgo de infecciones respiratorias en niños [38].

La presencia de altas concentraciones Dióxido de Azufre (SO_2) en la población expuesta produce problemas en la respiración, alteraciones en las defensas de los pulmones, produce enfermedades respiratorias, al mismo tiempo que agrava los problemas cardiovasculares que ya existían dentro del grupo. A su vez, existe un subgrupo el cual es más sensible a este contaminante, por lo cual les produce mayor afectación que al resto de la población; este se constituye de individuos asmáticos, personas con enfermedades cardiovasculares o aquellos que cuentan con problemas crónicos en los pulmones, ejemplo de esto serian las habitantes que sufren de bronquitis o enfisema [43, 44].

2.3.3 Normas Mexicanas: NOM-021-SSA1-1993 Monóxido de Carbono, NOM-020-SSA1-1993 Ozono, NOM-023-SSA1-1993 Dióxido de Nitrógeno, NOM-022-SSA1-1993 Dióxido de Azufre.

En la Norma mexicana NOM-021-SSA1-1993 del Monóxido de Carbono, se señala como la Ley General de Salud, Ley General del Equilibrio Ecológico y de Protección al Ambiente, tanto como el Programa Nacional para la Protección del Medio Ambiente 1990–1994, convergen en un mismo punto, el cual es reducir los efectos de este contaminante para así proteger a la vida humana y al ambiente en el país. Debido a la problemática se estableció la norma antes mencionada la cual señala el máximo valor para la concentración de CO es de 11.00 ppm como promedio móvil de 8 horas una vez al año [45].

La modificación de la norma mexicana NOM-020-SSA1-1993 del contaminante Ozono, fue creada por una colaboración de diversas Secretarías (Salud, Gobierno, Medio

Ambiente y Recurso Naturales), Gobiernos (Distrito Federal, México, Chihuahua, Nuevo León), Institutos (Mexicano del Petróleo, de Investigaciones Nucleares), Universidades (Autónoma de Xochimilco, Nacional Autónoma de México), entre otras organizaciones, todas unidas con el fin de establece los valores límites permitidos de O_3, los cuales fueron dictaminados en, el valor de 0.110 ppm en promedio de una hora, una vez al año en un periodo de tres años [46].

La norma mexicana NOM-023-SSA1-1993 referida al contaminante Dióxido de Nitrógeno, externa el punto de coincidencia al que han arribado la Ley General de Salud, Plan Nacional de Desarrollo 1989 - 1994, Programa Nacional para la Protección del Medio Ambiente 1990 - 1994, como en la Ley General de Equilibrio Ecológico y Protección al Ambiente; en el cual se muestra como las condiciones del ambiente afectan y dañan a la salud humana, por ello la necesidad de restablecer, prevenir, reducir los efectos del contaminante. Esta norma menciona el valor máximo permisible el cual es de 0.21 ppm en una hora una vez al año [47].

La norma mexicana NOM-022-SSA1-1993 del Dióxido de Azufre, indica la convergencia evidente en prevenir, restablecer, reducir y mantener la calidad del aire con el propósito de proteger la salud humana, dicha afinidad se encuentra expresada en la Ley General de Salud, Plan Nacional de Desarrollo 1989 - 1994, Programa Nacional para la Protección del Medio Ambiente 1990 - 1994, como en la Ley General de Equilibrio Ecológico y Protección al Ambiente. La norma señala los límites establecidos como protección a su población sensible; los cuales son [48]:

- Una concentraciones SO_2 máxima de 0.13 ppm en 24 hrs una vez al año.
- Un promedio anual de 0.03 ppm.

Capítulo 3

Imputación de Series de Tiempo

3.1 Imputación de datos en las series de tiempo

Lo más apropiado cuando se desea realizar estudios de series de tiempo es contar con toda la información completa. Desafortunadamente en la vida real ocurren diversos sucesos o factores que provocan carencia de datos en la series de tiempo, lo cual es cotidiano en estudios ecológicos, teniendo como consecuencia series de datos que presentan faltantes. Lamentablemente esta problemática es frecuente. Darse a la tarea de completar las carencias de datos dentro de la serie de tiempo sigue siendo un reto para los investigadores o estadísticos, los cuales le han denominado imputación a este ejercicio.

En el momento que se ejecuta la imputación de datos a la serie de tiempo con el fin de suprimir los huecos, se debe de tener cuidado para que dichos valores no provoquen sesgo en la nueva serie (serie imputada). De suceder el sesgo, este repercutiría directamente en todos los resultados o inferencias que se encuentren a partir de los modelos que se fabriquen con los datos, por lo cual se recomienda probar con diversos tratamientos de imputación para obtener la técnica o técnicas que mejor se ajusten a las series de tiempo originales, con el propósito de identificar cual será la serie estadísticamente más cercana a

la serie original, de esta manera tener la certeza de que los datos son adecuados, además de confiables para realizar modelos y conjeturas sobre los mismos.

Los métodos de imputación utilizados en el estudio son cuatro: grupo y subgrupo, promediado, "automatic arma" (AA), y finalmente el software Amelia II. Estos tres primeros fueron desarrollados en el transcurso de esta investigación y serán explicados durante este capítulo.

Con la intención de poder comparar los métodos entre ellos se estableció la serie de pasos mostrados en la figura 3.1.

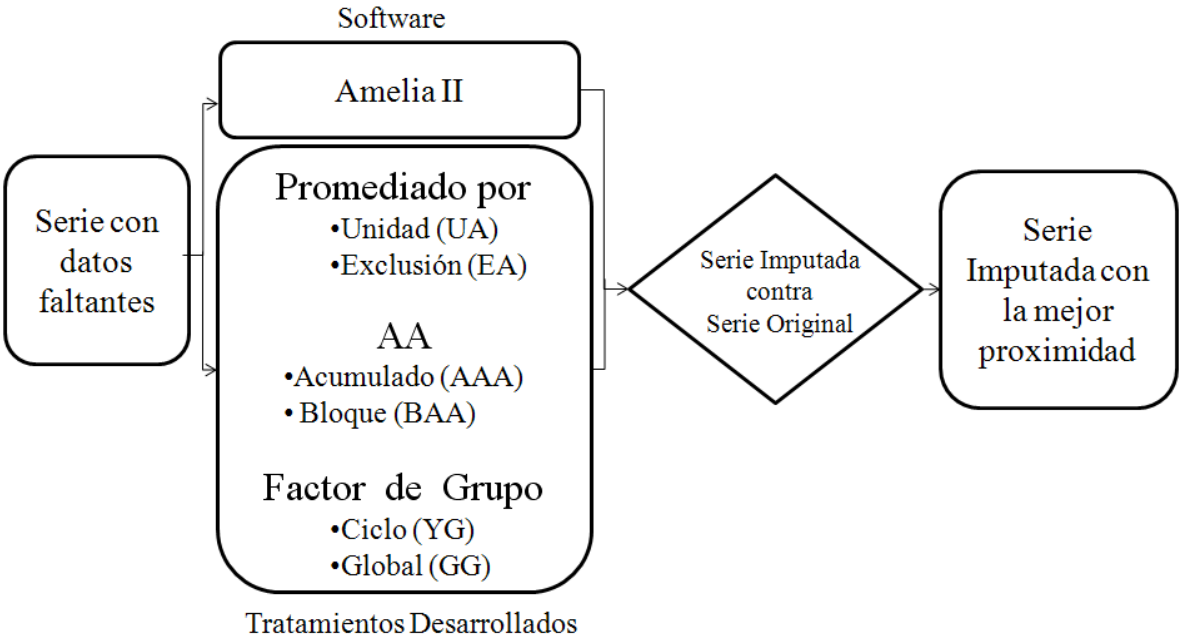


Figura 3.1. Diagrama de flujo para el proceso de imputación.

3.2 Imputación de factor de grupo

Dentro de las series de tiempo puede darse el caso de encontrar algunas que por su naturaleza presenten una estrecha relación; la cual se puede identificar como un factor, que establezca la relación entre las dos series tiempo. Para imputar los datos en dichas series se aprovechó una característica denominada factor de grupo, ésta se presenta cuando una serie se encuentra contenida dentro de otra, como es el caso del PM_{10}, el cual tiene como subgrupo al $PM_{2.5.}$ Estos contaminantes al ser uno subgrupo del otro se podría pensar en solo evaluar uno, pero debido a que las afectaciones producidas por cada contaminante en la salud son distintas, se opta por manejar a estos dos contaminantes por separado. Para plasmar el factor con el cual se vinculan matemáticamente las series de tiempo de grupo y subgrupo, se pueden utilizar los datos correspondientes del grupo, para llenar los huecos existentes en el subgrupo y viceversa. Al aplicar este tratamiento en la investigación, se denominó al contaminante PM_{10} como grupo y al $PM_{2.5}$ como subgrupo.

Para determinar el factor que enlaza las series de tiempo identificadas como grupo y subgrupo se utilizan dos técnicas, las cuales se han sido nombradas por el autor como:

- Factor de serie.
- Factor de ciclo.

3.2.1 Factor de Serie (GG)

Este factor consiste en obtener la mediana (Ma) de la serie de tiempo del grupo y del subgrupo como se observa en la figura 3.2, es decir se calcula la Ma para los valores que conforman a la serie de grupo como para los de subgrupo. Con las medianas establecidas se realizada el siguiente paso, el cual consiste en dividir los valores resultantes calculados a fin de poder obtener el factor que se desea. Para construir la división, se coloca el valor correspondiente a la mediana de subgrupo como numerador y como denominador se establece la cantidad correspondiente a la mediana del grupo.

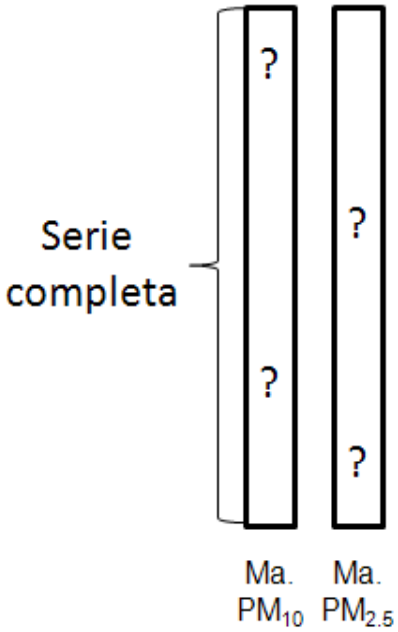


Figura 3.2. Cómo establecer la imputación por el método GG

Es de esta manera como se plantea el factor que vincula al subgrupo $PM_{2.5}$ con el grupo PM_{10}. La igualdad que describe matemáticamente esta relación es la Ecuación 3.1.

$$\left[\frac{Ma(PM_{2.5})}{Ma(PM_{10})}\right]_{GG} = GG \qquad \textbf{Ecuación 3.1}$$

Al analizar toda la serie como un conjunto solamente, se corre el riesgo de omitir tendencias que se pudiesen presentar durante cada ciclo transcurrido. Debido a esta incertidumbre, se opto por establecer un método paralelo con de fin de cubrir este punto, por lo que se planteo determinar el factor que une a grupo y subgrupo contemplado al conjunto de datos como conjunto de ciclos. En esta investigación cada ciclo fue definido por cada año registrado del contaminante.

3.2.2 Factor de Ciclo

Factor de ciclo o también nombrado (YG), plantea separar series de tiempo de los respectivos contaminantes por cada ciclo definido como un año en esta investigación tal como se visualiza en la figura 3.3. A su vez encontrar el factor presente en cada uno de los ciclos. La forma de obtener el factor es muy similar a la del factor por serie, se sigue respetando la igualdad matemática pero haciendo algunas modificaciones como se observa en la ecuación 3.2. En ella se establece que en lugar de contar con un solo factor, como en el caso del GG, se tendrá un factor por cada ciclo de estudio, que en este caso será un YG para cada año. La mediana de cada contaminante se realiza por cada ciclo, esto significa que se tendrá una calcularse una Ma por cada año en el estudio, en vez de contemplar a la serie como un conjunto. Se estudian cinco años por lo cual se obtendrán la misma cantidad de factores.

$$\left[\frac{\mathrm{Ma}(\mathrm{PM}_{2.5})}{\mathrm{Ma}(\mathrm{PM}_{10})}\right]_{\mathrm{YG}} = \mathrm{YG} \qquad \textbf{Ecuación 3.2}$$

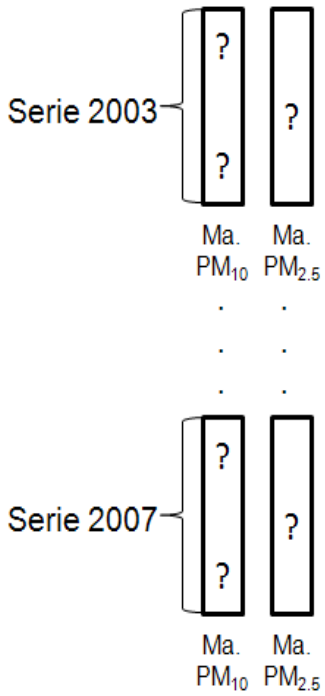


Figura 3.3. Representación de la imputación por YG

Una vez obtenidos los factores, se procede a imputar los datos en la series de tiempo, es necesario tener en cuenta que se utilizaron dos técnicas diferentes para los factores, por lo cual es vital que al momento de llenar los huecos en un contaminante se utilice la misma ecuación que se utilizó para calcular ese factor. Aunque las ecuaciones son semejantes es importante no mezclarlas ya que si esto sucediese se estaría incurriendo en un error. Es decir, al momento de imputar los datos por ejemplo del contaminante $PM_{2.5}$, se debe verificar con qué fórmula se obtuvo el factor; supóngase que fue con la ecuación 3.2 y se encuentra completando el primer año, se deberá utilizar el YG calculado del primer año y dejar por un lado todos los demás factores. Una vez identificado el factor a emplear se despeja la ecuación para el $PM_{2.5}$, de esta manera la igualdad quedará establecida por la multiplicación del YG por el dato que ese día mostró el contaminante PM_{10} tal como se observa en la tabla 3.1.

Tabla 3.1. Ejemplo de imputación de grupo con YG en el contaminante $PM_{2.5}$

Fecha	$PM_{2.5}$ Org	PM_{10} Org	YG	PM_{10}*YG	$PM_{2.5}$ Imp con PM_{10}
12-Ene	36.25	58.00	0.6719	38.97	36.25
13-Ene		52.50	0.6719	35.28	**35.28**
14-Ene		68.70	0.6719	46.16	**46.16**
15-Ene		72.60	0.6719	48.78	**48.78**
16-Ene		74.80	0.6719	50.26	**50.26**
17-Ene		89.60	0.6719	60.20	**60.20**
18-Ene		104.90	0.6719	70.48	**70.48**
19-Ene		71.90	0.6719	48.31	**48.31**
20-Ene		98.50	0.6719	66.18	**66.18**
21-Ene		105.80	0.6719	71.09	**71.09**
22-Ene	65.71	62.80	0.6719	42.20	65.71
23-Ene	42.42	69.90	0.6719	46.97	42.42

Imp: Imputado. Org: Original.

YG: Factor de ciclo. Ene: Enero.

$PM_{2.5}$: Partículas con diámetro menor a 2.5 Micrómetros.

PM_{10}: Partículas con diámetro menor a 10 Micrómetros.

Tabla 3.2. Ejemplo de imputación de grupo con GG en el contaminante PM_{10}

Fecha	$PM_{2.5}$ Org	PM_{10} Org	GG	$PM_{2.5}$/GG	PM_{10} Imp con $PM_{2.5}$
20-Ene		98.50	0.6719		98.50
21-Ene		105.80	0.6719		105.80
22-Ene	65.71	62.80	0.6719	97.80	62.80
23-Ene	42.42	69.90	0.6719	63.13	69.90
24-Ene	52.17	85.10	0.6719	77.64	85.10
25-Ene	56.42	86.60	0.6719	83.96	86.60
26-Ene	54.46	83.50	0.6719	81.05	83.50
27-Ene	49.46		0.6719	73.61	**73.61**
28-Ene	40.50		0.6719	60.28	**60.28**
29-Ene	46.96		0.6719	69.89	**69.89**
30-Ene	49.21		0.6719	73.24	**73.24**
31-Ene	42.13		0.6719	62.69	**62.69**
01-Feb	48.88		0.6719	72.74	**72.74**
02-Feb	59.75		0.6719	88.93	**88.93**

Imp: Imputado. Org: Original. Feb: Febrero.
GG: Factor de Serie. Ene: Enero.
$PM_{2.5}$: Partículas con diámetro menor a 2.5 Micrómetros.
PM_{10}: Partículas con diámetro menor a 10 Micrómetros.

Por otro lado se desea encontrar datos faltantes en el contaminante PM_{10}, suponiendo que estos sean del segundo año utilizando el GG, el cual es calculado con la serie completa, por lo cual será el mismo factor para cada uno de los años y no se tiene la necesidad de cambiar el factor en cada año.

Al despejar con el fin de encontrar los valores perdidos para la serie de tiempo del PM_{10}, se tiene que esta es calculada mediante la división de la concentración en el $PM_{2.5}$ entre el factor GG, como se aprecia en la tabla del ejemplo de la tabla 3.2.

Un dato importante es que el tratamiento facto de grupo puede ser combinado con cualquiera de los otros dos tratamientos (AA, Promediado).

3.3 Imputación de promediado

3.3.1 Promediado por unidad (UA)

Para completar los faltantes en los datos de las series de tiempo por este método, es necesario tomar la serie que contenga los faltantes y separarla. Contemplando que la descomposición de la misma debe estar basada en comportamientos cíclicos, para lo cual se seleccionaron los años (*i.e.* 2003, 2004,.., 2007) como dicha base. Esto es debido a que se aprecia que durante cada uno de los años se presentan comportamientos que modifican las concentraciones de los contaminantes y estos a su vez son repetidos cada año. Como son los festejos de las fiestas navideñas, temperaturas extremas típicas de un desierto, entre otras.

Una vez seccionada la base de tiempo en la cual se descompondrá la serie, se procede a determinar el promediado por unidad en este caso anual, dicha acción generará una nueva serie a la cual se le denomina patrón y se utilizará para encontrar los datos faltantes de la serie original.

Para crear la serie patrón (ver figura 3.4) se necesita colocar los datos de la base seleccionada (anuales) en paralelo y se efectúa el promediado de los años, el promediado es llevado a cabo contemplado cada uno de los días que examina la investigación.

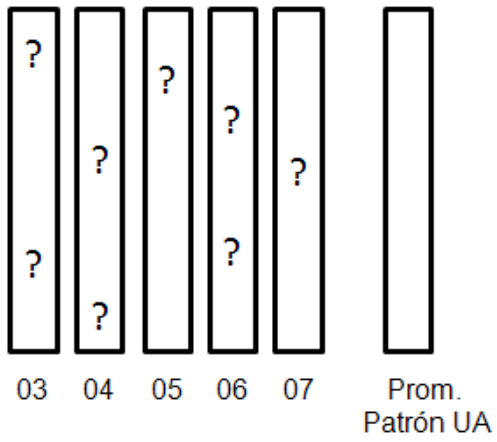


Figura 3.4. Cómo se ordena para el cálculo del patrón UA

Tabla 3.3. Organización de los datos para el cálculo del patrón en el método promedios por unidad

Fecha	CO 2003	CO 2004	CO 2005	CO 2006	CO 2007	CO Patrón 03-07
01-Ene	2.4071	3.2397	1.8413	1.8932	1.0119	2.0786
02-Ene	2.6675	2.3667	1.2565	0.4341	1.4333	1.6316
03-Ene	5.819	0.973	1.2	1.4246	2.2317	2.3297
04-Ene	7.0619	3.2302	0.7014	1.3818	1.0687	2.6888
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
28-Dic	3.0857	1.6317	2.6594	0.2913	1.5452	1.8427
29-Dic	2.4246	0.5675	2.7507	1.3841	1.9167	1.8087
30-Dic	2.6468	0.4571	2.3391	1.3357	1.6952	1.6948
31-Dic	3.3968	1.377	1.4511	1.7774	1.1063	1.8217

CO: Monóxido de Carbono. 03: Año 2003. 04: Año 2004.
05: Año 2005. 06: Año 2006. 07: Año 2007.
Ene: Enero. Dic: Diciembre.

La forma adecuada de desarrollarlo consiste en acomodar los datos de cada año empezando por el primero de Enero y culminando con el 31 de Diciembre, haciendo coincidir los días de cada año como se aprecia en la tabla 3.3.

Una vez alineados los datos se realiza el promedio por día, es importante respetar los faltantes en los datos que presenta cada año, para conservar los huecos solo se necesita dejar un espacio en blanco por cada dato faltante, de ninguna manera se deberán omitir estos espacios. Si se diera el caso de excluir los huecos esto podría producir una alteración en el resultado del promediado y el patrón obtenido puede ser erróneo.

Teniendo el patrón determinado se procede a reconstruir los huecos presentes en los determinados días de la serie, utilizando los datos diarios respectivos del patrón del contaminante. La metodología explicada se muestra mediante un simple ejemplo: supóngase que hiciese falta el dato que corresponde al día doce de Noviembre del 2007, entonces se toma el dato que corresponde a ese día en el patrón, en este caso el valor es

Tabla 3.4. Ejemplo de imputación utilizando el patrón en el método promediado por unidad

FECHA	CO Org 07	Patrón 03-07	Imp CO 07
10-Nov-07	0.7000	1.7603	0.7000
11-Nov-07	1.1667	1.6161	1.1667
12-Nov-07		1.2862	**1.2862**
13-Nov-07		1.4048	**1.4048**
14-Nov-07		2.3493	**2.3493**
15-Nov-07		1.1589	**1.1589**
16-Nov-07		1.7935	**1.7935**
17-Nov-07		2.6983	**2.6983**
18-Nov-07		2.8502	**2.8502**
19-Nov-07		2.9909	**2.9909**
20-Nov-07		2.5023	**2.5023**
21-Nov-07		1.9244	**1.9244**
22-Nov-07	1.6400	1.4535	1.6400
23-Nov-07	0.6091	1.6684	0.6091

Org: Original. Imp: Imputada.
03: Año 2003. 07: Año 2007.
Nov: Noviembre. CO: Monóxido de carbono.

1.2862, se coloca en el hueco de la serie, se recomienda ponerlo con un color diferente con el fin de poder identificar fácilmente cuales son los datos imputados, posteriormente se realiza esta acción con los de mas valores faltantes de la serie, como se aprecia en la tabla 3.4.

3.3.2 Imputación de promediado por exclusión (EA)

Al igual que los promedios por unidad, este tratamiento también requiere establecer ciclos dentro de la seria que se requiere imputar. Semejante a lo hecho anteriormente la serie de tiempo del contaminante es seccionada separando los años que se incluyen en el estudio. La diferencia de estas dos técnicas radica en la forma en la que se realiza el promediado, mientras la técnica anterior se concentraba en solo encontrar un patrón y estos valores los utilizaba todos los años, la técnica actual encuentra patrones por año, por lo cual

los valores a imputar en un año son diferentes a los valores que se utilizaran para completar otro año del mismo contaminante.

El método establece la realización del promediado del primer año, el cual empieza por excluir dicho año y solamente se promedian los restantes. Para ejemplificar mejor este procedimiento se utiliza la figura 3.5, en la cual se pretende encontrar la serie de tiempo con la cual se imputará el primer año que contempla el proyecto, o sea, el año 2003. Este será excluido y se promedian del 2004 al 2007 utilizando los huecos propios de cada año, al igual que los otros procedimientos. Es importante el manejo que se le dan a los faltantes de cada año dentro de la serie, es necesario respetar con un espacio en blanco cada vez que se presente alguno, por ningún motivo se deben de omitir los espacios o rellenar con datos promediados de otros años: La serie de tiempo resultante (ver tabla 3.5) será el patrón a utilizar para imputar el año en cuestión, en este caso se trata del 2003.

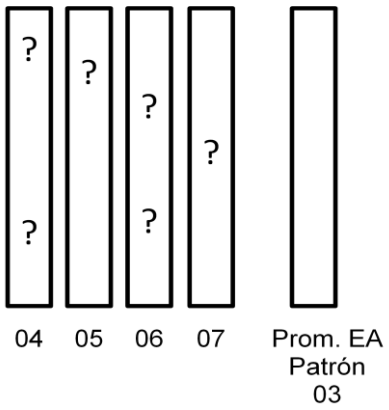


Figura 3.5. Esquema de cómo organizar los datos para el método EA.

Tabla 3.5. Ejemplo de cómo crear un patrón en el tratamiento promedio por exclusión (EA)

Fecha	CO 04	CO 05	CO 06	CO 07	Patrón CO 04-07
01-Ene	2.4071	3.2397	1.8413	1.8932	2.3453
02-Ene	2.6675	2.3667	1.2565	0.4341	1.6812
03-Ene	5.8190	0.9730	1.2000	1.4246	2.3542
04-Ene	7.0619	3.2302	0.7014	1.3818	3.0938
05-Ene	4.4429	2.7849	1.3068	1.0273	2.3905
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
27-Dic	1.4222	2.8349	1.6362	0.8952	1.6972
28-Dic	3.0857	1.6317	2.6594	0.2913	1.9170
29-Dic	2.4246	0.5675	2.7507	1.3841	1.7817
30-Dic	2.6468	0.4571	2.3391	1.3357	1.6947
31-Dic	3.3968	1.3770	1.4511	1.7774	2.0006

CO: Monóxido de Carbono. 04: Año 2004. 05: Año 2005.
06: Año 2006. 07: Año 2007. Ene: Enero.
Dic: Diciembre.

De presentarse la necesidad de obtener el patrón para otro año como por ejemplo el 2005, este será necesario omitirlo del promediado, por lo cual estarían interviniendo solamente los años 2003, 2004, 2006 y 2007. La serie obtenida de la anterior acción entrega el patrón a utilizar para el año 2005. Y siguiendo este lineamiento es como se debe de aplicar la técnica para cada uno de los años que cuente con datos perdidos y que a su vez sean necesarios en la investigación.

El siguiente paso es completar los faltantes de la serie original con los valores de la serie patrón, es importante tomar la serie patrón correspondiente del año que se desea e imputar los valores que corresponden exactamente al día donde se encuentra el hueco o los huecos, utilizando el procedimiento que se observó en la tabla 3.4. Es importante tanto en esta técnica como en la anterior que el día en el que se produce el hueco, sea el mismo día que se tome el valor del patrón, ya que de lo contrario se estará atrasando o adelantando los valores en la serie, lo cual incurriría en un error.

3.4 Automatic ARMA (AA)

El nombre de ARMA proviene de las siglas en inglés que quieren decir autoregresivo (AR-"Autoregressive") y medias móviles (MA -"Moving Average"). Este método consiste en crear varios modelos basados en las medias móviles y autogresiones encontrados a partir de la serie de tiempo que se pretende imputar, con el fin de pronosticar una cierta cantidad de datos subsecuentes de la misma. Para identificar cual de los modelos creados es el indicado, esta herramienta se guía por el comportamiento del análisis residual.

Resumiendo, se tiene que el instrumento crea varios modelos candidatos, luego hace comparaciones con sus residuos y el que presenta el mejor comportamiento es seleccionado. Es por ello que se le ha nombrado "Automatic" que en español significa automático, ya que selecciona por sí mismo un determinado modelo, automatizando esta decisión [49]. Esta es una herramienta que se puede encontrar dentro del programa estadistico NCSS.

3.4.1 Metodología utilizada por ARMA

El programa NCSS basa esta herramienta en el método Box-Jenkins, el cual plantea una serie de pasos con el propósito de construir un modelo que prediga la cantidad de datos que se desean o necesitan, los cuales serán utilizados para imputar las series de tiempo en el proyecto:

- Remover la tendencia de la serie de datos. Para lograrlo el software se apoya en el método planteado por Padit and Wu [50].
- Calcular los modelos. Para realizarlo es necesario determinar las autoregresiones y medias móvibles a partir de la serie de tiempo, El software parte del modelo más sencillo el cual lo construye con la primera autoregresión y media móvil, después solo sigue incrementándose la cantidad de medias móviles y autoregresiones que contemplará para la construcción del modelo. La ecuación 3.3 muestra los fundamentos del método:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + a_1 - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \qquad \textbf{Ecuación 3.3}$$

Donde

ϕ_p : son los parámetros de autoregresión a estimar.

θ_q : son las medias móviles.

X_t : son la serie original.

a_{t-q} : son los residuales.

- Determinar el modelo óptimo. El criterio que se sigue para escoger dentro de todos los modelos creados anteriormente, es escoger el modelo que contenga los valores más pequeños en la suma cuadrática de los residuales. En un caso ideal esta suma cuadrática expresa que sus residuales no son significativos y que la predicción elaborada es confiable. Pero se pueden observar casos en los que todos los modelos presenten valores residuales significativos; en esta situación el programa selecciona al que presente el mejor comportamiento, aunque incluya residuales significativos, pero al darse esta situación el modelo no será de utilidad, por lo cual es necesario que cada vez se construya un modelo y se corroboren sus residuales, con el fin de no incluir datos que pudieran en un momento determinado aportar un sesgo en la serie de tiempo.

La herramienta AA se utiliza para predecir los faltantes de la serie, el autor decide aplicar la herramienta utilizando dos métodos a los cuales nombró AA por bloques y AA acumulado, la razón de ser nombrados de esta manera obedece a la forma en la que cada respectivo método trabaja.

3.4.2 AA por bloques (BAA)

Se comienza por separar la serie de tiempo por unidad, de esta manera se tomó como unidad cada año estudiado, teniendo cinco años que corresponden al periodo del 2003 al 2007.

Dentro de cada año se toma el primer bloque de datos, el cual comienza el primer día del año y culmina en cuanto se encuentre con el primer dato faltante (observar figura 3.6). Una vez identificado el bloque se procede a determinar un modelo con los datos contenidos dentro de él. Utilizando el software NCSS con su herramienta AA se pronostican la cantidad de datos faltantes que radican después del bloque.

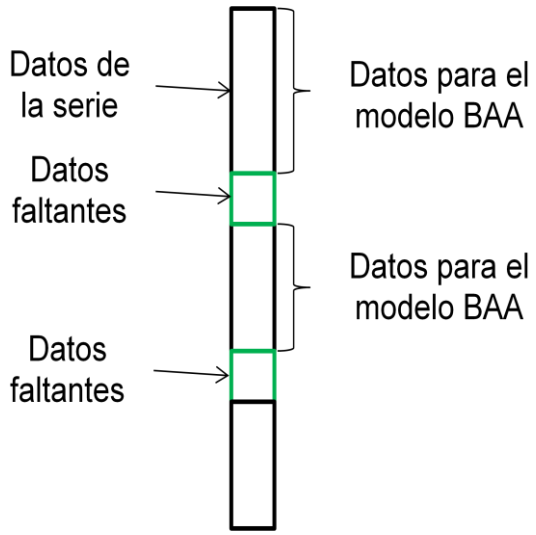


Figura 3.6. Orden de los datos para establecer los modelos predictivos de los faltantes con el tratamiento BAA.

Tabla 3.6. Ejemplo de selección de datos por BAA

Fecha	PM_{10} Original	PM_{10} Imp con YG y BAA
06-Abr-03		**115.3**
07-Abr-03	67.9	67.9
08-Abr-03	57.7	57.7
09-Abr-03	81.6	81.6
10-Abr-03	88.8	88.8
11-Abr-03	81.5	81.5
12-Abr-03	50.3	50.3
13-Abr-03	41.2	41.2
14-Abr-03	118.9	118.9
.	.	.
.	.	.
.	.	.
26-May-03	96.4	96.4
27-May-03	112.9	112.9
28-May-03		**119.1**
29-May-03		**89.4**
30-May-03		**48.3**
31-May-03		**100.1**
01-Jun-03		**75.7**
02-Jun-03		**115.3**
03-Jun-03	91.6	91.6
04-Jun-03	96.1	96.1

Imp: Imputado. YG: Factor de Ciclo.
Abr: Abril. May: Mayo.
Jun: Junio. BAA: Bloque Automático de ARMA.

Para plasmar la idea sobre este procedimiento es pertinente observar la tabla 3.6 en la cual se identifica el segundo bloque de datos, el cual es tomado a partir del primer dato que la serie original prosigue, después del último dato imputado por el primer bloque, y se selecciona la cantidad de datos encontrados antes de presentarse el siguiente faltante. Al tener delimitado el segundo bloque se continua con el mismo procedimiento indicado para el primer bloque, el cual consistía en la utilización de la herramienta “Automatic ARMA” y se pronostica la cantidad de datos requeridos para completar el segundo hueco que la serie

presenta. Así sucesivamente se seleccionan los bloques de datos e imputando los faltantes de la serie por año.

3.4.3 AA por acumulado (AAA)

Al igual que el procedimiento por bloques, en el acumulado también se separa la serie de tiempo por unidad, la cual fue determinada utilizando el mismo procedimiento teniendo cinco unidades las cuales corresponden a los cinco años que contempla el estudio.

Se selecciona el primer acumulado del año, el cual parte del primer valor que se presente en el año hasta el primer faltante registrado dentro del mismo; mediante el uso de la herramienta "AA" se pronostican los datos faltantes en el hueco contiguo al primer acumulado.

Identificación del segundo acumulado (seguir figura 3.7). Se parte del primer dato registrado en la serie durante el año en cuestión y se selecciona la cantidad de datos del primer acumulado más los datos predichos por el procedimiento anterior; además de agregar los datos presentes después del último dato pronosticado hasta toparse con el siguiente hueco de la serie tal como se observa en la tabla 3.7. Nuevamente se utiliza la herramienta ya mencionada y se pronostican los datos necesarios para cubrir los faltantes, de la misma forma se hacen los acumulados para los demás huecos dentro de un mismo año.

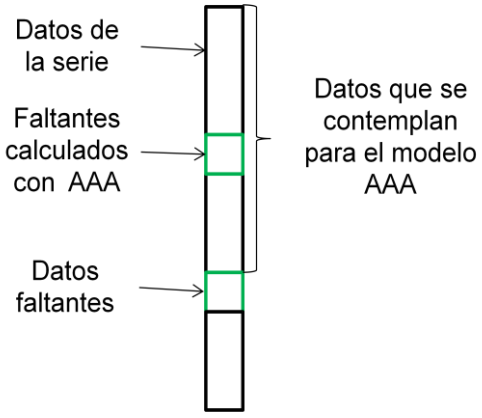


Figura 3.7 Esquema de cómo organizar los datos para el método AAA.

Tabla 3.7. Ejemplo de selección de datos por Acumulado AA

Fecha	PM_{10} Original	PM_{10} Imp con YG y BAA
06-Abr-03		**115.3**
07-Abr-03	67.9	67.9
08-Abr-03	57.7	57.7
09-Abr-03	81.6	81.6
10-Abr-03	88.8	88.8
11-Abr-03	81.5	81.5
12-Abr-03	50.3	50.3
13-Abr-03	41.2	41.2
14-Abr-03	118.9	118.9
.	.	.
.	.	.
.	.	.
26-May-03	96.4	96.4
27-May-03	112.9	112.9
28-May-03		**119.1**
29-May-03		**89.4**
30-May-03		**48.3**
31-May-03		**100.1**
01-Jun-03		**75.7**
02-Jun-03		**115.3**
03-Jun-03	91.6	91.6
04-Jun-03	96.1	96.1

Imp: Imputado. YG: Factor de Ciclo.
Abr: Abril. May: Mayo.
Jun: Junio. BAA: Bloque Automático de ARMA.

La diferencia entre estos método consiste en que el primero (bloques) no contempla los datos predichos por los anteriores modelos, solo contempla datos originales de la serie; en cambio el método de acumulado contempla no sólo los datos predichos si no también los bloques anteriores al hueco que se está trabajando.

3.5 Amelia

Amelia es un software diseñado para imputar datos faltantes dentro de las series de tiempo, la forma en la que este software funciona es diferente a las técnicas planteadas anteriormente. El programa emplea correlaciones cruzadas de diferentes series de tiempo que fueron suscitadas simultáneamente respecto a la serie que se pretende imputar. Es importante mencionar que el programa asume que los faltantes siguen un comportamiento aleatorio, dicho comportamiento es conocido por sus siglas en inglés como MAR (Missing At Random) [49].

El programa comienza su fundamentación al crear una matriz a la cual es nombrada Y con dimensiones n x p utilizando las series de tiempo, como lo muestra el esquema de la figura 3.8. Al establecer la matriz, las columnas tienen dimensión P y son constituidas por cada una de las serie de tiempo, mientras que los N renglones representan las unidades de observación, las cuales deben de ser independientes y contar con distribuciones idénticas, a este tipo de formato se le ha denominado como iid [51, 52].

Elabora una matriz R con dimensiones idénticas a la de Y, su diferencia radica en el contenido de la misma, ya que es una matriz de ceros y unos. Le proporciona el valor de uno si el dato que se encuentra en esa misma celda dentro de la matriz Y existe y a este determinado conjunto de datos existentes dentro de la matriz se les denomina Y_{obs}. En el caso contrario, es decir de encontrarse con un faltante en la matriz Y, la matriz R le otorgara el valor de cero a dicha celda, de igual manera al conjunto de faltantes dentro de de la matriz Y son conocidos como Y_{mis} [51, 53, 54].

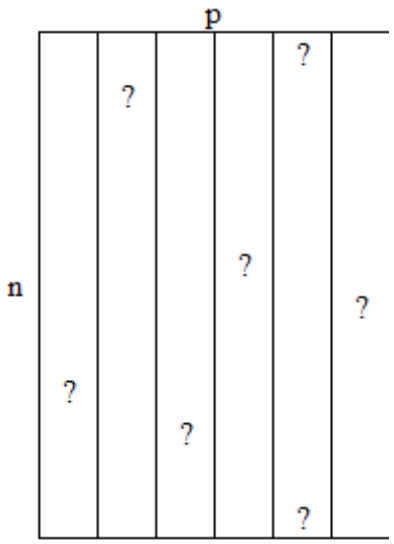


Figura 3.8. Matriz Y, empleada por Amelia II.

El análisis se maneja renglón por renglón apoyándose en la variable i, la cual se mueve entre los valores del uno hasta n. Debido a que los renglones cuentan con el supuesto iid, la función de probabilidad está descrita por la ecuación 3.4 [51, 53, 54].

$$P(Y \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta) \qquad \textbf{Ecuación 3.4}$$

Donde:

Y: Matriz que contiene los valores de las series de tiempo consideradas en el análisis.

f: Es la función de probabilidad para cada renglón.

θ: Es el vector parámetro.

y_i: Renglón de la matriz Y.

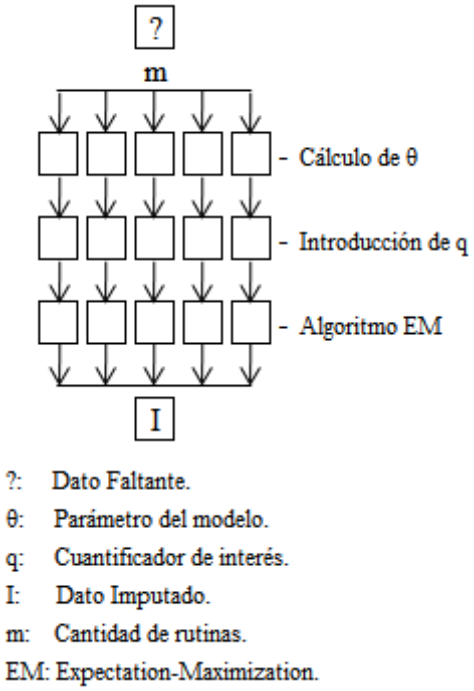


Figura 3.9. Rutina del Programa Amelia

Al momento de enfrentarse con un dato faltante dentro de la serie dependiente, se ejecuta un procedimiento, el cual cuenta con m rutinas en paralelo, donde la cantidad m es indicada por el investigador o analista. Como se aprecia en la figura 3.9 con el fin de otorgar un valor al hueco de la serie, lo cual es mejor conocido como imputación [51, 53, 55].

Debido a que la matiz R está relacionada con la matriz Y, es evidente que su probabilidad también la contemple P(R|Y,ξ). Por lo que la densidad de distribución de las variables R y Y queda expresada en la ecuación 3.5 [56].

$$f(Y,R \mid \theta,\xi) = f(Y,\theta) f(R \mid Y,\xi) \qquad \textbf{Ecuación 3.5}$$

Donde:

Y: Matriz que contiene los valores de las series de tiempo consideradas en el análisis.

R: Es la matiz de ceros y unos.

f: Es la función de probabilidad para cada renglón.

θ: Es el vector parámetro del modelo.

ξ: Parámetro desconocido, que representa la distribución del mecanismo de los datos faltantes.

La variable Y es descompuesta en Y_{obs} y Y_{mis}, se sabe que $f(Y \mid \theta) \equiv f(Y_{obs}, Y_{mis} \mid \theta)$, por lo tanto la probabilidad conjunta está expresada por la distribución de las variables Y_{obs} y Y_{mis}, pero también es conocido que se puede obtener una densidad de probabilidad marginal contemplando solamente los valores de Y_{obs} e integrando con respecto a Y_{mis}. Aplicando el anterior concepto, la ecuación 3.5 queda expresada como lo muestra la ecuación 3.6 [51, 56].

$$f(Y_{obs}, R \mid \theta, \xi) = \int P(R \mid Y_{obs}, \xi) P(Y_{obs}, Y_{mis} \mid \theta) dY_{mis}$$
$$= P(R \mid Y_{obs}, \xi) P(Y_{obs} \mid \theta) \qquad \textbf{Ecuación 3.6}$$

La verosimilitud es calculada utilizando dos factores, uno que depende del parámetro del modelo θ, y el otro que pertenece al parámetro de distribución de mecanismo de datos faltantes ξ. De acuerdo a lo establecido por Little and Rubin [56], los parámetros θ y ξ son distintos, y comprobaron que la verosimilitud puede ser calculada solamente conociendo el parámetro del modelo, Con este establecimiento montaron la ecuación permite calcularla evidenciada en la ecuación 3.7 [53].

$$L(\theta \mid Y_{obs}) \ \alpha \ P(Y_{obs} \mid \theta) \qquad \textbf{Ecuación 3.7}$$

θ se establece como el parámetro que proporciona μ y Σ para los renglones de la matriz Y (θ =(μ, Σ)), no solamente para los renglones Y_{obs}, esto es debido a que este análisis es basado en la inferencia de que los datos faltantes siguen un comportamiento de

tipo MAR [51, 56]. Con lo anterior se demuestra que es válido encontrar los datos perdidos de la matriz Y basándose en los datos observados dentro de la misma [55].

En cada imputación el programa intenta evitar el exceso de confianza, para hacer esto adiciona errores estándar (SE) generados por estructuras Bayesianas utilizando cuantificación de interés (q) en cada grupo m de imputación; este procedimiento es conocido como distribución posterior (posterior distribution, por su nombre en Inglés) [51, 54, 55]

Para elaborar el modelo que generará el dato estimado, se utiliza el algoritmo "Expectation-Maximization" mejor conocido como EM, el cual funciona escogiendo valores aleatorios entre los parámetros μ, Σ contemplando la distribución posterior de $p(p+3)/2$, donde p es el número de variables explicativas que son utilizadas en el modelo [53, 55].

El resultado generado por el software será una m cantidad de archivos, donde m representa la cantidad de imputaciones que fueron solicitadas por cada valor faltante. Dichos archivos contendrán las series de tiempo completas, los autores recomiendan promediar los datos por variable para generar la serie de tiempo que ellos consideran la más idónea [55].

3.6 Comparación

Una vez que se cuenta con las series imputadas es necesario saber cuál de ellas es la que estadísticamente es la más cercana a la serie original, con esto se pretende saber cuál fue el método o métodos que dieron lugar al mejor ajuste. La forma común de comparar dos series de tiempo es haciendo una correlación entre ellas.

Desafortunadamente este método no puede ser aplicado en este caso, ya que la serie de tiempo original cuenta con faltantes y al realizar la correlación estos renglones serán

omitidos. Debido a esta problemática se decide desarrollar unos métodos para cuantificar dicha comparación con el fin de contar con una base solida para poder tomar la decisión.

Los métodos son basados en la separación por ventiles. El primer ventil también es conocido como percentil 5, lo cual hace ver que los ventiles no son más que la separación de la serie de tiempo en 20 secciones, coincidiendo con el valor mostrado cada 5 percentiles [49, 57].

3.6.1 Comparación por Ventiles.

Es necesario segregar la serie original que presenta los faltantes, al igual que las series imputadas, utilizando el procedimiento de identificación de ventiles, tal como se observa en tabla 3.8. En la gráfica de distribución de ventiles (figura 3.10) se puede observar el seguimiento de la serie imputada sobre la serie original.

Una vez identificado los ventiles de cada una de las series, se procede a correlacionarlos, la finalidad de este procedimiento es identificar cual de las series imputadas presenta la mayor cercanía estadística a la serie original.

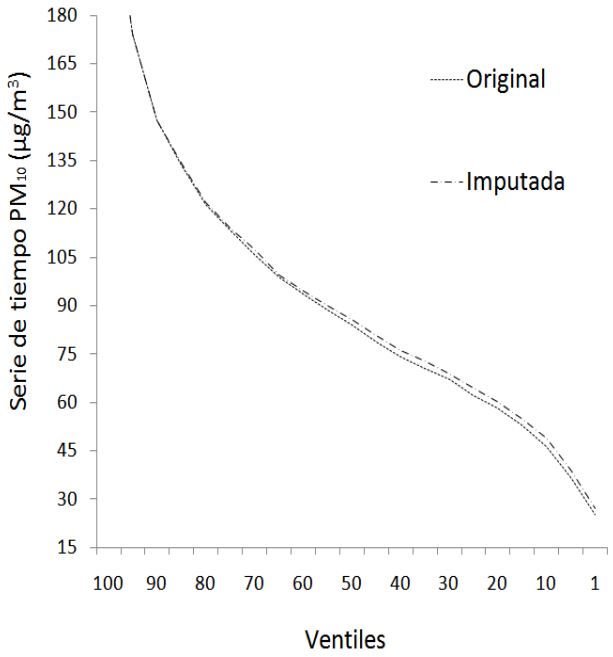


Figura 3.10. Gráfica de distribución de Ventiles

Tabla 3.8. Serie separada por ventiles

Ventil	VSO	VSI
100	226.61	229.49
95	174.36	174.11
90	147.42	147.50
85	133.50	134.40
80	121.30	121.90
75	113.50	114.00
70	105.88	107.39
65	99.00	99.60
60	93.64	94.71
55	88.52	89.80
50	84.00	85.65
45	78.88	80.78
40	74.10	76.19
35	70.40	72.70
30	67.00	68.80
25	62.30	64.60
20	58.10	60.10
15	53.06	55.01
10	46.34	49.00
5	36.84	39.04

VSO: Valor serie original.

VSI: Valor serie Imputada.

3.6.2 Comparación por Rangos Interventílicos.

Semejante a la comparación anterior, es necesario identificar los ventiles en las series que se desean comparar (ie. serie original y series imputadas), la diferencia en este procedimiento radica en la creación de los rangos interventílicos, los cuales son elaborados de una manera semejante al cálculo del rango de un cuartil; basándose en la figura 3.11 se toma el dato del último ventil de la serie (percentil 100) y se le resta el dato del antepenúltimo ventil de la misma (percentil 90), esto indica que siempre permanecerá un ventil en medio de esta resta. De la misma manera se sigue avanzando en los datos, al ventil 19 se le restará el ventil 17, así sucesivamente hasta llegar al tercer ventil, al cual se le restará el primero, el resultado será semejante al observado en la tabla 3.9.

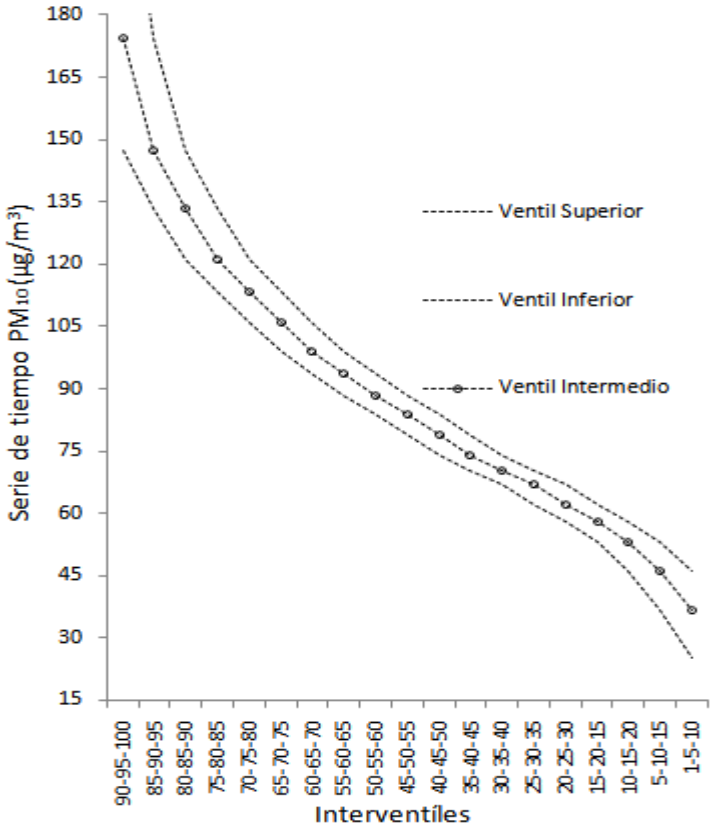


Figura 3.11. Orden de ventiles para crear los rangos interventílicos.

Ya calculados los Rangos Interventílicos de cada una de las series (original e imputada), el siguiente paso es correlacionarlos. La finalidad de este de realizar este procedimiento es tener un mecanismo que maneje una precisión más fina, con el fin de cuantificar cual de las series imputadas que presentan la mayor cercanía estadística a la serie original.

Las series de tiempo de los contaminantes fueron las que presentaron faltantes, es importante mencionar que hay contaminantes que podrán ser imputados con más métodos que otros, como fue el caso del PM_{10} y $PM_{2.5}$, los cuales fueron tratados con 13 técnicas diferentes. Mientras que los contaminantes CO, NO_2, O_3 y SO_2 fueron imputados con sólo cinco.

Tabla 3.9. Rangos Interventílicos

RIV	VSO	VSI
90-100	79.2	82.0
85-95	40.9	39.7
80-90	26.1	25.6
75-85	20.0	20.4
70-80	15.4	14.5
65-75	14.5	14.4
60-70	12.2	12.7
55-65	10.5	9.8
50-60	9.6	9.1
45-55	9.6	9.0
40-50	9.9	9.5
35-45	8.5	8.1
30-40	7.1	7.4
25-35	8.1	8.1
20-30	8.9	8.7
15-25	9.2	9.6
10-20	11.8	11.1
5-15	16.2	16.0
1-10	21.1	22.0

VSO: Valor serie original.

VSI: Valor serie Imputada.

RIV: Rango Interventílico.

Las comparaciones hechas mediante la metodología de ventil, dieron como resultado que todos los métodos manifestaron tener correlaciones perfectas de r=1, con una alta probabilidad.

Las comparaciones hechas con la metodología de interventil, manifestaron ser correlaciones muy altas y se alcanza apreciar que efectivamente hay diferencia entre los métodos utilizados en el estudio. Estos resultados pueden ser apreciados en el apéndice A

Capítulo 4

Asociación de los principales contaminantes del aire con el riesgo relativo de muerte en la población de Mexicali

4.1 Introducción

Tal y como se ha observado, la contaminación viene degradando la calidad de vida de la población; principalmente la que se encuentra en exposición directa. Este hecho se traduce en incrementos en las enfermedades, tales como infecciones respiratorias, asma, cáncer, e incluso llega a desembocar en muerte [58-60].

¿La contaminación aérea afecta a la población de Mexicali? La capital del estado de Baja California tiene una grave problemática acerca de los contaminantes PM_{10} y CO. Solamente durante el año 2004 el PM_{10} rebasó la norma 88 días posesionando a Mexicali en el deshonroso primer lugar dentro del país, otro contaminante que pone esta ciudad en la misma posición es el Monóxido de Carbono, aunado a esto, se tiene la problemática con las emisiones de NO_2, SO_2 y O_3 [11]. Estos contaminantes generan complicaciones en la salud de los habitantes de la localidad afectada y algunos estudios han comprobado su relación con la morbilidad [61, 62], entonces es posible encontrar en la literatura trabajos que vinculan y cuantifican, mediante correlación o regresión, algunos contaminantes presentes

en la ciudad, como por ejemplo PM_{10}, CO y O_3 con casos de asma, neumonía e infecciones respiratorias aguadas [63, 64].

Cuantificar cuánto contribuye cada uno de los contaminantes del aire al Riesgo Relativo (RR) de muerte de la población mexicalense, es un dato importante para comparar con los estándares del aire de la región.

4.2 Materiales.

4.2.1 Lugar de estudio.

Mexicali está situada dentro de los Estados Unidos Mexicanos, al extremo noroeste del país. Debe su nombre a la unión de las dos sílabas iniciales de México y California, esta última es un estado de la Unión Americana con la cual colinda. Su localización geográfica está delimitada por las latitudes 32° 43’ norte y 30° 52’ sur; a su vez por las longitudes 114º 42’ este y 115º 56’ oeste y la altitud del municipio completo está situada a 8 m sobre el nivel del mar [65].

Según el INEGI [66] la población total de la ciudad de Mexicali para el año 2010 era de 936,826 habitantes y la densidad demográfica de 64.3 habitantes por km^2; Además las principales actividades desarrolladas fueron las siguientes: agrícola, industria (aeroespacial, metalmecánica, electrónica, médica y automotriz), ganadera, a su vez cuenta con dos termoeléctricas, una geotérmica, además de contar con una flota vehicular antigua [66, 67].

4.2.1 Variables de estudio

Crear una base de datos para cada una de las variables a considerar en el estudio (i.e. clínicas, meteorológicas y contaminantes), la cual debe de contemplar los años del estudio que son del 2003 al 2007, además de que esta base de datos no debe tener faltantes, pero desafortunadamente la serie de contaminantes cuenta con datos no registrados, por lo cual fue tratada con las técnicas presentadas en el capítulo 3.

4.2.1.1 Series de tiempo clínicas

Los registros de las variables clínicas fueron cedidos por la Jefatura de Jurisdicción Sanitaria No1 ISESALUD de Mexicali. La base de datos está compuesta de dos series de tiempo clasificada de la siguiente manera [62]:

- Muertes: Está serie de tiempo es considerada como la variable respuesta y es representada por la suma de los fallecimientos naturales diarios. Para hacer su cálculo es necesario realizar una selección exclusiva de las defunciones identificadas como naturales, pues las mismas se ubican dentro de los códigos A00 hasta R99 quedándose excluidos los códigos que van del V01 en adelante, por tratarse de causas externas [68]. La nomenclatura utilizada se llama Clasificación Internacional de Enfermedades en la décima edición CIE-10, la cual es empleada tanto para enfermedades como para casos de muerte.

- Infecciones Respiratorias Agudas (IRAs): Los registros de esta variable en Mexicali son conteos semanales contemplando la suma de los casos ocurridos, por otro lado el estudio emplea casos diarios. Por esta razón se correlacionó la serie de tiempo IRAs de Mexicali, con las IRAs reportadas diariamente en el Valle Imperial, y se observó una alta correlación, $r = 0.8$, entre estas variables [62]. Como Mexicali emplea la décima Clasificación Internacional de Enfermedades (CIE-10) dentro de las enfermedades infecciosas del aparato respiratorio, los códigos que son considerados como IRAs son [68]:

  J00 - Rinofaringitis aguda (resfriado común).

  J01 - Sinusitis Aguda.

  J02.8 - Faringitis aguda debida a otros microorganismos especificados.

  J02.9 - Faringitis aguda, no especificados.

  J03.8 - Amigdalitis aguda debida a otros microorganismos especificados.

  J03.9 - Amigdalitis aguda, no especificados.

  J04 - Laringitis y Traqueítis agudas.

  J05 - Laringitis obstructiva aguda (crup) y epiglotitis.

J06 - Infecciones agudas de las vías respiratorias superiores de sitios múltiples y no especificados.

J20 - Bronquitis aguda.

J21 - Bronquitis aguda (con broncospasmo).

Las IRAs del Valle Imperial utilizan la Novena Clasificación Internacional de Enfermedades (CIE-9) y los códigos que se contemplaron fueron los del 460 al 466:

460 Resfriado común.

461.0 Sinusitis Aguda, maxilar.

461.1 Sinusitis Aguda, frontal.

461.9 Sinusitis aguda, NOS.

462 Faringitis aguda.

463 Amigdalitis aguda.

464.00 Laringitis Aguda, sin obstrucción.

464.30 Epiglottitis, aguda.

464.4 Crup.

465.9 Infecciones agudas de las vías respiratorias superiores, NOS.

466.0 Bronquitis aguda.

4.2.1.2 Series de tiempo de la contaminación

El organismo encargado de las estaciones de monitoreo en la ciudad de Mexicali es la Secretaria de Protección al Ambiente de Baja California [69], este órgano combinado con el Instituto Nacional de Ecología (INE) y con la Agencia de Recursos del Aire de California (California Air Resources Board - CARB) concedieron las series de tiempo de los contaminantes de Mexicali B.C., para la realización de este estudio.

La red de monitoreo atmosférico cuenta con seis estaciones dentro de la propia ciudad, como ilustra la figura 4.1. Estas estaciones surgieron gracias al programa Frontera XXI mediante el apoyo de la Agencia de Recursos del Aire de California, de la Agencia de Protección al Ambiente de Estados Unidos (USEPA, por su sigla en inglés) y de la

SEMARNAP, ahora conocido como SEMARNAT. Esta red tiene como finalidad medir las concentraciones de los contaminantes atmosféricos mediante las diversas estaciones como se muestra en la tabla 4.1 [11].

Para conocer la concentración diaria de contaminantes en el aire, inicialmente es necesario calcular el promedio de los datos de los contaminantes colectados en cada estación en una determinada hora, después obtener la media de los resultados de las estaciones en una misma hora y, finalmente, hacer la media de los valores calculados de cada hora.

Los principales contaminantes, aquellos con mayores concentraciones en Mexicali según lo establecido por el Instituto Nacional de Ecología, son los estudiados en este trabajo [11]:

- Partículas con diámetro menor a 10 micrómetros (PM_{10}) – medida en microgramos por metro cubico de aire ($\mu g \cdot m^{-3}$).
- Partículas con diámetro menor a 2.5 micrómetros ($PM_{2.5}$) – medida en $\mu g \cdot m^{-3}$.
- Monóxido de Carbono (CO), su medición está dada en partículas por millón (ppm).
- Ozono (O_3) medida en ppm.
- Dióxido de Nitrógeno (NO_2) medida en ppm.
- Dióxido de Azufre (SO_2) medida en ppm.

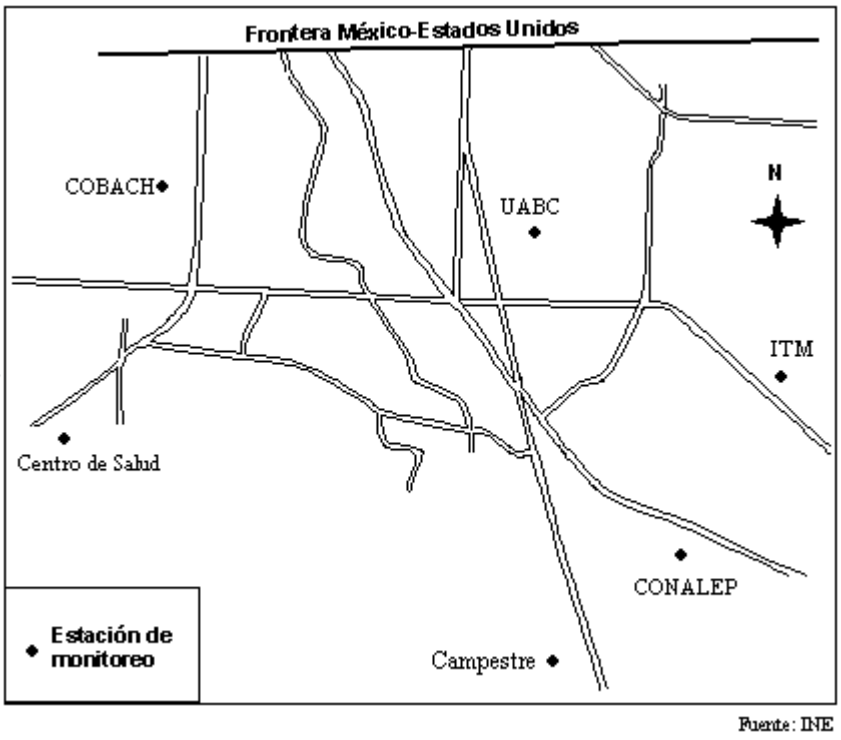


Figura 4.1. Distribución de Estaciones de Monitoreo en Mexicali BC.

Tabla 4.1. Contaminantes monitoreados por estación

Estación	Clave	PM_{10}*	O_3	CO	SO_2	NO_2
Campestre***	CAMP	✓	✓	✓		
Colegio de Bachilleres	Cobach	✓	✓	✓		✓
Universidad Autónoma de Baja California	UABC	✓**	✓	✓	✓	✓
Instituto Tecnológico	ITM	✓	✓	✓		✓
Centro de Salud	Progreso	✓				
CONALEP	CONALEP	✓				

* Equipo manual

** La estación UABC también tiene equipo automático para $PM_{2.5}$ y PM_{10}.

*** La estación Campestre empezó a medir las PM_{10} a partir de abril de 2004, y el resto de los contaminantes a partir de julio de 2005.

Fuente: Tercer Almanaque INE.

4.2.1.3 Series de tiempo meteorológicas.

Las series de tiempo meteorológicas analizadas fueron Temperatura (Temp) y Humedad Relativa (Hr). La primera es medida en grados centígrados o Celsius (ºC) y la otra es dada en porcentaje (%). Estos datos fueron proporcionados por el Departamento de Meteorología del Instituto de Ingeniería de la Universidad Autónoma de Baja California, el cual genera reportes semanales [62].

Tanto para la serie de Hr como para la serie de Temp fue realizado un promediado de los datos colectados en un mismo día con la finalidad de obtener el valor diario, pero para los datos de Temp se consideró también los datos mínimos registrados diariamente, pues algunos autores indican ser la mejor manera de manejar las series de tiempo para mejorar la predicción [60, 62].

4.2.1.4 Variables Categóricas.

También conocidas como variables ficticias o "dummy", son variables construidas de manera artificial con la finalidad de mostrar la presencia o ausencia de cierta característica, entonces su valor puede solamente ser cero o uno. Un ejemplo de estos tipos de variables serían las temporadas de frío y calor, pues en la temporada de frío esta variable asume el valor de uno, mientras la otra variable (temporada calor) es cero.

En este trabajo se realizó la identificación de dos temporadas (cálida e invernal) durante los cinco años del estudio. Se optó por identificar estas temporadas porque esta ciudad está geográficamente situada en un desierto, por lo cual presenta un clima extremoso, o sea, son bajas la temperatura en el invierno y altas en el verano; además en muchas de estas áreas, las estaciones de transición (primavera y otoño) sólo pueden ser observadas durante unos cuantos días, entonces en lugar de tener cuatro estaciones en esta región, solamente hay dos. Eso está de acuerdo con lo expresado en algunos artículos [62, 70, 71], en los cuales se hace mención a la división de las estaciones del año en dos temporadas. Al emplear estas variables en el modelo de regresión se puede logra una explicación más razonable del mismo.

Para determinar la manera de cómo los contaminantes pueden afectar la salud dependiendo de la época de estudio (i.e., invierno o verano), se recurre a la introducción de dos variables binarías, también conocidas como variables categóricas o de control. Estas variables dividen al año en dos temporadas [60], las cuales son las temporadas cálida y fría. La época identificada como cálida o verano se establece a partir del primer domingo del mes de abril hasta el último sábado del mes de octubre, durante este periodo la variable cálida tendrá un valor de uno; en caso contrario, el valor de ella será cero. Por el contrario, la temporada de frío o invernal se considera que inicia el último domingo del mes de octubre y culmina el primer sábado del mes de abril, en dicho periodo la variable contará con un valor de uno, al finalizar el periodo la variable cambiará su valor a cero, a este desarrollo también se le denomina términos de interacción [62, 64, 70].

Cuando las variables presentan más de dos categorías, como por ejemplo, los días de la semana, es necesario crear n-1 variables ficticias (en este caso se construirán seis variables dummy), donde el valor n indica la cantidad de categorías que se quieren describir (i.e., siete categorías, debido a que son siete los días de la semana). Estas variables se manejan en grupo, con esto se quiere decir que aunque una de las variables "dummy" no presente significancia estadística no puede ser retirada del modelo, a menos que sea retirado su grupo [72].

4.2.1.5 Colinealidad entre variables independientes.

El concepto de colinealidad está relacionado con la existencia de correlación entre variables independientes, además puede ser investigada por medio de la matriz de correlaciones de las variables ya mencionadas.

La existencia de colinealidad provoca problemas cuando se desea construir un modelo, pues la misma produce inestabilidades dentro del modelo, principalmente en el coeficiente de regresión y esta inestabilidad afecta directamente al error estándar del modelo [73].

Una manera sencilla de detectar la existencia de colinealidad dentro de un modelo, es colocar las series de tiempo de las variables independientes en una matriz, la cual se

correlaciona entre ella tomando en cuenta sus retardos. Para determinar la existencia de colinealidad el valor de la correlación debe ser igual al 0.8 o mayor.

4.3 Método.

Se utilizó el software Microsoft Excel (versión 2007) para ordenar los datos y el software estadístico NCSS 2007 (versión 7.15, NCSS, Kaysville, Utah, EUA) para analizar las series de tiempo, o sea, para crear modelos, establecer la estadística descriptiva y hacer correlaciones entre las variables.

Para calcular el riesgo relativo de muerte (RR) de la población es necesario identificar el tipo de modelo que pueda proporcionar una mayor explicación a la variabilidad de la serie de tiempo de muertes.

Schwartz *et al*. [60] mencionan que es necesario tener un modelo no lineal cuando en el estudio se contemplen más de un año, con la finalidad de incrementar la probabilidad del modelo; a su vez al crear el modelo utilizando la Regresión de Poisson se logra reducir la distancia entre los valores observados por la serie de tiempo respuesta " Y " y el valor esperado por el modelo "E(Y)" o en otras palabras se mejora la heterocedasticidad que presenta el modelo comparado con otros; además de establecer como unidad de observación al día [60, 70].

La variable muertes presenta una distribución parecida a la de Poisson pero con un efecto denominado sobre dispersión, o sea, la varianza es mayor a la media. Este fenómeno de sobre dispersión es atribuido principalmente a los factores externos (como temperatura y la estacionalidad) más que un factor derivado del número de defunciones [71].

La probabilidad de los valores observados en la variable respuesta (muertes) es definida como:

$$\text{prob}\left(\frac{y}{\lambda}\right) = \frac{e^{-\lambda}\lambda^{y}}{y!} \qquad \textbf{Ecuación 4.1}.$$

Donde la variable λ también conocida como $E(Y)$ expresa el número estimado de muertes diarias calculadas mediante el modelo. El empleo del método Poisson asume que el riesgo calculado es homogéneo para toda la población que interviene en el estudio, pero este no es estacionario en el tiempo, además se sabe que ocurre una alteración dependiendo de las variables predictivas contempladas [60].

El procedimiento empleado indica que inicialmente se debe establecer un modelo considerando exclusivamente las posibles variables de confusión, las cuales pueden ser: estacionalidad, tendencia, variables meteorológicas y de calendario; el modelo generado es denominado modelo basal. Posteriormente, al modelo basal se le agrega cada una de las series de contaminantes (variables explicativas), con el fin de identificar la contribución de cada contaminante sobre la variable respuesta muertes [71, 74].

4.3.1 Crear Modelo Básico o Basal.

La distribución de la variable respuesta se caracteriza por estar formada por valores enteros positivos (i.e. discretos) y sus eventos son independientes unos de otros, además de que la cantidad de defunciones en la población total suele ser pequeña comparada con la población total existente, la presencia de estos factores es básica para sospechar que la variable respuesta sigue una distribución de Poisson, entonces el modelo más apropiado para trabajar con estas variables es la regresión de Poisson, la cual es expresada en la ecuación 4.2 [60, 75].

$$\text{Log}(E(Y)) = \beta_0 + \beta_b X_b \qquad \textbf{Ecuación 4.2}.$$

Donde:

Y : Número de muertes observadas en un día.

$E(Y)$: Cantidad de muertes estimadas por día.

X_b: Son los datos de las series de tiempo consideradas como variables de control y/o explicativas en el modelo basal

β_b: Coeficientes de regresión asociadas a las variables de control y/o explicativas.

Una condición necesaria para emplear la regresión de Poisson es que la media sea igual a la varianza, pero desafortunadamente esta situación suele presentarse muy pocas veces en estos estudios, ya que en la mayoría del tiempo se encuentran valores de varianza mayor a la media, y ese fenómeno es conocido como sobre dispersión. Para enmendar el rechazo de esta condición se utiliza un factor de corrección (ϕ) sobre los errores estándar (ecuación 4.3) [75].

$$\phi = \frac{1}{g}\sum_{i=1}^{n}\frac{(y_i - \lambda_i)^2}{\lambda_i} \qquad \textbf{Ecuación 4.3}.$$

Donde:

ϕ: Factor de corrección para sobre dispersión.

g: Grados de libertad del modelo.

λ: Conocida como E(Y). Se refiere a la cantidad de muertes estimadas por día.

Siguiendo el método planteado por Schwartz *et al.* [60] y Saez *et al*. [70] las posibles variables confusoras "X_b" de la ecuación 4.2 que serán contempladas para realizar la construcción del modelo básico son:

- Estacionalidad.
- Tendencia.
- Meteorológicas.
- IRAs.
- Años analizados (i.e. 2003, 2007).
- Días de la semana.

- Días festivos.

Según el método mencionado arriba, las variables meteorológicas e IRAs, se tomarán en cuenta junto con sus respectivos retardos, los cuales son hasta de orden siete. En otras palabras, el máximo desfase o retardo se encuentra en el séptimo día después de sucedida una determinada concentración [60, 70].

La estacionalidad es calculada mediante el apoyo del software NCSS y para realizar el cálculo se utiliza la función llamada "decomposition forecasting". Esta función descompone la serie de tiempo y extrae varios valores, entre ellos se encuentra la estacionalidad y la tendencia que presenta determinada serie de tiempo. La función extrae los valores mediante el cálculo de una nueva serie, la cual es determinada mediante la división de la serie original entre su media; una vez obtenida la nueva serie, se determinan las medias móviles. El siguiente paso es dividir cada uno de los valores de la nueva serie entre las medias móviles con el propósito de determinar la estacionalidad, la cual estará aunada a un factor de ruido [76]. Para aislar la estacionalidad del ruido es necesario aplicar una ventana de suavizado, pero se recomienda que la ventana sea mayor a dos meses para evitar eliminar patrones de corto plazo dentro de las series [77]. Para el caso de esta investigación, cada dato representa un día de estudio, para lograr hacer un suavizado de dos meses por ejemplo, fue necesario contemplar que dos meses representan 60 días y que esto a su vez simboliza aproximadamente el tres por ciento de la serie total (calculado mediante una regla de tres simple), por lo cual en el programa NCSS se deberá seleccionar la ventana de suavizado "Loess" mayor al porcentaje ya mostrado.

Dentro de las variables utilizadas para el modelo básico se encuentran las meteorológicas (i.e., temperatura y humedad relativa), su consideración dentro del modelo es importante, debido a que presenta una correlación significativa con la variable respuesta y no es de sorprenderse ya que esta asociación es común en varios de los estudios de contaminación. Por otro lado se tiene el conocimiento de que estas variables expresan las características propias del hábitat de la región; además de que las variables meteorológicas

contienen información sobre los efectos a corto plazo de la mortalidad, según Schwartz esta aseveración es válida siempre que exista asociación entre ellas [60].

En la lectura se observa que al introducir en el modelo, las variables de los días de la semana, como las fechas conmemorativas, pueden producir un impacto en los coeficientes de correlación, por lo cual se debe de identificar si el impacto es de consideración o no para el modelo [60, 70, 74, 77].

Varios estudios recomiendan insertar al modelo basal la variable de influenza por considerarla una variable de interés [70, 71], pero desafortunadamente en la ciudad de Mexicali no se cuenta con los conteos de dicha variable, por lo que se utilizó la variable IRAs [62].

Retardos en las variables o como se les conoce en inglés "lags", deben de ser incluidos en los modelos y la principal razón es debida a que cierta exposición puede traer repercusiones en el mismo día, pero también podría ser que la secuela se presente dentro de dos, tres o más días [60].

4.3.2 Método de selección de variables.

Una vez identificadas las variables que se utilizarán para construir el modelo, se aplica el método "Forward Selection and Switching" con el fin de solo incluir al modelo las variables que cuenten con la significancia estadística del 95% [75]:

1. Se comienza por no tener variables en el modelo.
2. En la primera corrida identifica cual es la variable que proporciona una mayor explicación del modelo, para esta decisión es tomada en base al "pseudo R-squared" que proporciona el ajuste de cada una de las variables contra la variable respuesta muertes.
3. El siguiente paso es ir agregando una a una las variables que se van a contemplar para el modelo. Estas variables deben de seguir la siguiente

estructura, a partir de la segunda corrida, cada vez que se requiera incrementar una variable en el modelo, es necesario encontrar la variable que proporcione el valor más elevado de "pseudo R-squared"; una vez encontrada, se conmutan una por una las variables que ya habían sido seleccionadas para estar dentro del modelo, si se encuentra alguna variable que mejore el "pseudo R-squared" se procede a hacer el cambio, y nuevamente se hace la comparación. Cuando se encuentra que ningún candidato logra mejorar este parámetro, se procede a incluir la siguiente variable y se reanuda el proceso ya mencionado.

El pseudo R-squared, es calculado mediante la ecuación 4.4, en ella se compara la verosimilitud del modelo [49, 78]. Esta fórmula es una comparación de la verosimilitud que presenta el modelo incluyendo a los valores predictivos del mismo, con la verosimilitud máxima para ese modelo.

$$R^2 = \frac{LL_{fit} - LL_0}{LL_{max} - LL_0} \qquad \textbf{Ecuación 4.4}.$$

Donde:

LL_0: Es la verosimilitud del intercepto del modelo β_0.

LL_{fit}: Es la verosimilitud del modelo actual.

LL_{max}: Es la verosimilitud que puede tener el modelo actual.

4.3.3 Elaborar Modelo con variables explicativas.

Una vez definido el modelo basal, el siguiente paso es ajustar este modelo a cada uno de los contaminantes con sus respectivos retardos, para eso se realiza un ajuste aplicando el método "Forward Selection and Switching" visto en la Sección 4.3.2.

Como ya se ha mencionado, los principales contaminantes que se encuentran presentes en la ciudad de Mexicali, PM_{10}, CO, $PM_{2.5}$, NO_2, O_3 y SO_2, fueron seleccionados

para hacer el estudio contemplando cinco años. Estos mismos contaminantes fueron utilizados para identificar su comportamiento durante la época de inverno y verano, así como detectar cuál o cuáles contaminantes producen riesgo a la población expuesta y en qué periodos.

En el capítulo 3 fue visto que los contaminantes PM_{10} y $PM_{2.5}$ cuentan con 13 diferentes técnicas de imputación de datos, las cuales generan la misma cantidad de series de tiempo, y todas ellas serán ajustadas, individualmente, con el modelo básico. Por otro lado, los contaminantes restantes (CO, NO_2, O_3, SO_2) poseen cinco técnicas diferentes para imputar los datos faltantes.

Es importante que cada una de las series imputadas sea introducida al modelo basal y se determine el RR de muertes por periodos, con el fin de identificar cuáles son los métodos que realizaron la mejor imputación.

Las variables categóricas, específicamente las variables de temporada, son utilizadas para identificar la variación que presentan los contaminantes en determinado semestre. Esto se logra siguiendo los pasos de los términos de interacción, también conocidos como análisis estratificado.

Para determinar el efecto de algún contaminante en la temporada invernal se incluye en el modelo la interacción de la variable cálida con el contaminante (esto vendría siendo la variable del semestre de calor multiplicada por el respectivo contaminante), sin olvidar añadir al modelo las variables independientes calor y el contaminante en cuestión; de manera que la contribución del contaminante en el semestre frio se presenta cuando la variable cálida es igual a cero y es medido en el coeficiente asociado al termino independiente del contaminante [71].

De manera contraria, se realiza el cálculo del efecto del contaminante en el semestre cálido, donde se deberá incluir la multiplicación de las variables frío y el respectivo

contaminante en el modelo, de igual manera no se debe olvidar incluir las variables sin interacción del frío y el contaminante en cuestión. El efecto de la temporada de calor es calculado de manera semejante al de la temporada de frío. Para obtener la contribución del contaminante se observa el coeficiente asociado a la variable contaminación [71].

Para el cálculo del riesgo relativo, RR, a partir de un coeficiente obtenido de la regresión de Poisson ($\hat{\beta}_c$, se necesita seguir la ecuación 4.5 en la cual al coeficiente de regresión de la variable explicativa, se le asocia una unidad de incremento del contaminante (, que en la investigación se identificó como rango inter cuartílico (RIC) [79].

$$RR = e^{(\hat{\beta}_c)(\Delta)} \qquad \textbf{Ecuación 4.5}$$

El RR de muerte con valor menor a uno se considera como "factor sin riesgo", Pero de superar este valor el RR de muerte se considera de riesgo para la población afectada [80].

4.3.4 Pasos para la toma de decisión de los Métodos de imputación a utilizar.

Para determinar qué método explica mejor el comportamiento de cada contaminante se analizan principalmente tres procedimientos:

- Seleccionar los valores que presenten una alta correlación y que sean estadísticamente significativas, para lograr esto es necesario realizar correlaciones donde se involucre la serie de tiempo imputada y la serie original con faltantes tal como se marca en el procedimiento de la Sección 3.6.2.
- Elaborar una tabla por contaminante que muestre los principales riesgos relativos por periodo, los cuales fueron destacados por su significancia estadística y calculados con las distintas series de tiempo imputadas mediante las diferentes técnicas.
- Identificar la temporada del contaminante que contiene las mayores concentraciones. Para hacer este reconocimiento es necesario tomar en cuenta el

valor de la mediana de cada uno de los contaminantes, lo cual podrá ser visualizado en diagramas de cajas y ejes; además cada diagrama debe contemplar la serie completa del contaminante y las temporadas que intervienen en el estudio. Esto con la finalidad de facilitar una toma de decisión al respecto de cual método emplear.

Para tomar la decisión, primero se identifica en la tabla del RR de muerte de un mismo contaminante, cuales son los métodos que se encuentran en las primeras posiciones, después se elige el método que coincide en el primer lugar para todos los periodos.

Por otro lado, se tiene el caso de que un mismo contaminante presente un mejor ajuste con determinada técnica para una temporada y este método no sea el adecuado en otra temporada, entonces se tendrá que ver cuál temporada maneja las más altas concentraciones (con el apoyo del diagrama de caja y ejes del contaminante) y optar por el método que exprese una mayor significancia (valor p, de la tabla que muestra los valores RR) en esa temporada, esto es, que al consultar el diagrama del contaminante se identifique qué temporada es la que maneja los valores más elevados en concentraciones del contaminante e identificar la técnica sobresaliente de esa temporada. De esta manera se selecciona el método.

Si se tiene que una misma técnica sea la mejor para dos periodos y a su vez uno de esos periodos sea el de mayor concentración, de darse este caso aumenta la seguridad de haber escogido la técnica adecuada.

Es importante mencionar la razón por la cual se incluyó la temporada con una mayor concentración del contaminante (mediana mayor), en la toma de decisión de la técnica, pues el método que maneje una mayor significancia estadística dentro de esa temporada, es el método que se puede definir como el que imputa mejor la serie. Entonces, la razón mencionada es que al incrementarse la concentración del contaminante, es de esperarse que se eleve el número de casos afectados en salud. Por lo consiguiente si el contaminante es el responsable de provocar cierta cantidad de muertes, al incrementarse su concentración también debe de aumentarse su efecto.

4.4 Resultados.

Antes de comenzar con la construcción de los modelos, se realizaron análisis de los datos con la finalidad de identificar la tendencia central (media y mediana), valores mínimos, máximos, desviación estándar de los datos, así como la distribución de los percentiles 10, 25, 75, 90 y su rango interventílico (Apéndice B). A la par, se construyeron diagramas de caja y ejes (Apéndice C), los cuales muestran el percentil 25 mediante la línea inferior de la caja, el percentil 50, también llamado mediana, indicado por la línea intermedia de la caja y el percentil 75, el cual es representado por la línea superior de la caja. Las figuras en forma de T que se encuentran tanto en la parte superior, como inferior de la caja son denominadas ejes, los cuales son calculados contemplando el percentil 75 más 1.5 veces el RIC, para el caso del eje superior y en el caso del eje inferior se considera el percentil 25 menos 1.5 veces el RIC. De la misma manera se encuentran los valores extremos, los cuales se caracterizan por salirse del rango de valores delimitado por los ejes. Si el valor extremo es severo, éste será identificado con un color rojo. Los valores extremos severos son aquéllos que sobrepasan el valor percentil indicado 75 y 25, en más de 3 veces el RIC o en menos 3 veces, respectivamente.

Como se explica en la Sección 4.2.1.5, es importante que las variables que serán utilizadas para integrar el modelo básico, no poseen colinealidad entre ellas, pues eso afecta el desempeño del modelo. Entonces las correlaciones y sus valores son mostrados en el Apéndice D, señalando que las variables que interactúan en la construcción del modelo basal, presentan valores menores a 0.61 con significancia estadística dentro de la correlación. Debido a que este valor no superó el valor establecido, no hay existencia de una dependencia lineal cercana y como consecuencia se suprime la presencia de colinealiadad entre las variables [81, 82].

4.4.1 Modelo Basal

Inicialmente se debe calcular la estacionalidad, para ello se utiliza el procedimiento descrito en la sección 4.3.1, en él se indica la descomposición de la variable respuesta y se

extrae la estacionalidad mezclada con ruido. En este trabajo se utiliza el filtro de cuatro meses (7%) para aislar la estacionalidad.

Un buen indicador de cuales variables son las que deban de intervenir en el modelo basal es: que las series de tiempo al presentar su correlación con la variable respuesta muestren tener asociación y que la misma muestre significancia estadística, la cual se ilustra en la figura 4.2 con el valor de la letra p. Al mismo tiempo se observa, que las series de tiempo meteorológicas e IRAs efectivamente presentan una correlación significativa, lo que implica que estas variables contribuyen a la explicación de la variabilidad de las muertes, por eso deben ser agregadas en el modelos basal.

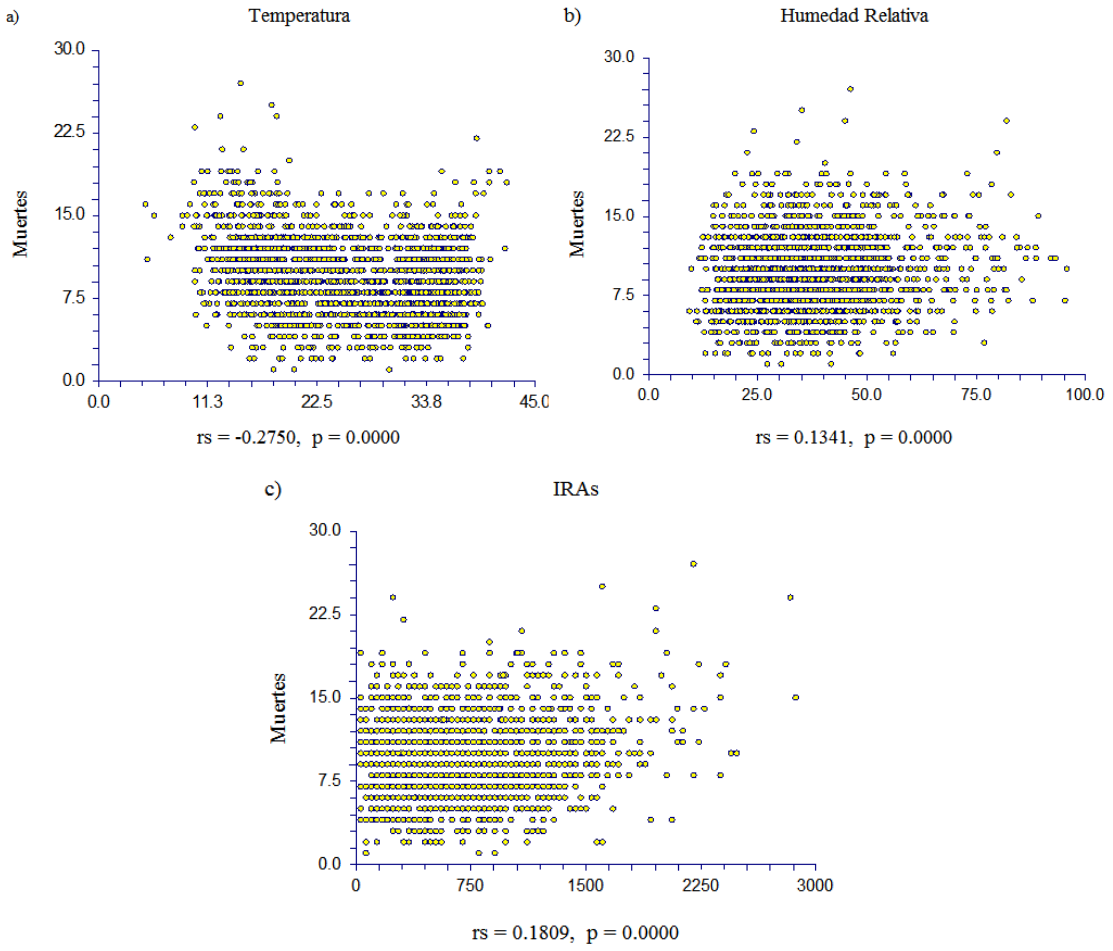


Figura 4.2. Correlaciones entre la variable respuesta muertes y variables independientes del modelo básico, a) serie de tiempo de temperatura, b) humedad relativa y c) IRAs.

Para construir el modelo basal es necesario tener identificadas las variables, posteriormente se utiliza el método "Forward Selection and Switching", el cual marca que las variables deben ser incluidas de una por una. Además no olvidar que las variables meteorológicas e IRAs deben de incluirse con sus respectivos lags.

Finalmente se estableció un modelo basal considerando cada una de las variables que al ser insertadas en el modelo mostraron nivel de significancia estadística $p \leq 0.05$ dentro del modelo. A partir de eso, para construir el modelo se escogieron las siguientes variables:

- Meteorológicas
    - Temperatura junto con los retardos tres y cinco.
    - Temperatura al cuadrado con su retardo cuatro.
    - Humedad relativa con su retardo de dos días.
    - El retardo dos de la humedad relativa al cuadrado.
- Clínicas
    - El retardo tres de las infecciones respiratorias agudas.
- Estacionalidad
    - La ventana de cuatro meses fue la que proporcionó una depuración con mayor claridad de esta variable.

Con la finalidad de validar el modelo basal se analizaron los residuales del mismo. Siguiendo la teoría de que los residuales de un modelo deben de comportarse de manera normal, se corroboró si el modelo es adecuado para explicar el comportamiento de determinados datos. Entonces se aplicaron cinco pruebas de normalidad, a partir de las cuales se concluyó que no se puede rechazar la normalidad (ver tabla 4.2).

Tabla 4.2. Pruebas de normalidad de los residuales del modelo basal

Prueba de Normalidad - Modelo Basal					
Nombre de la Prueba	Valor Prueba	Nivel Prob.	Valor Crítico 10%	Valor Crítico 5%	Decisión -5%
Shapiro-Wilk W	1.00	0.55			No se pudo rechazar normalidad
Anderson-Darling	0.41	0.34			No se pudo rechazar normalidad
Martinez-Iglewicz	0.99		1.00	1.00	No se pudo rechazar normalidad
Kolmogorov-Smirnov	0.01		0.02	0.02	No se pudo rechazar normalidad
D'Agostino Skewness	0.31	0.76	1.65	1.96	No se pudo rechazar normalidad
D'Agostino Kurtosis	0.79	0.43	1.65	1.96	No se pudo rechazar normalidad
D'Agostino Omnibus	0.72	0.70	4.61	5.99	No se pudo rechazar normalidad

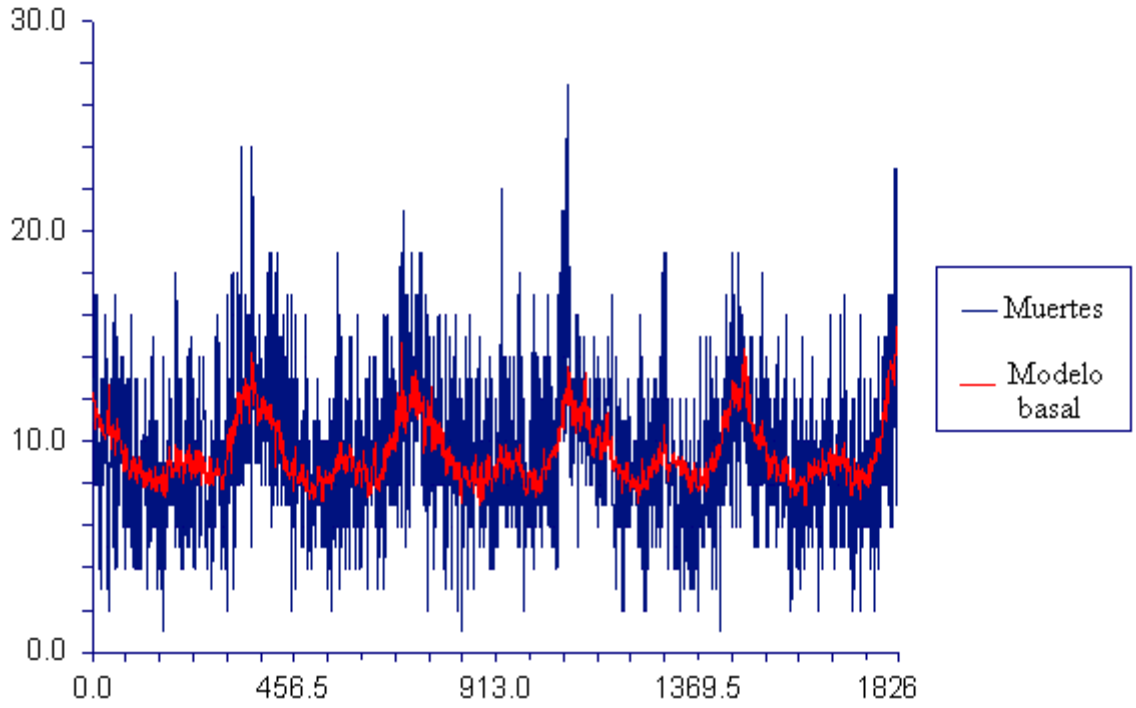


Figura 4.3. Comportamiento diario mostrado por la variable respuesta muertes y valores pronosticados por el modelo basal.

En la figura 4.3 se observa en la línea azul el comportamiento diario mostrado por la variable muertes, dentro de la misma figura resaltan con el color rojo los valores predichos por el modelo basal ajustado en esta investigación.

La variable determinística días de la semana y la tendencia no cumplen con la significancia estadística establecida, por esta razón fueron descartadas del modelo basal.

4.4.1 Modelos con ajuste por contaminante.

La manera, en cómo se empleó el ajuste del contaminante al modelo basal, es la señalada en la Sección 4.3.2. Siguiendo el procedimiento descrito en esta sección, se calcularon los RR para cada contaminante con cada una de las series de tiempo imputadas que mostraron tener alta correlación con la serie original (Apéndice A), como fue explicado en las Secciones 4.3.3 y 4.3.4.

El Apéndice A muestra claramente que todos los métodos en todos los contaminantes presentaron altas correlaciones con la serie original, por lo cual, se prosiguió con la construcción de las tablas de RR por contaminante.

El RR de muerte por efecto del contaminante PM_{10} es presentado en la tabla 4.3, donde es posible observar que de los cuatro métodos empleados, el método BAAYG tiene mayor significancia estadística tanto para el periodo de cinco años como para la temporada de frío. Aunado a esto al observar el diagrama del Apéndice C, que hace posible corroborar que esta temporada es la que presenta mayores concentraciones. Por lo tanto, el método BAAYG es el que mejor expresa el contaminante PM_{10}.

Tabla 4.3. RR de muerte por efecto del contaminante PM_{10}

PM_{10}														
Frío					Calor					Cinco años				
Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor
BAAYG	5	67.0	1.033	0.0355	EAYG	5	37.3	1.016	0.2165	BAAYG	5	50.2	1.023	0.0217
AAAYG	5	63.7	1.029	0.0612	UAYG	5	37.3	1.016	0.2165	UAYG	5	49.4	1.022	0.0291
EAYG	5	66.5	1.030	0.0739	AAAYG	5	38.0	1.013	0.3290	EAYG	5	49.4	1.022	0.0295
UAYG	5	66.5	1.030	0.0739	BAAYG	5	38.7	1.012	0.3514	AAAYG	5	48.6	1.021	0.0368

BAAYG = Automatic ARMA por Bloques con Factor de Ciclo
AAAYG = Automatic ARMA por Acumulado con Factor de Ciclo
EAYG = Promediado por Exclusión con Factor de Ciclo
UAYG = Promediado por Unidad con Factor de Ciclo
Lag = Retardo en la serie de tiempo
RIQ = Rango Inter Cuartílico
RR = Riesgo Relativo
p valor = Significacia estadística
PM_{10} cuenta con 22.07% de faltantes

Tabla 4.4. RR de muerte por efecto del CO

CO														
Frío					Calor					Cinco años				
Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor
BAA	5	1.71	1.059	0.0012	AMELIA II	1	0.40	1.015	0.0863	BAA	5	1.14	1.036	0.0018
AAA	5	1.71	1.059	0.0012	EA	5	0.38	1.009	0.4112	AAA	5	1.14	1.036	0.0018
EA	5	1.73	1.056	0.0026	AAA	5	0.38	1.004	0.7201	AMELIA II	1	1.10	1.032	0.0030
UA	5	1.73	1.056	0.0026	BAA	5	0.38	1.004	0.7202	EA, UA	5	1.15	1.034	0.0036

BAA = Automatic ARMA por Bloques
AAA = Automatic ARMA por Acumulado
EA = Promediado por Exclusión
UA = Promediado por Unidad

Lag = Retardo en la serie de tiempo
RIQ = Rango Inter Cuartílico
RR = Riesgo Relativo
p valor = Significacia estadística

CO cuenta con 0.55% de faltantes

La tabla 4.4 exhibe el RR de muerte por efecto del Monóxido de Carbono, por medio de ella se puede_ver una situación semejante a la del PM_{10}, en la cual una misma técnica sobresale en la temporada invernal y en el periodo de cinco años, aunado a esto se tiene que el diagrama del Apéndice C muestra que el CO maneja sus concentraciones más elevadas en la época de frío, por lo anterior, se opta por tomar los métodos BAA y AAA como los adecuados para el contaminante en cuestión.

Al prestar atención en las tablas 4.5 y 4.6 es notable que caen el mismo caso de los dos contaminantes que los antecedieron, el cual es, una misma técnica que sobresale en dos de tres periodos, contemplado dentro de estos dos periodos a la temporada de mayor concentración del contaminante. Por lo que es evidente que se tiene que llegar a un resultado semejante. El cual es que: la técnica apropiada para el contaminante $PM_{2.5}$ es la utilización de software Amelia II. Por su parte el método BAA es el adecuado para el contaminante NO_2.

Tabla 4.5. RR de muerte por efecto del $PM_{2.5}$

$PM_{2.5}$														
Frío					Calor					Cinco años				
Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor
AMELIA II	5	48.1	1.041	0.0045	EA	5	17.135	1.02463	0.0216	AMELIA II	5	33.5	1.028	0.0017
UAYG	5	42.845	1.01772	0.2537	AMELIA II	5	21.5	1.017	0.1403	UAYG	5	31.2075	1.01478	0.1276
EAYG	5	42.845	1.01772	0.2538	UAYG	5	21.7925	1.01228	0.3072	BAAYG	5	31.1175	1.01064	0.2373
BAAYG	5	43.795	1.01234	0.4027	EAYG	5	21.7925	1.01228	0.3125	UAGG	5	31.0675	1.00999	0.3144

UAYG = Promediado por Unidad con Factor de Ciclo
EAYG = Promediado por Exclusión con Factor de Ciclo
BAAYG = Automatic ARMA por Bloques con Factor de Ciclo

Lag = Retardo en la serie de tiempo
RIQ = Rango Inter Cuartílico
RR = Riesgo Relativo
p valor = Significacia estadística

$PM_{2.5}$ cuenta con 40.85% de faltantes

Tabla 4.6. RR de muerte por efecto del NO_2

NO_2														
Frio					Calor					Cinco años				
Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor
BAA	5	0.01305	1.03073	0.1127	AMELIA II	0	0.009	1.037	0.0158	BAA	5	0.0151	1.03756	0.0397
UA	5	0.01303	1.03041	0.1159	BAA	5	0.0092	1.01886	0.2672	UA	5	0.0151	1.03601	0.0485
AAA	5	0.01303	1.03037	0.1164	UA	5	0.0092	1.01664	0.3281	AAA	5	0.0151	1.03548	0.0514
EA	5	0.01303	1.03032	0.117	AAA	5	0.00928	1.0159	0.3509	EA	5	0.01503	1.03532	0.0516

BAA = Automatic ARMA por Bloques
UA = Promediado por Unidad
AAA = Automatic ARMA por Acumulado
EA = Promediado por Exclusión

Lag = Retardo en la serie de tiempo
RIQ = Rango Inter Cuartilico
RR = Riesgo Relativo
p valor = Significacia estadística

NO_2 cuenta con 1.53% de faltantes

El comportamiento de los métodos con el ozono difiere de los anteriores, pues en este caso, se tiene una clara aceptación del contaminante por una misma técnica (BAA) para todos los periodos, mostrando que entre todos los métodos este enseña los valores más significativos, aunque, desafortunadamente, no se pueda expresar que esta técnica cumpla con el nivel de significancia estadística establecido en este estudio.

El contaminante SO_2 fue imputado solamente con tres técnicas, pues presentó problemas de residuales significativos con los métodos AAA y BAA, entonces ellos fueron excluidos. Analizando la tabla 4.8 se puede ver claramente que este contaminante se adecua de muy buena manera al método EA, ya que esta técnica se destaca entre los tres periodos, además de mostrar significancia estadística en dos de ellos. Con esta evidencia se identifica al método como el adecuado para este contaminante.

Tabla 4.7. RR de muerte por efecto del O_3

O_3														
Frio					Calor					Cinco años				
Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor
BAA	7	0.011	0.9845	0.4848	BAA	7	0.0149	0.97642	0.1507	BAA	7	0.0161	0.97403	0.103
AAA	7	0.011	0.9862	0.5350	AAA	7	0.0147	0.97788	0.1896	AAA	7	0.01573	0.97605	0.1374
EA	7	0.011	0.9886	0.6084	EA	7	0.0144	0.97943	0.2161	EA	7	0.0155	0.97808	0.1707
UA	7	0.011	0.9888	0.6123	UA	7	0.0144	0.97944	0.2163	UA	7	0.0155	0.97813	0.1717

BAA = Automatic ARMA por Bloques
AAA = Automatic ARMA por Acumulado
EA = Promediado por Exclusión
UA = Promediado por Unidad

Lag = Retardo en la serie de tiempo
RIQ = Rango Inter Cuartilico
RR = Riesgo Relativo
p valor = Significacia estadística

O_3 cuenta con 2.19% de faltantes

Tabla 4.8. RR de muerte por efecto del SO_2

SO_2														
Frío					Calor					Cinco años				
Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor	Método	Lag	RIQ	RR	p valor
EA	1	0.0047	1.03843	0.0102	EA	1	0.0028	1.01275	0.3451	EA	1	0.0038	1.02688	0.0081
UA	1	0.0047	1.03838	0.0103	UA	1	0.0028	1.01259	0.352	UA	1	0.00373	1.02627	0.0083
AMELIA II	1	0.005	1.029	0.0509	AMELIA II	1	0.003	0.997	0.8253	AMELIA II	1	0.004	1.023	0.0277

EA = Promediado por Exclusión
UA = Promediado por Unidad

Lag = Retardo en la serie de tiempo
RIQ = Rango Inter Cuartílico
RR = Riesgo Relativo
p valor = Significacia estadística

SO_2 cuenta con 18.57% de faltantes

Resumiendo, los valores de los coeficientes de la regresión de Poisson son expresados en la tabla 4.9. Con él se determinaron los RR para cada una de las variables explicativas como se muestra en las tablas 4.10.

Tabla 4.9. Coeficientes de regresión mostrados por periodos
(frío, calor y cinco años) con los métodos que mostraron tener la mejor imputación

		Frío					Calor					
Variables	Lag	Método de Imputación	B	SE	Wald's χ^2	(p-valor)	95% IC	Método de Imputación	β	SE	Wald's χ^2	(p-valor)	95% IC
PM10	5	BAAYG	0.0005	0.0002	4.42	0.0355	0.0000 - 0.0010	BAAYG	0.00032	0.00035	0.87	0.3514	-0.0004 - 0.001
CO	5	BAA	0.0336	0.0104	10.45	0.0012	0.0132 - 0.0540	BAA	0.00936	0.02613	0.13	0.7202	-0.0419 - 0.06058
PM2.5	5	AMELIA II	0.0008	0.0003	8.08	0.0045	0.0003 - 0.0014	AMELIA II	0.0008	0.0005	2.2	0.1403	-0.0003 - 0.0018
NO2	5	BAA	2.32	1.4622	2.52	0.1127	-0.55 - 5.19	BAA	2.03	1.83	1.23	0.2672	-1.56 - 5.62
O3	7	BAA	-1.39	1.9968	0.49	0.4848	-5.31 - 2.52	BAA	-1.60	1.11	2.06	0.1507	-3.79 - 0.58
SO2	1	EA	8.0242	3.1242	6.60	0.0102	1.90 - 14.15	EA	4.52	4.79	0.89	0.3451	-4.87 - 13.92

B = Coeficiente asociado al contaminante
SE = Error Estandar

IC = Intervalo de Confianza
p-valor = probabilidad asociada al RR

		Cinco años					
Variables	Lag	Método de Imputación	β	SE	Wald's χ^2	(p-valor)	95% IC
PM10	5	BAAYG	0.0005	0.0002	5.27	0.0217	0.0001 - 0.0008
CO	5	BAA	0.0309	0.0099	9.74	0.0018	0.0115 - 0.0503
PM2.5	5	AMELIA II	0.0008	0.0003	9.9	0.0017	0.0003 - 0.0013
NO2	5	BAA	2.44	1.19	4.23	0.0397	0.12 - 4.77
O3	7	BAA	-1.63	1.00	2.66	0.1030	-3.60 - 0.33
SO2	1	EA	6.98	2.63	7.02	0.0081	1.82 - 12.14

B = Coeficiente asociado al contaminante
SE = Error Estandar

IC = Intervalo de Confianza
p-valor = probabilidad asociada al RR

Tabla 4.10 . Valores del RR de muerte mostrados para los tres periodos del estudio, con métodos de imputación sobresalientes

		Frío					Calor				
		Método de	Incremento				Método de	Incremento			
Variables	Lag	Imputación	RIC	RR	(p-valor)	95% IC	Imputación	RIC	RR	(p-valor)	95% IC
PM10	5	BAAYG	67.0	1.033	0.0355	1.0026 - 1.0651	BAAYG	38.7	1.012	0.3514	0.9859 - 1.0397
CO	5	BAA	1.7	1.059	0.0012	1.0229 - 1.0970	BAA	0.4	1.004	0.7202	0.9843 - 1.0231
PM2.5	5	AMELIA II	48.1	1.041	0.0045	1.0127 - 1.0696	AMELIA II	21.5	1.017	0.1403	0.9945 - 1.0398
NO2	5	BAA	0.013	1.031	0.1127	0.9929 - 1.0700	BAA	0.009	1.019	0.2672	0.9858 - 1.0531
O3	7	BAA	0.011	0.984	0.4848	0.9423 - 1.0286	BAA	0.015	0.976	0.1507	0.9452 - 1.0087
SO2	1	EA	0.005	1.038	0.0102	1.0090 - 1.0688	EA	0.003	1.013	0.3451	0.9865 - 1.0397

RIC = Rango Inter Cuartilico
RR = Riesgo Relativo
IC = Intervalo de Confianza
p-valor = probabilidad asociada al RR

		Cinco años				
		Método de	Incremento			
Variables	Lag	Imputación	RIC	RR	(p-valor)	95% IC
PM10	5	BAAYG	50.2	1.023	0.0217	1.0029 - 1.0432
CO	5	BAA	1.1	1.036	0.0018	1.0131 - 1.0588
PM2.5	5	AMELIA II	33.5	1.028	0.0017	1.0101 - 1.0452
NO2	5	BAA	0.015	1.038	0.0397	1.0017 - 1.0747
O3	7	BAA	0.016	0.974	0.1030	0.9437 - 1.0053
SO2	1	EA	0.004	1.027	0.0081	1.0069 - 1.0472

RIC = Rango Inter Cuartilico
RR = Riesgo Relativo
IC = Intervalo de Confianza
p-valor = probabilidad asociada al RR

En la figura 4.3 se muestra el riesgo relativo de los principales contaminantes de la ciudad de Mexicali con sus respectivos intervalos de confianza, donde ICS muestra el intervalo de confianza superior e ICI el intervalo de confianza inferior.

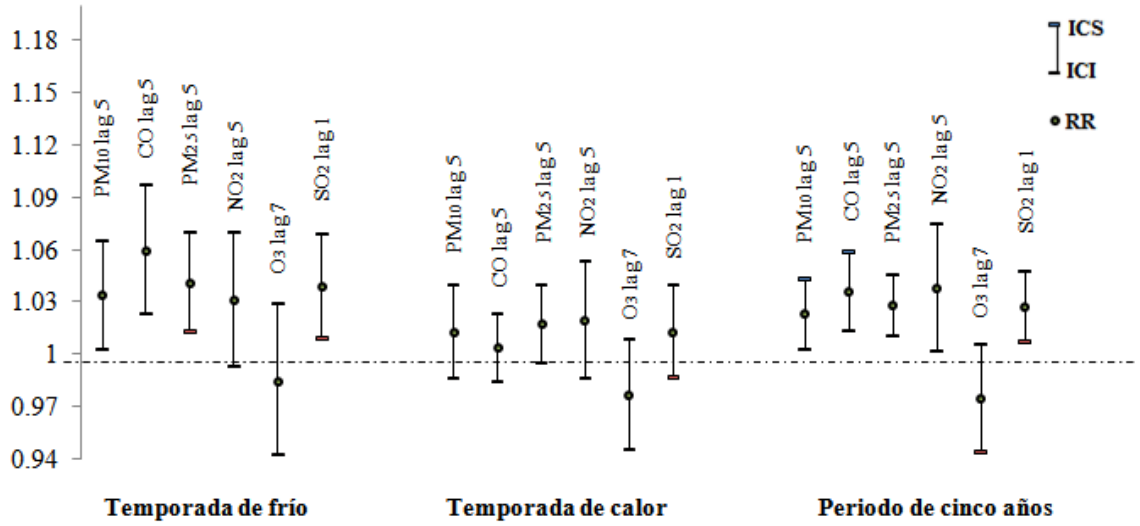


Figura 4.4. Riesgo relativo por temporadas de la ciudad de Mexicali BC

4.5 Discusiones

Tomando como base la tabla de estadística descriptiva del Apéndice B, es notorio que la variable repuesta muertes, presenta una tendencia central mayor en la época de invierno, de manera semejante a la mayoría de las variables explicativas con excepción del contaminante O_3.

La importancia de la no existencia de colinealidad entre las variables independientes consideradas para crear el modelo de este estudio y la relación lineal de las mismas con la variable respuesta fue mostrada en la publicación [83].

La información sobre el desarrollo de los métodos de imputación creados en este estudio fue reportada en "Assessment of procedures for air pollution time-series reconstruction" [84], por otro lado se tiene que la identificación del metodo que proporciona el mejor ajuste entre la variable imputada con la variable original del contaminante fue publicado en "Evaluación de los Procedimientos para la Reconstrucción de la serie Tiempo de la Contaminación del aire" [85].

Por otra parte se encontró que el método que ajusta mejor a una mayor cantidad de series de tiempo fue el BAA, el cual sobresalió en tres de las series de tiempo (CO, NO_2, O_3) y una variación de este método (BAAYG) fue el que realizó un mejor ajuste del contaminante PM_{10}.

De la misma forma se encontró que la serie de tiempo que presenta un número mayor de faltantes ($PM_{2.5}$, 40.85%) obtuvo su mejor resultado con el método AMELIA II. Otro dato sobresaliente fue que las series que mostraron faltantes alrededor del 20% o menos, manejaron un mejor ajuste con el método BAA y sus variantes, a excepción del contaminante SO_2, el cual no pudo ser imputado con este método ya que mostró residuales significativos en la elaboración del modelo, pero dicho contaminante tuvo un muy buen comportamiento con el método EA.

Durante los cinco años del estudio se observaron las concentraciones diarias de los contaminantes y se compararon con las normas mexicanas, las cuales se encuentran resumidas en las secciones 2.2.3 y 2.3.3 del Capítulo 2. Con ese análisis se percibió que el contaminante PM_{10} sobrepasó la norma (NOM-025-SSA1-1993 en su modificación indica que las concentraciones en promedio de 24 horas deben ser menores a 120 $\mu g \cdot m^{-3}$) 295 veces; a su vez, esta norma indica que las concentraciones promedio permisibles por año las cuales son de 50 $\mu g \cdot m^{-3}$, lo cual también fue sobrepasado en todos los años. La misma norma anterior muestra que la concentración para el $PM_{2.5}$ debe de ser menor a 65 $\mu g \cdot m^{-3}$ en promedios de 24horas, esta indicación también fue violada 311 veces y para el promedio anual también excedieron todos los años el valor de 15 $\mu g \cdot m^{-3}$. La NOM-021-SSA1-1993 es la norma para el contaminante CO y ésta indica que no debe de sobrepasar el valor de 11.00 ppm en promedios móviles diarios de ocho horas, esta normatividad fue sobrepasada 84 veces. La norma del NO_2 es NOM-023-SSA1-1993, la cual indica que este contaminante no puede exceder el valor de 0.21 ppm en promedios de una hora, habiéndose cumplido con la norma. El contaminante O_3 es regulado por la NOM-020-SSA1-1993, en la que se indica que este contaminante no debe de manejar concentraciones mayores a 0.110 ppm en promedios de una hora por un periodo de tres años, el estudio abarca casi dos periodos ya que son cinco años dentro de los cuales observaron 17 violaciones. Los valores de los promedios de 24 horas del contaminante SO_2 respetan el promedio de 0.13 ppm establecido por la NOM-022-SSA1-1993 y en el caso del promedio por año la norma establece un valor de 0.03 ppm, lo cual también fue respetado en todos los años.

Los periodos que presentaron efectos significativos en el RR de muerte son la temporada de frío y el análisis para los cinco años, en los cuales los coeficientes asociados a los contaminantes de los modelos mostraron una significancia estadística del 95% para la mayoría de los contaminantes (CO, $PM_{2.5}$, SO_2, PM_{10}) y en el caso del contaminante NO_2 se manejó la misma significancia estadística para los cinco años. Por otro lado, el único contaminante que no tuvo significancia estadística fue el O_3.

El tercer almanaque del Instituto Nacional de Ecología muestra que los contaminantes predominantes en la ciudad de Mexicali son el PM_{10} y CO, los cuales presentaron significancia estadística (p=0.005) en el riesgo relativo de muerte en el periodo de estudio de los cinco años, así como para la temporada invernal.

En el contaminante PM_{10} los RR observados fueron de 1.033 con intervalo de confianza (IC) entre 1.0026 y 1.0651 en el invierno, por otro lado en el estudio de cinco años se obtuvo un RR de 1.023 con un intervalo de confianza entre 1.0029 y 1.0432. Es importante enfatizar que los dos periodos mencionados fueron calculados con un retardo de cinco días, al igual que el CO, el cual mostró su mejor desempeño en el mismo retardo (cinco días), además de identificar un RR de 1.059 con IC entre 1.0229 y 1.0970 en la temporada de frío, ya para el periodo de cinco años se consiguió un RR de 1.036 con un intervalo de confianza entre 1.0131 y 1.0588. Las temporadas de calor no mostraron tener RR significativo para la población para ningún contaminante.

Los contaminantes $PM_{2.5}$ y $S0_2$ mostraron la misma significancia estadística (p=0.05) en el RR de muerte que los contaminantes anteriores, tanto para la temporada de frío como para el periodo de cinco años, pero el $PM_{2.5}$ obtuvo un RR de 1.041 con un IC de 1.0127 a 1.0696 en la temporada invernal y un RR de 1.028 con IC entre 1.0101 y 1.0452 para los cinco años; además estos resultados se determinaron con un retardo de cinco días. El RR del contaminante SO_2 fue de 1.038 y su IC de 1.0090 a 1.0688 en el periodo de frío, no obstante, para el análisis de los cinco años se identificó un RR de 1.027 y un IC de 1.0069 a 1.0472. Para eso, el cálculo de los RR fue obtenido con un retardo de un día para este contaminante.

El contaminante NO_2 tiene significancia estadística para el período de cinco años, RR de 1.038 con un IC de 1.0017 a 1.0747, los valores fueron determinados con retardo de cinco días en el contaminante.

Como en este estudio se contempló el ajuste por separado de cada contaminante con el modelo basal, pero sin relacionar a las demás variables explicativas, tomando en cuenta lo anterior sería conveniente realizar un nuevo estudio que involucre a los principales contaminantes y de esta forma identificar las contribuciones de los mismos, y a su vez se podría calcular el riesgo relativo de muerte de esta comunidad con mayor precisión.

4.6 Conclusión.

Los contaminantes CO, $PM_{2.5}$, SO_2, PM_{10}, NO_2 presentaron asociación significativa además de convertirse en un factor de riesgo relativo de muerte en la población en los cinco años de estudio, esto fue observado en tanto en el RR de muerte como el de ICI ya que ambos supera el valor de uno, esto considerando el incremento del rango intercuartílico. En la comparación de estos factores con los valores limites de las normas de los contaminantes, se ha podido observar que los contaminantes que exceden las normas en México presentan riesgo relativo a la población (CO, $PM_{2.5}$, PM_{10}, NO_2), de la misma manera se demuestra que hay contaminantes que no sobrepasan la norma, como por ejemplo, el SO_2, pero también representa riesgo relativo a la población.

La temporada que exhibió un riesgo mayor fue la temporada de invierno, debido que los contaminantes (CO, $PM_{2.5}$, SO_2, PM_{10}) demostraron tener significancia estadística en el riesgo relativo de muerte, por lo cual esta temporada es la que manifiesta tener un mayor riesgo en la población Mexicalense.

Con la información ya recabada para el contaminante SO_2, se propone bajar el valor límite considerado en la norma NOM-022-SSA1-1993 para la región de Mexicali, ya que los valores de este contaminante prevalece por debajo la norma pero los mismos demostraron ser un factor de RR de muerte para la población mexicalense. De la misma manera se propone que los valores límites de las normas NOM-021-SSA1-1993, NOM-025-SSA1-1993, NOM-022-SSA1-1993 de los respectivos contaminantes CO, $PM_{2.5\ y\ 10}$, SO_2, disminuyan para la temporada invernal ya que en estudio se observo que esta es la temporada en la cual se incrementa el RR de muerte para los habitantes de Mexicali.

Refrencia Bibliográfica

[1] National Aeronautic and Space Administration (NASA), 2010, Earth's Atmosphere, Disponible en: http://www.nasa.gov/audience/forstudents/9-12/features/912_liftoff_atm.html. Consultado en: 4 June 2011.

[2] L.E.Prim, *Ciencias de la Tierra y del Medio Ambiente*. Editorial Teide, 1998.

[3] P.Warneck, *Chemistry of the natural atmosphere*. Academic Press, 1988.

[4] The Columbia Center for Children's Enviromnental Health (CCCEH), 2011, The Truth about Air Pollution. Don't just breathe it all in!, West Harlem Environmental Action. Disponible en: http://www.epa.gov/ncer/childrenscenters/outreach_docs/AN000337.pdf. Consultado en: 14 September 2011.

[5] U.S.Environmental Protection Agency (EPA), 2010, What Are the Six Common Air Pollutants?, Disponible en: http://www.epa.gov/air/urbanair/. Consultado en: 4 June 2011.

[6] U.S.Environmental Protection Agency (EPA), "National Ambient Air Quality Standards for Particulate Matter", *Federal Register,* vol. 71, pp. 61144-61233, October 2006.

[7] U.S.Environmental Protection Agency (EPA) Office of Mobile Sources, "Automobiles and Carbon Monoxide," *Ann Arbor*,EPA 400-F-92-005, Jan.1993.

[8] U.S.Consumer Product Safety Commission, "Carbon Monoxide Questions and Answers," Bethesda,CPSC Document #466, September, 2008.

[9] NIOSH, CDPEHE, CPSC, OSHA and EPA, "Preventing Carbon Monoxide Poisoning from Small Gasoline-Powered Engines and Tools", *DHHS (NIOSH) Publication,* December, 1996.

[10] U.S.Environmental Protection Agency (EPA), 2011, Health and Environment Effects of Ground-Level Ozone, Disponible en: http://www.epa.gov/region7/air/quality/o3health.htm. Consultado en: 13 September 2011.

[11] M.Zuk, M.G.T.Cervantes and L.R.Brach, *Tercer almanaque de datos y tendencias de la calidad del aire en nueve ciudades mexicanas*., Ciudad de México: Instituto Nacional de Ecología (INE-Semarnat), 2007.

[12] F.B.Díez, M.S.Zafra, M.E.A.Fustel, M.T.Trunk, J.M.O.Iriarte and I.A.Ontoso, "El proyecto EMECAM: Estudio Multicéntrico Español sobre la relación entre la Contaminación Atmosférica y la Mortalidad. Antecedentes, Participantes,

Objetivos y Metodología," *Revista Española de Salud Pública,* vol. 73, no. 2, pp. 165-175, April, 1999.

[13] F.B.Díez, M.S.Zafra, S.Pérez-Hoyos, A.D.Codina, J.B.B.Blasco and Á.C.Martínez, "El proyecto EMECAM: Discusión de los resultados en las ciudades participantes," *Revista Española de Salud Pública*, vol. 73, no. 2, pp. 303-314, April, 1999.

[14] National Aeronautics and Space Administration (NASA), 2011, 45 Environment Pollution, Disponible en: http://www.sti.nasa.gov/sscg/45.html. Consultado en: 4 June 2011.

[15] Aspen Global Change Institute, 2011, The Atmosphere, Disponible en: http://www.agci.org/classroom/atmosphere/index.php. Consultado en: 4 June 2011.

[16] World Health Organization (WHO), 2008, Air quality and health, Disponible en: www.who.int/mediacentre/factsheets/fs313/en/index.html. Consultado en: 4 June 2011.

[17] Instituto Nacional de Ecología, 2009, Contaminantes primarios y secundarios, Disponible en: http://www.ine.gob.mx/calaire-informacion-basica/525-calaire-cont-primarios-secundarios. Consultado en: 4 June 2011.

[18] R.Nixon, 2011, Reorganization Plan No. 3 of 1970, U.S. Environmental Protection Agency (EPA). Disponible en: http://www.epa.gov/aboutepa/history/org/origins/reorg.html. Consultado en: 13 September 2011.

[19] A.Peters and D.W.Dockery, "Air pollution and health effects: Evidence from epidemiologic studies," in *Lung Biology in Health and Diseases* Oxford: Taylor and Francis Group, 2005.

[20] Instituto Nacional de Ecología, 2009, Contaminantes criterio, Instituto Nacional de Ecología. Disponible en: http://www.ine.gob.mx/calaire-informacion-basica/523-calaire-cont-criterio%3E%20[consulta:%2027%20de%20Septiembre%20del%202009. Consultado en: 20 August 2011.

[21] WHO Regional Publications, *Air Quality Guidelines for Europe,* 2 ed., Copenhagen: World Health Organization Regional Office for Europe Copenhagen, 1987.

[22] WHO Regional Publications, *Air Quality Guidelines for Europe,* 2 ed., Copenhagen: World Health Organization Regional Office for Europe Copenhagen, 2000.

[23] U.S.Enviromental Protection Agency (EPA), 2011, National Ambient Air Quality Standards (NAAQS), Disponible en: http://www.epa.gov/air/criteria.html. Consultado en: 12 September 2011.

[24] Megan, 2009, Carta de Plinio sobre la erupción del Vesubio, phpBB Group. Disponible en: http://www.abretelibro.com/foro/viewtopic.php?f=6&t=32233&start=0&st=0&sk=t&sd=a&sid=66ab9ec57b7ee3143893d1dc2f4c2bbf#p921418. Consultado en: 10 September 2011.

[25] Plinio el Joven, 2008, La erupción del Vesubio que destruýo a Pompeya, Elmestizo. Disponible en: http://elmestizo.wetpaint.com/page/La+erupci%C3%B3n+del+Vesubio+que+destru%C3%BDo+a+Pompeya. Consultado en: 11 September 2011.

[26] M.L.Bell and D.L.Davis, "Reassessment of the Lethal London Fog of 1952: Novel Indicators of Acuteand Chronic Consequences of Acute Exposure to Air Pollution", *Environmental Health Perspectives,* vol. Vol. 109, 2001.

[27] Ministry of public Health, "Mortality and Morbidity during the Fog of December 1952", *Public Health and medical subjects*, Her Majesty's Stationary Office, September 1954.

[28] Coordinación General de Comunicación Social, 2011, ¿Qué es la SEMARNAT?, SEMARNAT. Disponible en: http://www.semarnat.gob.mx/conocenos/Paginas/quienessomos.aspx. Consultado en: 13 September 2011.

[29] M.A.Danzon, "Criteria used in establishing guideline values," in *Air Quality Guidelines,* 2nd ed Copenhagen: World Health Organization , 2000, pp. 11-31.

[30] G.Corey, 2011, Evaluación del riesgo asociado a contaminantes del aire, Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente de la Organización Panamericana de la Salud. Disponible en: http://www.bvsde.ops-oms.org/bvsci/e/fulltext/riesgo/index.html. Consultado en: 24 September 2011.

[31] U.S.Environmental Protection Agency (EPA), 2011, What is PM?, Disponible en: http://www.epa.gov/region1/airquality/pm-what-is.html. Consultado en: 13 September 2011.

[32] C.Rice, J.Levine, M.Nichols, T.Agardy, I.A.Quiroz, J.Barton, R.P.Gil, E.Phipps, J.Reaser, E.Shoch and K.Thornton, *El mosaico de América del Norte panorama de los problemas ambientales más relevantes*., Montreal: Comisión para la Cooperación Ambiental, 2008.

[33] E.E.Rubio, 2005, MODIFICACION a la Norma Oficial Mexicana NOM-025-SSA1-1993, Salud ambiental. Criterios para evaluar la calidad del aire ambiente, con respecto a material particulado. Valor de concentración máxima de material particulado para partículas suspendidas totales PST, partículas menores de 10 micrómetros PM_{10} y partículas menores de 2.5 micrómetros $PM_{2.5}$ en el aire ambiente como medida de protección a la salud de la población, para quedar como Norma Oficial Mexicana NOM-025-SSA1-1993, Salud ambiental. Criterios para evaluar el valor límite permisible para la concentración de material particulado. Valor límite permisible para la concentración de partículas suspendidas totales PST, partículas menores de 10 micrómetros PM_{10} y partículas menores de 2.5 micrómetros $PM_{2.5}$ de la calidad del aire ambiente. Criterios para evaluar la calidad del aire, Secretaría Salud. Disponible en: http://201.159.134.50/Federal/PE/APF/APC/SSA/Modificaciones/26092005(1).pdf. Consultado en: 9 September 2011.

[34] P.Aldunate, O.Paz and K.Halvorsen, "Los efectos de la contaminación atmosférica por PM10 sobre la salud ciudad de La Paz – Bolivia (3650 m.s.n.m.)", vol. 3, pp. 422-442, June, 2006.

[35] L.Diego, "PM10 and Mortality." Maestría Facultad de Medicina de la Universidad de Salamanca, 2005.

[36] J.A.A.Florido, 2011, Contaminación medioambiental por ozono troposférico, Protección Civil Andalucia. Disponible en: http://www.proteccioncivil-andalucia.org/Documentos/Ozono.htm. Consultado en: 12 September 2011.

[37] L.Brown, *Química de la Ciencia*, 7th ed. Prentice Hall, 1998.

[38] U.S.Environmental Protection Agency (EPA), 2010, An Introduction to Indoor Air Quality (IAQ), Disponible en: http://www.epa.gov/iaq/no2.html#Health%20Effects%20Associated%20with%20Nitrogen%20Dioxide. Consultado en: 13 September 2011.

[39] C.D.S.Merino, F.L.Gutiérrez, A.G.Torres, C.I.R.Tejeda and A.L.A.Marco, "Patrones de Aglomeracion de la Contaminacion por Ozono y Oxidos de Nitrogeno en la Ciudad de México," in: *XIX Congreso de Investigación Cuam-ACMor,* 2008.

[40] WHO Working Group, *Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide* ., Bonn: WHO Regional Office for Europe, 2003.

[41] U.S.Environmental Protection Agency (EPA), 2011, Sulfur Dioxide, Disponible en: http://www.epa.gov/air/sulfurdioxide/. Consultado en: 13 September 2011.

[42] Air and Radiation, "Ozone and your health", *U.S.Environmental Protection Agency (EPA),* vol. EPA-452/F-99-003, 2009.

[43] U.S.Environmental Protection Agency (EPA), 2011, Health Effects of Pollution, Disponible en: http://www.epa.gov/region7/air/quality/health.htm. Consultado en: 13 September 2011.

[44] Sistema de Monitoreo de la Ciudad de México, 1996, Dióxido de Azufre (SO2), Disponible en: http://148.243.232.112/simat/pnso2.htm. Consultado en: 27 October 2010.

[45] F.P.Duarte, 1994, Norma Oficial Mexicana NOM-021-SSA1-1993. "Salud Ambiental Criterio para Evaluar la Calidad del Aire Ambiente con Respecto al Monoxido de Arbono (CO). Valor Permisible para la Concentracion de Monoxido de Carbono (CO) en el Aire Ambiente, como Medida de Proteccion a la Salud de la Poblacion", Secretaría de Salud. Disponible en: http://www.salud.gob.mx/unidades/cdi/nom/021ssa13.html. Consultado en: 11 September 2011.

[46] E.E.Rubio, 2002, MODIFICACION a la Norma Oficial Mexicana NOM-020-SSA1-1993, Salud ambiental. Criterios para evaluar la calidad del aire ambiente con respecto al ozono (O_3). Valores normados para la concentración de ozono (O_3) en el aire ambiente como medida de protección a la salud de la población, para quedar como Norma Oficial Mexicana NOM-020-SSA1-1993, Salud ambiental. Criterio para evaluar el valor límite permisible para la concentración de ozono (O_3) de la calidad del aire ambiente. Criterio para evaluar la calidad del aire., Secretaría de Salud. Disponible en: http://www.salud.gob.mx/unidades/cdi/nom/m020ssa13.html. Consultado en: 5 September 2011.

[47] F.P.Duarte, 1994, NORMA Oficial Mexicana NOM-023-SSA1-1993, Salud ambiental. Criterio para evaluar la calidad del aire ambiente con respecto al bióxido de nitrógeno (NO2). Valor normado para la concentración de bióxido de nitrógeno (NO2) en el aire ambiente como medida de protección a la salud de la población., Secretaría de Salud. Disponible en: http://www.cofepris.gob.mx/work/sites/cfp/resources/LocalContent/590/2/023ssa1.pdf. Consultado en: 13 September 2011.

[48] F.P.Duarte, 1994, Norma Oficial Mexicana NOM-022-SSA1-1993. "Salud Ambiental. Criterio para Evaluar la Calidad del Aire Ambiente, con Respecto al Bioxido de Azufre (SO_2). Valor Normado para la Concentracion de Bioxido de Azufre (SO_2) en el Aire Ambiente, como Medida de Proteccion a la Salud de la Poblacion", Secretaría de Salud. Disponible en: http://www.salud.gob.mx/unidades/cdi/nom/022ssa13.html. Consultado en: 8 September 2011.

[49] J.L.Hintze, *NCSS Help System*, Kaysville: NCSS, 2007.

[50] S.M.Pandit and S.M.Wu, *Time Series and System Analysis with Applications*., New York: John Wiley and Sons, 1983.

[51] J.L.Schafer, *Analysis of Incomplete Multivariate Data*, London: Chapman & Hall/CRC, 1997.

[52] J.L.Schafer and M.K.Olsen, "Multiple imputation for multivariate missing-data problems: A data analyst's perspective", vol. 33, pp. 545-571, March 1998.

[53] G.King, J.Honaker, A.Joseph and K.Scheve, "Analyzing incomplete political science data: An alternative algorithm for multiple imputation," *American Political Science Review*, vol. 95, no. 1, pp. 49-69, March 2001.

[54] J.Honaker, G.King and M.Blackwell, 2011, Amelia II: A Program for Missing Data, Disponible en: http://gking.harvard.edu/amelia/. Consultado en: 10 June 2011.

[55] J.Honaker and G.King, "What to Do about Missing Values in Time-Series Cross-Section Data", vol. 54, pp. 561-581, April 2010.

[56] R.J.A.Little and D.B.Rubin, *Statistical analysis with missing data*., New York: Wiley, 1987.

[57] Vitutor, 2011, Percentiles, Disponible en: http://www.vitutor.com/estadistica/descriptiva/a_13.html. Consultado en: 10 June 2011.

[58] A.L.Williams, "Poor Air Quality Threatens Human Health," in ***Is air pollution a serious threat to health?***, Andrea C.Nakaya, Ed. San Diego: Thomson Gale, 2005, pp. 20-24.

[59] C.A.Pope III, R.T.Burnett, M.J.Thun, E.E.Calle, D.Krewski, K.Ito and G.D.Thurston, "Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution," *Journal of the American Medical Association*, vol. 287, no. 9, pp. 1132-1141, March, 2002.

[60] J.Schwartz, C.Spix, G.Touloumi, L.Bachárová, T.Barumamdzadeh, A.Le Tertr, T.Piekarks, A.Ponce De Leo, A.Pönk, G.Ross, M.Sae and J.P.Schoute, "Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions," *Journal of Epidemiology and Community Health*, vol. 50, no. SUPPL. 1, pp. S3-S11, April, 1996.

[61] M.Quintero, M.A.Reyna-Carranza, K.Collins, S.Guzmán, B.Powers and A.Mendoza, "Issues Related to Air Quality and Health in the California-Baja California Border Region," in *The U.S.- Mexican border enviromet: Binational Air Quality Management* San Diego State University Press, 2006, pp. 1-46.

[62] M. A. Reyna-Carranza, "Un sistema de vigilancia Binacional como indicador de los efectos de la contaminación aérea sobre la salud ambiental de Mexicali y Valle Imperial," Academic and Professional Programs for the Americas (LASPAU), México, Informe Final, 2009.

[63] M. A. Reyna-Carranza, M. Quintero and K. Collins, "Correlation Study of the Association of PM10 with the Main Respiratory Diseases in the Populations of Mexicali, Baja California and Imperial County, California", vol. 26, pp. 22-36, March, 2005.

[64] M. A. Reyna-Carranza and H. G. Arriola, "Estudio basado en la regresión de Poisson, para conocer el grado de asociación que presentan las principales enfermedades respiratorias con los contaminantes del aire que rebasan los estándares en Mexicali," Academic and Professional Programs for the Americas (LASPAU), México, Informe Final, 2006.

[65] Gobierno del Estado de Baja California, "Mexicali," in *Enciclopedia de los Municipios de México Estado de Baja California* Instituto Nacional para el Federalismo y el Desarrollo Municipal, 2005.

[66] Instituto Nacional de Estadística y Geografía, 2011, INEGI. Censo de Población y Vivienda 2010, Disponible en: http://www.inegi.org.mx/sistemas/mexicocifras/default.aspx?src=487&ent=02. Consultado en: 13 September 2011.

[67] A.G.Alcocer, V.H.Celada, J.C.Lillo, J.A.G.Fernández and E.Provencio, *Programa para mejorar la calidad del aire de Mexicali 2000-2005*., México: Gobierno del Estado de Baja California; Gob Municipal de Mexicali; SEMARNAP; SSA; INE, 1999.

[68] Organización Panamericana de la Salud, *Clasificación Estadística Internacional de Enfermedades y Problemas Relacionados con la Salud*, 10 ed., Washington: Organización Mundial de la Salud, 1995.

[69] Camara de Diputados del H. Congreso de la Union, "**Distribución de Competencias y Coordinación**," in *Ley General del Equilibrio Ecológico y la Protección al Ambiente*. Ciudad de México: Camara de Diputados del H. Congreso de la Union, 1988.

[70] M.Saez, S.Pérez-Hoyos, A.Tobias, C.Saurina, M.A.Barceló and F.Ballester, "Métodos de Series Temporales en el Estudio de Epidemiologicos sobre Contaminación Atmosférica", vol. 73, pp. 133-143, April, 1999.

[71] S.Pérez-Hoyos, M.S.Zafra, M.A.Barceló, C.Cambra, A.F.Guzmán, J.M.Ordóñez, F.G.Grima, R.Ocaña, J.Bellido, L.C.Suárez and A.A.López, "Protocolo

EMECAM: análisis del efecto a corto plazo de la contaminación atmosférica sobre la mortalidad", vol. 73, pp. 177-185, April, 1999.

[72] I.M.Peláez, "Modelos de Regresión: lineal simple y regresión logística," in *Métodos Estadísticos para Enfermería Nefrológica.* Sociedad Española de Enfermería Nefrológica (SEDEN), 2006, pp. 195-214.

[73] J.O.Rawlings, S.G.Pantula, and D.A.Dickey, "Collinearity," in *Applied Regression Analysis* New York: Springer-Verlag, 2011, pp. 433-459.

[74] B.Ostro, *A Methodology for Estimating Air pollution Health Effects*., Geneva: Office of Global and Integrated Environmental Health. World Health Organization (WHO), 1996.

[75] J.L.Hintze, "Poisson Regression," in *NCSS Help System* Kaysville: NCSS, 2007, pp. 1-45.

[76] J.L.Hintze, "Decomposition Forcasting ," in *NCSS Help System* Kaysville: NCSS, 2007, pp. 1-12.

[77] K. Katsouyanni, G. Touloumi, E. Samoli, A. Gryparis, A. Le Tertre, Y. Monopolis, G. Rossi, D. Zmirou, F. Ballester, A. Boumghar, H. R. Anderson, B. Wojtyniak, A. Paldy, R. Braunstein, J. Pekkanen, C. Schindler and J. Schwartz, "Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 project," *Epidemiology*, vol. 12, no. 5, pp. 521-531, July, 2001.

[78] J.F.Hair Jr, R.E.Anderson, R.L.Tatham and W.C.Black, "Multiple Regression Analysis," in *Multivariate data Analysis,* 5 ed Prentice Hall, 1998.

[79] A.Tobías and M.Saez, "Time-series regression models to study the short-term effects of environmental factors on health," in: *Working Papers of the Department of Economics,* Girona, April, 2004.

[80] S.Parodi and E.Bottarelli, "Poisson Regression Model in Epidemiology - An Introduction," *Annali della Facoltàdi Medicina Veterinaria di Parma,* vol. XXVI, pp. 25-44, 2006.

[81] J.O.Rawlings, S.G.Pantula and D.A.Dickey, "Geometry of Least Squares," in *Applied Regression Analysis*, 2nd ed. New York: Springer-Verlag, 1998, pp. 197-201.

[82] A.Pulido and J.R.Vález, "¿Afecta la multicolinealidad al análisis empírico de la productividad de las infraestructuras regionales?," in: *Working Paper*, Instituto Lawrence R. Klein: Universidad Autónoma de Madrid, 2006.

[83] M. A. Reyna-Carranza, A. P. Moreno-Flores and R. Lopez-Avitia, "Acute respiratory infections and its spearman correlation with meteorological and air pollutants data from Mexicali, BC, Mexico," in: *Pan American Health Care Exchanges, PAHCE 2010,* Lima, p. 17, March, 2010.

[84] A. P. Moreno and M. A. Reyna, "Assessment of procedures for air pollution time-series reconstruction," in: *Pan American Health Care Exchanges, PAHCE 2011,* Rio de Janeiro, p. 18, March, 2011.

[85] A.P.Moreno-Flores, M.A.Reyna-Carranza and H.G.Arriola, "Evaluación de los Procedimientos para la Reconstrucción de la serie Tiempo de la Contaminación del aire," in: *Congreso Regional de Bioingeniería,* UABC-Valle de las Palmas, p. 36, April, 2011.

APÉNDICE

APÉNDICE A. TABLA QUE INDICA LOS VALORES DE CORRELACIÓN DE LAS SERIES IMPUTADAS COMPARADAS CONTRA LA SERIE ORIGINAL

Tabla A.1 Resultados de las comparaciones de las series de tiempo imputadas con los diferentes métodos, comparados con la serie original

Métodos	PM_{10}		$PM_{2.5}$		CO		O_3		NO_2		SO_2	
	rs	p	rs	p	rs	p	rs	p	rs	p	rs	p
EAYG	0.9886	0.0000	0.9034	0.0000								
EAGG	0.9798	0.0000	0.9073	0.0000								
UAYG	0.9886	0.0000	0.9007	0.0000								
UAGG	0.9803	0.0000	0.9073	0.0000								
BAAYG	0.9658	0.0000	0.9201	0.0000								
BAAGG	0.9838	0.0000	0.9363	0.0000								
AAAYG	0.9574	0.0000	PMSR									
AAAGG	0.9873	0.0000	0.9068	0.0000								
EA	0.9759	0.0000	0.8718	0.0000	1.0000	0.0000	0.9661	0.0000	0.9947	0.0000	0.9723	0.0000
UA	0.9759	0.0000	0.8718	0.0000	1.0000	0.0000	0.9661	0.0000	0.9943	0.0000	0.9568	0.0000
BAA	PMSR		MBFY		1.0000	0.0000	0.9881	0.0000	0.993	0.0000	PMSR	
AAA	PMSR		MBFY		1.0000	0.0000	0.9903	0.0000	0.9969	0.0000	PMSR	
AMELIA II	0.9719	0.0000	0.9289	0.0000	0.986	0.0000	0.9881	0.0000	0.9917	0.0000	0.9828	0.0000

MBFY = Faltantes al Principio del Primer Año (Missing at the Beginning of First Year)

PMSR = Modelo Predictivo con Residuales Significativos (Predictive model with significant residual)

EAYG = Promediado por Exclusión con Factor de Ciclo

EAGG = Promediado por Exclusión con Factor de Serie

UAYG = Promediado por Unidad con Factor de Ciclo

UAGG = Promediado por Unidad con Facto de Serie

BAAYG = Automatic ARMA por Bloques con Factor de Ciclo

☐ Tratamientos no aplicable a contaminante

BAAGG = Automatic ARMA por Bloques con Factor de Serie

AAAYG = Automatic ARMA por Acumulado con Factor de Ciclo

AAAGG = Automatic ARMA por Acumulado con Factor de Serie

EA = Promediado por Exclusión

UA = Promediado por Unidad

BAA = Automatic ARMA por Bloques

AAA = Automatic ARMA por Acumulado

APÉNDICE B. ESTADÍSTICA DECRIPTIVA DE LAS VARIABLES CLÍNICAS, METEOROLÓGICAS Y CONTAMINANTES

Tabla B.1. Estadística descriptiva de las variables Clínicas, Meteorológicas y Contaminantes, por el periodo de cinco años, además de las temporadas de invierno y calor.

Variable	Min	Max	Rango	Media	SE	Mediana	Percentiles					RIC
							10%	25%	50%	75%	90%	
Muertes	1	27	26	9	4	9	5	7	9	12	14	5
Temporada invernal	1	27	26	11	4	10	6	8	10	13	16	5
Temporada cálida	1	22	21	8	3	8	5	6	8	10	13	4
IRAs Semanales	2	3371	3369	478	366	432	190	271	432	591	763	321
Temporada invernal	248	3371	3123	698	440	598	384	560	598	723	1023	217
Temporada cálida	2	746	744	312	157	303	86	213	303	422	512	209
Temperatura	4.9	42.1	37.2	25.2	8.1	24.7	14.7	18.0	24.7	32.9	35.9	14.9
Temporada invernal	4.9	29.4	24.5	17.8	4.1	17.5	12.8	14.9	17.5	20.3	23.7	5.4
Temporada cálida	15.9	42.1	26.2	30.9	5.4	32.1	22.9	26.8	32.1	35.1	36.9	8.3
Humedad Relativa	9.5	95.8	86.3	37.4	14.7	35.8	20.0	26.8	35.8	45.2	56.1	18.4
Temporada invernal	13.0	95.8	82.8	42.2	16.2	40.1	23.4	30.0	40.1	50.5	66.0	20.6
Temporada cálida	9.5	78.5	69.0	33.7	12.1	33.1	18.8	24.0	33.1	41.6	49.6	17.6
PM10	15.3	265.2	249.9	91.7	41.7	84.0	46.3	62.3	84.0	113.5	147.4	51.2
Temporada invernal	15.3	265.2	249.9	101.9	49.0	95.0	43.3	65.5	95.0	133.1	171.2	67.5
Temporada cálida	18.1	230.9	212.8	84.3	33.6	78.5	49.5	61.1	78.5	100.0	126.3	38.9
CO	0.1	7.1	7.0	1.1	1.1	0.7	0.3	0.4	0.7	1.5	2.7	1.1
Temporada invernal	0.1	7.1	7.0	1.8	1.3	1.6	0.5	0.8	1.6	2.5	3.5	1.7
Temporada cálida	0.1	4.4	4.3	0.6	0.5	0.5	0.2	0.3	0.5	0.7	1.3	0.4
PM2.5	11.0	260.2	249.2	58.0	30.0	50.8	30.4	38.6	50.8	69.7	95.1	31.1
Temporada invernal	11.0	260.2	249.2	70.5	38.5	64.5	26.8	40.8	64.5	90.0	120.1	49.2
Temporada cálida	12.5	237.0	224.5	51.0	20.8	47.4	31.2	37.6	47.4	58.4	74.5	20.9
NO2	0.003	0.055	0.053	0.021	0.010	0.019	0.009	0.013	0.019	0.028	0.035	0.015
Temporada invernal	0.006	0.055	0.049	0.027	0.009	0.027	0.015	0.020	0.027	0.033	0.038	0.013
Temporada cálida	0.003	0.043	0.040	0.016	0.007	0.014	0.008	0.010	0.014	0.020	0.026	0.009
O3	0.001	0.066	0.065	0.025	0.011	0.024	0.012	0.017	0.024	0.033	0.041	0.016
Temporada invernal	0.001	0.058	0.057	0.019	0.008	0.018	0.010	0.013	0.018	0.024	0.030	0.011
Temporada cálida	0.004	0.066	0.062	0.030	0.010	0.030	0.017	0.023	0.030	0.037	0.044	0.015
SO2	0.001	0.030	0.029	0.005	0.003	0.004	0.001	0.002	0.004	0.006	0.009	0.004
Temporada invernal	0.001	0.030	0.029	0.006	0.004	0.004	0.002	0.002	0.004	0.008	0.011	0.005
Temporada cálida	0.001	0.015	0.014	0.004	0.003	0.003	0.001	0.002	0.003	0.005	0.007	0.003

SE = Desviación Estandar Min = Valor Mínimo de la Serie RIC = Rango Inter Cuartílico

IRAs = Infecciones Respiratorias agudas casos diarios. Max = Valor Máximo de la Serie Muertes = Muertes por todas las causas menos lasexternas registradas diariamente

APÉNDICE C. DIAGRAMA DE CAJA Y EJES DE LOS CONTAMINANTE

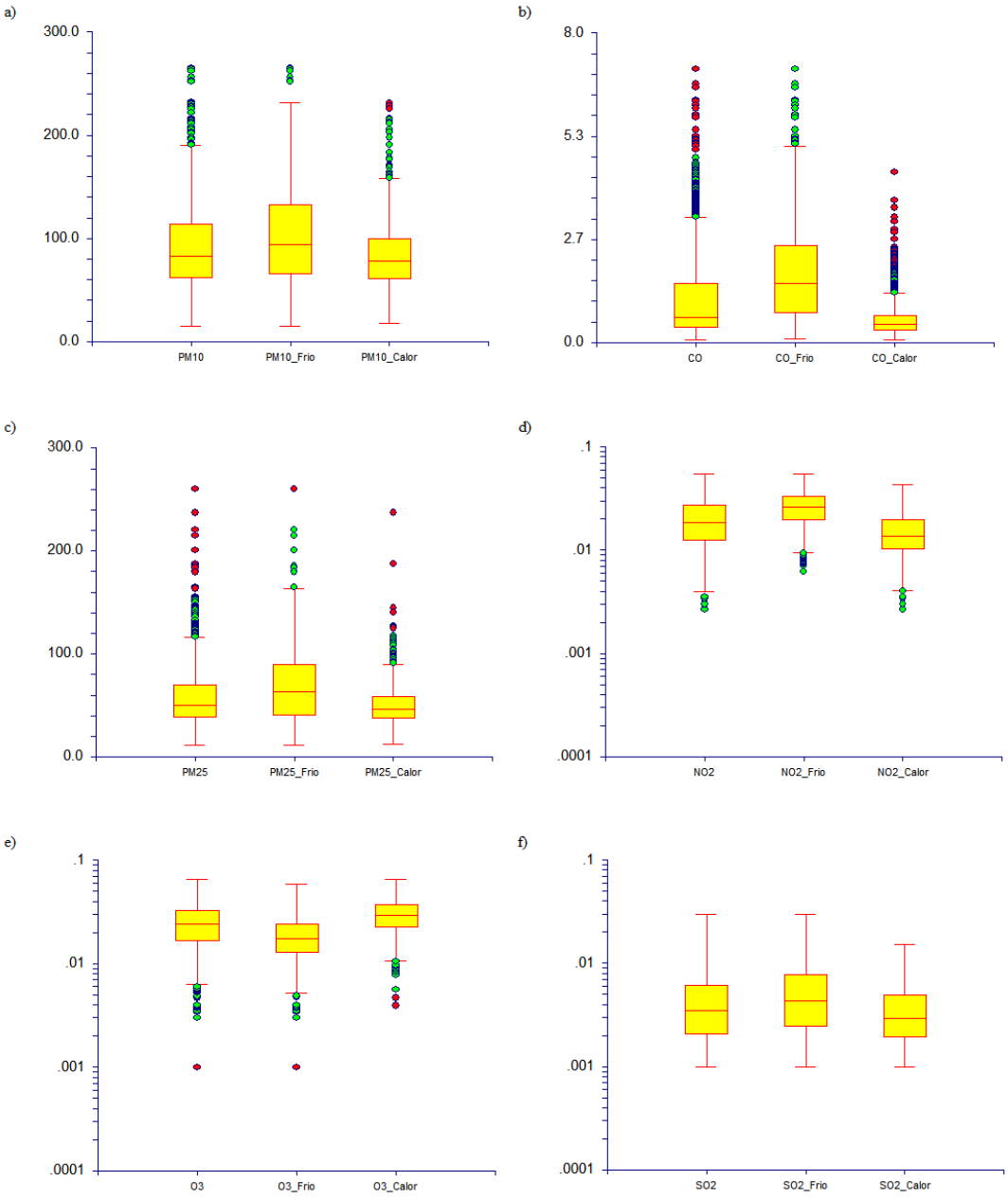


Figura C.1 Diagramas de caja y ejes de los contaminantes. Donde a) muestra el contaminante PM10, b) CO, c) PM2.5, d) NO2, O3 y SO2 en los incisos e y f. Dentro de cada diagrama se encuentra que la primera gráfica corresponde a valores de los cinco años del contaminante, la siguiente muestra la temporada de frío y la última, la de calor.

APÉNDICE D. CORRELACION ENTRE VARIABLES INDEPEDIENTES

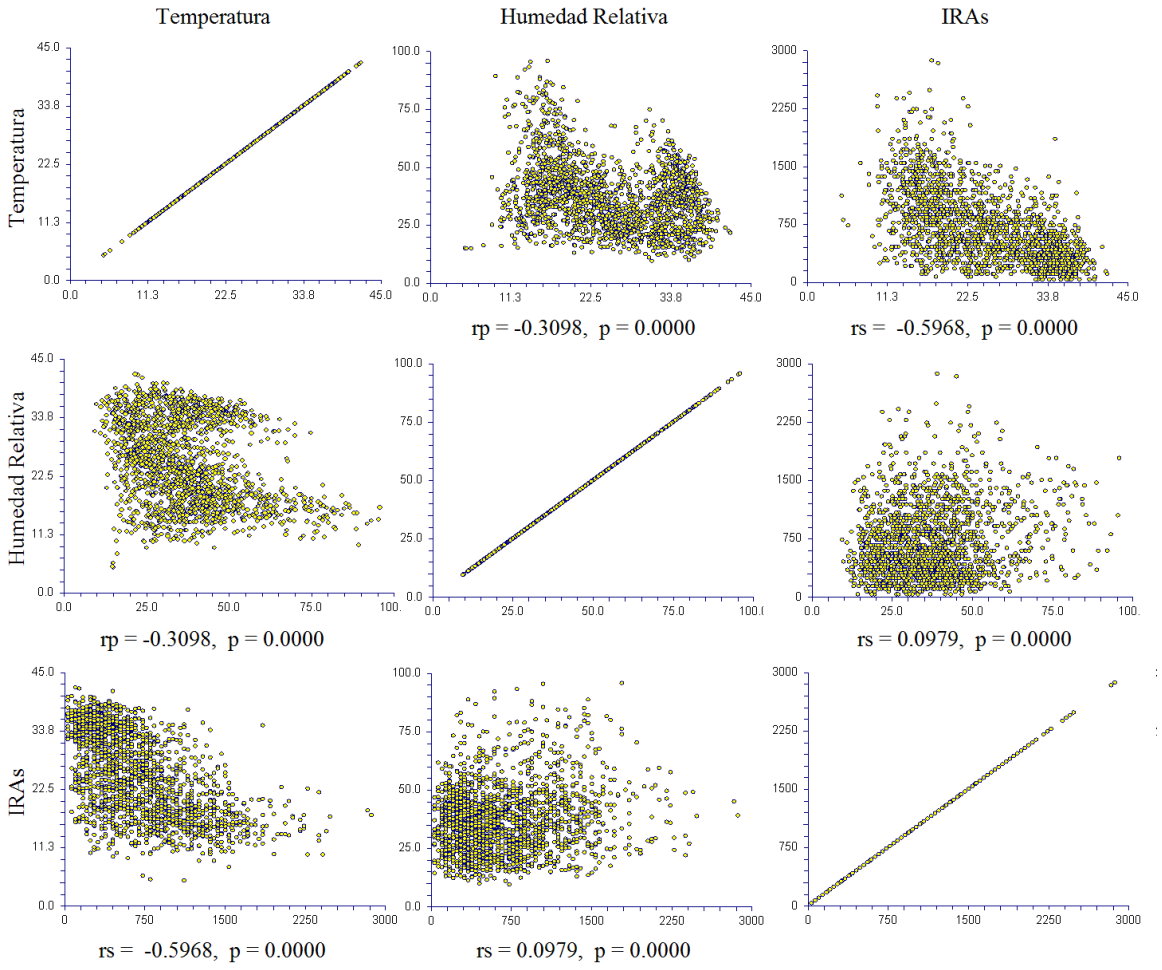


Figura D.1 Matriz de correlaciones entre variables del modelo basal.

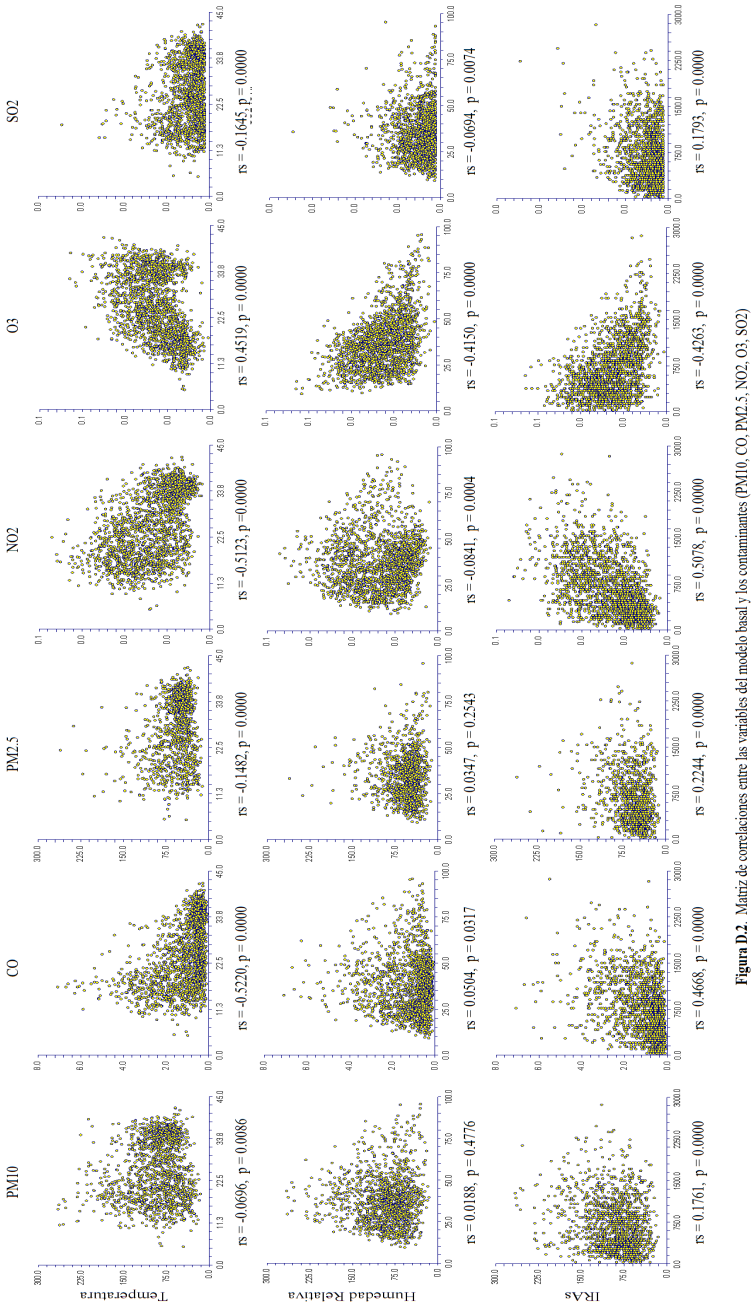


Figura D.2. Matriz de correlaciones entre las variables del modelo basal y los contaminantes (PM10, CO, PM2.5, NO2, O3, SO2)