



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Estudios de Postgrado  
Maestría en Estadística Aplicada

**DISEÑO DE UN MODELO DE SERIES TEMPORALES APLICADO A LA ESTIMACION DE  
LAS CONCENTRACIONES DE PARTÍCULAS SUSPENDIDAS EN EL AIRE (PM10, PM2.5 Y  
PM1) EN UNA INTERSECCIÓN DE CARRETERA**

**Ing. Kevin Emanuel Itzep Mendoza**

Asesorado por el Mtro. William Eduardo Fagiani Cruz

Guatemala, febrero de 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE UN MODELO DE SERIES TEMPORALES APLICADO A LA ESTIMACION DE  
LAS CONCENTRACIONES DE PARTÍCULAS SUSPENDIDAS EN EL AIRE (PM10, PM2.5 Y  
PM1) EN UNA INTERSECCIÓN DE CARRETERA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**ING. KEVIN EMANUEL ITZEP MENDOZA**  
ASESORADO POR EL MTRO. WILLIAM EDUARDO FAGIANI CRUZ

AL CONFERÍRSELE EL TÍTULO DE

**MAESTRO EN ESTADÍSTICA APLICADA**

GUATEMALA, FEBRERO DE 2023

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando Jose Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANA	Inga. Aurelia Anabela Cordova Estrada
DIRECTOR	Mtro. Edgar Darío Alvarez Cotí
EXAMINADOR	Mtro. William Eduardo Fagiani Cruz
EXAMINADORA	Dra. Aura Marina Rodríguez
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

## **HONORABLE TRIBUNAL EXAMINADOR**

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**DISEÑO DE UN MODELO DE SERIES TEMPORALES APLICADO A LA ESTIMACION DE LAS CONCENTRACIONES DE PARTÍCULAS SUSPENDIDAS EN EL AIRE (PM10, PM2.5 Y PM1) EN UNA INTERSECCIÓN DE CARRETERA**

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha 10 de noviembre 2020.

**Ing. Kevin Emanuel Itzep Mendoza**

Decanato  
Facultad de Ingeniería  
24189101- 24189102  
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.241.2023

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Estudios de Posgrado, al Trabajo de Graduación titulado: **DISEÑO DE UN MODELO DE SERIES TEMPORALES APLICADO A LA ESTIMACION DE LAS CONCENTRACIONES DE PARTÍCULAS SUPENDIDAS EN EL AIRE (PM10, PM2.5 Y PM1) EN UNA INTERSECCIÓN DE CARRETERA**, presentado por: **Ing. Kevin Emanuel Itzep Mendoza**, que pertenece al programa de Maestría en artes en Estadística aplicada después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

  
Inga. Aurelia Anabela Cordova Estrada  
Decana



Guatemala, febrero de 2023

AACE/gaoc



**Guatemala, febrero de 2023**

LNG.EEP.OI.241.2023

En mi calidad de Director de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor, verificar la aprobación del Coordinador de Maestría y la aprobación del Área de Lingüística al trabajo de graduación titulado:

**“DISEÑO DE UN MODELO DE SERIES TEMPORALES APLICADO A LA ESTIMACION DE LAS CONCENTRACIONES DE PARTÍCULAS SUSPENDIDAS EN EL AIRE (PM10, PM2.5 Y PM1) EN UNA INTERSECCIÓN DE CARRETERA”**

presentado por **Ing. Kevin Emanuel Itzep Mendoza** correspondiente al programa de **Maestría en artes en Estadística aplicada** ; apruebo y autorizo el mismo.

Atentamente,

“Id y Enseñad a Todos”



**Mtro. Ing. Edgar Darío Álvarez Cotí**  
Director

**Escuela de Estudios de Postgrado**  
**Facultad de Ingeniería**



Guatemala 12 de febrero 2022.

**M.A. Edgar Darío Álvarez Cotí**  
**Director**  
**Escuela de Estudios de Postgrado**  
**Presente**

**M.A. Ingeniero Álvarez Cotí:**

Por este medio informo que he revisado y aprobado el Informe Final del trabajo de graduación titulado **“DISEÑO DE UN MODELO DE SERIES TEMPORALES APLICADO A LA ESTIMACION DE LAS CONCENTRACIONES DE PARTICULAS SUSPENDIDAS EN EL AIRE (PM10, PM2.5 Y PM1) EN UNA INTERSECCION DE CARRETERA”** del estudiante **Kevin Emanuel Itzep Mendoza** quien se identifica con número de carné **2588455620101** del programa de Maestría en Estadística Aplicada.

Con base en la evaluación realizada hago constar que he evaluado la calidad, validez, pertinencia y coherencia de los resultados obtenidos en el trabajo presentado y según lo establecido en el *Normativo de Tesis y Trabajos de Graduación aprobado por Junta Directiva de la Facultad de Ingeniería Punto Sexto inciso 6.10 del Acta 04-2014 de sesión celebrada el 04 de febrero de 2014*. Por lo cual el trabajo evaluado cuenta con mi aprobación.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.

Atentamente,

**MSc. Ing. Edwin Adalberto Bracamonte Orozco**  
**Coordinador**  
**Maestría en Estadística Aplicada**  
**Escuela de Estudios de Postgrado**

Guatemala, 31 de julio de 2021.

M.A. Ing. Edgar Darío Álvarez Cotí

Director

Escuela de Estudios de Postgrado

Presente

Estimado M.A. Ing. Álvarez Cotí

Por este medio informo a usted, que he revisado y aprobado el Trabajo de Graduación y el Artículo Científico: **“DISEÑO DE UN MODELO DE SERIES TEMPORALES APLICADO A LA ESTIMACION DE LAS CONCENTRACIONES DE PARTICULAS SUSPENDIDAS EN EL AIRE (PM10, PM2.5 Y PM1) EN UNA INTERSECCION DE CARRETERA”** del estudiante **Kevin Emanuel Itzep Mendoza** del programa de Maestría en **ESTADÍSTICA APLICADA**, identificado con número de carné: **999003441**.

Agradeciendo su atención y deseándole éxitos en sus actividades profesionales me suscribo.



**INGENIERO QUÍMICO  
WILLIAM EDUARDO FAGIANI CRUZ  
COL. 1734**

M.A. Ing. William Eduardo Fagiani Cruz

Colegiado No. 17,734

Asesor

## **ACTO QUE DEDICO A:**

- Mi madre** Marleny Mendoza (q. e. p. d.), por ser una influencia muy valiosa en mi vida, ya que sin ella no sería el profesional que soy hoy. Te amo madre.
- Mi padre** Manuel Itzep, por sus consejos y su apoyo incondicional.
- Mis hermanos** Claudia y Roberto Itzep, por el apoyo y cariño que me brindan.
- Mi sobrina** Damari Chacón Itzep, por su amor, ya que cada día con ella es una nueva aventura.

## **AGRADECIMIENTOS A:**

<b>Universidad de San Carlos de Guatemala</b>	Por darme la oportunidad de formarme en la carrera.
<b>Facultad de Ingeniería</b>	Por darme la oportunidad de continuar con mi desarrollo profesional.
<b>Mis amigos de la Facultad</b>	A todos mis amigos, tanto del Departamento de Matemática, como de la carrera, ya que han sido de gran apoyo para mi persona.
<b>Mis amigos</b>	Gersson Guoz, Cesar Revolorio, Juan Carlos Martini, Pablo Chirix y Cesar Lopez, por la confianza y ayuda que me han brindado.
<b>Mi asesor</b>	William Fagiani, por su valiosa ayuda y enseñanza en este proceso.
<b>Mi amigo de la maestría</b>	Jonathan Álvarez por ser un excelente compañero.



2.1.3.4.	Proceso de diferenciación (d) .....	19
2.1.3.5.	Prueba Dickey-Fuller aumentada (DFA) .....	20
2.1.3.6.	Criterios de información para validación del modelo .....	22
2.1.3.6.1.	Akaike AIC .....	22
2.1.3.6.2.	Bayesiano Schwarz (BIC/SIC) .....	22
2.1.3.6.3.	MAPE .....	23
2.1.3.6.4.	ECM .....	23
2.1.3.7.	Ajuste del modelo .....	24
2.1.3.7.1.	Prueba de normalidad de <i>Kolmogorov Smirnov</i> .....	24
2.1.3.7.2.	Prueba de Normalidad Shapiro-Wilk .....	26
2.2.	Intervalos de confianza .....	26
2.2.1.	Intervalos de confianza con $\sigma$ desconocido y $n < 30$ .....	27
2.3.	Partículas en suspensión PM .....	27
2.3.1.	Clasificación de las partículas en suspensión .....	28
2.3.2.	Medición de las partículas en suspensión .....	28
2.3.3.	Legislación nacional e internacional de las partículas en suspensión .....	31
3.	PRESENTACION DE RESULTADOS.....	33
3.1.	Partículas PM.....	33
3.1.1.	Distribución de concentraciones PM por muestras .....	33
3.1.2.	Suavizamiento por medias móviles .....	34

3.1.3.	Serie de tiempo.....	35
3.1.4.	Estacionalidad de la serie PM.....	38
3.1.5.	Diferenciación (d).....	40
3.1.6.	ACF – PACF.....	42
3.1.7.	ARIMA.....	46
3.1.7.1.	Elección del mejor modelo ARIMA.....	46
3.1.8.	Normalidad de los datos.....	49
3.2.	Intervalos de confianza.....	59
3.2.1.	Intervalo de confianza PM1.....	59
3.2.2.	Intervalo de confianza PM2.5.....	60
3.2.3.	Intervalo de confianza PM10.....	61
4.	DISCUSIÓN DE RESULTADOS.....	63
4.1.	Análisis interno.....	63
4.1.1.	Modelo ARIMA PM1, PM2.5 y PM10.....	63
4.1.2.	PM1.....	63
4.1.3.	PM2.5.....	64
4.1.4.	PM10.....	64
4.1.5.	Normalidad de los datos.....	65
4.1.6.	Intervalos de confianza ARIMA.....	65
4.2.	Análisis externo.....	66
	CONCLUSIONES.....	67
	RECOMENDACIONES.....	69
	REFERENCIAS.....	71
	APENDICES.....	75



# ÍNDICE DE ILUSTRACIONES

## FIGURAS

1.	Fases para establecer un modelo ARIMA .....	10
2.	Correlograma simple y parcial AR .....	16
3.	Correlograma simple y parcial MA.....	18
4.	Prueba de Dickey – Fuller .....	21
5.	Prueba de Kolmogorov-Smirnov.....	25
6.	Sensor PMS3003 .....	29
7.	Composición del sensor de polvo .....	30
8.	Diagrama de muestreo y almacenamiento de datos.....	30
9.	Concentración PM1, PM2.5 y PM10 por muestras .....	34
10.	Concentración PM1, PM2.5 y PM10 – suavizamiento por medias móviles .....	35
11.	Serie PM .....	36
12.	Componente estacional PM1.....	37
13.	Componente estacional PM2.5 .....	37
14.	Componente estacional PM10.....	38
15.	Diferenciación PM1 .....	41
16.	Autocorrelación – PM1, PM2.5 y PM10 .....	43
17.	Autocorrelación parcial PACF – PM1, PM2.5 y PM10 .....	44
18.	Primera diferencia ACF – PM1, PM2.5 y PM10.....	44
19.	Primera diferencia PACF – PM1.....	45
20.	Pronóstico ARIMA (12, 1, 19) – PM1.....	47
21.	Pronóstico ARIMA (9, 1, 19) – PM2.5.....	48
22.	Pronóstico ARIMA (9, 1, 19) – PM10.....	49

23.	Q-Q Plot - PM1 .....	50
24.	Q-Q Plot – PM2.5.....	50
25.	Q-Q Plot – PM10 .....	51
26.	Correlograma de residuos – PM1 .....	54
27.	Histograma de residuos – PM1 .....	55
28.	Correlograma de residuos – PM2.5 .....	56
29.	Histograma de residuos – PM2.5.....	56
30.	Correlograma de residuos – PM10 .....	57
31.	Histograma de residuos – PM10 .....	58

## TABLAS

I.	Variables del estudio.....	XXII
II.	Formato de la base de datos .....	XXIV
III.	Serie Tiempo RStudio® - Script.....	7
IV.	ARIMA RStudio® - Script.....	12
V.	ACF - PACF RStudio® - Script .....	18
VI.	Medias móviles RStudio® - Script.....	19
VII.	Diferenciación RStudio® - Script.....	20
VIII.	Dickey-Fuller RStudio® - Script .....	21
IX.	Pines del sensor PMS3003.....	29
X.	Calidad del aire .....	31
XI.	Calidad del aire OMS.....	32
XII.	Prueba de Dickey-Fuller PM1 .....	39
XIII.	Prueba de Dickey-Fuller PM2.5 .....	39
XIV.	Prueba de Dickey-Fuller PM10 .....	40
XV.	Prueba de Dickey-Fuller PM1 diferenciada .....	41
XVI.	Prueba de Dickey-Fuller PM2.5 diferenciada .....	42
XVII.	Prueba de Dickey-Fuller PM10 diferenciada .....	42

XVIII.	ARIMA – PM1 .....	46
XIX.	ARIMA – PM2.5.....	47
XX.	ARIMA – PM10.....	48
XXI.	Test de ajuste de normalidad PM1 .....	52
XXII.	Test de normalidad PM1 .....	52
XXIII.	Test de ajuste de normalidad PM2.5 .....	52
XXIV.	Test de normalidad PM2.5 .....	53
XXV.	Test de ajuste de normalidad PM10 .....	53
XXVI.	Test de normalidad PM10 .....	53
XXVII.	Pronóstico PM1 .....	59
XXVIII.	Intervalo de confianza PM1 .....	60
XXIX.	Pronóstico PM2.5 .....	60
XXX.	Intervalo de confianza PM2.5 .....	61
XXXI.	Pronóstico PM10 .....	61
XXXII.	Intervalo de confianza PM10 .....	62



## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
<b>A</b>	Aleatoriedad
$\phi$	Coefficiente autorregresivo
$\theta$	Coefficiente media móvil
$\alpha$	Confiabilidad
<b>s</b>	Desviación Estándar Muestral
$\sigma$	Desviación Estándar Poblacional
$t_0$	Distribución t de Student
$H_a$	Hipótesis alterna
$H_0$	Hipótesis nula
$x_t$	Media móvil
$\bar{x}$	Media muestral
$\mu$	Media poblacional
<b>m</b>	Metro
$m^3$	Metro cúbico
$\mu g$	Microgramo
<b>mm</b>	Milímetro
<b>nm</b>	Nanómetro
<b>B</b>	Operador de retardo
$Y_t$	Serie temporal discreta
<b>T</b>	Tendencia
<b>C</b>	Variación cíclica
<b>E</b>	Variación Estacional



## GLOSARIO

<b>ACF</b>	Función de autocorrelación
<b>AIC</b>	Criterio de información Akaike
<b>ARIMA</b>	Modelo auto regresivo integrado de media móvil
<b>BIC</b>	Criterio de información Bayesiano
<b><i>Dickey-Fuller</i></b>	Prueba de estacionalidad de la serie
<b>ECM</b>	Error cuadrático medio
<b>ICA</b>	Índice de calidad del aire
<b>MAPE</b>	Error absoluto medio porcentual
<b>PACF</b>	Función de autocorrelación parcial
<b>PM</b>	Material particulado
<b>PM1</b>	Material particulado de $1\mu\text{m}$ (micrómetro de diámetro).
<b>PM10</b>	Material particulado de $10\mu\text{m}$ (micrómetro de diámetro).

**PM2.5**

Material particulado de  $2.5\mu\text{m}$  (micrómetro de diámetro).

***Shapiro-Wilks***

Prueba de normalidad

## RESUMEN

El propósito de este trabajo fue inferir un modelo ARIMA que estime a futuro las concentraciones PM, así mismo conocer el nivel de confiabilidad del modelo, para ello recomendar el uso de protección y con ello lograr la prevención de enfermedades respiratorias.

El objetivo principal fue identificar el modelo ARIMA que mejor se ajusta a los datos, a través de un análisis de series temporales, se emplea la metodología Box-Jenkins, análisis de residuos y se hace uso de los criterios de información.

Para cumplir con el propósito planteado la metodología consistió en identificar 4 modelos para cada tipo de concentración y mediante los criterios de información se seleccionó el que mejor se ajustan para cada nivel PM, así mismo se estimó el intervalo del nivel de concentración que estará presente en la intersección de carretera. El diseño realizado tuvo un alcance de tipo descriptivo adicionalmente es un estudio de tipo cuantitativo.

Se identificó que el modelo para cada concentración, para PM1 ARIMA (12, 1, 19), PM2.5 ARIMA (9, 1, 19) y PM10 ARIMA (9, 1, 19). Se descartaron los otros modelos ya que no se ajustan adecuadamente.

Se concluye que el modelo ARIMA, es una metodología que se adecua satisfactoriamente a la predicción del comportamiento del material particulado, además emplea los intervalos de confianza para estimar el rango promedio de dichas concentraciones. Por último, es probable que PM2.5 y PM1 sean las partículas que poseen un nivel dañino para grupos sensibles y moderado,

mientras PM10 se considera que tiene un nivel bueno. Es recomendable reducir el tiempo de muestreo para evitar fluctuaciones entre las muestras y aumentar el periodo de muestreo.

## PLANTEAMIENTO DEL PROBLEMA

- Contexto general

Cada día el ser humano está expuesto a un entorno aéreo que contiene una variedad de partículas en suspensión (PM, sus siglas en inglés *Particulate Matter*) lo cual representa un riesgo medioambiental y para la salud humana, tanto en el exterior como en interiores. Esta compleja mezcla incluye partículas orgánicas e inorgánicas, tal como el polen, polvo, humo, etc., que varían en tamaño, composición y origen. Dichas partículas tienen un amplio rango de tamaños entre ellos PM10, PM2.5 y PM1 los cuales poseen un diámetro de 10, 2.5, y 1 micrómetro ( $\mu\text{m}$ ) respectivamente.

Según la Organización Mundial de la Salud [OMS] (2018) en un artículo publicó que se estima que la contaminación ambiental en el aire, tanto en las ciudades, como en las zonas rurales, fue causa de 4,2 millones de muertes prematuras en todo el mundo por año. Adicional en el año 2016 la OMS estima que aproximadamente el 58 % de muertes prematuras, relacionadas con la contaminación atmosférica, se debieron a cardiopatías isquémicas y accidentes cerebrovasculares, el 18 % de las muertes se debió a la enfermedad pulmonar obstructiva crónica e infecciones respiratorias agudas, y el 6 % de las muertes se debieron al cáncer de pulmón, donde uno de los factores principales fue el de las partículas en suspensión, PM.

Por tanto, es necesario llevar un análisis de las partículas PM en los rangos anteriormente descritos, con el fin de observar cuál de los tres rangos está más presente e inferir en qué nivel se está expuesto según The World Air

Quality Index (AQICN), así mismo tomar acciones para poder disminuir el riesgo al ser expuesto ante las partículas en suspensión.

- Descripción del problema

Una evaluación realizada por la Centro Internacional de Investigaciones sobre el Cáncer de la OMS en el año 2013 determinó que la contaminación del aire exterior es carcinógena para el ser humano, estrechamente relacionado con la creciente incidencia del cáncer, especialmente el de pulmón. Adicionalmente el humo en interiores representa un grave riesgo para la salud, abarcando a unos 3000 millones de personas que cocinan, calientan sus hogares con combustibles fósiles o carbón.

Las PM conforman un indicador comúnmente representativo de la contaminación del aire, afectando a más personas que cualquier otro contaminante, donde sus principales componentes son, nitratos, amoniacos, cloruro de sodio, los polvos minerales y agua, que conforman una mezcla compleja de partículas líquidas y sólidas de sustancias inorgánicas y orgánicas que están suspendidas en el aire. Se pueden categorizar como partículas gruesas a las que están conformadas por PM10 y como partículas finas las que sean menor o igual que las PM2.5 al incluirse dentro de esta categoría a las PM1.

Debido a que las partículas en suspensión no son visibles al ojo humano, es necesario realizar un estudio que permita conocer los niveles de concentraciones PM y a la vez permita estimar los valores de las concentraciones a través del tiempo, mediante estadística descriptiva o estadística paramétrica, haciendo uso de series temporales que por medio de mediciones cuantitativas y considerar el efecto de los cambios en el clima al transcurrir del día (madrugada, mañana, tarde y noche) los cuales pueden afectar o variar las concentraciones

de las partículas en suspensión PM, en un punto concurrido por personas y vehículos (como una intersección de carretera), con el propósito de conocer los rangos de exposición según AQI (Air Quality Index). Por tanto, se desconoce cuál es el modelo con mejor ajuste sobre estos datos, para lo cual es necesario la caracterización de concentraciones antes mencionadas.

- Formulación del problema

En este apartado se planteó la pregunta central de investigación y las preguntas auxiliares que sirvieron de guía en el diseño del experimento.

- Pregunta central

¿Cuál debe de ser el modelo de series temporales que permita estimar las concentraciones PM y a su vez conocer su nivel de validez?

- Preguntas auxiliares

- ¿Cuál debe ser el modelo de series temporales que permita estimar las concentraciones de las partículas PM?
- ¿Cuál es el nivel de confiabilidad del modelo de predicción construido?
- ¿Cuáles son los promedios de las concentraciones PM presentes en una intersección de carretera durante la noche?

- Delimitación del problema

El problema se abordará en una intersección de carretera, donde se limitará el área de estudio a la perturbación de la composición de partículas en suspensión en la atmosfera. El área geográfica de estudio es la interconexión Naranjo-Minerva localizada en 14.672816 de latitud, -90.554793 de longitud, Zona 4 de Mixco, ciudad de Guatemala. Debido a que esta intersección presenta afluencia vehicular tanto en la madrugada y durante la noche, pero con interés específicamente en la noche, se obtendrán muestras de las concentraciones de partículas en suspensión en las tres categorías: PM10, PM2.5 y PM1. El plan de muestreo se realizará en el horario de 18:00 a 19:00 durante esta hora se tomarán 45 muestras en intervalos de 1 minuto durante los primeros 10 días del mes de octubre del año 2020.

## **OBJETIVOS**

### **General**

Diseñar un modelo que permita estimar los niveles de concentraciones PM por medio del análisis de series temporales para conocer su nivel de confiabilidad caracterizando los valores diarios observados en una intersección de carretera en función del tiempo.

### **Específicos**

1. Identificar cual debe de ser el modelo de series temporales, mediante la identificación de tendencias y un método de suavizado, que permita estimar las concentraciones de partículas suspendidas.
2. Estimar el nivel de confiabilidad, mediante un análisis de validez para el modelo de predicción seleccionado.
3. Estimar cuales son los promedios de las concentraciones de partículas suspendidas (PM<sub>10</sub>), por medio de intervalos de confianza, presentes en una intersección de carretera de alta afluencia vehicular.



## RESUMEN DEL MARCO METODOLÓGICO

A continuación, se presenta las características del estudio cuantitativo, las variables a emplear y las fases que se llevarán a cabo en la aplicación del estudio.

- Características del estudio

El enfoque del estudio propuesto fue cuantitativo ya que pretende estimar a futuro las concentraciones de los niveles PM<sub>10</sub>, PM<sub>2.5</sub> y PM<sub>1</sub>, se hizo uso de series temporales.

El alcance fue descriptivo, dado que se pretende diseñar un modelo que permitió estimar las concentraciones de los niveles PM en un tiempo específico durante la noche.

El diseño experimental fue observacional, pues la información sobre las partículas PM que se analizaron en su estado original sin ninguna manipulación; además fue longitudinal pues se estudió durante un periodo de tiempo las tendencias de los valores PM, y el comportamiento de las muestras que fueron obtenidas para generar conocimiento a dicho estudio. El modelo por utilizar fue un ARIMA ya uno de los enfoques más implementados en estudios relacionados con variabilidad climática dado que se considera registros estacionarios.

- Unidades de análisis

La población en estudio estaba formada por las partículas en suspensión PM (*Particulate Matter*) donde el muestreo se realizó en una intersección de

carretera en un área urbana, la cual se encuentra dividida en subpoblaciones dadas por PM1, PM2.5 y PM10, de la cual se extrajeron muestras por medio de un sensor óptico que tiene la capacidad de medir en los tres rangos PM descritos con anterioridad, que serán estudiadas en su totalidad.

- Variables

Tabla I. **Variabes del estudio**

<b>Variable</b>	<b>Definición teórica</b>	<b>Definición operativa</b>
Partículas suspendidas ( <i>PM1, PM2.5, PM10</i> )	Partículas sólidas y líquidas que se encuentran suspendidas en el aire. Variable cuantitativa continua.	Número en $\mu\text{g}/\text{m}^3$ , representa la concentración del material particulado en un volumen. Escala de razón.
Tiempo ( <i>t</i> )	Periodo determinado durante el que se realiza un acontecimiento. Variable cuantitativa continua.	Representa una secuencia de eventos. La unidad de medida es días. Escala de intervalo.

Fuente: elaboración propia.

- Fases del estudio
  - Fase 1: revisión de literatura
- Fundamentos sobre las partículas en suspensión o material particulado

Se elaboró una recopilación y análisis de información que será de uso para el estudio, con la finalidad de resaltar los puntos importantes que consideraron

dentro del análisis de las partículas en suspensión, así mismo indicar los efectos nocivos producidos a la exposición.

- Fundamentos estadísticos

Se elaboró una recopilación metodológica que proporcione información sobre el análisis de series temporales, se analizó cual fue mejor modelo ARIMA que pueda estimar las concentraciones PM a futuro, adicional se recopilará información sobre los criterios de información AIC, BIC, MAPE y ECM que permita definir el nivel de confiabilidad del modelo. Adicional se empleó el uso de gráficos para determinar la normalidad de la serie mediante Q-Q plot, histograma de residuos y correlograma. Por último, se aplicará la prueba de Kolmogorov-Smirnov y la prueba de Dickey-Fuller donde se observó la estacionalidad de la serie.

- Fase 2: gestión o recolección de la información

- Selección del instrumento de la información

El instrumento que se utilizó fue un sensor óptico PMS3003, ya que se consideró el factor económico y la precisión, adicional se agregó un módulo TTL ya que ayudará a la comunicación entre el sensor y la computadora. Por último, se desarrolló un script hecho en Python, que tuvo la función de controlar la información proporcionada con el sensor y de convertir la información de base hexadecimal a base decimal.

- Creación de la base de datos

Con la finalidad de almacenar los datos provenientes del sensor PMS3003, se procedió a utilizar un sistema de gestión de base de datos relacional en este caso se utilizó MySQL, por su facilidad de conexión y rápida lectura de datos.

- Conexión de la base de datos con el sensor

La conexión entre MySQL y PMS3003 se realizó mediante un script hecho en Python. Que tuvo la función de escribir y leer los datos muestreado por el sensor, adicional se transformaron los datos guardados en un archivo CSV (*comma-separated values*), para posteriormente ser analizados.

- Fase 3: recolección de la información

Al recabar las muestras en una intersección de calle, mediante el uso del sensor PMS3003, en un tiempo específico durante la noche, con el fin de que los datos posean un formato, los cuales se guardaron en la base de datos, lo que tendrá la siguiente estructura.

Tabla II. **Formato de la base de datos**

No. de Muestra	Hora	Tamaño de Muestra		
		PM1	PM2.5	PM10
N	...	....	....	....
.				
.				
N+1				

Fuente: elaboración propia.

Se realizaron 45 repeticiones por día para cada tamaño de muestra en un intervalo de tiempo de 10 días.

- Fase 4: análisis de la información

En el análisis de información se utilizó un software en específico que facilitó los cálculos estadísticos.

- Análisis de series temporales

Para el análisis de series temporales los datos se recopilaron en un intervalo de tiempo, el cual fue en días y en una hora específica de tiempo y de manera visual se observó el comportamiento entre el tiempo y las tres categorías de las partículas en suspensión PM. Adicional se observaron las componentes que conforman una serie temporal, para su posterior diseño del modelo. De manera que se plantearon las componentes, que son necesarias para diseñar el modelo que permitió estimar a futuro los niveles de concentraciones.

- Tendencia: comportamiento o evolución a través del tiempo.
- Variación estacional: patrón que se repite en un corto periodo de plazo.
- Variación cíclica: fluctuación en forma de onda alrededor de la tendencia.
- Aleatoriedad: También llamado ruido de carácter errático no muestra ninguna regularidad.

- Análisis gráfico

Se realizará a través de gráficos que representen el comportamiento de las concentraciones de las partículas en suspensión, de ser necesario se

utilizaron tipo de gráficos como los de dispersión, histograma y un gráfico de barras.

- Fase 5: interpretación de información

Con los resultados obtenidos, se buscará encontrar un modelo de series temporales que sea capaz de estimar las concentraciones de las partículas en suspensión presentes en una bifurcación de carretera.

- Fase 6: elaboración del informe final

Los resultados obtenidos fueron analizados e interpretados en el informe y agregados en este informe final.

## INTRODUCCIÓN

El presente trabajo fue una sistematización de un análisis experimental, ya que se buscó estimar el modelo ARIMA para cada concentración PM.

La investigación se encuadra en la línea de pronósticos, debido a que se requirió un modelo para hacer predicciones, en el análisis de las concentraciones de partículas en suspensión (PM) en el aire, detectadas en una intersección de carretera. Dichas partículas son la causa de infinidad de enfermedades, especialmente de tipo respiratorio.

Para analizar esta problemática es necesario mencionar sus principales causas de origen. Una de ellas son las que proceden de manera natural originadas debido a la descomposición mineral u orgánica y la segunda a partir de una actividad antropogénica (causada por el hombre). Con el presente estudio se pretende observar y pronosticar los niveles de concentraciones, con la finalidad de brindar una solución que buscará contribuir en la disminución de enfermedades producidas por la exposición.

El modelo se construyó a partir de inferir su tendencia, efecto estacional y componente aleatoria, donde las primeras dos son determinísticas y la última es aleatoria más conocida como ruido blanco. Así mismo se aplicó un método de proyección de tendencias, donde se buscó identificar la tendencia por medio de un ajuste por regresión y método de suavizado por medias móviles.

Con los resultados obtenidos, fue posible brindar recomendaciones sobre el uso de protección, tanto para los peatones, como para las personas que viajan

en motocicleta, buses urbanos y brindar información sobre la importancia de realizar los servicios mecánicos necesarios a todo tipo de vehículo para disminuir el nivel de las concentraciones del material particulado. La metodología que se empleó para este estudio se basa principalmente en identificar, estimar y diagnosticar un modelo de series temporales.

El informe final está estructurado en 4 capítulos, el primer capítulo se desarrolló el marco referencial donde se explicaron algunos estudios previos relacionados a los efectos producidos por la exposición y la importancia de un modelo de series temporales en el ámbito de pronósticos ya que proporciona información relevante para la toma de decisiones en futuro con cierta confiabilidad.

En el segundo capítulo se expuso la metodología sobre el análisis estadístico que se emplearon para la construcción de un modelo y se encuentra dividida en tres secciones, la primera sección está integrada desde estacionalidad, la normalidad, el análisis de sus residuos de la serie y los criterios de información.

La segunda sección está integrada por los conceptos estadísticos referentes a los intervalos de confianza y la tercera sección se desarrolló toda la teoría referente respecto al material particulado.

En el tercer capítulo se presentaron los resultados de la investigación, donde gráficamente se observaron la distribución de los datos observados contra predichos. A partir de ello se seleccionaron los modelos que mejor se ajuste a los datos observados, la metodología que se utilizó para la selección del modelo fue por medio de los criterios de información.

El cuarto capítulo es una discusión detallada de los resultados obtenidos con los que se realizaron las conclusiones y recomendaciones.



## 1. MARCO REFERENCIAL

El diseño de predicción hace referencia a un modelo simplificado que tenga sentido desde la perspectiva ambiental, que se ajuste estrechamente a los datos disponibles y a la vez que aporte predicciones validas al aplicarlo a datos independientes. Debido a la exposición prolongada en ambientes contaminados, se tiene consecuencias que afectan la calidad y a las condiciones esenciales de la vida causando un incremento en el riesgo de sufrir trastornos respiratorios, tales como la neumonía, o enfermedades crónicas, como el cáncer.

García *et al.*, (2011) establece un sistema de control de calidad estadística para una tasa de incidencia acumulada semanal de intoxicaciones atendidas en urgencias de un hospital, donde empleó las notificaciones diarias de casos atendidos por intoxicación del Hospital Nuestra Señora de Candelaria, entre enero 2005 y diciembre 2009, haciendo uso una serie temporal. Empleando la metodología ARIMA (modelos autorregresivos integrados de medias móviles). Finalmente se predijo el valor endémico esperando una tasa con un periodo de antelación suficiente como para tomar medidas de intervención a partir del comportamiento observado. Esto hace constar la importancia de la metodología ARIMA en las series temporales, como un instrumento adecuado y fiable.

García *et al.*, (2016) realizaron un estudio observacional, descriptiva y aplicada hacia los pacientes hipertensos de Santiago de Cuba, empleando un análisis de series temporales, que abarco desde enero 2001 hasta diciembre 2014, las variables analizadas fueron la edad y el año de ocurrencia del proceso. Se usó el método de alisamiento exponencial con dos parámetros aplicado en la serie de tiempo y su tendencia. Finalmente se observó un incremento de

mortalidad y morbilidad en el pronóstico en cuanto a los indicadores con un intervalo de confianza del 95 % en 2015. Esto hace constar la importancia sobre el uso de los intervalos de confianza aplicado al análisis de series temporales.

Jiménez *et al.* (2006) comenta que las series temporales pueden ser agrupados en dos grandes categorías: métodos cualitativos y métodos cuantitativos. El primero se emplea cuando las observaciones pasadas no proporcionan conocimiento directamente sobre el problema y el segundo para obtener la mayor información posible contenida en los datos. (Law, 2000) Las series temporales se pueden descomponer en tendencia, factor cíclico, estacionalidad y componente irregular (Uriel y Muñiz, 1993). Finalmente se concluye que las series temporales son el resultado de la integración de esos cuatro componentes. Esto hace constar la importancia de incluir estos factores para el diseño de un modelo de series temporales.

Contreras (2015) establece que es importante controlar de manera eficiente los niveles de contaminación atmosférica hasta su eliminación o hasta conseguir la reducción a niveles aceptables en el área de Valencia. Se emplearon modelos predictivos a los que se le suministra ciertos datos de entrada (datos del emisor, del receptor y meteorológicos) los cuales generan unos datos de salida (concentraciones horarias, días entre otros). Se hizo del concepto minería de datos, se utilizaron modelos descriptivos, el uso de la correlación y modelos predictivos, como el uso de regresión. Se concluye que la técnica de Random Forest es capaz de generar las mejores predicciones. Esto sugiere que existen otro tipo de métodos que se puede emplear para el estudio de series temporales.

Martínez (2005) investigó sobre las partículas PM10, PM2.5 o partículas en suspensión en Madrid. Hizo uso de series temporales que incluían la media simultánea de PM10 y de PM2.5 en entornos característicos con la caracterización química de las partículas en ambas fracciones de tamaño, mediante el análisis de sus principales componentes e iones. Finalmente empleó un banco de datos de los años 1996 - 2000, se observó un aumento de partículas de las fuentes antropogénicas, como el método de combustión, donde se observa una mayor concentración de carbono. Esto sugiere que existe un aumento significativo a través del tiempo de forma alarmante.

Alonso *et al.*, (2005), realizaron un estudio sobre las partículas en suspensión con un tamaño menor de 10µm (PM10). Se observaron que los indicadores de mortalidad de salud en el área de Barcelona, Bilbao y Valencia presentaron problemas respiratorios y problemas cardiovasculares debido a la exposición. En este estudio se realizó el cálculo del impacto de los efectos de la contaminación en un periodo de 40 días. Hizo uso de series de temporales que estiman la probabilidad de enfermar o morir prematuramente a causa de la exposición de la contaminación atmosférica. Finalmente, los resultados muestran que los niveles PM10 sobrepasan los 50 µg/m<sup>3</sup> y son responsables de 1.4 muertes prematuras por 100,000 habitantes al año debido a sus efectos a corto plazo y 2.8 muertes/100,000 en un periodo de hasta 40 días tras la exposición. Esto sugiere la importancia de las pruebas preliminares durante el diseño de series temporales.

Vázquez *et al.* (2012) realizaron un análisis que aborda la contaminación causada por las partículas en suspensión que poseen un diámetro menor a 10µm y que poseen una concentración superior a 60 µg/m<sup>3</sup> (límite establecido sobre la buena calidad de salud del aire) utilizó tres lugares de monitoreo localizado en la cercanía en el Valle de México. El estudio abarca tres meses del año febrero,

agosto y noviembre, la información fue recolectada por día, incluye fines de semana y tuvo en cuenta los días de pago de quincenas, con el fin de estimar el comportamiento. Finalmente se observó que las épocas del año y la frecuencia de transitar aportan una influencia significativa en los niveles PM10, donde se estima que los días en que se observa una mayor contaminación son los jueves y viernes. Esto sugiere la importancia de un muestreo en un periodo de tiempo mayor a un mes.

AQICN es un sitio en internet donde se observan los niveles en tiempo real sobre el material particulado PM10 (10  $\mu\text{m}$ ), PM2.5 (2.5  $\mu\text{m}$ ) y PM1 (1  $\mu\text{m}$ ), emplea el sensor PMS3003 en una determinada región, donde se actualiza la gráfica en un periodo de tiempo determinado. Adicional presenta una serie de rangos de concentración con una breve descripción en cada una. Esto sugiere que se pueden comparar los valores estimados del modelo en un futuro.

Ballester *et al.* (1999) realizaron un estudio donde aplicaron series temporales, agregan que son las más utilizadas con el propósito de medir los efectos o cambios producidos en un periodo corto de tiempo sobre la contaminación ambiental y atmosférica. Ya que se analizaron las variaciones a través del tiempo sobre la exposición del contaminante y el indicador de la salud, siendo estudiado en una población específica. Los estudios ecológicos tienen como objetivo medir las exposiciones ambientales y cuáles son los efectos en poblaciones humanas representativas. Adicionalmente se considera que el rango de exposición está presente en el ambiente, como las diversas combinaciones de los factores ambientales y socioculturales, ya que pueden determinar el acontecimiento en los efectos sobre la salud. Esto sugiere la importancia de que las etapas del día pueden influir en las concentraciones.

WHO (2018), desarrolla y elabora directrices sobre la calidad del aire donde recomienda los límites máximos de exposición a las principales emisiones que contaminan el aire. Según la OMS realiza evaluaciones sanitarias periódicamente ante las diferentes clases de contaminantes presentes en el ambiente, incluyendo las partículas, el carbono negro, el ozono, etc. La OMS desarrolló una herramienta llamada AirQ+, que permite evaluar las repercusiones sanitarias de varios contaminantes.

Rodríguez (2007), utilizó los intervalos de confianza 95 % aplicado a la fórmula eritrocítica sabiendo que la fórmula varía a causa de factores como: edad, altura, actividad muscular, sexo, medicación, alimentación entre otros. El estudio fue observacional y comparativo, realizado en un Municipio de Comitán, que se encuentra localizado en el estado de Chiapas, México. Con base en los intervalos de confianza se determina que los parámetros hematológicos directos no fueron iguales entre sexos, e iguales en los indirectos. Esto sugiere la importancia de utilizar intervalos de confianza para el estudio del material particulado, ya que es de suma importancia conocer cuáles serán las concentraciones de ese día.

Por último, se concluye que un diseño de pronóstico para este estudio se hará utilizando un modelo de series temporales ARIMA ya que posee una gran aplicabilidad y se considera que es una gran herramienta esencial capaz de tomar una decisión ante cualquier proceso, ya que por su uso permite obtener un patrón de comportamiento de una variable mediante la observación de los datos recolectados a través del tiempo. Proyecta estos patrones para obtener una estimación para el futuro.



## 2. MARCO TEÓRICO

A continuación, se establece las referencias teóricas estadísticas y sobre el campo de aplicación del estudio.

### 2.1. Análisis de series temporales

Las series temporales, según Carrión (2001), son una secuencia de valores señalados a una misma variable correspondiente sobre distintos intervalos de tiempo. Mauricio (2007) afirma:

“Una secuencia de  $N$  observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o sobre varias características (serie multivariante o vectorial) de una unidad observable en diferentes momentos” (p.1).

En la siguiente tabla se definirá la serie y así mismo se asignará la frecuencia, en este caso se utilizará 45, ya que muestrearan 45 muestras al día.

Tabla III. **Serie Tiempo RStudio® - Script**

Comando en RStudio®
PM1 = ts (na. omit (pm1), frequency = 45)

Fuente: elaboración propia.

### 2.1.1. Proceso estocástico

Una definición del proceso estocástico, indica que “Un proceso estocástico es una secuencia de variables aleatorias, ordenadas y equidistantes cronológicamente, referidas a una (proceso univariante o escalar) o a varias (proceso multivariante o vectorial) características de una unidad observable en diferentes momentos” (Mauricio, 2007, p. 9).

Gras (2001) menciona que una serie temporal es continua cuando las observaciones son tomadas continuamente en el tiempo y es simboliza por  $Y(t)$ . Para este estudio las observaciones fueron registradas en intervalos predeterminados y de igual longitud. En este caso, los datos o valores forman una serie temporal discreta que esta simbolizada por  $Y_t$ :

$$Y_1, Y_2, Y_3, \dots, Y_{t-1}, Y_t \tag{Ec.01}$$

Para ello es necesario definir si las observaciones pertenecen a un proceso estocástico no estacionario o estacionario.

#### 2.1.1.1. Proceso estocástico estacionario

El proceso estocástico se define como estacionario ( $Y_t$ ) cuando la distribución de la probabilidad varia de forma aproximadamente constante en un periodo de tiempo. “Es estacionario cuando las propiedades estadísticas de cualquier secuencia finita  $Y_{t1}, Y_{t2}, \dots, Y_{tn}$  ( $n \geq 1$ ) de componentes ( $Y_t$ ), son semejantes a las de la secuencia  $Y_{t1+h}, Y_{t2+h}, \dots, Y_{tn+h}$  para cualquier número entero  $h = \pm 1, \pm 2, \dots$ ” (Mauricio, 2007, p. 11).

### **2.1.1.2. Proceso estocástico no estacionario**

Un proceso estocástico no estacionario ( $Y_t$ ) es cuya distribución de probabilidad varia de forma no constante. Indica que “es no estacionario cuando las propiedades estadísticas de al menos una secuencia finita  $Y_{t1}, Y_{t2}, \dots, Y_n (n \geq 1)$  de componentes de ( $Y_t$ ), son diferentes de las de la secuencia  $Y_{t1+h}, Y_{t2+h}, Y_{n+h}$ ” (Mauricio, 2007, p. 12).

### **2.1.2. Componentes asociadas a una serie temporal**

La inspección de una serie temporal a analizar que está conformada por cuatro componentes:

- Tendencia (T): comportamiento o evolución en el tiempo.
- Variación Estacional (E): patrón que se repite en un corto periodo de plazo.
- Variación Cíclica (C): fluctuación en tipo de onda en torno de la tendencia.
- Aleatoriedad (A): también llamado ruido de carácter errático no muestra ninguna regularidad.

### **2.1.3. Modelo ARIMA**

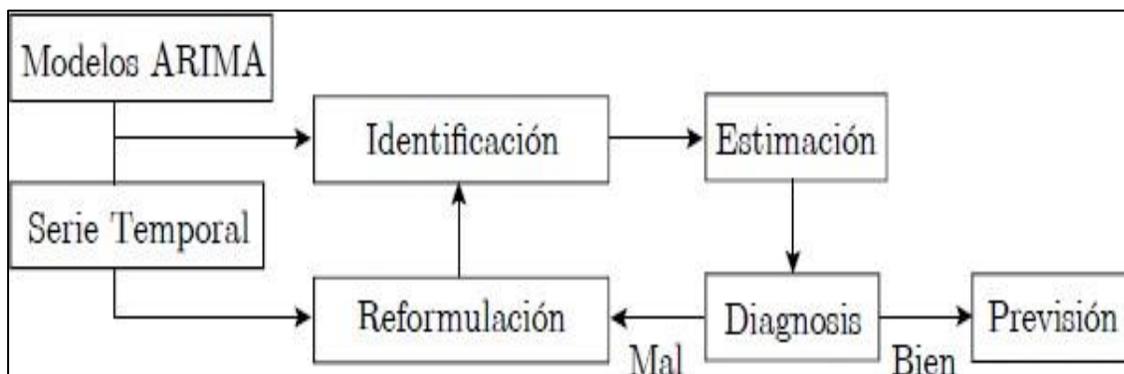
Un modelo autorregresivo de promedio móvil o ARIMA, requiere variaciones y regresiones sobre los datos estadísticos con la única finalidad de poder encontrar ciertos patrones para realizar una estimación hacia los eventos futuros. Fue un modelo desarrollado por Box y Jenkins en 1976.

Para definir el modelo que se ajuste al estudio ha de ser necesario tomar en cuenta las etapas requeridas para establecer o definir un modelo, Mauricio

(2007) argumenta sobre 3 etapas que son necesarias para la establecer de un modelo.

- Fases para establecer un modelo univariante:
  - Identificación
  - Estimación
  - Diagnósis

Figura 1. Fases para establecer un modelo ARIMA



Fuente: Mauricio. (2007). *Introducción al Análisis de Series Temporales*.

Para definir el modelo ARIMA, es necesario tener ciertas consideraciones que son indispensables para el presente estudio.

Según Carrión (2001) al representar el impacto de hechos pasados respecto al presente (tomando en cuenta el resultado en un tiempo venidero) de un proceso estocástico, se consideran diferentes expresiones:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + a_t$$

(Ec.02)

La función autorregresiva está definida por la ecuación 02, ya que tiene características de modelo de regresión sobre un proceso en sí mismo.

Adicional Carrión (2001) agrega que se debe de considerar la aleatoriedad dentro del modelo para poder asegurar su carácter estocástico:

$$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (\text{Ec.03})$$

La función de medias móviles está definida por la ecuación 03.

Por último. “Un elemento que resulta de gran interés en estudio de estos modelos procesos lineales es el llamado operador retardo  $B$ . Tal operador actúa sobre un término de un proceso estocástico reduciendo el índice temporal en una unidad” (Carrión, 2001, p. 41).

$$B x_t = x_{t-1} \quad (\text{Ec.04})$$

Se definirán dos conceptos importantes que serán de mucha utilidad para el estudio, tanto que es la Invertibilidad y el coeficiente de correlación.

Previamente calculado los coeficientes (p, d, q) del modelo ARIMA, en la siguiente tabla se observa el script que se empleara tanto para obtener el modelo ARIMA y para realizar el pronóstico con un 95 % de confianza.

Tabla IV. **ARIMA RStudio® - Script**

---

<b>Comando en RStudio®</b>
PM1_arima = arima (PM1_dce, order = c (0,0,0))
PM1_forecast = forecast (PM1_arima, h=45, level = 95)

---

Fuente: elaboración propia.

### **2.1.3.1. Ruido blanco**

Ruido blanco se define como “Un proceso de ruido blanco univariante es una secuencia ( $A_t$ ) de variables aleatorias escalares idénticas e independientes distribuidas con media 0 y varianza  $\sigma^2$ , representado de otra manera como  $A_t \sim IDD (0, \sigma^2)$ ” (Mauricio, 2007, p. 13).

Si  $\{A_t\}$  posee una distribución normal Mauricio (2007) lo denomina como Normal o Gaussiano y se representa como  $A_t \sim NIDD(0, \sigma_A^2)$ . En dado caso sea un proceso no estacionario  $\{Y_t\}$ , lo define que posee una aleatoriedad y está definida como:

$$Y_t = \mu + Y_{t-1} + A_t \quad \forall \quad t = \pm 0, \pm 1, \dots, \quad (\text{Ec.05})$$

Donde:  $\{\mu\}$  son parámetros, en ocasiones vale cero,  $\{A_t \sim IDD (0, \sigma^2)\}$  posee esta distribución. Por último, estas consideraciones tomadas en cuenta para modelo de ARIMA (0, 1, 0).

### 2.1.3.2. Modelo autorregresivo (AR)

Un modelo autorregresivo, es una representación de un proceso aleatorio y donde la variable de interés depende especialmente de los valores de las observaciones previas.

Es una representación de un proceso al azar o aleatorio, donde la variable u objeto de interés depende únicamente de sus observaciones pasadas. Carrión (2001) redefine la ecuación 02 y hace uso del operador de retardo definido en la ecuación 04 se expresa el siguiente polinomio:

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \quad (\text{Ec.06})$$

Donde:  $\{\phi\}$  es el coeficiente autorregresivo y  $\{B\}$  es el operador de retardo.

De la ecuación 07 se le llama polinomio auto regresivo ya que el proceso estocástico posee un comportamiento que va acorde a la expresión la cual es denominada Autorregresivo que posee un orden  $p$ , se abrevia AR( $p$ ).

Se plantea la ecuación para un modelo ARIMA (1,0,0):

$$Y_t = \mu + \phi_1 Y_{t-1} + A_t \quad \forall \quad t = \pm 0, \pm 1, \dots, \quad (\text{Ec.07})$$

Donde:  $\{\mu, \phi\}$  son los parámetros,  $\{|\phi_1| < 1\}$  que definen a la condición de la estacionariedad y  $\{A_t \sim IDD(0, \sigma_A^2)\}$ . Por último, estas consideraciones tomadas en cuenta para modelo de ARIMA (1, 0, 0).

Adicional Mauricio (2007) define a la auto covarianza de orden  $k$  ( $k > 0$ ) sobre un proceso  $\{Y_t\}$ . Estacionario representado por  $\gamma_k$  y está definido como  $\gamma_k \equiv Cov [Y_t, Y_{t+k}]$ .

$$Y_t = \mu + \phi_1 Y_{t-1} + A_t \quad \forall \quad t = \pm 0, \pm 1, \dots, \quad (\text{Ec.08})$$

Se define la media como:

$$\mu_Y \equiv E[Y_t] \quad (\text{Ec.09})$$

Una vez definida la media en términos de la esperanza de la ecuación 09 este autor define la varianza:

$$\sigma_Y^2 \equiv VAR[Y_t] \equiv E[(Y_t - \mu_Y)^2] \quad (\text{Ec.10})$$

Y la auto covarianza de orden  $k$  de la siguiente manera:

$$\gamma_k \equiv Cov [Y_t, Y_{t+k}] \equiv E[(Y_t - \mu_Y) * (Y_{t+k} - \mu_Y)] \quad (k = 1, 2, \dots, ) \quad (\text{Ec.11})$$

Adicionalmente establece que la autocorrelación simple, que posee un orden  $k$  ( $k > 0$ ) es considerado como un proceso  $\{Y_t\}$  que es estacionario, representado como:

$$\rho_k \equiv \frac{Cov[Y_t, Y_{t+k}]}{Var[Y_t + Y_{t+k}]^{\frac{1}{2}}} \equiv \frac{\gamma_k}{\gamma_0} \quad (\text{Ec.12})$$

Donde  $\rho_k$  establece el nivel de asociación lineal de las dos variables cuantitativas.

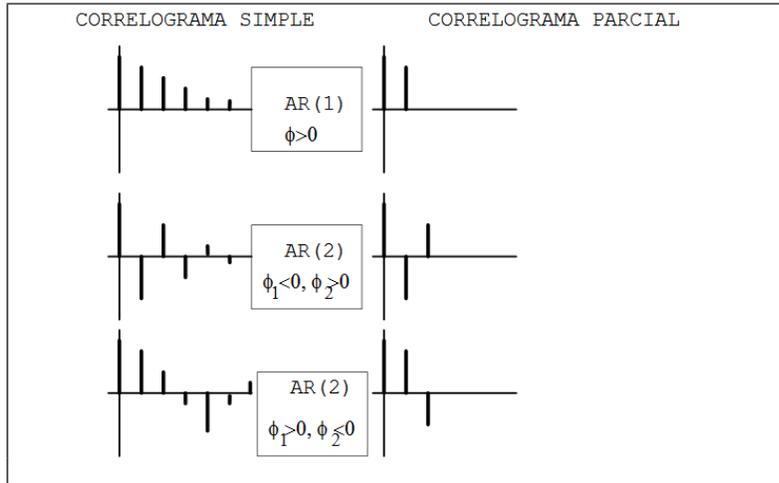
#### 2.1.3.2.1. Propiedades AR(p)

A continuación, se describe las propiedades AR(p):

- Estacionariedad: Carrión (2001) describe que la serie debe de ser estacionaria.
- Correlograma Simple: Un correlograma simple se formula como: Un proceso  $AR(p)$  está conformado por infinitos coeficientes de correlación no nulos, decreciendo según mezcla de exponenciales y sinusoidales amortiguadas.
- Correlograma Parcial: Asimismo, el correlograma parcial se establece como: Un proceso  $AR(p)$  tiene no nulos lo  $p$  primeros coeficientes de correlación  $\rho'_k$  (desde  $k = 1$  hasta  $k = p$ ), siendo nulos a partir de  $k=p+1$ . (Carrión, 2001, p. 42)

De manera que se adjunta una imagen de los frecuentes correlogramas parcial y simple de un proceso AR.

Figura 2. **Correlograma simple y parcial AR**



Fuente: Carrión. (2001). *Análisis de Series Temporales, técnicas y prevención*.

### 2.1.3.3. **Proceso de medias móviles (MA)**

Rodó (2020) establece que un modelo de media móvil será siempre estacionario. Los términos de error de períodos anteriores son independientes de los términos de error actuales y tienen la misma distribución (idéntica) de probabilidad con media 0 y varianza condicional.

Parte de la ecuación 03 y haciendo uso del operador de retardo definido en la ecuación 04. Carrión (2001) expresa el siguiente polinomio:

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \tag{Ec.13}$$

Donde:  $\{\theta\}$  es el coeficiente de media móvil y  $\{B\}$  es el operador de retardo.

Se plantea la ecuación para un modelo ARIMA (0,0,1):

$$Y_t = \mu + A_t - \theta_1 Y_{t-1} \quad \forall \quad t = \pm 0, \pm 1, \dots, \quad (\text{Ec.14})$$

Donde:  $\{\mu, \theta\}$  son los parámetros,  $\{|\theta_1| < 1\}$  que definen a la condición de la invertibilidad y  $\{A_t \sim IDD(0, \sigma_A^2)\}$ . Por último, estas consideraciones tomadas en cuenta para modelo de ARIMA (0, 0, 1).

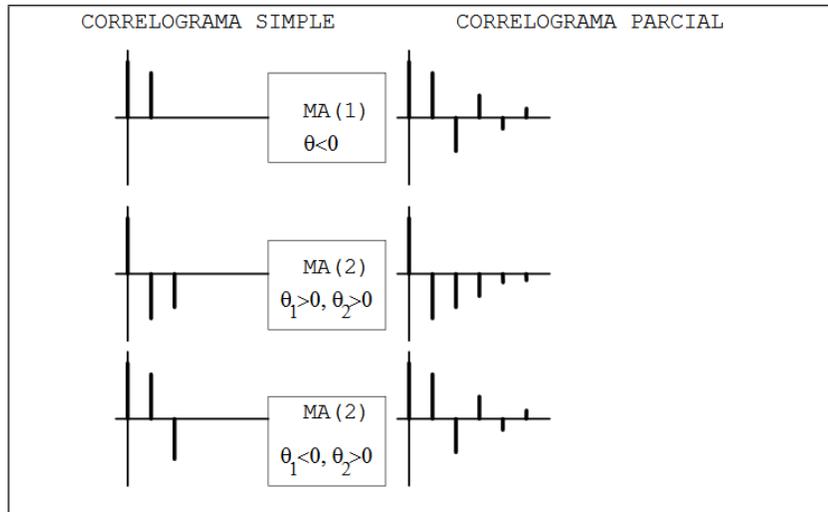
#### 2.1.3.3.1. Propiedades MA(q)

A continuación, se describen las propiedades MA(q):

- Estacionariedad: la serie debe de ser estacionaria.
- Invertibilidad: existe invertibilidad si todas las raíces de la serie poseen un valor por encima del círculo unitario.
- Correlograma simple: los primeros valores de los coeficientes  $\rho_k$ , indica el orden de MA(q).
- Correlograma parcial: Asimismo se define que: el correlograma parcial de un proceso MA(q) consta de infinitos coeficientes de correlación no nulos, decreciendo según una mezcla de exponenciales y sinusoidales amortiguadas. (Carrión, 2001, p. 44)

De manera que se adjunta una imagen de los frecuentes correlogramas simples y parciales de procesos MA.

Figura 3. **Correlograma simple y parcial MA**



Fuente: Carrión. (2001). *Análisis de Series Temporales, técnicas y prevención*.

Para realizar el gráfico de correlograma tanto para modelo AR y MA, se empleará el siguiente script.

Tabla V. **ACF - PACF RStudio® - Script**

Comando en RStudio®
Acf (PM1, main='ACF - PM1')
Pacf (PM1, main='PACF - PM1', col="blue")

Fuente: elaboración propia.

### 2.1.3.3.2. Suavizamiento de media móvil

La característica principal de este método consiste en realizar un promedio en un periodo más largo, el suavizamiento se aplica a todos los datos disponibles pasados. Lo que significa que permite reducir las fluctuaciones a corto plazo que

se pudieren observar en los datos, ya que este método toma información de periodos pasados.

$$x_t = \frac{z_{t-k} + z_{t-k+1} + \dots + z_{t-1}}{k} \tag{Ec.15}$$

En la siguiente tabla, se observa el Script que se utilizará en el software R.

Tabla VI. **Medias móviles RStudio® - Script**

Comando en RStudio®
pm1 <- ma (pm1\$dato.PM1, order = 5)

Fuente: elaboración propia.

#### 2.1.3.4. **Proceso de diferenciación (d)**

Si la sucesión de datos no fuera estacionaria, se llevará a cabo el proceso de diferenciación, el cual está definido como: “Un proceso estocástico  $Y_t$  es integrado de orden  $d$  ( $d \geq 0$  entero) o  $I(d)$ , sí solo si el proceso  $\nabla^2 Y_t$  sigue un modelo  $ARMA(p, q)$  estacionario e invertible” (Mauricio, 2007, p. 13).

En la siguiente tabla se hará uso para aplicar la diferenciación a la serie, si en dado caso la serie no es estacional.

Tabla VII. **Diferenciación RStudio® - Script**

---

<b>Comando en RStudio®</b>
<code>d = diff (PM1, differences = 1)</code>

---

Fuente: elaboración propia.

### **2.1.3.5. Prueba Dickey-Fuller aumentada (DFA)**

Para determinar la estacionalidad de la serie, se empleará mediante la prueba de Dickey-Fuller. La prueba de Dickey-Fuller (1981) es una prueba de raíz única que tiene como propósito detectar estadísticamente si la sucesión de datos es estacional mediante una prueba de hipótesis.

La hipótesis por tener presente para la prueba de Dickey-Fuller:

$$H_0: \varphi = 1$$

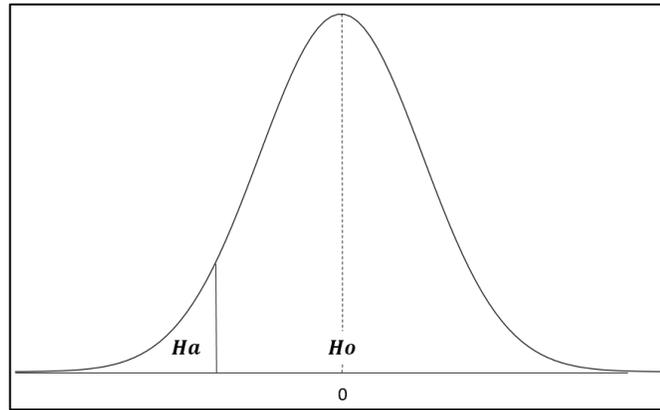
$$H_a: \varphi < 1$$

Si  $\varphi$  es muy cercano a 1 señala que la sucesión de datos es estacional. Para esta prueba se considerará p valor como característica de decisión para comprobar si la sucesión de datos es estacional o no, mediante una prueba de hipótesis la cual es la siguiente:

$$H_0: \text{No estacionaria}$$

$$H_a: \text{Estacionaria}$$

Figura 4. Prueba de Dickey – Fuller



Fuente: elaboración propia, realizado con RStudio®.

Dado que p valor  $> 0.05$ , se acepta  $H_0$ , lo cual indicará que la serie de datos no es estacionaria.

En la posterior tabla se hará uso del siguiente Script para la prueba de Dickey-Fuller.

Tabla VIII. Dickey-Fuller RStudio® - Script

Comando en RStudio®
<code>adf.test (PM1, alternative = "stationary")</code>

Fuente: elaboración propia.

### **2.1.3.6. Criterios de información para validación del modelo**

Para poder considerar el modelo adecuado para este estudio es necesario analizar los criterios de información:

#### **2.1.3.6.1. Akaike AIC**

Mauricio (2007) define que para un modelo de  $m$  parámetros y  $n$  observaciones, el criterio de información Akaike es calculado de la siguiente ecuación:

$$AIC = -\frac{2L^*}{n} + \frac{2m}{n} \quad (\text{Ec.16})$$

Donde:  $\{L^*\}$  logaritmo neperiano que está asociada al modelo estimado.

Una vez determinado el valor de AIC para cada estimación del modelo se debe de considerar aquel cuyo valor sea el mínimo.

#### **2.1.3.6.2. Bayesiano Schwarz (BIC/SIC)**

Amaris (2017) el criterio de información Bayesiano / Schwarz (BIC/SIC). BIC o SIC es considerado como una medida de bondad de ajuste de un modelo estadístico. Mauricio (2007) define que un modelo que posee  $n$  observaciones y posee  $m$  variables, el criterio bayesiano es calculado de acuerdo con la siguiente ecuación:

$$BIC = -\frac{2L^*}{n} + \frac{m * Ln(n)}{n}$$

(Ec.17)

Donde:  $\{L^*\}$  logaritmo que está asociada al modelo estimado.

### 2.1.3.6.3. MAPE

Amaris (2017) indica otro factor a considerar es el error porcentual absoluto medio (MAPE). Ya que MAPE evalúa la exactitud de una metodología para la elaboración ajustada de valores de la sucesión de datos en el tiempo.

$$MAPE = \frac{100}{n} * \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right|$$

(Ec.18)

Donde:  $\{x_i\}$  observación actual,  $\{\hat{x}_i\}$  observación estimada o pronosticada y  $\{n\}$  numero de datos no faltantes.

### 2.1.3.6.4. ECM

Finalmente, Amaris (2017) agrega un último criterio a considerar, el error cuadrático medio (ECM). ECM es la diferencia entre el valor observado y el valor de estimado o pronosticado. Es considerado el ECM como una función de riesgo que corresponde al valor esperado de la pérdida del error al cuadrado.

$$ECM = \frac{1}{n} * \sum_{i=1}^N (\hat{Y}_i - Y_i)$$

(Ec.19)

Donde:  $\{Y_i\}$  observación actual y  $\{\hat{Y}_i\}$  observación estimada.

### **2.1.3.7. Ajuste del modelo**

El ajuste del modelo de una serie temporal, Amaris (2017) indica que la verificación del análisis de residuos de la serie diferenciada, para observar el comportamiento aproximado del modelo planteado respecto hacia los valores reales de las observaciones.

#### **2.1.3.7.1. Prueba de normalidad de *Kolmogorov Smirnov***

La prueba de normalidad para este estudio determinará la bondad de ajuste de dos distribuciones de probabilidad entre sí, con la intención de conocer si se posee una distribución normal. Y se define como: “La prueba de Kolmogorov-Smirnov de una muestra es otra prueba de la bondad de ajuste. Está interesada en el grado de acuerdo entre la distribución de un conjunto de valores muestreados (puntuaciones observadas) y alguna distribución teórica específica” (Siegel, 1995, p. 73-74).

Para llevar a cabo la prueba es necesario determinar la desigualdad entre los valores  $S_n(x)$  con relación a los valores de la distribución normal  $N(x)$ . La prueba se enfoca sobre las desviaciones mas grandes. El valor absoluto más grande se conoce como la máxima desviación de D:

$$D_n = \max|S_n(x) - N(x)| \tag{Ec.20}$$

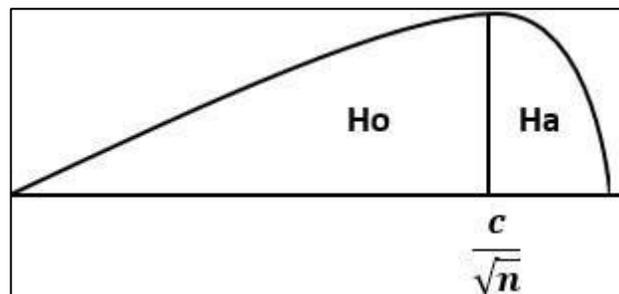
La distribución estadística  $D_n$ , se observará la región de densidad probabilística con un nivel de significancia  $\alpha$ , se encuentra definida como:

$$P\left(D_n > \frac{c}{\sqrt{n}}\right) = \alpha \quad (\text{Ec.21})$$

Donde:  $\{c\}$  valor cuya función es cuantificar el estadístico de prueba, donde la diferencia sea máxima

Cuando la diferencia sea menor al valor crítico, se considerará que los residuos son normales. Donde la decisión se presenta en el siguiente gráfico:

Figura 5. **Prueba de Kolmogorov-Smirnov**



Fuente: elaboración propia, realizado con RStudio®.

La hipótesis para tomar en cuenta para la prueba de Kolmogorov-Smirnov:

$$H_0: \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$$

$$H_a: \varepsilon_{ijk} \neq N(0, \sigma_\varepsilon^2)$$

### 2.1.3.7.2. Prueba de Normalidad Shapiro-Wilk

La prueba de *Kolmogorov-Smirnov* se contrastará con la prueba *Shapiro-Wilk*. Al calcular las diferencias entre las muestras en un orden de primero con el último, el segundo con el penúltimo y así sucesivamente, se corrige con los coeficientes tabulados por *Shapiro y Wilk*. El estadístico de prueba es:

$$W = \frac{D^2}{nS^2} \quad (\text{Ec.22})$$

Donde *D* es la suma de las diferencias corregidas, la hipótesis de normalidad es si el estadístico *W* es menor que el valor crítico. La hipótesis para tomar en cuenta para la prueba de Shapiro-Wilk:

*Ho: La serie presenta una distribución Normal*

*Ha: La serie no presenta una distribución Normal*

## 2.2. Intervalos de confianza

Con la finalidad de estimar cual serán los niveles de las concentraciones de las partículas o material particulado en el aire, se procede a hacer uso de los intervalos de confianza aplicados hacia los valores adquiridos del modelo ARIMA (*p*, *d*, *q*) en las concentraciones PM1, PM2.5 y PM10.

### 2.2.1. Intervalos de confianza con $\sigma$ desconocido y $n < 30$

Para ello definiremos que: “Una muestra aleatoria a partir de una distribución normal, entonces la variable aleatoria tiene una distribución  $t$  de Student” (Walpole, 2012, p. 274).

$$P\left(\bar{X} - t_{\frac{\alpha}{2}} * \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2}} * \frac{\hat{s}}{\sqrt{n}}\right) = 1 - \alpha$$

(Ec.23)

Donde  $s$  es la desviación estándar de la muestra,  $n$  tamaño de la muestra,  $\bar{X}$  media de muestral y  $t_{\frac{\alpha}{2}}$  es la distribución  $t$  de Student con  $n - 1$  grados de libertad. La desviación estándar de una muestra está definida como:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_j - \bar{x})^2}{n - 1}}$$

(Ec.24)

### 2.3. Partículas en suspensión PM

Para poder definir las partículas o material particulado según Mészáros (1999) se consideran uno de principales contaminantes a nivel atmosférico más estudiados mundialmente, ya que es definido como un grupo de partículas que poseen una composición sólida y líquida presentes en el ambiente y en la atmósfera.

Su origen puede provenir de fuentes antropogénicas o naturales. Según Chen, Ying y Kleeman (2009) la presencia que se observa en la atmósfera sobre este contaminante ocasiona diversos impactos entorno a los seres vivos, por

ejemplo, el declive visual en la atmósfera, que es causada por la absorción y la dispersión de la luz.

### **2.3.1. Clasificación de las partículas en suspensión**

Las partículas en suspensión PM están clasificadas por su de diámetro, para este estudio:

- Partículas Gruesas (PM10 – PM2.5)
- Partículas Finas (PM1)

### **2.3.2. Medición de las partículas en suspensión**

La magnitud de las partículas en suspensión PM cuantifica la concentración mediante las unidades de  $\mu\text{g}/\text{m}^3$ . El equipo utilizado es conocido como un medidor de partículas. Adicional se detallan las partes del equipo a continuación:

- Unidad óptica: encargado de cuantificar las concentraciones de dicho material particulado que atraviesan la unidad óptica.
- Ventilador: encargado de introducir aire a ser muestreado dentro de la recamara interna del dispositivo donde se localiza la unidad óptica.
- Circuito interno: posee un microcontrolador interno encargado de cuantificar por categoría, los niveles de concentraciones PM que ingresan dentro de la recamara, adicional es el encargado de la conexión serial.
- Conector hembra de 8 pines: son los pines de conexión que alimentan el sensor, transmiten y reciben información atreves de ellos.

Tabla IX. **Pines del sensor PMS3003**

<b>No. PIN</b>	<b>Significado</b>	<b>Comentario</b>
PIN1	VCC	Alimentación Positiva 5V
PIN2	GND	Alimentación Negativa
PIN3	SET	Set pin TTL 3.3V, Alto nivel funcionamiento normal y bajo nivel modo reposo
PIN4	RX	Puerto serial recibiendo
PIN5	TX	Puerto serial enviando
PIN6	RESET	Reinicia el sensor
PIN7/8	NC	

Fuente: elaboración propia.

El sensor se muestra a continuación, se puede observar el ventilador, conector hembra de 8 pines y recamara que dentro de esta se encuentra localizado la unidad óptica y circuito interno del dispositivo.

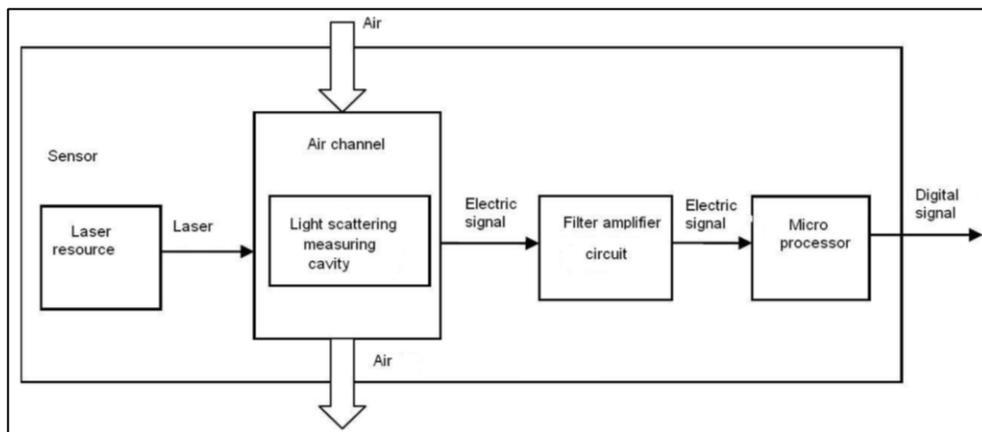
Figura 6. **Sensor PMS3003**



Fuente: [Fotografía de Kevin Itzep]. (Mixco, aldea Lo de Fuentes, Guatemala. 2021). Colección particular. Guatemala.

Adicional se presenta la estructura interna que conforma el circuito interno del sensor PMS3003.

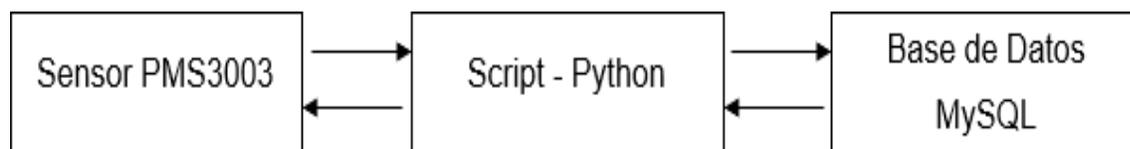
Figura 7. **Composición del sensor de polvo**



Fuente: Plantower Technology. (2016). *Digital universal particle concentration sensor*. Consultado el 24 de septiembre de 2020. Recuperado de <https://www.dynamoelectronics.com/descargas/PMS3003.pdf>.

Por último, se observa, en la siguiente imagen, el diagrama del procedimiento que se llevó a cabo que incluye el muestreo que se realizó haciendo uso del sensor y siendo almacenando en la base de datos, ambas partes siendo controladas mediante el script hecho en Python.

Figura 8. **Diagrama de muestreo y almacenamiento de datos**



Fuente: elaboración propia, realizado con PowerPoint 365.

### 2.3.3. Legislación nacional e internacional de las partículas en suspensión

Actualmente se encuentran en curso diferentes normas a nivel internacional que definen que la concentración máxima de exposición tanto en media anual como media de 24 horas.

En la siguiente tabla se observan los índices de calidad del aire (ICA) tanto en su media anual y media de 24 horas.

Tabla X. **Calidad del aire**

ICA	Color	Clasificación	PM <sub>10</sub> 24h µg/m <sup>3</sup>	PM <sub>2.5</sub> 24h µg/m <sup>3</sup>
0 - 50	Verde	Buena	0 – 54	0.0 – 15.4
51 – 100	Amarillo	Moderada	55 – 154	15.5 – 40.4
101 – 150	Naranja	Dañina para grupos sensibles	155 – 254	40.5 – 65.4
151 – 200	Rojo	Dañina para la salud	255 – 354	65.5 – 150.4
201 – 300	Purpura	Muy dañina	355 – 424	150.5 – 250.4
301 – 400	Marrón	Peligrosa	425 – 504	250.5 – 350.4
401 – 500	Marrón	Peligrosa	505 – 604	350.5 – 500.4

Fuente: elaboración propia.

Según la OMS (2005) establece que el aire limpio debe de ser un requisito básico y principal sobre la salud y por el bienestar de los seres humanos. Donde la contaminación representa una amenaza sumamente importante hacia la salud a nivel mundial. Más aún, el impacto ocasionado o generado hacia la salud son muy amplios, dado que las partículas al ingresar en los sistemas respiratorios y cardiovascular, han demostrado que el riesgo de diversos efectos aumenta por el tiempo de exposición. El manual de la calidad del aire de esta institución establece los siguientes valores de máximos de concentración en media anual y media de 24 horas.

Tabla XI. **Calidad del aire OMS**

<b>Tipo de Concentración</b>	<b>Concentración</b>
MP <sub>2.5</sub>	10 µg/m <sup>3</sup> media anual
	25 µg/m <sup>3</sup> media de 24 horas
MP <sub>10</sub>	20 µg/m <sup>3</sup> media anual
	50 µg/m <sup>3</sup> media de 24 horas

Fuente: elaboración propia.

En Guatemala el Decreto Número 68-86 Ley de Protección y Mejoramiento del medio ambiente, Título III, Capítulo I del Sistema Atmosférico en el Artículo 14 establece que emitirá los reglamentos correspondientes y además dictará las disposiciones que sean necesarias para la regulación sobre la calidad del aire. Por último, en el mismo artículo del inciso (a) hasta el inciso (g), se describe cuáles serán las acciones por considerar para la reducir las emisiones contaminantes.

### **3. PRESENTACION DE RESULTADOS**

Se presentan los siguientes resultados sin dejar de tomar en cuenta los objetivos propuestos del diseño experimental.

Para facilitar la comprensión de los resultados y diferenciar el nivel de concentración, se empleó distintos colores en las gráficas, para referenciar de que concentración se está presentando los resultados, PM1 se aplicó el color azul, PM2.5 se aplicó el color verde y para PM10 fue el color rojo.

- Objetivo 1. Identificar cual debe de ser el modelo de series temporales, mediante la identificación de tendencias y un método de suavizado, que permita estimar las concentraciones de partículas suspendidas.

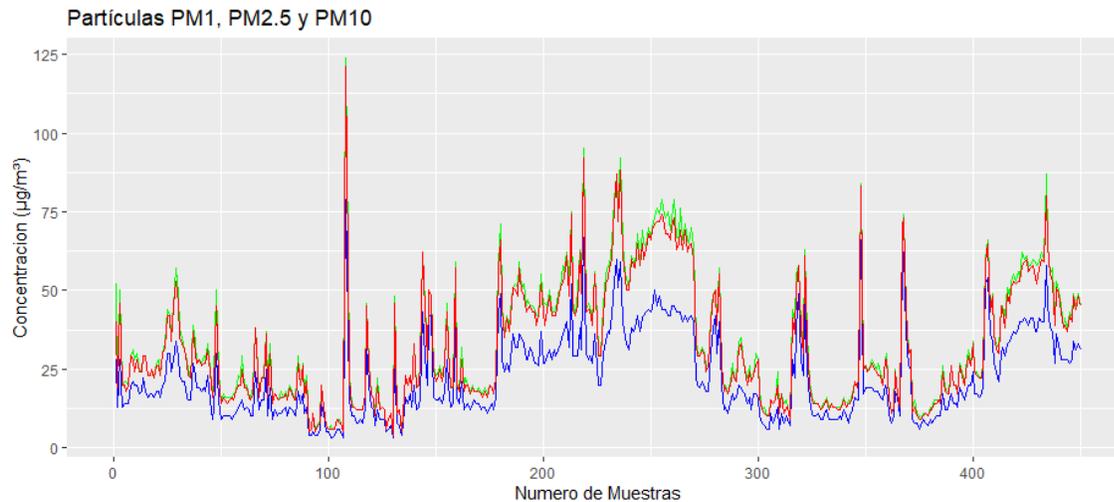
#### **3.1. Partículas PM**

Se presentan los resultados obtenidos del modelo ARIMA, con la finalidad de estimar a futuro las concentraciones (PM1, PM2.5 Y PM10).

##### **3.1.1. Distribución de concentraciones PM por muestras**

Se realizó un gráfico preliminar para cada tipo de concentración PM para observar la distribución de las partículas en relación con el número de muestras. Adicionalmente se observa que cada concentración posee similar distribución, a diferencia que la concentración para el caso de PM1 es menor que los niveles de PM2.5 y PM10. En contraste, se observó que los niveles de PM1 al ser valores de concentraciones bajos indica que no existe un mayor riesgo al exponerse.

Figura 9. **Concentración PM1, PM2.5 y PM10 por muestras**



Fuente: elaboración propia, realizado con RStudio®.

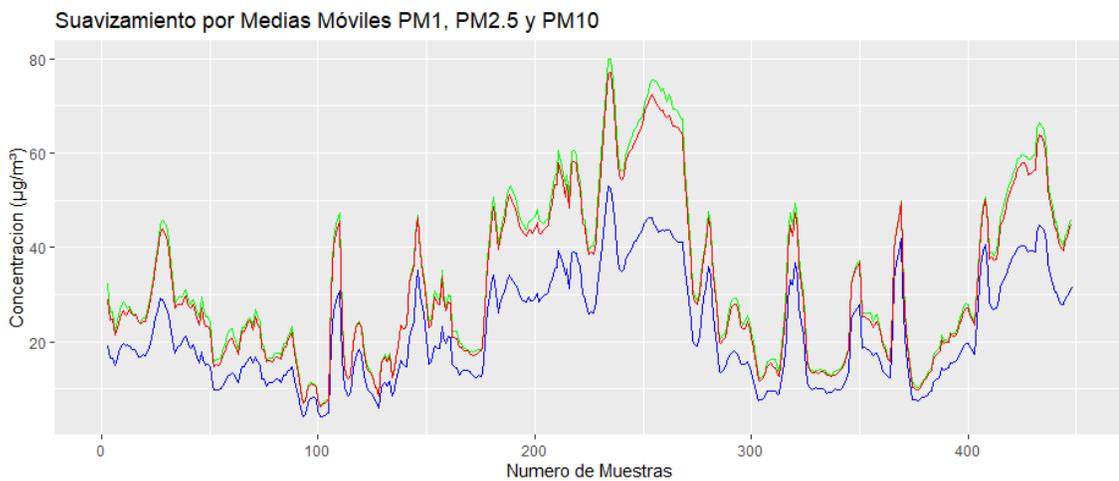
### 3.1.2. **Suavizamiento por medias móviles**

Se empleo el método de suavizamiento por medias móviles, ya que se deseaba conocer cuál sería el promedio por 5 días, así mismo eliminar las fluctuaciones a corto plazo lo que permitió observar la tendencia de las concentraciones de partículas en el aire. Para este caso se decidió utilizar un orden 5 para realizar el suavizamiento por medias.

Con ayuda del software RStudio®, se realizó el procedimiento del suavizamiento por medias móviles donde se obtuvo el resultado siguiente tanto de manera gráfica con el resultado de la operación para los datos de PM1 se adjuntan en Anexo A, para PM2.5 se adjuntan en Anexo B y para PM10 se adjuntan en Anexo C, se obtuvieron los resultados en el cual se observó que al utilizar el método de suavizamiento por medias móviles perdemos  $k/2$  términos en cada extremo de la serie.

Se observó una diferencia significativa de los datos originales, se aplicó el suavizamiento por medias móviles para las concentraciones de PM1, PM2.5 y PM10, donde se calculó los promedios de todos los grupos de k términos consecutivos de la serie, al aplicar este método se anula las variaciones aleatorias. El grafico siguiente corresponde al método empleado por medias móviles (ver figura 10).

Figura 10. **Concentración PM1, PM2.5 y PM10 – suavizamiento por medias móviles**



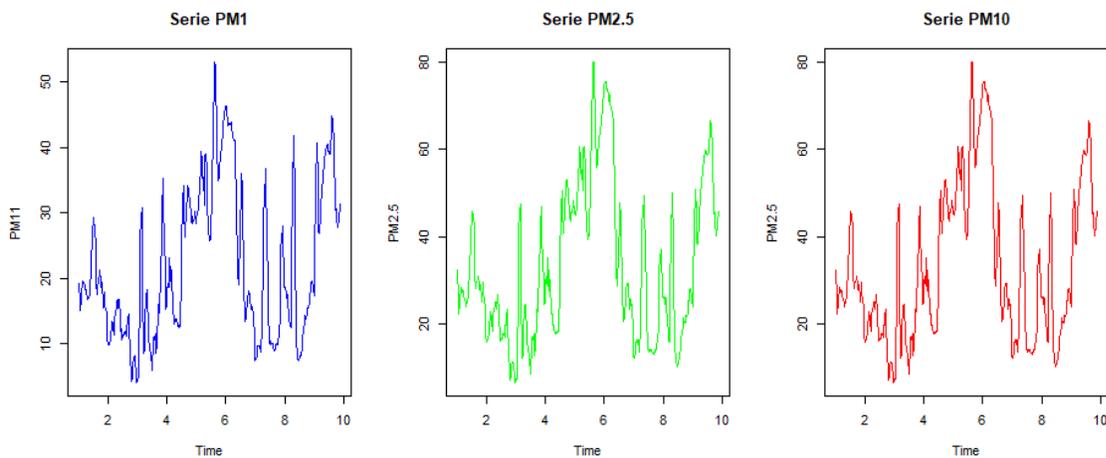
Fuente: elaboración propia, realizado con RStudio®.

### 3.1.3. Serie de tiempo

Se realizó la transformación de los datos a partir del resultado obtenido del suavizamiento de medias móviles, donde se indicó la frecuencia entre 45 ya que se realizará por semana:

El grafico siguiente señala el comportamiento de la sucesión de datos, hasta este punto no se realizó ninguna transformación que modificara a la sucesión de datos respecto al grafico anterior, únicamente se modificó el eje x que corresponde al valor del número de días que se obtuvieron las muestras (ver figura 11).

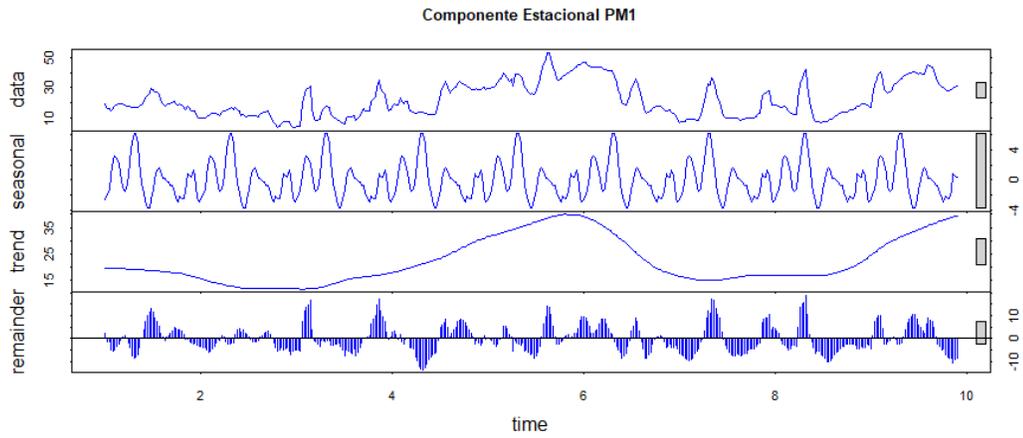
Figura 11. **Serie PM**



Fuente: elaboración propia, realizado con RStudio®.

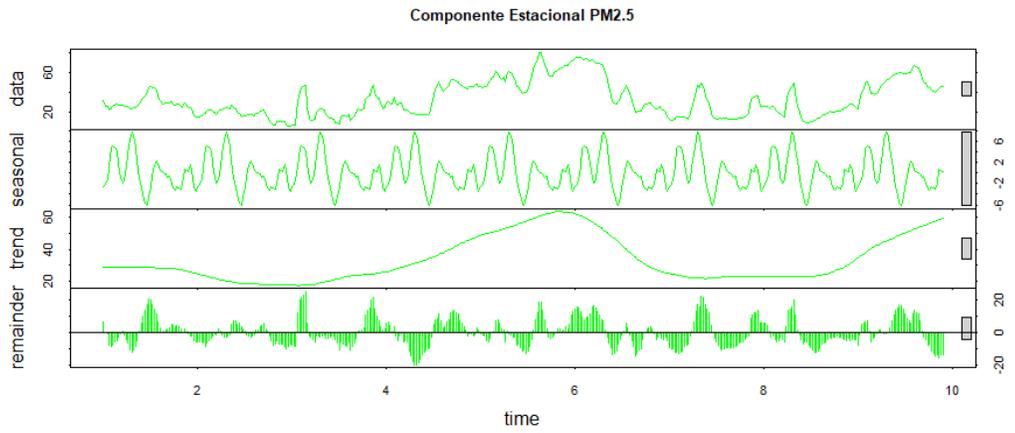
En las siguientes gráficas, se observó que la serie es posiblemente estacional ya que representa un movimiento periódico y no posee una clara tendencia ya que se observó que los valores crecen y decrecen (ver figura 12, 13, 14).

Figura 12. **Componente estacional PM1**



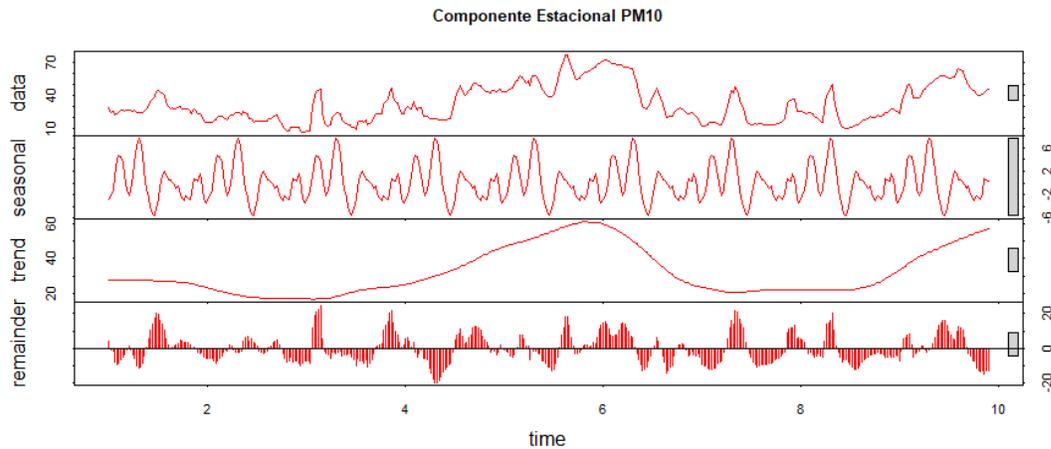
Fuente: elaboración propia, realizado con RStudio®.

Figura 13. **Componente estacional PM2.5**



Fuente: elaboración propia, realizado con RStudio®.

Figura 14. **Componente estacional PM10**



Fuente: elaboración propia, realizado con RStudio®.

#### 3.1.4. **Estacionalidad de la serie PM**

De manera gráfica se observó que para PM1, PM2 y PM10 las series de datos son posiblemente estacionales, lo cual se verificó mediante la prueba de *Dickey-Fuller*. La hipótesis que se probó, mediante el contraste de áreas (p-valor) que proporciona Rstudio® es la siguiente.

*Ho: No estacionaria*

*Ha: Estacionaria*

A partir de la prueba de *Dickey-Fuller* se aceptó la hipótesis nula para PM1, PM2.5 y PM10, es decir que las series no son estacionarias, los resultados de las pruebas se observan en las siguientes tablas:

Tabla XII. Prueba de Dickey-Fuller PM1

<b><i>Augmented Dickey-Fuller Test</i></b>	
Data	PM1
Dickey-Fuller	-2.8946
Lag order	7
p_value	0.1995
Debido a $P > 0.05$ , se acepta $H_0$	

Fuente: elaboración propia.

Donde el p valor para PM1 es mayor a la significancia, se establece que la sucesión de datos no es estacional.

Tabla XIII. Prueba de Dickey-Fuller PM2.5

<b><i>Augmented Dickey-Fuller Test</i></b>	
Data	PM2.5
Dickey-Fuller	-2.645
Lag order	7
p_value	0.305
Debido a $P > 0.05$ , se acepta $H_0$	

Fuente: elaboración propia.

Donde el p valor para PM2.5 es mayor a la significancia, se determinó que la serie no es estacional.

Tabla XIV. **Prueba de Dickey-Fuller PM10**

<b><i>Augmented Dickey-Fuller Test</i></b>	
Data	PM10
Dickey-Fuller	-2.6605
Lag order	7
p_value	0.2985
Debido a $P > 0.05$ , se acepta $H_0$	

Fuente: elaboración propia.

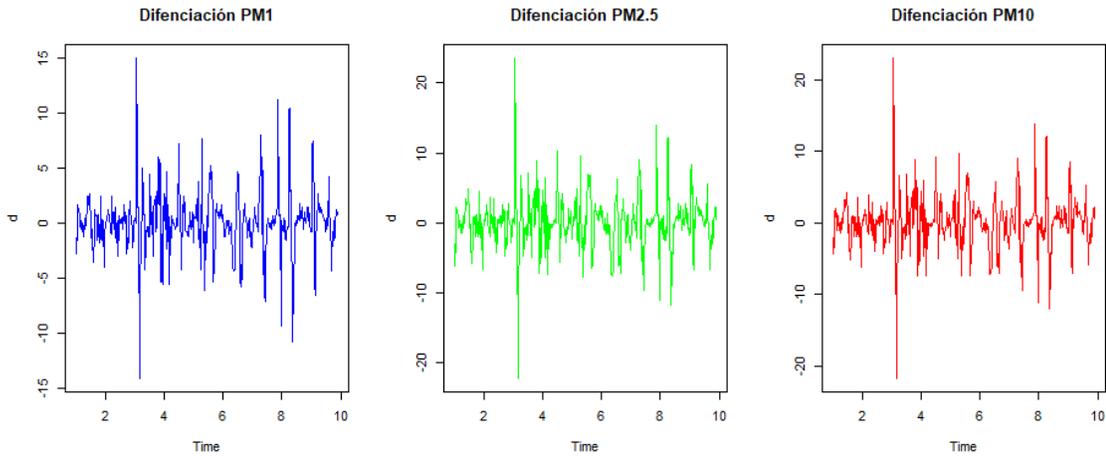
Por último, el p valor para PM10 es mayor a la significancia, lo que indica que la serie probablemente no es estacional.

### **3.1.5. Diferenciación (d)**

Debido a que la serie PM1, PM2 y PM10 no son estacionales, se procedió a realizar el método diferenciación de medias.

En la siguiente gráfica se observó un patrón oscilante alrededor de 0, además no se observó una tendencia visible, quiere decir que la diferenciación de orden 1 es suficiente y este resultado sugirió que debe de incluirse en el modelo. Donde será de orden 1 en el modelo ARIMA.

Figura 15. **Diferenciación PM1**



Fuente: elaboración propia, realizado con RStudio®.

Al existir un patrón oscilante y sin una tendencia visible, se realizó nuevamente la prueba de Dickey-Fuller. Se rechazó la hipótesis nula de la serie PM1, PM2.5 y PM10, lo que señala es que la sucesión de datos es estacionaria y se comprueba que de manera gráfica la diferenciación de orden 1 es suficiente y debe de incluirse en el diseño del modelo.

Tabla XV. **Prueba de Dickey-Fuller PM1 diferenciada**

<b><i>Augmented Dickey-Fuller Test</i></b>	
Data	d
Dickey-Fuller	-7.8183
Lag order	7
p_value	.01

Debido a  $P < .05$ , se rechaza  $H_0$

Fuente: elaboración propia.

Tabla XVI. **Prueba de Dickey-Fuller PM2.5 diferenciada**

<b>Augmented Dickey-Fuller Test</b>	
Data	d
Dickey-Fuller	-7.6104
Lag order	7
p_value	.01
Debido a $P < .05$ , se rechaza $H_0$	

Fuente: elaboración propia.

Tabla XVII. **Prueba de Dickey-Fuller PM10 diferenciada**

<b>Augmented Dickey-Fuller Test</b>	
Data	d
Dickey-Fuller	-7.7028
Lag order	7
p_value	.01
Debido a $P < .05$ , se rechaza $H_0$	

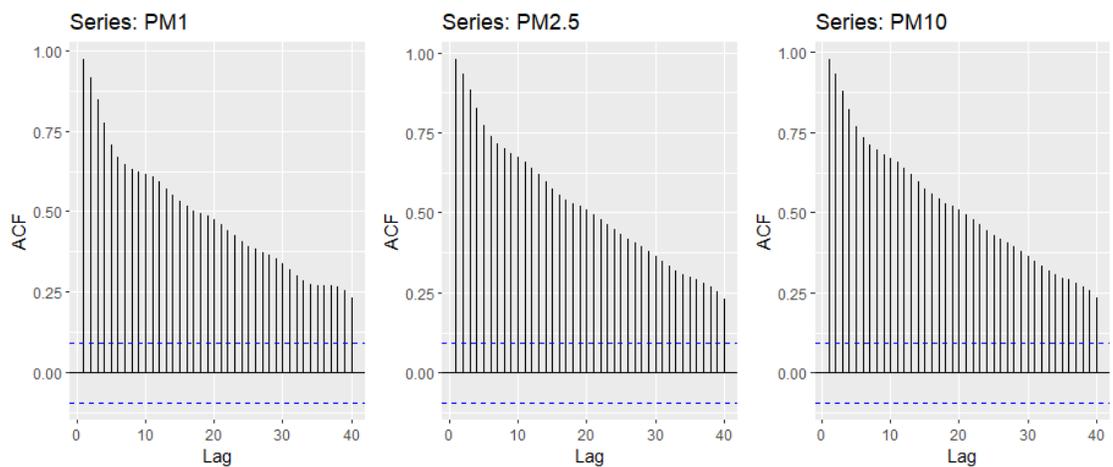
Fuente: elaboración propia.

### 3.1.6. ACF – PACF

Se obtuvo de forma gráfica se derivó los resultados de los ACF y PACF, ya que la autocorrelación brinda información sobre como una observación influye en las siguientes y la autocorrelación parcial proporciona información directa existente entre inspecciones separadas por  $k$  retardos. Con el propósito de seleccionar  $p$  y  $q$  para el modelo ARIMA.

La función ACF se usó para identificar el proceso MA y la función PACF se usó para identificar el proceso AR, los gráficos contienen bandas con un intervalo de confianza del 95 %. Para ACF se omitió la primera barra ya que corresponde a la correlación con la misma observación.

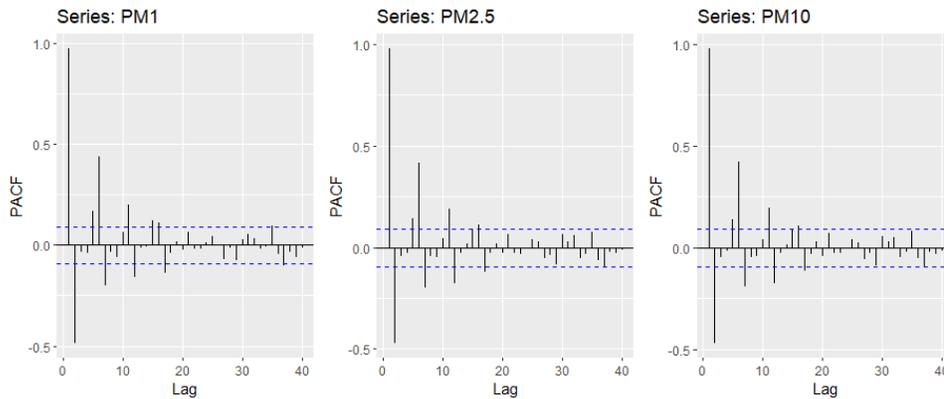
Figura 16. **Autocorrelación – PM1, PM2.5 y PM10**



Fuente: elaboración propia, realizado con RStudio®.

Se observó que, en la gráfica anterior, que la autocorrelación de manera sucesiva está disminuyendo a medida de aumenta los retrasos, quiere decir que hace parecer que no existe al menos una relación lineal entre todas las observaciones separados por retrasos los más grandes. Y se determinó que  $q = 0$ .

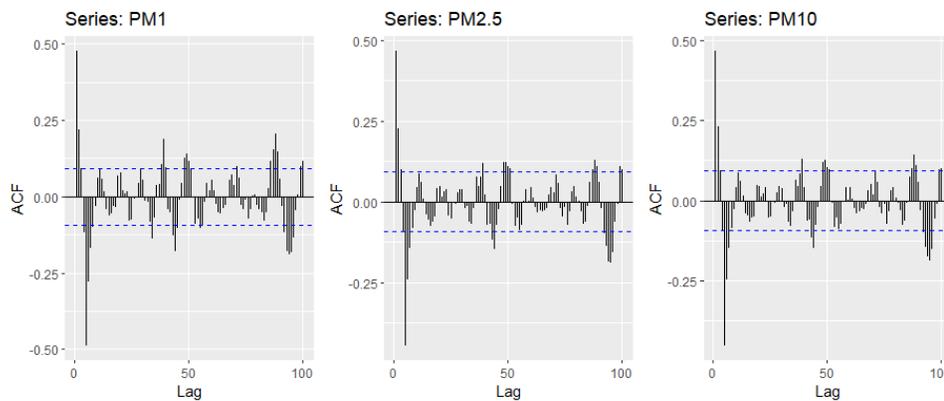
Figura 17. Autocorrelación parcial PACF – PM1, PM2.5 y PM10



Fuente: elaboración propia, realizado con RStudio®.

Se observó en el gráfico anterior donde se identificó el número de coeficientes de auto regresión AR, para PM1 se obtuvo un valor de  $p = 8$ , para PM2.5 se obtuvo un valor de  $p = 9$  y para PM10 se obtuvo un valor de  $p = 9$ , (ver figura 17).

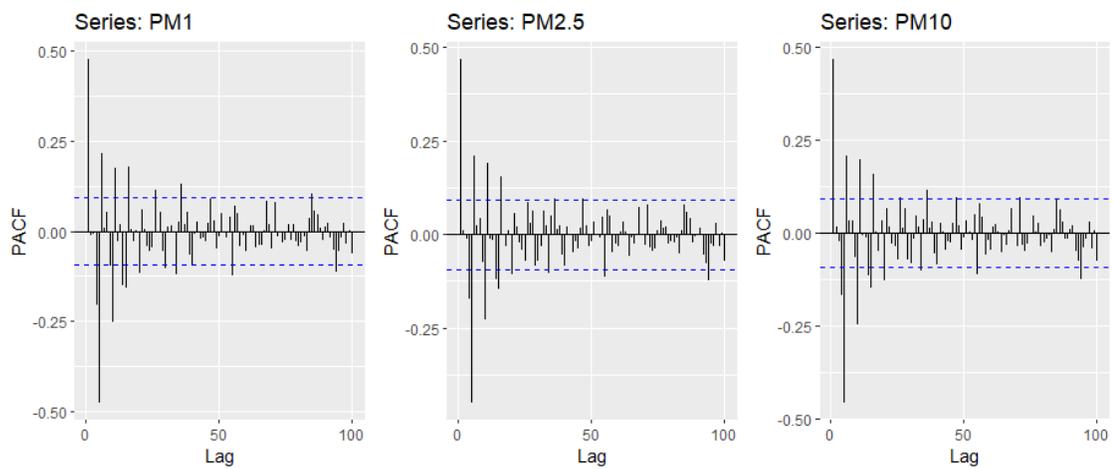
Figura 18. Primera diferencia ACF – PM1, PM2.5 y PM10



Fuente: elaboración propia, realizado con RStudio®.

En el gráfico anterior se observó que ACF por diferencia de medias es significativa donde se identificó el coeficiente MA, para PM1 se obtuvo un valor de  $q = 19$ , para PM2.5 se obtuvo un valor de 19 y para PM10 se obtuvo un valor de 19, (ver figura 18).

Figura 19. **Primera diferencia PACF – PM1**



Fuente: elaboración propia, realizado con RStudio®.

En el gráfico anterior se observó que PACF por diferencia de medias disminuye en forma oscilatoria y se identificó el coeficiente AR, para PM1 se obtuvo un valor de  $p = 12$ , para PM2.5 se obtuvo un valor de  $p = 12$  y para PM10 se obtuvo un valor de  $p=12$ .

- Objetivo 2. Identificar cual debe de ser el nivel de confiabilidad, mediante un análisis de validez para el modelo de predicción.

### 3.1.7. ARIMA

Se obtuvo el modelo ARIMA, basándose en los valores anteriormente obtenidos en donde se variaron los parámetros donde se obtuvo el mejor modelo.

#### 3.1.7.1. Elección del mejor modelo ARIMA

Se determinó, la elección del mejor modelo hizo uso de los criterios de información para seleccionar el mejor modelo y se presentan en la siguiente tabla, a partir de los datos obtenidos anteriormente.

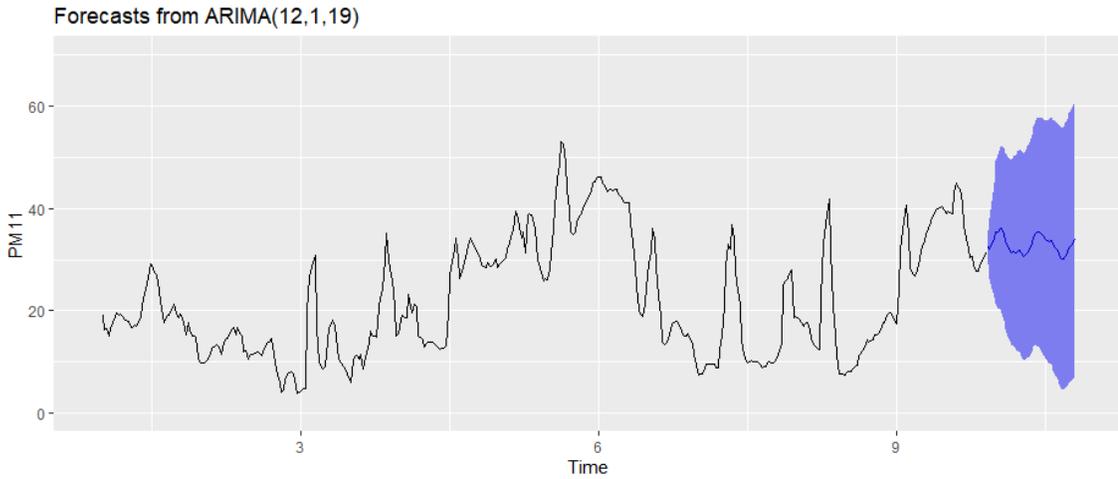
Tabla XVIII. ARIMA – PM1

ARIMA (p, d, q)	AIC	BIC	ECM	MAPE
ARIMA (8,1,0)	1895.23	1932.137	1.987605	6.692896
ARIMA (8,1,19)	1781.39	1896.195	1.634388	5.327885
ARIMA (12,1, 0)	1854.23	1907.529	1.878708	6.366335
<b>ARIMA (12,1, 19)</b>	<b>1776.16</b>	<b>1907.370</b>	<b>1.602427</b>	<b>5.343395</b>

Fuente: elaboración propia.

Se estimó que el óptimo modelo, se ajusta de mejor manera a la serie de datos para PM1 fue ARIMA (12, 1, 19) (ver tabla XVIII). En la gráfica siguiente se observó el pronóstico del modelo un día después.

Figura 20. Pronóstico ARIMA (12, 1, 19) – PM1



Fuente: elaboración propia, realizado con RStudio®.

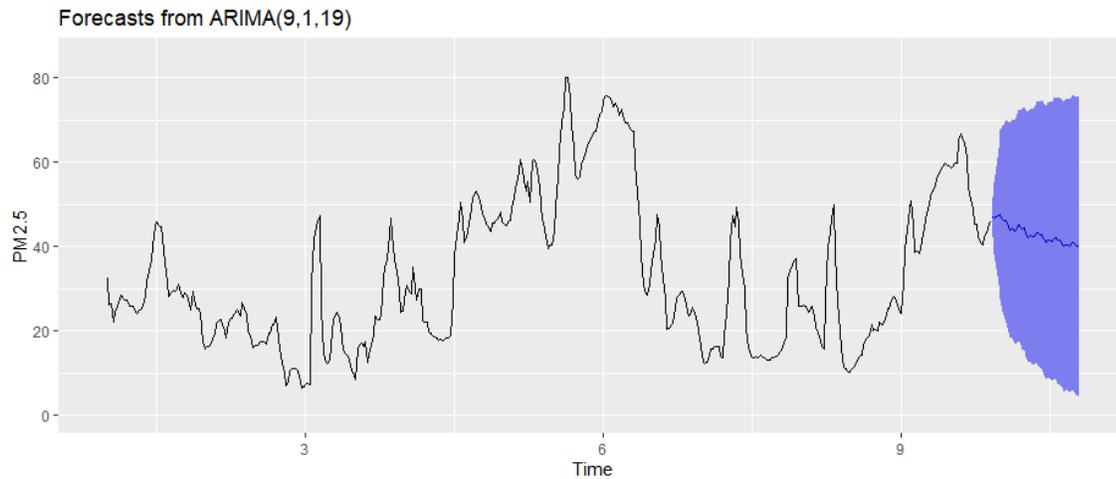
Tabla XIX. ARIMA – PM2.5

ARIMA (p, d, q)	AIC	BIC	ECM	MAPE
ARIMA (9,1,0)	2196.43	2237.438	2.782218	6.550607
<b>ARIMA (9,1,19)</b>	<b>2085.71</b>	<b>2204.616</b>	<b>2.288787</b>	<b>5.214188</b>
ARIMA (12,1,0)	2158.04	2211.341	2.643669	6.225834
ARIMA (12,1,19)	2092.78	2223.991	2.289992	5.205423

Fuente: elaboración propia.

Así mismo se estimó que el mejor modelo que se ajusta de mejor manera a los datos para PM2.5 fue ARIMA (9, 1, 19). En el gráfico siguiente se observó el pronóstico del modelo un día después.

Figura 21. Pronóstico ARIMA (9, 1, 19) – PM2.5



Fuente: elaboración propia, realizado con RStudio®.

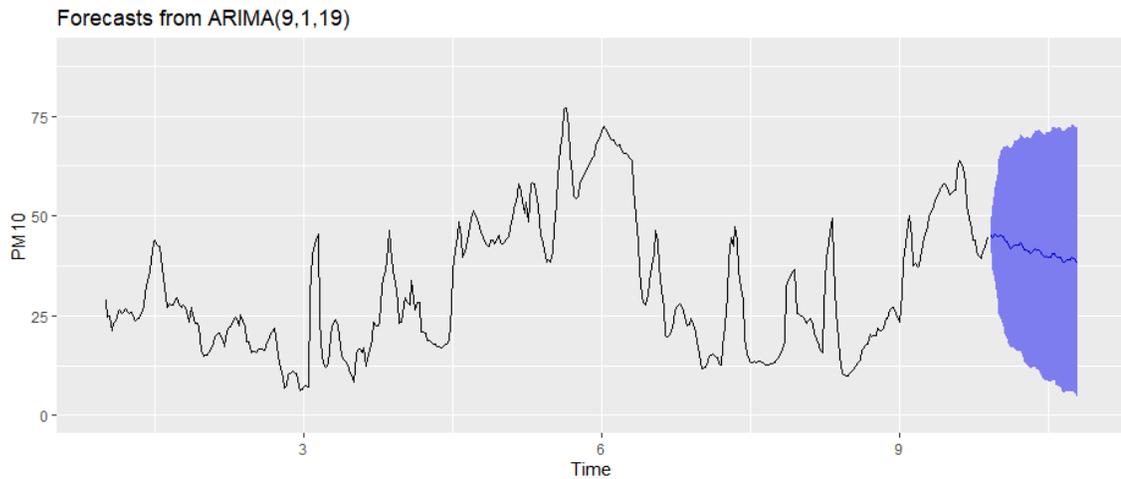
Tabla XX. ARIMA – PM10

ARIMA (p, d, q)	AIC	BIC	ECM	MAPE
ARIMA (9,1,0)	2166.72	2207.726	2.690951	6.49995
<b>ARIMA (9,1,19)</b>	<b>2053.57</b>	<b>2172.484</b>	<b>2.207681</b>	<b>5.189115</b>
ARIMA (12,1,0)	2125.26	2178.561	2.547975	6.178438
ARIMA (12,1,19)	2054.92	2179.126	2.255367	5.196264

Fuente: elaboración propia.

Por último, se estimó que el mejor modelo que se ajusta de mejor manera a los datos para PM10 fue ARIMA (9, 1, 19). En el gráfico siguiente se observó el pronóstico del modelo un día después.

Figura 22. **Pronóstico ARIMA (9, 1, 19) – PM10**

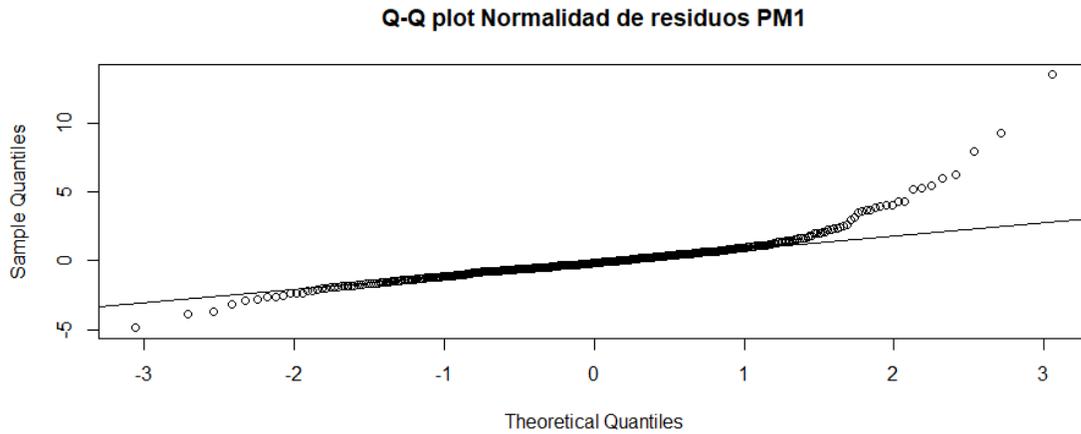


Fuente: elaboración propia, realizado con RStudio®.

### 3.1.8. Normalidad de los datos

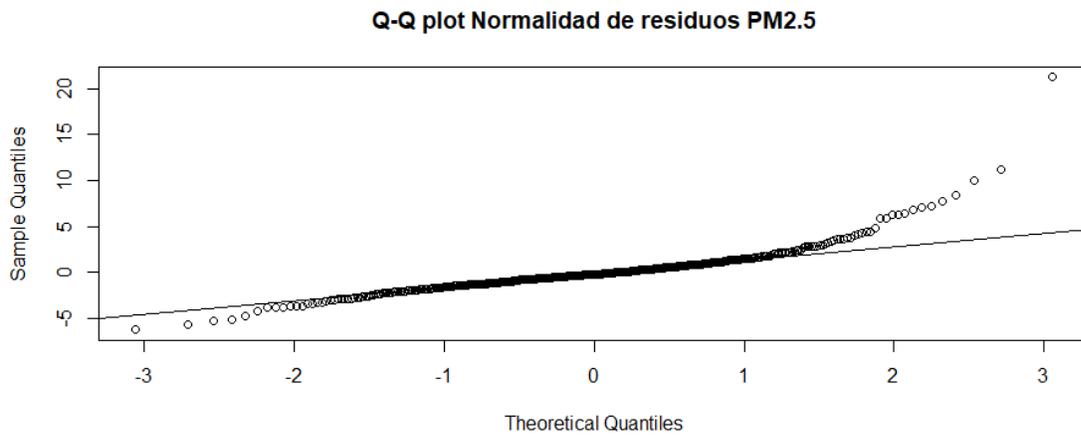
Se realizó el gráfico de Q-Q plot hacia los residuos de los modelos ARIMA, anteriormente estimados, con la finalidad de comprobar la normalidad. Ya que se observó cuan cerca esta la distribución de la serie respecto a la normal. Se observaron que la mayoría de los puntos se localizaron sobre la recta. De manera gráfica se observó una posible distribución normal.

Figura 23. Q-Q Plot - PM1



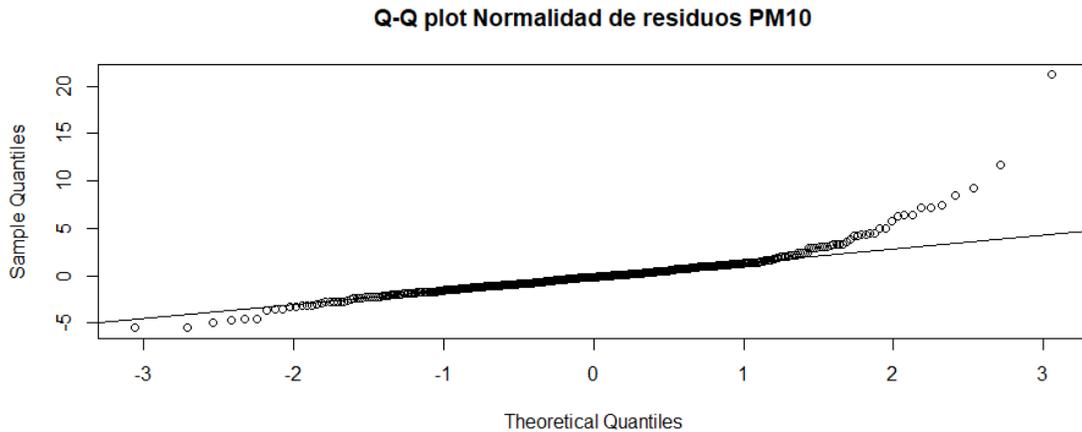
Fuente: elaboración propia, realizado con RStudio®.

Figura 24. Q-Q Plot – PM2.5



Fuente: elaboración propia, realizado con RStudio®.

Figura 25. Q-Q Plot – PM10



Fuente: elaboración propia, realizado con RStudio®.

Para corroborar la normalidad de los residuos a los mejores modelos seleccionados, se aplicó la prueba de Kolmogorov-Smirnov, además para contrastar dichos resultados se aplicó el test de Shapiro-Wilk con la finalidad de determinar si la normalidad de la serie de datos. La hipótesis por plantear para ambas pruebas fue el siguiente:

*H<sub>0</sub>: Posee una distribución Normal*

*H<sub>a</sub>: No posee una distribución Normal*

Si  $p \text{ valor} > 0.05$ , se acepta la hipótesis nula para ambas pruebas, dieron por aceptado que los datos poseen una distribución normal (ver tabla XXI, XXIII, XXV) para Kolmogorov-Smirnov y Shapiro-Wilk.

Tabla XXI. **Test de ajuste de normalidad PM1**

---

***Kolmogorov-Smirnov test***

---

data:		PM1_arima\$residuals
D		.13846

p\_value = 7.484e-6

---

Debido a  $P < .05$ , se rechaza  $H_0$

---

Fuente: elaboración propia.

Tabla XXII. **Test de normalidad PM1**

---

***Shapiro-Wilk test***

---

data:		PM1_arima\$residuals
W		.84513

p\_value = 2.2e-16

---

Debido a  $P < .05$ , se rechaza  $H_0$

---

Fuente: elaboración propia.

Tabla XXIII. **Test de ajuste de normalidad PM2.5**

---

***Kolmogorov-Smirnov test***

---

data:		PM2.5_arima\$residuals
D		.12002

p\_value = 5.25e-06

---

Debido a  $P < .05$ , se rechaza  $H_0$

---

Fuente: elaboración propia.

Tabla XXIV. **Test de normalidad PM2.5**

---

***Shapiro-Wilk test***

---

data:	PM1_arima\$residuals
W	.84231

p\_value = 2.3e-16

---

Debido a  $P < .05$ , se rechaza  $H_0$

---

Fuente: elaboración propia.

Tabla XXV. **Test de ajuste de normalidad PM10**

---

***Kolmogorov-Smirnov test***

---

data:	PM10_arima\$residuals
W	.81066

p\_value = 2.283e-07

---

Debido a  $P < .05$ , se rechaza  $H_0$

---

Fuente: elaboración propia.

Tabla XXVI. **Test de normalidad PM10**

---

***Shapiro-Wilk test***

---

data:	PM1_arima\$residuals
W	.79607

p\_value = 2.2e-16

---

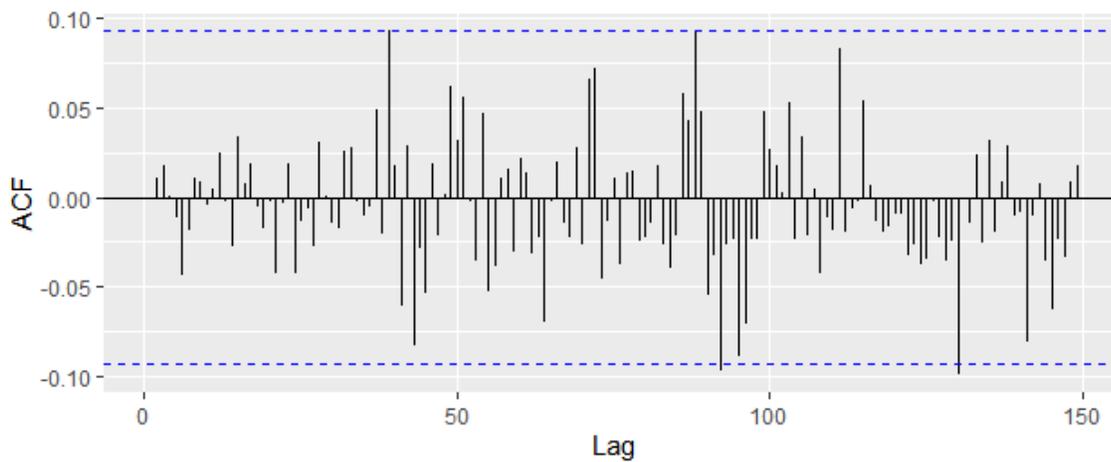
Debido a  $P < .05$ , se rechaza  $H_0$

---

Fuente: elaboración propia.

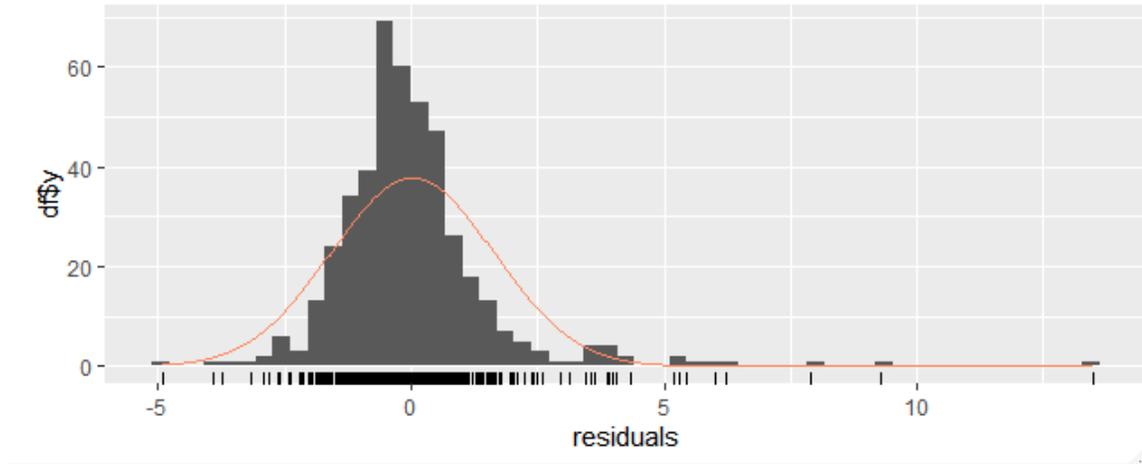
Con base en los resultados anteriores, obtenidos mediante la prueba de Kolmogorov-Smirnov y el test de Shapiro-Wilk se infiere que los modelos ARIMA que fueron estimados no poseen una distribución normal ya que la probabilidad fue menor a la significancia, cabe resaltar que otro factor que se consideró fue el error cuadrático medio ya que tiene como función estimar la desigualdad existente entre el valor real y el valor estimado. Y se consideró una tendencia a la normalidad más que otros modelos.

Figura 26. **Correlograma de residuos – PM1**



Fuente: elaboración propia, realizado con RStudio®.

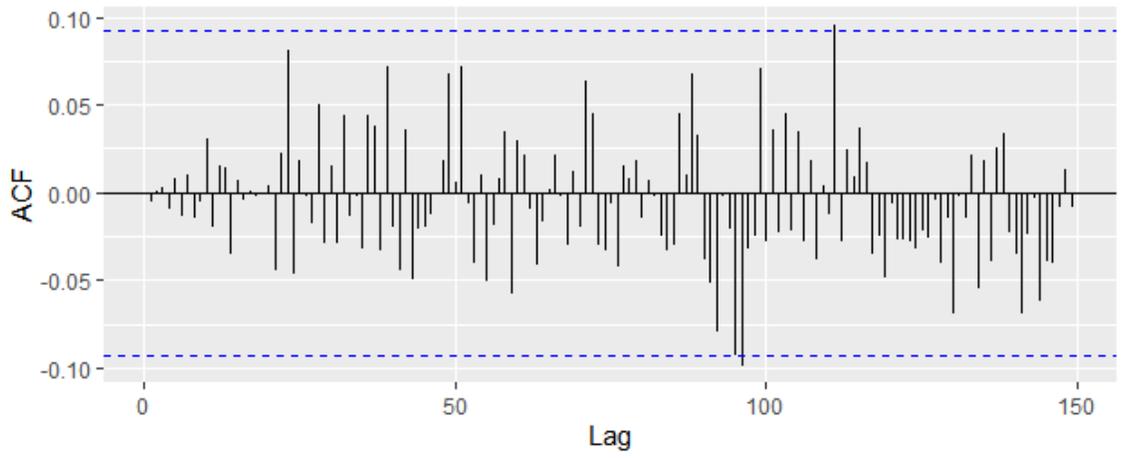
Figura 27. **Histograma de residuos – PM1**



Fuente: elaboración propia, realizado con RStudio®.

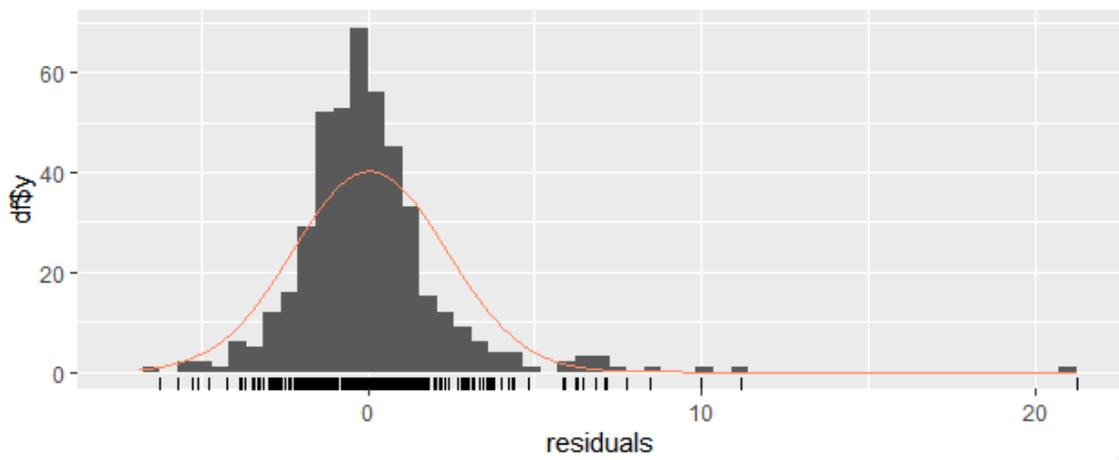
En el gráfico anterior se observa los gráficos de los residuos para el modelo ARIMA – PM1 (ver figura 26). Donde se observa aleatoriedad de los datos como se observa en el gráfico de autocorrelación que se satisface el supuesto de que los residuos son independientes, sin embargo existe 2 correlaciones significativas, estos desfases se atribuyen al error aleatorio y por lo cual no representan una señal de que no se satisface el supuesto, por el último se observó el histograma de residuos (ver figura 27) muestra una distribución para todas las observaciones, donde se percibe que la totalidad de residuos posee una media cercana a cero.

Figura 28. **Correlograma de residuos – PM2.5**



Fuente: elaboración propia, realizado con RStudio®.

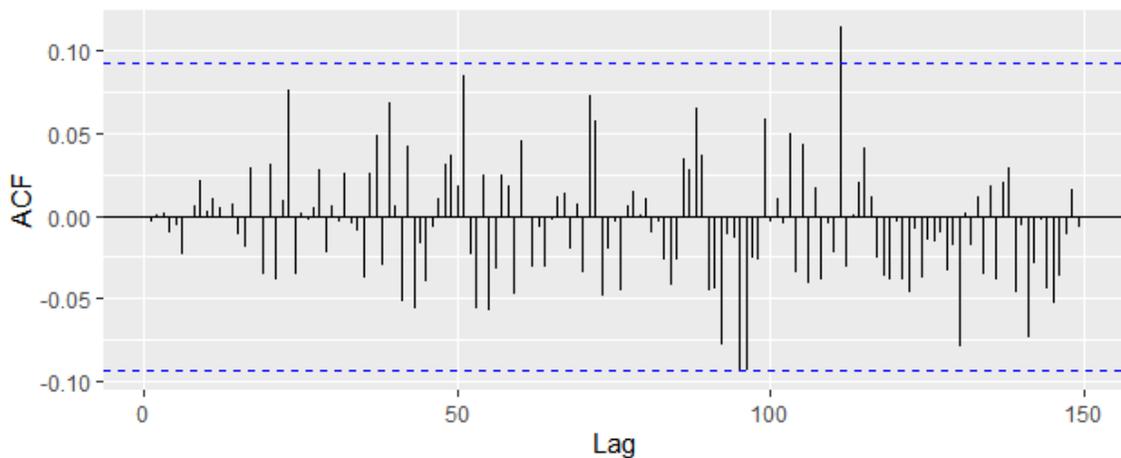
Figura 29. **Histograma de residuos – PM2.5**



Fuente: elaboración propia, realizado con RStudio®.

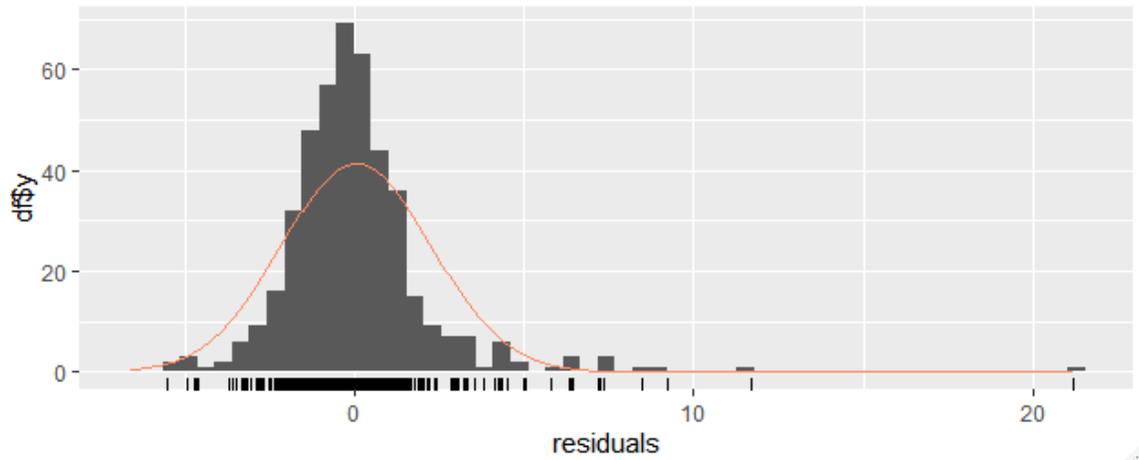
En el gráfico anterior los gráficos de los residuos para el modelo ARIMA – PM2.5. Donde se observa aleatoriedad de los datos como se observa en el gráfico de residuos, la autocorrelación se observó que satisface el supuesto de que los residuos son independientes, sin embargo existe 3 correlaciones son significativas, estos desfaces puede atribuirse al error aleatorio y por lo cual no representan una señal de que no se satisface el supuesto, por el último se observó el histograma de residuos muestra una distribución para todas las observaciones, donde se observa que la mayoría de residuos posee una media cercana a cero.

Figura 30. **Correlograma de residuos – PM10**



Fuente: elaboración propia, realizado con RStudio®.

Figura 31. **Histograma de residuos – PM10**



Fuente: elaboración propia, realizado con RStudio®.

En el gráfico anterior los residuos para el modelo ARIMA – PM10 (ver figura 30). Donde se observa la aleatoriedad de los datos como se observa en el gráfico de residuos, la autocorrelación se observó que satisface el supuesto de que los residuos son independientes, sin embargo existe 2 correlaciones son significativas, estos desfases puede atribuirse al error aleatorio y por lo cual no representan una señal de que no se satisface el supuesto, por el último se observó el histograma de residuos (ver figura 31) muestra una distribución para todas las observaciones, donde se observa que la mayoría de residuos posee una media cercana a cero.

- Objetivo 3. Estimar cuales son los promedios de las concentraciones PM, por medio de intervalos de confianza, presentes en una intersección de carretera de alta afluencia vehicular.

### 3.2. Intervalos de confianza

Con base en las inferencias del modelo ARIMA, los valores anteriormente obtenidos, se calculó el intervalo de confianza para cada concentración.

Se estimó cuáles fueron los niveles de las concentraciones a futuro con base en el pronóstico que se obtuvo con el mejor ARIMA para las concentraciones PM1, PM2.5 y PM10. Para ello se estimaron 45 valores a futuro y con base en estos valores se eligieron las observaciones impares ya que se empleó un intervalo de confianza *t-Student* donde  $\sigma$  es desconocida y  $n < 30$  con una confiabilidad del 95 %.

Donde  $\alpha = .05$ ,  $n = 23$ ,  $v = 22$  y  $t_{.025} = 2.0739$ .

#### 3.2.1. Intervalo de confianza PM1

Se procedió a calcular el intervalo para la concentración de PM1.

Tabla XXVII. Pronóstico PM1

No.	Valor								
1	32.42287	10	33.64813	19	30.48824	28	34.81361	37	31.02848
2	32.08268	11	32.6539	20	31.00645	29	34.55615	38	30.17148
3	32.94919	12	31.83917	21	31.35833	30	33.96068	39	30.14264
4	33.8694	13	31.37951	22	31.93689	31	33.64888	40	30.6487
5	35.28837	14	31.61445	23	33.01799	32	33.49846	41	31.30513
6	35.4688	15	31.39894	24	34.47169	33	33.77268	42	32.17803
7	35.97138	16	31.66801	25	35.16941	34	32.93791	43	32.75153
8	36.01509	17	31.69924	26	35.41473	35	32.30496	44	33.4953
9	35.1217	18	31.33069	27	35.36952	36	31.67633	45	33.83622

Fuente: elaboración propia.

La desviación estándar muestral sobre el pronóstico PM1 fue obtenida a partir de la tabla anterior. La desviación estándar fue  $s = 1.715652$ . El intervalo de confianza sobre el pronóstico de PM1 (ver tabla XXX).

Tabla XXVIII. **Intervalo de confianza PM1**

<b>Intervalo de Confianza concentración PM1</b>
$P\left(32.1987 \frac{\mu g}{m^3} \leq \mu \leq 33.6825 \frac{\mu g}{m^3}\right) = .95$

Fuente: elaboración propia.

### 3.2.2. Intervalo de confianza PM2.5

Se procedió calcular el intervalo para la concentración de PM2.5.

Tabla XXIX. **Pronóstico PM2.5**

No.	Valor								
1	47.02078	10	43.90118	19	42.20078	28	40.88352	37	39.95509
2	46.72894	11	44.25362	20	42.65297	29	41.41124	38	40.33399
3	46.89464	12	43.58887	21	42.40821	30	41.38594	39	40.35652
4	47.18625	13	43.78255	22	42.41337	31	41.12205	40	40.08759
5	47.59185	14	45.11902	23	43.37209	32	41.92741	41	40.84236
6	46.53885	15	44.51675	24	43.09277	33	41.99757	42	41.01549
7	46.09722	16	44.00058	25	42.57348	34	41.38406	43	40.26938
8	46.18273	17	44.35224	26	42.83058	35	41.38232	44	40.18446
9	44.61179	18	43.23704	27	41.87825	36	40.74758	45	39.91222

Fuente: elaboración propia.

La desviación estándar muestral sobre el pronóstico PM2.5 fue obtenida a partir de la tabla anterior. La desviación estándar fue  $s = 2.345872$ . El intervalo de confianza sobre el pronóstico de PM2.5 (ver tabla XXXI).

Tabla XXX. **Intervalo de confianza PM2.5**

<b>Intervalo de Confianza concentración PM2.5</b>
$P\left(41.977 \frac{\mu g}{m^3} \leq \mu \leq 44.005 \frac{\mu g}{m^3}\right) = .95$

Fuente: elaboración propia.

### 3.2.3. Intervalo de confianza PM10

Se procedió calcular el intervalo para la concentración de PM10.

Tabla XXXI. **Pronóstico PM10**

No.	Valor								
1	45.25333	10	42.8563	19	41.35178	28	39.61545	37	38.72047
2	44.48492	11	41.80811	20	41.08145	29	40.03492	38	38.50323
3	45.37422	12	42.32564	21	40.52017	30	39.77357	39	39.06932
4	44.85926	13	42.8503	22	41.26113	31	39.75318	40	39.08363
5	44.82748	14	42.70556	23	41.79974	32	40.41522	41	38.9807
6	45.30997	15	42.79458	24	41.37918	33	40.69675	42	39.57256
7	44.80049	16	43.39275	25	41.38492	34	39.89563	43	39.28987
8	43.81804	17	42.32083	26	41.30036	35	39.78286	44	38.5353
9	43.57901	18	41.45526	27	40.34028	36	39.58509	45	38.32089

Fuente: elaboración propia.

La desviación estándar muestral sobre el pronóstico PM10 fue obtenida a partir de la tabla anterior. La desviación estándar fue  $s = 2.193124$ . El intervalo de confianza sobre el pronóstico de PM10.

Tabla XXXII. **Intervalo de confianza PM10**

---

**Intervalo de Confianza concentración PM10**

---

$$P\left(40.5148 \frac{\mu g}{m^3} \leq \mu \leq 42.4116 \frac{\mu g}{m^3}\right) = .95$$

---

Fuente: elaboración propia.

## **4. DISCUSIÓN DE RESULTADOS**

A continuación, se expone la discusión de resultados, los que anteriormente fueron obtenidos de los 3 diseños experimentales y que se presentaron en el apartado anterior. Se seleccionaron los modelos con base en los criterios de información, los modelos propuestos se analizaron los residuos y no satisficieron por medio de los supuestos de normalidad, cabe resaltar que se infiere que tiende a una distribución de normalidad con base en el error cuadrático medio ya que el error es muy pequeño.

### **4.1. Análisis interno**

En los siguientes incisos, se describe el análisis interno realizado.

#### **4.1.1. Modelo ARIMA PM1, PM2.5 y PM10**

Con base en las gráficas (ver figura 20, 21 y 22) del pronóstico para observar qué modelo se ajusta mejor a los datos. Se procede a interpretar cada uno de los modelos ARIMA.

#### **4.1.2. PM1**

Para PM1, se estima que el modelo fue un ARIMA (12, 1, 19), donde la AR (12) es la parte autorregresiva de orden 12, MA (19) corresponde a la media móvil de orden 19 y para I (1) es la parte integrada. Se observó la gráfica (ver figura 20) donde no presenta alguna tendencia lo que significa que el nivel de las concentraciones para este modelo se mantienen en un rango constante, según

el resultado obtenido y al compararlo con la tabla de calidad del aire al no tener un rango específico para PM1 se contrastará con los resultados de PM2.5, el rango según ICA representa el color amarillo y su clasificación es moderada, quiere decir que el nivel de concentraciones en el periodo de 6PM a 7PM se consideran moderadas, se sugiere utilizar algún equipo protector al transitar por esta intersección de carretera.

#### **4.1.3. PM2.5**

Para PM2.5, se estima que el modelo fue un ARIMA (9, 1, 19), donde la AR (9) es la parte autorregresiva de orden 9, MA (19) corresponde a la media móvil de orden 19 y para I (1) es la parte integrada. Se observó el grafico (ver figura 21) donde presenta alguna tendencia negativa lo que significa que el nivel de las concentraciones para este modelo variarán dependiendo del día de la semana, según el resultado obtenido y al compararlo con la tabla de calidad del aire, el rango según ICA representa el color naranja y su clasificación es dañina para grupos sensibles, quiere decir que el nivel de concentraciones en el periodo de 6PM a 7PM se consideran dañinas, se sugiere utilizar algún equipo protector al transitar por esta intersección de carretera.

#### **4.1.4. PM10**

Para PM10, se estima que el modelo fue un ARIMA (9, 1, 19), donde la AR (9) es la parte autorregresiva de orden 9, MA (19) corresponde a la media móvil de orden 19 y para I (1) es la parte integrada. Se observo el grafico (ver figura 22) donde presenta alguna tendencia negativa lo que significa que el nivel de las concentraciones para este modelo variarán, esto dependerá del día de la semana, según el resultado obtenido y al compararlo con la tabla de calidad del aire, el rango según ICA representa el color verde y su clasificación es buena, el

nivel de concentraciones en el periodo de 6PM a 7PM se consideran normales, y que no representa ningún riesgo a la salud, se aconseja el uso de algún tipo de protección ya que están presentes los otros dos tipos de concentraciones en la intersección de carretera.

Para este estudio la cifra de las partículas PM10 se mantiene en un rango aceptable por otro lado poseen una media simultánea con las partículas PM2.5 al compararlos con los resultados de Martínez.

#### **4.1.5. Normalidad de los datos**

El resultado del modelo ARIMA para las concentraciones PM1, PM2.5 y PM10, en los residuos de los modelos se observó una tendencia a una distribución normal. En los gráficos de residuos Q-Q *plot* (ver gráficos 23, 24 y 25) se observó que los residuos tienden a una distribución normal, no obstante, se determinó mediante la prueba estadística de normalidad *Kolmogorov-Smirnov* (ver tablas XXI, XXIII y XXV) y test *Shapiro-Wilk* (ver tablas XXII, XXVI y XXVI), p valor para ambas pruebas es menor que la confiabilidad establecida, por lo tanto, los datos no poseen una distribución normal. Por último, se estimó que los datos tienden a la normalidad ya que el ECM es muy bajo.

#### **4.1.6. Intervalos de confianza ARIMA**

Una de las limitaciones que se obtuvo para este estudio fue la manera en que se muestrearon los datos ya que era necesario tener la computadora a la vista junto con el sensor en una intersección de carretera y esto a su vez representaba cierto peligro. Al no ser modular el sistema de muestreo, no permitía obtener un mayor número de muestras en un tiempo mayor a una hora.

## 4.2. Análisis externo

Se podría afirmar, que las partículas PM10 al compararlos con los resultados obtenidos son cercanos a 50  $\mu\text{g}/\text{m}^3$  según los resultados de Alonso *et al.*, (2005) estos resultados son alarmantes ya que representan 1.4 muertes prematuras por cada 100,000 habitantes al año y sus efectos a corto plazo representan 2.8 muertes por cada 100,000 en un periodo de hasta 40 días. Teniendo esto en cuenta se puede decir que las concentraciones PM10 posee un resultado similar a excepción que este tipo de partículas son mayormente dañinas a la exposición y las concentraciones PM2.5 son de mayor alerta ya que su concentración es catalogada como dañina para grupos sensibles.

Los resultados obtenidos mediante el intervalo de confianza (ver tabla XXX, XXXII y XXXIV), se está de acuerdo con lo encontrado con la investigación de Contreras (2015) ya que es necesario controlar de manera eficiente los niveles de concentración, se sugiere replicar el estudio obteniendo un mayor número de muestras para obtener un mejor resultado aplicado a diferentes horas del día.

El resultado obtenido mediante intervalos de confianza es similar a lo encontrado en resultados previos con la página AQICN, podría ser considerado realizar un muestreo de un minuto cada medio segundo o incluso cada 15 segundos para observar el comportamiento de las concentraciones en el aire. Ya que representa una contaminación medio ambiental y dañina a la salud de los individuos que transitan por este punto.

## CONCLUSIONES

1. Para las partículas PM1 el modelo que mejor se ajusta es un ARIMA (12,1, 19), para PM2.5 un ARIMA (9, 1, 19) y para PM10 ARIMA (9, 1, 19).
2. Según los criterios de información AIC, BIC, ECM, MAPE y la evaluación del mejor modelo para cada concentración, la autocorrelación y la autocorrelación parcial, los modelos seleccionados presentan un nivel de confiabilidad aceptable.
3. La intersección de carretera presenta, con un 95 % de confianza, una concentración de  $32.19 \mu\text{g}/\text{m}^3 - 33.68\mu\text{g}/\text{m}^3$  de partículas PM1, de  $41.97 \mu\text{g}/\text{m}^3 - 44.01 \mu\text{g}/\text{m}^3$  para PM2.5 y de  $40.51 \mu\text{g}/\text{m}^3 - 42.41 \mu\text{g}/\text{m}^3$  de partículas PM10. Son los niveles catalogados como moderada, dañina para grupos sensibles y buena, respectivamente.
4. Se diseñó un modelo que se ajusta mejor a la serie de datos para cada nivel de concentración PM mediante las gráficas de autocorrelación y autocorrelación parcial, se determinaron los posibles modelos y los criterios para la selección del modelo.



## RECOMENDACIONES

1. Realizar un muestreo con reducción en el tiempo entre cada muestra, para observar cambio gradualmente para evitar las fluctuaciones entre datos, además aumentar el tiempo de muestreo antes y después de la hora indicada para observar el comportamiento de las partículas en el lugar establecido.
2. Introducir un diseño experimental, la covariable de temperatura para considerar si es un factor que influye y la covariable del tiempo de cada semáforo ya que al haber una aglomeración de vehículos puede afectar los niveles de cada concentración.
3. Realizar un intervalo de confianza haciendo uso de estadística no paramétrica para observar si existe algún cambio en los niveles PM.
4. Emplear otro tipo de pronóstico, un modelo SARIMA, Holt para realizar una comparación entre modelos y determinar cuál se ajusta mejor.



## REFERENCIAS

1. Acuerdo Gubernativo Número 68-86. Ley de protección y mejoramiento del medio ambiente. Diario de Centroamérica. Guatemala. 28 de noviembre de 1986.
2. Amarís, G., Ávila, H. y Guerrero, T. (enero, 2017). Aplicación de modelo ARIMA para el análisis de series de volúmenes anuales en el río Magdalena. *Revista Tecnura*, 21(52), 88-101. Recuperado de [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-921X2017000200007](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-921X2017000200007).
3. Ballester, F., Tenias, M. y Pérez, S. (marzo, 1999). Efectos de la contaminación atmosférica sobre la salud: una introducción. *Rev. Esp. Salud Pública*, 73(2), 109-121. Recuperado de <https://www.scielosp.org/article/resp/1999.v73n2/109-121/>
4. Box, J., Jenkins, G. y Reinsel, G. (2008). *Time Series Analysis: Forecasting and Control*. Estados Unidos: Wiley.
5. Carrión, A. (2001). *Análisis de series temporales, técnicas y prevención*. España: Universidad Politécnica de Valencia.
6. Chen, J., Ying, Q., y Kleeman, M. (enero, 2009). Source apportionment of visual impairment during the California regional PM10/PM2.5 air quality study. *Atmospheric Environment*, (43), 6136-6144. Recuperado de <https://pubs.acs.org/doi/pdf/10.1021/es0349305>.

7. Contreras, L. (2015). *Predicción e interpolación dinámica de los niveles de contaminación atmosférica mediante datos de intensidad de tráfico y dirección del viento* (Tesis de maestría). Universidad Politécnica de Valencia, España. Recuperado de <https://1library.co/document/yeo9n07q-prediccion-interpolacion-dinamica-contaminacion-atmosferica-mediante-intensidad-direccion.html>.
8. Dickey, D. y Fuller, W. (junio, 1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 76, (366) 427-431. Recuperado de <https://www.jstor.org/stable/2286348>.
9. Alonso, E., Martínez, T., Cambra, K., Lopez, L., Boldo, E., Zorrilla, B., Daponte, A., Aguilera, I., Toro, S., Iñiguez, C., Ballester, F., García, F., Plasencia, A., Artazcoz, L. y Medina, S. (marzo, 2005). Evaluación en cinco ciudades españolas del impacto en salud de la contaminación atmosférica por partículas proyecto europeo APHEIS. *Revista Española de Salud Pública*, 79 (2), 297-308. Recuperado de <https://www.redalyc.org/articulo.oa?id=17079217>.
10. García, M., Arias, A., Rodríguez, C., Morcillo, A. y Aguirre, A. (febrero, 2011). Análisis de series temporales aplicado a intoxicaciones atendidas en urgencias hospitalarias. *Revista de la Sociedad Española de Medicina de Urgencias y Emergencias*. 100 (3), 193-199.
11. García, M., Prusakov, A., Martínez, E. y Carbonell, I. (enero, 2016). Tendencias y pronósticos de la hipertensión arterial en la provincia

de Santiago de Cuba. *Medisan*. 20 (4), 433 – 443. Recuperado de <https://www.redalyc.org/articulo.oa?id=368445189003>.

12. Gras, J. (2001). *Diseño de series temporales: técnicas de análisis*. España: Editorial Edicions Universitat de Barcelona.
13. Hacker, R. y Hatemi, A. (2010). *The Properties of Procedures Dealing with Uncertainty about Intercept and Deterministic Trend in Unit Root Testing*. Suecia: CESIS.
14. Hamilton, J. (1994). *Time Series Analysis*. Reino Unido: Princeton University Press.
15. Jiménez, J., Gázquez, J. y Sánchez, R. (febrero, 2006). La capacidad predictiva en los métodos Box-Jenkins y Holt-Winters: una aplicación al sector turístico. *Revista Europea de Dirección y Economía de la Empresa*, 15 (3), p. 185-198.
16. Martínez, P. (2005). *Caracterización de la contaminación atmosférica producidas por partículas en suspensión en Madrid*. (Tesis doctoral). Universidad Complutense de Madrid, España. Recuperado de <https://eprints.ucm.es/id/eprint/5402/1/T27355.pdf>.
17. Mauricio, J. (2007). *Introducción al Análisis de Series Temporales*. España: Universidad Complutense de Madrid.
18. Montero, J. (2007). *Estadística Descriptiva*. Madrid, España: Editorial Thomson.

19. Organización Mundial de la Salud. (2005). *Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y dióxido de azufre*. Suiza: Autor. Recuperado de [https://apps.who.int/iris/bitstream/handle/10665/69478/WHO\\_SDE\\_PHE\\_OEH\\_06.02\\_spa.pdf;jsessionid=47D693FB4C842C598D4F252415F0F1BB?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/69478/WHO_SDE_PHE_OEH_06.02_spa.pdf;jsessionid=47D693FB4C842C598D4F252415F0F1BB?sequence=1)
  
20. Siegel, S. y Castellan, N. (1995). *Estadística no paramétrica aplicada a las ciencias de la conducta*. México: Editorial: Trillas.
  
21. Vázquez, C., Beristain, F. y Audelo, E. (julio, 2012). Caracterización del patrón de comportamiento de las partículas en suspensión en el Valle de México. *Información Tecnológica*, 23 (3), 3-10. Recuperado de [https://scielo.conicyt.cl/scielo.php?pid=S0718-07642012000300002&script=sci\\_arttext](https://scielo.conicyt.cl/scielo.php?pid=S0718-07642012000300002&script=sci_arttext).
  
22. Walpole *et al.*, (2012). *Probabilidad y estadística para ingeniería y ciencias*. México: Editorial Pearson.

## APENDICES

### Apéndice 1. **Matriz de coherencia para los objetivos**

Elementos del problema	PREGUNTAS DE INVESTIGACIÓN
01. Se desconoce los niveles de las concentraciones PM presentes en una intersección de carretera durante la noche.	01. ¿Cuáles son los niveles de las concentraciones PM presentes en una intersección de carretera durante la noche?
02. Se desconoce si existe un modelo que permita estimar las concentraciones de partículas PM	02. ¿Cuál debe de ser el modelo de series temporales que permita estimar las concentraciones de las partículas PM?
03. Se desconoce el nivel de validez y confiabilidad del modelo.	03. ¿Cuál debe de ser el nivel de confiabilidad del modelo de pronóstico?
Se desconoce un modelo y el nivel de validez que permita estimar los niveles de las concentraciones PM presentes en una intersección de carretera durante la noche,	¿Cuál debe de ser el modelo de series temporales que permita estimar las concentraciones PM y a su vez conocer el nivel de validez?

Fuente: elaboración propia.

## Apéndice 2. Suavizamiento por medias móviles PM1

No.	Valor														
	1 – 15	NA	NA	19.2	16.4	16.6	15	16.6	17.8	19	19.6	19.2	19.4	19	18.2
16 – 30	18	16.8	16.8	17.2	17	17.6	18.6	21	23.4	25	26.6	29.2	29	27.6	27
31 – 45	25.6	22	19.2	17.6	18.8	19.2	19.8	20.6	21.2	19.6	18.6	18.8	19.6	17.8	16.4
46 – 60	35	34	31.2	31.4	31	10.8	9.8	9.8	9.8	10	10.6	11.4	12.6	13	13.4
61 – 75	13.2	12.4	11.4	13.8	14.6	14.6	15.6	16.6	16.8	15.4	16.8	15.6	15	12	12.2
76 – 90	10.6	11.2	11.4	11.4	11.8	12	11.8	11.2	12.8	13	14	14	14.6	11.8	9.8
91 – 105	7.4	6	4.2	4.6	6.6	7.8	8	8.2	7.6	5.4	4	4.2	4.4	4.8	4.8
106 – 120	35.6	42.4	44.2	45.2	46.6	16.6	10.2	9	8.6	9.2	14.2	16.6	17.4	18.2	17
121 – 135	12.8	10.6	9.8	9.2	8.8	7.6	7.2	6	10.4	11	11.2	10.6	11.6	8.6	9.8
136 – 150	11.8	13.6	16	15.2	15	14.8	20.8	22.6	24.2	30	35.2	29.8	27	26	20.8
151 – 165	15.2	15.6	18.2	19.2	18.6	18.6	23.2	20.8	19.6	21.2	20.8	15.2	14.8	14.4	13
166 – 180	13.6	13.8	13.8	13.8	13.6	13.2	12.6	12.4	12.8	12.6	13.2	19.6	26.8	29.6	32
181 – 195	34.2	30.4	26.2	27.8	29.4	30.4	32.8	34.2	33.6	32.8	32	31.2	30.2	29	28.8
196 – 210	28.4	29.4	28.8	28.8	29.2	30.2	28.4	29.2	29.6	30	30.4	32	33.2	35.4	35.6
211 – 225	39.4	38	36.4	34.2	35.4	31.2	52.4	52.6	52.4	50.6	49.8	30	29.2	27.6	25.8
226 – 240	26.4	26	28.2	31.4	35.8	40.4	45.6	48.4	53	52.6	49	43.6	39.6	35.4	34.8
241 – 255	35.4	37.2	38.4	39	40	41	41.4	42.6	43.2	44.8	45.2	46	46.2	46.2	45
256 – 270	44.4	43.2	43.4	43.6	43.4	43.6	43.8	42.8	42.2	41.6	41.2	41	41	37.2	33.2
271 – 285	28.8	24.6	20.4	19.4	19	21.2	24.6	29.2	31.6	36	34	28.8	23	19.2	14
286 – 300	13.4	14	14.6	16.4	17.4	17.6	18	17.6	16.8	15.4	15.2	15.2	15.6	14.4	13.6
301 – 315	11.6	9.4	7.4	7.6	7.6	8.2	9.6	9.6	9.6	9.6	9.6	8.8	9	12.4	15.2
316 – 330	20.8	28.8	33	32	36.8	34.2	27.4	24	21.4	14.2	11.4	10.2	9.8	10	10.4
331 – 345	10	10	10	9.6	9	9.2	9.2	9.8	10	9.8	9.8	10.4	11.2	12.2	13.6
346 – 360	24.8	25.8	26.4	27.2	28	18.6	18.8	18.6	18.4	17.8	17	17.4	17.6	16.2	14.6
361 – 375	13.6	13.2	12.6	12.4	34.4	44.8	48.6	50.6	53.6	31.8	21	15.2	11.2	7.6	7.6
376 – 390	7.6	7.4	8	8.2	8.2	8.6	9.2	9.4	11.2	11.8	12.2	13	14.4	13.8	14.2
391 – 405	14.4	15.4	15.4	15.6	16.2	17.6	17.8	19.2	19.6	19.6	18.8	18.2	17.4	34.6	42
406 – 420	46	49.4	50.8	35	28.4	27.2	26.8	27.2	29	31.6	32.6	33.6	35	36	37.2
421 – 435	38.2	39	39.8	40	40.2	40.4	39.6	39	39.4	39.2	39	43.2	44.8	44.2	43.6
436 – 450	41	36.6	34.2	32.2	30.4	30.6	29	27.8	27.8	29	29.6	30.6	31.4	NA	NA

Fuente: elaboración propia.

### Apéndice 3. Suavizamiento por medias móviles PM2.5

No.	Valor														
1 – 15	NA	NA	32.4	26	26.4	22	24.4	26	27.6	28.4	27.6	27.2	27.4	26	25.8
16 – 30	25.8	24	24	25	25	26	28.2	32	34.4	37	40	44.8	45.8	45	44.8
31 – 45	42	36	31.4	28	29	30	29.4	30	31	29	27.8	29	28.4	26.4	25
46 – 60	29.4	28	25.4	25	24.2	17	15.8	16	16.2	16.6	18	19.2	21.8	22.4	22.8
61 – 75	21.4	20	18.2	22	23	23	24.2	25	24.8	23.4	26.8	25.2	24	19.8	19
76 – 90	16	16	16.6	17	17.6	18	17.6	17	19.6	19.6	21.4	21.6	23.2	18.4	15.6
91 – 105	12.4	10	7	7.8	10.6	11	11.2	11	10.6	7.8	6.4	6.8	7.4	7.6	7.4
106 – 120	31	42	44.6	46	47.4	25	14.8	13	12.2	13	19.8	22.8	23.8	24.4	23.2
121 – 135	18.4	15	14.6	14	13	10	10.2	8.4	15.6	16.6	17.2	16.2	17.6	12.6	14.4
136 – 150	16.6	19	23.4	23	22.8	24	32.6	35	36.8	42.8	46.8	39.6	35	34.2	29.4
151 – 165	24.4	25	29.2	31	29.4	29	35.2	30	27.4	30	29.8	22.4	22.2	22.2	19.8
166 – 180	19.6	19	18.6	18	17.8	18	17.8	18	18.4	18.4	19	27.4	37.6	42.8	46.6
181 – 195	50.6	46	40.8	42	45.2	47	51.2	53	53	51.6	50.8	48	46.8	45.2	44.6
196 – 210	43.6	46	45.6	46	46.6	48	45.8	45	45	45.8	46	49	51.6	55.2	55.8
211 – 225	60.6	58	55.6	53	55.4	51	60.2	61	59.8	56.2	54	46.2	45.8	42.4	39.4
226 – 240	40.4	40	43.4	49	56.2	63	69.4	73	80	80	75.2	68.6	63.4	57.2	56
241 – 255	56.6	59	61	63	63.8	65	65.4	67	67.2	68.8	71.2	72	74.4	75.6	75.6
256 – 270	75.4	74	73.2	74	72.6	71	72.4	71	69.4	69.2	68.8	67.6	67.2	59.8	52.8
271 – 285	45.6	38	31.2	30	28.6	31	34.4	38	41.4	47.6	44.6	38.4	32.4	27.6	20.4
286 – 300	20.6	21	21.8	25	28	29	29.2	29	27.6	24.8	23.4	24.4	25.6	24	22.2
301 – 315	20.2	16	12.8	12	12.6	13	15.8	16	16.2	16.2	16.2	14	13.6	19	24.2
316 – 330	32.2	41	47.2	45	49.4	46	38.2	34	30.2	20.4	16	14.2	13.8	13.6	14
331 -345	13.8	14	14	14	13.2	13	13	14	13.6	14	14.2	15	16	17.2	18.6
346 - 360	32.6	34	35.4	37	37.2	26	26	26	26.2	26	24.4	24.8	25.8	23.6	21
361 - 375	19.6	19	16.4	16	27.6	40	43.6	46	50	38.4	26.6	20	15.8	11.4	11.2
376 - 390	10.4	10	10.8	11	12.2	13	14	14	17	17.8	18.4	19	21.4	20.2	20.4
391 - 405	20.2	22	21.6	22	22.6	25	25.4	27	28.2	28	26.2	25.4	24	31.2	39.6
406 - 420	43.8	49	50.8	45	38.6	39	38.2	39	43	46.2	47.8	49.8	52	53	54.2
421 - 435	56	57	58.6	59	59.8	59	59.2	59	58.8	59.6	59.6	65.2	66.6	65.8	64.8
436 - 450	60.8	54	51	49	45.2	45	42.4	41	40.2	42.4	43.2	45.4	45.8	NA	NA

Fuente: elaboración propia.

Apéndice 4. Suavizamiento por medias móviles PM10

No.	Valor														
	NA	NA	29	25	24.8	21	23	24	25.8	26.6	25.6	25.8	26.8	25.8	25.6
1 – 15															
16 – 30	25.8	24	23.8	24	24.4	25	27.4	31	33.8	36	39	43.2	44	42.6	42.4
31 – 45	39.6	34	29.8	27	28	28	27.8	29	29.6	27.6	27	27.8	27	25	23.4
46 – 60	27	25	23.2	23	22.4	16	14.8	15	15	15.6	16.8	17.8	19.6	20.4	20.6
61 – 75	19.8	19	17.2	21	22.2	22	23.6	25	24.2	22.4	25.2	23.6	22.6	18.6	18.4
76 – 90	15.6	16	16	16	16.6	17	16.6	16	18.6	18.8	20.6	20.8	22	17.6	14.8
91 – 105	11.8	9.6	6.8	7.4	10.2	11	11	11	10.4	7.6	6.2	6.6	7	7.4	7.2
106 – 120	30.2	40	43	44	45.6	24	14.4	12	12.2	13	19.6	22.4	23.4	24	22.6
121 – 135	17.6	15	14	13	12.6	10	10.2	8.4	15.2	16.2	16.8	15.6	17	12.4	14.2
136 – 150	16.4	19	23.4	23	22.6	23	32.2	34	36.2	42.2	46.4	39	34.4	33.4	28.4
151 – 165	23	23	27.8	29	28.2	28	33.8	29	26.6	28.4	28.2	20.8	20.8	20.6	18.8
166 – 180	18.8	18	18	18	17.4	17	17	17	17.8	18	18.8	27	36.2	41.2	44.8
181 – 195	48.6	45	39.6	41	44	45	49.4	51	50.6	49.2	48.4	46	44.6	43.6	42.8
196 – 210	42.4	44	43.8	43	44	45	43.2	43	43.6	44.2	44.8	47	49.4	53.2	53.4
211 – 225	58	56	53.8	51	53.4	48	58	58	58	54.4	52.4	45	44.4	41.4	38.4
226 – 240	39	38	41.2	47	53	60	67	71	76.8	77	72.4	65	60.6	55	54.4
241 – 255	55	58	59.6	61	61	62	62.8	64	65.2	67.4	69.2	70	71.6	72.4	71.8
256 – 270	71	70	69	69	68	67	68	67	65.8	65.6	65.4	64.6	64	56.8	50.2
271 – 285	43.2	36	30.2	28	27.8	30	33.4	38	40.6	46.4	43.6	37.6	31.4	27	20
286 – 300	19.6	20	20.8	24	26.4	27	28	28	26.2	23.8	22.6	22.8	24.4	22.8	21
301 – 315	18.6	15	11.8	12	12	13	14.8	15	15.4	14.8	14.4	13	12.6	17.4	22.4
316 - 330	30	39	44.6	42	47.4	44	36.4	32	28.8	19.4	15.4	13.8	13.4	13.4	13.6
331 -345	13.2	14	13.6	13	12.8	13	12.6	13	13	13.4	13.8	14.4	15.6	16.8	18.4
346 - 360	32.2	34	35	36	36.6	25	25.4	25	24.8	24.4	23	23.6	24.4	23	20.6
361 - 375	19.2	18	16.2	16	27.4	39	43.4	46	49.6	37.8	25.8	19	14.8	10.4	10.2
376 - 390	9.8	9.8	10.4	11	11.8	12	13.4	14	16.2	17	17.6	18	20.4	19.6	20
391 - 405	20	22	21.2	21	22	24	24.4	26	27.2	27	25.6	24.8	23.4	30.4	38.8
406 - 420	43	48	50.2	45	37.6	38	37.2	38	41.2	44.4	45.6	47.4	50.2	51	52.4
421 - 435	53.8	55	56.6	57	58	58	57	55	55.6	56.2	56.4	61.6	64	63.6	62.2
436 - 450	58.2	52	49.4	47	44	44	41.2	40	39.2	41.2	42.2	44.4	45	NA	NA

Fuente: elaboración propia.