

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemática

DETERMINACIÓN DE UN MODELO DE REGRESIÓN NO
PARAMÉTRICA MÚLTIPLE PARA ESTIMAR LA CONCENTRACIÓN
HORARIA DE MATERIAL PARTICULADO $PM_{2.5}$ y PM_1 EN LA CIUDAD
DE GUAYAQUIL.

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Ingeniería en ESTADÍSTICA

Presentado por:

Mariuxi Marianela Rizzo Calderón

GUAYAQUIL - ECUADOR

Año: 2019

DEDICATORIA

El presente proyecto se lo dedico a mi hija que ha sido mi fuente de motivación e inspiración para poder superarme cada día más. A mi madre, que con sus palabras de aliento no me dejaba decaer, que me impulsa siempre a seguir adelante y me motiva a que cumpla mis ideales, siempre guiados de la mano de Dios. A mi familia en general por apoyarme en todo momento, a mis amigos que compartieron conmigo momentos de alegría y tristeza, en general a todas las personas que han estado apoyándome para hacer este sueño realidad.

AGRADECIMIENTOS

Mi más sincero agradecimiento primero a Dios por guiarme y acompañarme en todo momento, gracias a mi familia por apoyarme en cada decisión tomada a lo largo de este camino, a mis compañeros y amigos que me impulsaban a seguir adelante. Agradezco especialmente a mis formadores, que con sus enseñanzas me han permitido desarrollarme profesionalmente y me han brindado sus conocimientos, un merecido agradecimiento a mi Tutora Mgtr. Heydi Roa que ha sido una pieza fundamental en este proyecto. Gracias a todos los que de una u otra manera han puesto un granito de arena por brindarme su apoyo incondicional y por estar siempre ahí.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; *Mariuxi Rizzo Calderón* doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



Mariuxi Rizzo Calderón

EVALUADORES



PHD. Sandra García
PROFESOR DE LA MATERIA



Mgtr. Heydi Roa
PROFESOR TUTOR

RESUMEN

Entre octubre de 2016 y marzo del 2017, en la ciudad de Guayaquil, Ecuador, se recolectaron en tiempo real muestras de material particulado correspondientes a $PM_{2.5}$ y PM_1 (PM por sus siglas en inglés) usando un equipo gravimétrico llamado HAZ DUST EPAM 5000 de alta sensibilidad que mide un tamaño de partícula a la vez. La información fue recolectada de acuerdo a puntos de muestreo seleccionados para la toma de las mediciones de $PM_{2.5}$ y PM_1 , los mismos que fueron: Sector Industrial, Centro Ciudad, Sector Cementeras Vía La Costa y Sector Residencial, caracterizados por tener alta demanda comercial, tráfico vehicular, así como actividades comercial y residencial. A la vez se consideraron para la medición las variables meteorológicas como: temperatura, dirección y velocidad del viento, la precipitación como variable dicotómica en caso de existencia o no, y variables de actividad antropogénica como el tráfico. El objetivo del proyecto integrador es obtener un modelo multivariante para la predicción de $PM_{2.5}$ y PM_1 a partir de variables ambientales, meteorológicas y eventualidades físicas en la ciudad de Guayaquil. Utilizando la información del muestreo y de fuentes secundarias se construyó una matriz de datos con variables ambientales, meteorológicas y antropogénicas, capaces de predecir el comportamiento de las partículas finas ($PM_{2.5}$ y PM_1) a partir de otros tamaños de partículas, temperatura, histórico de precipitación, etc.

Finalmente, se construyó un modelo suavizado de regresión múltiple para estimar concentraciones horarias promedio de $PM_{2.5}$ y PM_1 y establecer la influencia de las variables meteorológicas y antropogénicas, La varianza de PM_1 se explica en más del 17,4% de los casos a partir de esas variables ($R^2 = 0.158$; $p < 0.000277$, $GCV = 80.245$) y la de $PM_{2.5}$, se explica en 15.8% de los casos a partir de ($R^2 = 0.143$; $p < 0.00000833$, $GCV = 143.44$). Se validó el modelo verificando la normalidad y varianza constante de los residuos del modelo y se añade el diagnóstico sobre el grado de suavizado del modelo.

ABSTRACT

Between October 2016 and March 2017, in the city of Guayaquil, Ecuador, samples of particulate material corresponding to $PM_{2.5}$ and PM_1 were collected in real time using a gravimetric device called HAZ DUST EPAM 5000 High sensitivity that measures one particle size at a time. The information was collected according to the sampling points selected for taking the measurements of $PM_{2.5}$ and PM_1 , which were: Industrial Sector, City Center, Cement Sector and Residential Sector, characterized by having high commercial demand, vehicular traffic, as well as commercial and residential activities. At the same time, meteorological variables such as temperature, direction and wind speed, precipitation as a dichotomous variable in case of existence or not, and anthropogenic activity variables such as traffic were considered for the measurement. The objective of the integrating project is to obtain a multivariate model for the prediction of $PM_{2.5}$ and PM_1 from environmental, meteorological variables and physical eventualities in the city of Guayaquil. Using data from sampling and secondary sources, a data matrix was constructed with environmental, meteorological and anthropogenic variables, capable of predicting the behavior of fine particles $PM_{2.5}$ and PM_1 from other sizes of particles, temperature, precipitation history, etc.

Finally, a smoothed multiple regression model was constructed to estimate average hourly concentrations of $PM_{2.5}$ and PM_1 , and establish the influence of the meteorological and anthropogenic variables. The variance of PM_1 is explained in more than 17.4 % of cases from these variables ($R^2 = 0.158$; $p < 0.000277$, $GCV = 80.245$) and that of $PM_{2.5}$, is explained in 15.8% of cases from ($R^2 = 0.143$; $p < 0.00000833$) The model was validated verifying the normality and constant variance of the model residues and the diagnosis on the degree of smoothing of the model is added..

ÍNDICE GENERAL

EVALUADORES	¡Error! Marcador no definido.
RESUMEN	6
ABSTRACT.....	7
ÍNDICE GENERAL.....	8
ABREVIATURAS	11
SIMBOLOGÍA	12
ÍNDICE DE FIGURAS.....	13
ÍNDICE DE TABLAS	15
CAPÍTULO 1.....	16
1. Introducción.....	16
1.1 Descripción del problema.....	18
1.2 Justificación del problema	19
1.3 Objetivos.....	22
1.3.1 Objetivo General.....	22
1.3.2 Objetivos Específicos	22
1.4 Marco teórico	23
1.4.1 Meteorología y climatología	23
1.4.2 Factores antropogénicos	26
1.4.3 Contaminación Atmosférica	26
1.4.4 Impacto de la contaminación atmosférica	27
1.4.5 Material particulado (PM).....	27
1.4.6 Contaminación del aire por Material Particulado	28

1.4.7	Efectos sobre la salud	29
1.4.8	Fuentes de emisión del material particulado.....	29
1.4.9	Estándares Nacionales de Calidad del Aire	32
1.4.10	Medición de la Calidad del aire	33
1.4.11	Conceptos Estadísticos	34
CAPÍTULO 2.....		45
2.	Metodología	45
2.1	Área de estudio.....	45
2.1.1	Sector Industrial:.....	46
2.1.2	Sector Cementera:	47
2.1.3	Sector Centro	47
2.1.4	Sector Residencial:.....	48
2.2	Variables meteorológicas.....	49
2.2.1	Temperatura	49
2.2.2	Precipitación	50
2.2.3	Viento	50
2.2.4	Humedad.....	52
2.3	Protocolo de muestreo	53
2.4	Métodos estadísticos	53
2.4.1	Validación de la base de datos	54
2.4.2	Estadística Descriptiva	54
2.4.3	Análisis Exploratorio de la serie de datos	55
2.4.4	Comparaciones múltiples	55
2.4.5	Correlación Cruzada:.....	55

2.4.6	Diseño del modelo	56
CAPÍTULO 3.....		59
3.	Resultados Y ANÁLISIS	59
3.1	Descriptiva	59
3.2	Análisis exploratorio de la serie de datos	65
3.2.1	Diagramas de dispersión	65
3.3	Comparaciones múltiples	69
3.4	Correlaciones.....	73
3.5	Correlación cruzada	78
3.6	Diseño del modelo	84
3.6.1	Construcción del modelo	87
3.6.2	Diagnóstico del modelo	90
3.7	Implementación del modelo	94
3.8	Análisis costo beneficio	95
CAPÍTULO 4.....		96
CONCLUSIONES Y RECOMENDACIONES.....		96
BIBLIOGRAFÍA.....		99

ABREVIATURAS

ESPOL	Escuela Superior Politécnica del Litoral
OMS	Organización Mundial de la Salud
PM	Particulate Material
COV	Compuestos Orgánicos Volátiles
GCV	Validación Cruzada Generalizada
PNCA	Plan Nacional de Calidad del Aire
EPA	Environmental Protection Agency
TULSMA	Texto Unificado de la Legislación Secundaria del Ministerio del Ambiente
MLR	Regresión Lineal Múltiple
ACF	Función de Auto correlación

SIMBOLOGÍA

um	Micrómetro
mm	Milimetro
m	Metro
CO	Monóxido de carbono
SO ₂	Dióxido de azufre
O ³	Oxigeno
RH	Humedad relativa

ÍNDICE DE FIGURAS

Figura 1.1: Fuentes estacionarias, representación de las industrias	30
Figura 1.2: Fuentes de áreas, representación de las actividades	31
Figura 1.3: Fuentes móviles, emisiones por tipo de combustible.....	31
Figura 2.1: Cartografía digital de la ciudad de Guayaquil	45
Figura 2.2: Zona Industrial Inmaconsa-Inductroc	46
Figura 2.3: Zona Cementera Puerto Azul	47
Figura 2.4: Zona Centro (Colón y Chimborazo).....	48
Figura 2.5: Zona Residencial (Ciudadela el Paraíso)	48
Figura 2.6: Temperatura promedio anual	49
Figura 2.7: Precipitación promedio anual	50
Figura 2.8: Viento promedio anual	51
Figura 2.9: Humedad anual de Guayaquil	52
Figura 3.1 Concentración de PM1 y PM2.5 por día y lugar	59
Figura 3.2 Concentración de PM1 y PM2.5 en época lluvia y sequía por día y lugar	60
Figura 3.3 Concentración de PM1 en época lluvia y sequía por sectores.....	61
Figura 3.4 Concentración de PM2.5 en época lluvia y sequía por sectores.....	63
Figura 3.6 Concentración de Material particulado PM1 y PM2.5 por hora	63
Figura 3.7 Concentración de Material particulado PM1 y PM2.5 por hora	64
Figura 3.8 Diagramas de Dispersión de la concentración PM1 vs las variables meteorológicas.....	65
Figura 3.9 Diagramas de Dispersión de la concentración PM2.5 vs las variables meteorológicas.....	66
Figura 3.10 Panel de diagramas de dispersión de variables meteorológicas y concentración PM1	67
Figura 3. 11 Gráfica de la serie de tiempo de PM ₁ y PM _{2.5}	68

Figura 3. 12 Gráfica de la serie de tiempo de temperatura, velocidad, dirección del viento y humedad tanto para PM1 como para PM2.5.....	68
Figura 3. 13 Gráfica test de Bonferroni para la concentración de PM1	71
Figura 3. 14 Gráfica test de Bonferroni para la concentración de PM _{2.5}	73
Figura 3. 15 Gráfica Matriz de Correlación PM ₁	75
Figura 3. 16 Gráfica Matriz de Correlación PM _{2.5}	76
Figura 3. 17 Gráfica Matriz de Correlación PM ₁	77
Figura 3. 18 Gráfica Matriz de Correlación PM _{2.5}	78
Figura 3. 19 Diagramas de dispersión entre las variables PM ₁ y PM _{2.5}	79
Figura 3. 20 Autocorrelograma de la serie PM ₁	80
Figura 3. 21 Autocorrelograma de la serie PM _{2.5}	81
Figura 3. 22 Autocorrelograma de la serie PM ₁ & PM _{2.5}	81
Figura 3. 23 Autocorrelogramas de variables meteorológicas de PM ₁ & PM _{2.5}	83
Figura 3. 24 Relaciones entre PM ₁ y variables meteorológicas	86
Figura 3. 25 Relaciones entre PM _{2.5} y las variables meteorológicas	86
Figura 3. 26 Gráfica de modelo ajustado.....	90
Figura 3. 27 Gráfica de pares.....	90
Figura 3. 28 Gráficos de diagnóstico del modelo.....	91
Figura 3. 29 Gráficos de diagnóstico del modelo ajustado	92
Figura 3. 30 Modelo Resultante de concentración PM ₁	92
Figura 3. 31 Gráfica del modelo ajustado PM _{2.5}	93
Figura 3. 32 Diagnóstico del modelo ajustado PM _{2.5}	93
Figura 3. 33 Modelo final ajustado para PM _{2.5}	94

ÍNDICE DE TABLAS

Tabla 1.1: Estándares establecidos por la Legislación Ecuatoriana (TULSMA)	
.....	32
Tabla 1.2: Estándares establecidos por la OMS	32
Tabla 2.1: Cronograma de Monitoreo de Material Particulado	53
Tabla 3.1: Resumen Modelo Lineal PM1~Época	69
Tabla 3.2: Resumen Modelo Lineal PM1~Época	69
Tabla 3.3: Resumen Modelo Lineal PM1~día	70
Tabla 3.4: Tabla de análisis de varianza	70
Tabla 3.5: Resumen Modelo Lineal PM2.5~Época	71
Tabla 3.6: Resumen Modelo Lineal PM2.5~Lugar	72
Tabla 3.7: Resumen Modelo Lineal PM2.5~Días	72
Tabla 3.8: Tabla de análisis de varianza	72
Tabla3.9 Tabla de valores de coeficiente de correlación para PM1	74
Tabla3.10 Tabla de valores de coeficiente de correlación para PM2.5	75
Tabla 3.9 Autocorrelaciones PM1 & PM2.5	82

CAPÍTULO 1

1. INTRODUCCIÓN

La contaminación atmosférica en la ciudad de Guayaquil es un problema que afecta a más de dos millones y medio de habitantes. Considerando además que en los últimos años ha ido en constante crecimiento este problema debido a factores como: aumento del tráfico vehicular y el problema del transporte en general, el desarrollo no planificado de las industrias, el uso de tecnologías obsoletas en la producción, la mala calidad del saneamiento básico, entre otras.

Además cabe anotar que la ciudad de Guayaquil sufre anualmente durante la época seca, la quema de bosques en áreas rurales o quemas de desechos y otros, que agravan la situación. También sabemos que los vientos elevan gran cantidad de polvo en las avenidas de mayor tráfico vehicular, en las calles sin pavimentar y en las zonas descubiertas. Todos esos factores generan importantes cantidades de contaminación atmosférica y contribuyen al problema ambiental y deterioro de la calidad de vida de sus habitantes, convirtiéndose en uno de los mayores inconvenientes de la ciudad puesto que estas sustancias pueden ser transportadas a través del aire y afectar la salud de las personas.

Este proyecto tiene como finalidad establecer una asociación existente entre las concentraciones de material particulado (PM_1 y $PM_{2.5}$) con las variables meteorológicas que se midieron en las estaciones de muestreo ubicados en puntos estratégicos, con el fin de que abarquen toda la ciudad de Guayaquil

Cabe recalcar también que se trabajara con el contaminante de Material Particulado que se presenta en mayores concentraciones y que se encuentran suspendidos en la atmósfera, que tienen origen natural o antropogénico, y que son los principales contaminantes que afectan a las vías respiratorias especialmente en los niños.

Estas partículas vienen de tamaño menor a 1 micrómetro y gran contenido de carbono negro, y las que pueden penetrar a las vías con las menores a 2 micrómetros.

Para identificar los niveles de contaminación de la ciudad de Guayaquil y sentar una línea base en cuanto a la calidad del aire que respira la población se realiza este proyecto con el fin de crear un modelo estadístico de regresión no paramétrica, que nos indique la relación existente entre la concentración del material particulado y otras variables que pueden afectar los niveles de contaminación.

Por medio de este estudio se pretende estimar mediante un modelo de regresión no paramétrica múltiple, los niveles de concentración promedio de material particulado tanto para PM_1 como $PM_{2.5}$ en la ciudad Guayaquil, que permita ser fuente de información para posteriores estudios asociados a la salud de los ciudadanos y cambio climático. El trabajo se realiza con los datos rescatados en un muestreo que se hizo entre octubre del 2016 y enero del año 2017, y que intentó identificar la posible remoción de material particulado en la atmósfera debido al fenómeno natural de la lluvia que se presenta en la estación invernal.

1.1 Descripción del problema

Uno de los principales problemas ambientales por los que las ciudades se ven afectadas, es la contaminación atmosférica. Esta es una preocupación latente, que las autoridades o gobiernos seccionales deberían contemplar dentro de sus programas de administración, debido a que el incremento de las emisiones de contaminantes provenientes principalmente por el tráfico vehicular, las actividades industriales, las actividades agrícolas entre otras, afectan notablemente la salud de los habitantes, sin embargo son pocas las autoridades que tienen dentro de sus prioridades el estudio de calidad del aire, pero muchas veces estos estudios son muy dispersos.

El Ministerio de Ambiente del Ecuador ha creado desde el año 2010 el Plan Nacional de la Calidad del Aire con el fin de alcanzar una gestión ambiental adecuada de la calidad del recurso aire para proteger la salud humana, los recursos naturales y el patrimonio cultural, contribuyendo así al mejoramiento de la calidad de vida de la población ecuatoriana. (PNCA, 2010).

La contaminación atmosférica causa alrededor de 2 millones de muerte prematuras al año en todo el mundo. Como ya muchos conocen, la exposición a emisiones de material particulado (PM) van de la mano con la salud de la población. En la actualidad son varios los estudios epidemiológicos periódicos realizados a personas, los mismos que han dado como resultado el aumento del riesgo a presentar enfermedades relacionadas a las vías respiratorias. Así también, hay estudios donde la disminución a la exposición de material particulado ha generado mejoras en la salud. Las enfermedades más comunes por la exposición continua a altas concentraciones de material particulado son: irritación de garganta y mucosas, que ataca particularmente a los que sufren de asma.

También cabe recalcar que la interacción del material particulado con ciertas variables meteorológicas produce impactos sobre los ecosistemas, el clima, el ciclo hidrológico y la agricultura, como por ejemplo la limitación de la productividad en los ecosistemas oceánicos y terrestres.

En el país, se cuenta con muy poca información de estudios de la calidad del aire. Ciudades como Quito y Cuenca son referentes en este tipo de estudios; sin embargo, en la ciudad de Guayaquil, siendo una de las más grandes, son escasos los estudios que abarcan con mayor profundidad esta problemática.

En Guayaquil, desde el año 2016 se anunció la creación de 5 estaciones fijas y tres móviles. Estas entraron en funcionamiento, pero debido a los altos costos que implica la instalación de los equipos, no se ha completado aún la red de monitoreo. Actualmente, la única estación funcionando desde abril de 2018, se ubica en un predio de la universidad IDE Business School, en el kilómetro 13,5 de la vía a la costa y analiza el material particulado (PM) 10 y 2,5 (un tipo de medición).

Esta estación de monitoreo ha arrojado resultados preliminares en la ciudad de Guayaquil cuyos datos en promedio establecen que la calidad de aire de la ciudad está en un rango moderado, pero estos resultados no se demuestran concluyentes por lo que se necesitaría un análisis de las condiciones meteorológicas de la ciudad con las épocas con las que cuenta Guayaquil, que son: periodos de lluvia y periodos de sequía. Por tal motivo se necesita trabajar con datos de al menos un periodo de tiempo que a través de un tratamiento estadístico permita encontrar una tendencia en las emisiones de calidad del aire en el sector estudiado, creando un modelo que utilice el contaminante de interés como la variable dependiente (PM_{10} y $PM_{2.5}$), además las características del uso de suelo, tráfico, densidad de población, las condiciones físicas y geográficas entre otras como variables independientes.

1.2 Justificación del problema

Aproximadamente un 78% de nitrógeno, 21% de oxígeno y 1% de gases, componen una mezcla gaseosa denominada aire puro. Una alteración de estas propiedades químicas o físicas del aire por medio de cualquier sustancia como monóxido de carbono, óxidos de nitrógeno, óxidos de azufre, hidrocarburos, partículas, entre otras, se llama contaminación atmosférica.

Las partículas que producen la contaminación atmosférica son de cualquier forma, tamaño y pueden ser sólidas o líquidas. Se dividen en grupos y difieren por su forma y tamaño, yendo desde las más grandes como PM_{10} hasta la más pequeña como PM_1 .

Los componentes de estas partículas son muy diversos como: metales pesados, silicatos y sulfatos, entre otros, así mismo, las fuentes de emisión de las partículas pueden ser naturales, como: erosión del suelo, material biológico, incendios forestales, etc. o antropogénicas como: combustión del petróleo, procesos industriales y quemas en los campos agrícolas.

Con el fin de mitigar todos los efectos causantes de la contaminación atmosférica en las ciudades, se requiere dar soluciones para mejorar la calidad de aire que se respira en las zonas urbanas de la ciudad para proteger la salud de la población.

No existen grandes estudios de material particulado fino con diámetros inferiores a 1 micrómetro sin embargo la USEPA adoptó nuevos NAAQS para las partículas $PM_{2.5}$ que se encuentran en reacciones químicas o productos de combustión, ya que su pequeño tamaño los hace fácilmente inhalables y pueden ingresar a las vías respiratorias, causando un grave daño a la salud.

En referencia a las partículas más finas PM_1 los estudios y las normativas son casi nulos en la mayoría de las regiones, pero se sabe a ciencia cierta que son indicadores que afectan significativamente a la salud causando morbilidad y mortalidad en la mayoría de los casos.

EL estudio realizado por Investigadores de la Universidad de Washington, Seattle, USA y publicado en la revista de análisis de exposición y epidemiología ambiental (2005) 15, de los efectos a corto plazo de cuatro fracciones de tamaño de material particulado (PM_1 , $PM_{2.5}$, PM_{10} y fracción gruesa PM), y CO en recuentos diarios de visitas a urgencias, ingresos hospitalarios y mortalidad. Basado en estudios previos, encontraron asociaciones más fuertes con CO y las fracciones de menor tamaño de PM que hicieron con PM_{10} o $PM_{10-2.5}$, pero no influyeron en los efectos de la salud,

ya que en los hospitales no registraban como causa de la enfermedad el estar expuestos a estos materiales particulado.

Por ello este trabajo de investigación buscará caracterizar tanto el comportamiento de $PM_{2.5}$ como el PM_1 , para lo cual se evaluarán los datos de mediciones realizadas en cuatro sectores de la ciudad de Guayaquil durante siete días entre el período de octubre 2016 y enero de 2017, donde se consideró tanto período de lluvia como de sequía.

Los resultados que arroje el presente estudio podrán convertirse como soporte para la toma decisiones en lo que respecta a control por parte de los entes reguladores y autoridades municipales, así como también podrá servir como fuente base para otros estudios inherentes a problemas climáticos, epidemiológicos, entre otros.

1.3 Objetivos

1.3.1 Objetivo General

Determinar un modelo de regresión no paramétrica múltiple que explique la influencia de variables ambientales, meteorológicas y antropogénicas en el comportamiento horario de la concentración de material particulado PM_{2.5} y PM₁, considerando información proveniente de cuatro sectores de la ciudad de Guayaquil entre octubre de 2016 y marzo de 2017.

1.3.2 Objetivos Específicos

- 1) Validar la base de datos proporcionada para el estudio para que se garantice un apropiado análisis estadístico descriptivo e inferencial, haciendo uso del software estadístico R.
- 2) Establecer la asociación entre las variables meteorológicas, ambientales y antropogénicas para la correcta explicación de la concentración promedio de PM_{2.5} y PM₁ en los cuatro sectores de Guayaquil.
- 3) Determinar la presencia de relación entre las concentraciones de PM_{2.5} y PM₁ mediante la aplicación de correlación cruzada.

1.4 Marco teórico

1.4.1 Meteorología y climatología

La Meteorología etimológicamente se deriva del griego meteoro, que quiere decir que se mueve en el aire, y de logo que es un estudio o tratado, por eso la meteorología es la ciencia que estudia la atmósfera, sus propiedades y los fenómenos o procesos que tienen lugar en ella como: la presión. La temperatura, los vientos, la lluvia, etc. los cuales varían tanto en el espacio como en el tiempo.

1.4.1.1 Parámetros meteorológicos

Temperatura

Es una magnitud que se utiliza mayormente para describir el estado de la atmósfera, su variación entre el día y la noche y entre una estación y otra, así como, entre una ubicación geográfica y otra.

La temperatura es una magnitud que se relacionan con la rapidez del movimiento de las partículas que constituyen la materia, se la mide en base a propiedades de la materia que se ven alteradas cuando esta cambia, se realiza la medición a través de un termómetro.

En meteorología es habitual hablar de las temperaturas máximas y mínimas, altas y bajas que se registran en un periodo de tiempo.

Vientos

EL viento es el movimiento del aire de una zona a otra, este se origina cuando entre dos puntos en donde se establece una cierta diferencia de presión; que es cuando tienden a moverse de la zona de alta presión a la de la baja presión, o una diferencia térmica; que ocurre cuando una masa de aire adquiere una temperatura superior a la de su entorno, su volumen aumenta, lo cual hace disminuir su densidad.

El viento se mide en base a su **velocidad y dirección**; para medir la velocidad horizontal se utiliza un anemómetro de cazoletas en Km/h o m/s. Para medir la dirección se utilizan veletas que indican la procedencia geográfica del viento.

Humedad

El aire contiene una cantidad de vapor de agua que se la conoce como humedad, dicha cantidad dependerá de diversos factores como la lluvia, si se está cerca del mar o si hay plantas. La humedad en la atmósfera puede ser: absoluta, específica y razón de mezcla, sin embargo la medida de humedad que más se utiliza es la relativa que se expresa en porcentaje y se calcula por el contenido de vapor de la masa de aire versus su máxima capacidad de almacenamiento.

La humedad relativa nos indica que cerca esta una masa de aire de alcanzar la saturación.

La humedad se mide con un instrumento denominado psicrómetro, que consiste en dos termómetros iguales, uno de ellos es seco; que mide la temperatura del aire y el otro es húmedo; que consiste en humedecer una telilla por medio de una mecha. Existe otro instrumento más preciso que es el aspirópsicrómetro.

Precipitación

EL cambio de estado de vapor de agua de una masa de aire, que al ascender en la atmósfera se enfrían y se saturan, produce que una nube se cargue de gotas minúsculas y cristales de hielo y se encuentran en suspensión dentro de ellas por las corrientes ascendentes. Estas gotas actúan fundamentalmente debido a dos fuerzas; una la de arrastre y la otra es el peso, cuando este sea más grande para vencer la fuerza de arrastre, estas gotitas caerán al suelo en forma de lluvia. Dependiendo del tamaño con el que caigan las gotas hay diferentes tipos de precipitación líquida: llovizna (gotas pequeñas) y lluvia fuerte (gotas un poco más grandes y que caen de forma violenta).

En la climatología, podemos afirmar que el clima es la síntesis del tiempo, es decir, es el conjunto de estados de tiempo atmosféricos que se producen en una determinada región.

En el Ecuador se diferencian nueve tipos de climas que son agradables sin exagerar: seco, tres tropicales, tres mesotémicos y el páramo. Cabe recalcar que los factores que determinan el clima son la latitud, el relieve y las corrientes oceánicas.

1.4.1.2 Parámetros climáticos

Clima seco

El primer tipo de clima denominado seco se caracteriza por las pocas precipitaciones anuales, más o menos 500 mm. Hay una sola época lluviosa que va desde enero hasta abril. La temperatura promedio de este tipo de clima es de 24°C. Este clima se da sobre todo en la península de Santa Elena y el cabo San Lorenzo, siendo la ciudad de Salinas donde hay menos precipitaciones anuales (126mm) y una temperatura media de 23,4°C.

Clima tropical

El clima tropical de sabana se presenta al norte y este de la península de Santa Elena. Abarcaría la zona más occidental de las provincias de Guayas, exceptuando la península de Santa Elena, Manabí y Esmeraldas. Lluve aproximadamente entre 500 y 1.000 mm al año concentrándose las lluvias entre diciembre y mayo. Los bosques secos y de ceibos son particulares de esta zona climática. Las temperaturas medias son de unos 26°C.

Entre los climas tropicales tenemos: el húmedo. monzónico y de sabana; el húmedo tiene precipitaciones anuales elevadas que superan los 3000mm, la vegetación siempre es verde, permite el crecimiento de selvas tropicales, etc.

El monzón, en cambio, se extiende desde Esmeraldas a Guayaquil y ocurren principalmente entre diciembre y mayo con precipitaciones de entre 1000 y 2000 mm.

1.4.2 Factores antropogénicos

Los efectos, procesos o materiales que producen las acciones humanas como actividades agrícolas, las industrias, etc. los denominamos factores antropogénicos, que pueden dar lugar a la erosión del suelo y rocas, producción de gases invernadero entre otras.

Estas alteraciones que el clima sufre a partir de las actividades cotidianas del hombre dan lugar al cambio climático antropogénico.

Consideraremos como factores antropogénicos a la industrialización, la deforestación, las emisiones humanas, las detonaciones nucleares atmosféricas, el tráfico vehicular entre otras.

Para nuestro proyecto de Investigación citaremos al tráfico vehicular, que es en lo que nos enfocaremos para verificar la calidad del aire.

1.4.2.1 Tráfico vehicular

El tráfico vehicular se define como el flujo excesivo de vehículos en una vía, calle o autopista, este fenómeno se presenta con similitudes en otros fenómenos como el flujo de partículas y el de peatones. Lo que se resalta es que el tráfico vehicular tiene la connotación negativa por que suele ser abundante y hasta interminable.

El tráfico vehicular produce una serie de contaminantes que ocasionalmente afectan a las personas, debido al exceso de carros, de smoke, de personas, etc.

1.4.3 Contaminación Atmosférica

La contaminación atmosférica es emitida por los fenómenos naturales como las actividades humanas y consiste en una alteración de la composición natural de la

atmósfera terrestre causada por gases, partículas líquidas o sólidas añadidas a la atmósfera. Esto provoca consecuencias graves en las personas ya que les puede afectar la salud y desequilibrar el medio ambiente.

Los contaminantes del aire cuando se encuentran en concentraciones bajas y largos periodos de exposición producen afecciones crónicas, estos contaminantes pueden ser: Material particulado fino, óxido de azufre, monóxido de carbono, ozono, óxidos de nitrógeno y plomo, siendo estos los más comunes y que se encuentran en las zonas urbanas donde la población se concentra.

1.4.4 Impacto de la contaminación atmosférica

El impacto de la contaminación atmosférica se ve afectado de acuerdo al área de contaminación ya sea por Hidrocarburos volátiles en instalaciones petroleras, material particulado generado por las fábricas de cemento, por incendios forestales, lluvias ácidas entre otras.

Generalmente la presencia de un contaminante atmosférico se emiten en conjunto con otros contaminantes, la concentración de ellos tiene que ser reportados de manera periódica para que permita establecer la calidad del aire de las ciudades.

Como notamos la definición de contaminantes atmosféricos indican que éstos producen un efecto negativo sobre los seres vivos.

Existen un gran número de investigaciones realizadas para verificar los efectos producidos por estos contaminantes que tiene valores límites establecidos según Normas emitidas por La OMS (Organización Mundial de la Salud), basados en estudios epidemiológicos que describen los niveles de concentración de contaminantes en las zonas urbanas.

1.4.5 Material particulado (PM)

El material particulado, consiste en una mezcla de partículas sólidas y líquidas, de sustancias orgánicas e inorgánicas, suspendidas y dispersas en el aire. Las

partículas como el polvo, la suciedad o el humo son visibles a simple vista, sin embargo existen partículas tan pequeñas que solo pueden detectarse mediante un microscopio. Estas partículas que son el principal factor contaminante se han clasificado en términos de diámetro aerodinámico, ya que presentan diferentes tamaños y formas, por ello, tenemos a las partículas inhalables que tienen un diámetro de 10 μm (micrómetros), las inhalables finas que tienen un diámetro de 2,5 micrómetros o menores y las ultra finas que tienen 1 μm . La selección del tamaño del PM, se determina de acuerdo al diámetro, según el rango aerodinámico específico, ya que cada uno de estos tiene una especial relevancia en la inhalación y depósito, fuentes o toxicidad (Salinas, 2012).

La mayoría de las partículas se forman en la atmósfera como resultado de reacciones complejas de químicos, como el dióxido de azufre y los óxidos de nitrógeno, que son contaminantes emitidos por centrales eléctricas, industrias y automóviles. (EPA, 2017).

1.4.6 Contaminación del aire por Material Particulado

El material particulado, el cual contiene sólidos muy pequeños y gotas de líquido pequeños, pueden inhalarse y provocar graves enfermedades ya sea a los pulmones, al sistema respiratorio, etc.

Las partículas que suponen un mayor problema de salud son las menores de 10 micrómetros de diámetro, ya que pueden llegar hasta la sangre.

El ambiente se va contaminando por la volatilidad de estas partículas, tomando en cuenta que están en el aire y considerando que en promedio un adulto inhala entre 10,000 y 12,000 litros de aire. Una vez que estas partículas ingresan en el cuerpo de las personas se adhieren a las vías respiratorias, alojándose en los pulmones. Estos

Ya en el interior del cuerpo, las partículas se adhieren a las paredes de las vías respiratorias y algunas llegan a alojarse en el interior de las paredes de los

pulmones, provocando fuertes enfermedades de tipo respiratorio y en algunos casos hasta la muerte, por la exposición a estos contaminantes.

1.4.7 Efectos sobre la salud

Se han publicado varias revisiones de los estudios que han analizado la relación entre la exposición a largo plazo de las PM a la contaminación atmosférica y a los efectos de salud. (Chen, Goldberg, & Villeneuve, 2008). Chen y colaboradores realizaron una revisión sistemática de los estudios publicados con este enfoque, entre 1950 y 2007. Después de evaluar la heterogeneidad, estimaron efectos conjuntos para contaminantes específicos y realizaron un análisis de sensibilidad. En relación con las PM, concluyeron que por cada 10 $\mu\text{g}/\text{m}^3$ de incremento de la exposición media a largo plazo a las PM_{2,5}, se aumenta el riesgo de morir por causas no accidentales en un 6%, con independencia de la edad, el género y la región geográfica. La misma exposición se asoció a un incremento en el riesgo de morir por cáncer de pulmón, en un 21% y por causas cardiovasculares, en un 14%. Además, los mismos autores concluyeron que vivir cerca de zonas con intenso tráfico, parece estar asociado con elevados riesgos de los tres efectos comentados. Existen muchos estudios epidemiológicos de la contaminación por Material Particulado, que muestran el incremento de mortalidad por enfermedades respiratorias.

Por ello las partículas que son de mayor interés investigativo son las menores a 10 μm , ya que son fácilmente respirables y penetran con facilidad a los pulmones, incluyendo las nano partículas.

1.4.8 Fuentes de emisión del material particulado

Las fuentes de emisión del Material Particulado se generan por las actividades humanas en particular tales como: automóviles, construcciones, quemas de cualquier tipo, provenientes de las industrias, etc.

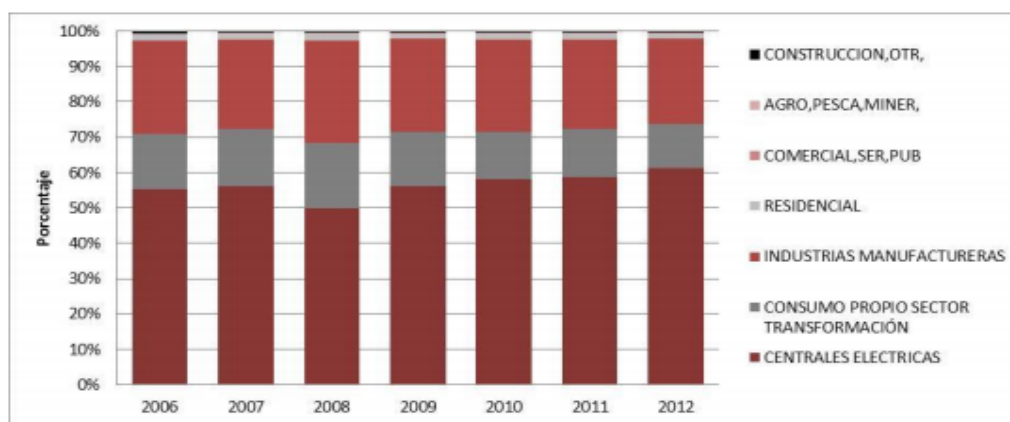
Tomando en cuenta tales emisiones consideramos a unas fuentes como: estacionarias, de área, y móviles.

1.4.8.1 Fuentes estacionarias:

Estas emisiones se las conoce como no móviles que incide en el total de emisiones que ocurren a partir de una instalación contaminante contigua más que a emisiones de puntos de emisión en una instalación.

Entre estas emisiones tenemos a las actividades de combustión que generan partículas de las plantas generadoras de electricidad y de las Industrias y empresas comercializadoras. Los procesos industriales en general provocan una fuerte emisión de estas partículas netamente por sus actividades, también la evaporación de materia volátil que se presentan en los hornos de vidrio.

Figura 1.1: Fuentes estacionarias, representación de las industrias



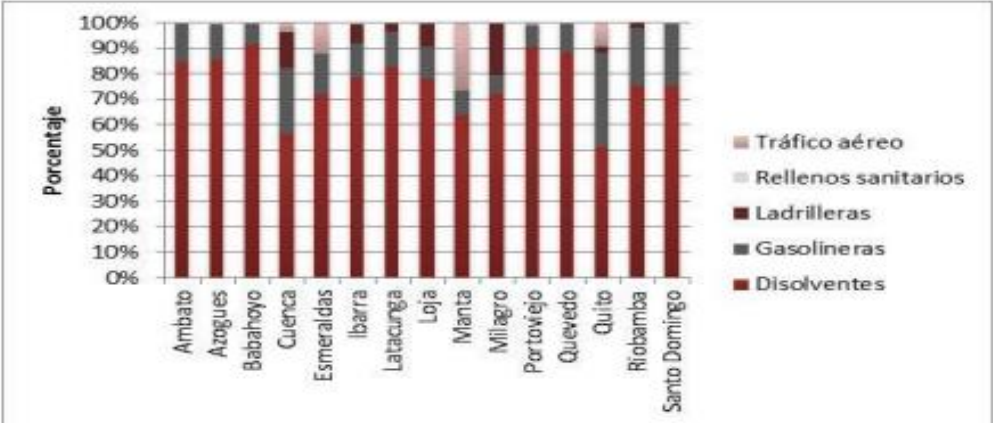
Fuente: SCAN, Ministerio del Ambiente Ecuador

1.4.8.2 Fuentes de área:

Estas fuentes son dispersas y numerosas que en conjunto emiten muchos contaminantes uno de los principales es COV. Esta emisión se da por el uso de disolventes, gases, actividades de construcción, quemas agrícolas, polvo, tráfico aéreo, rellenos sanitarios, ladrilleras, etc. En un estudio realizado por Parra (2012), indico que el uso de disolventes (22.200 toneladas), representan el 72% de COV y

el 66% del total de gases y material particulado (PM₁₀, PM_{2.5}). En segundo lugar están las gasolineras que solo participan en la emisión de COV (8.024 toneladas) y emiten el 24% del total de gases, mientras que las ladrilleras emiten todos los tipos de gases criterio y significan el 4% de las emisiones totales.

Figura 1.2: Fuentes de áreas, representación de las actividades



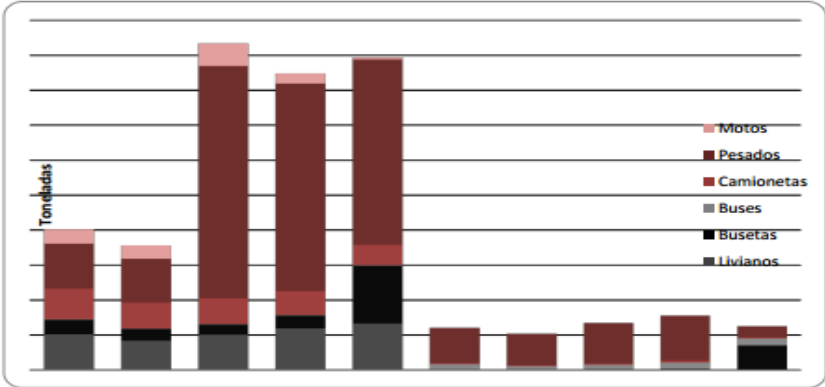
Fuente: SCAN, Ministerio del Ambiente Ecuador

1.4.8.3 Fuentes móviles:

La emisión de este tipo de fuentes son todas las motorizadas como automóviles, maquinarias de uso agrícola, locomotoras, embarcaciones, maquinarias de construcción, etc.

Los vehículos particulares son los que contribuye en mayor medida las emisiones, que en promedio sería 1611 mil toneladas de gases contaminantes al año.

Figura 1.3: Fuentes móviles, emisiones por tipo de combustible



Fuente: SCAN, Ministerio del Ambiente Ecuador

1.4.8.4 Fuentes naturales:

Estas emisiones se dan específicamente por contaminación atmosférica sin la participación de actividades humanas, se clasifican en biogénicas y geogénicas se da exclusivamente por los aerosoles marinos, la erosión y re suspensión, los incendios naturales, la actividad volcánica y la reproducción de bacterias y virus.

1.4.9 Estándares Nacionales de Calidad del Aire

En Ecuador, los estándares permisibles para la máxima concentración de contaminantes de Partículas, están establecidos por la Legislación Ecuatoriana (TULSMA), como se detalla en la Tabla 1.1, y el nivel máximo de concentración sugerido por la OMS (2005) en 24 horas de PM2.5 es de 25 $\mu\text{g}/\text{m}^3$, así como lo indica la Tabla 1.2.

Tabla 1.1: Estándares establecidos por la Legislación Ecuatoriana (TULSMA)

Contaminante	Tiempo de exposición	Máxima concentración permitida
Partículas Sedimentables	30 días	1mg/cm ² /mes
PM10	1 año	50 $\mu\text{g}/\text{m}^3$
	24 horas	150 $\mu\text{g}/\text{m}^3$
PM2,5	1 año	15 $\mu\text{g}/\text{m}^3$
	24 horas	65 $\mu\text{g}/\text{m}^3$

Fuente: (MAE, 2012). Valores tomados del anexo 4 del Libro VI del TULSMA.

Tabla 1.2: Estándares establecidos por la OMS

Contaminante	Tiempo de exposición	Máxima concentración permitida
Partículas Sedimentables	30 días	0,5mg/cm ² /mes
PM10	1 año	20 $\mu\text{g}/\text{m}^3$
	24 horas	50 $\mu\text{g}/\text{m}^3$
PM2,5	1 año	10 $\mu\text{g}/\text{m}^3$
	24 horas	25 $\mu\text{g}/\text{m}^3$

Fuente: Organización Mundial de la Salud

1.4.10 Medición de la Calidad del aire

La medición del Aire se lo puede realizar a través de muestreos por medio de la forma de muestras y de forma discontinua, lo que se requiere es determinar la concentración de las partículas suspendidas en sus diferentes diámetros, para luego ser sometidas a análisis y poder detectar la concentración y correlación con otras variables.

El muestreo se puede dar por recolección física o química, que sería el método activo en el caso de las partículas son impactadores, el más utilizado es el de alto volumen "High-Vol" para PM_{10} , $PM_{2.5}$ y PM_1 .

También existe el método automático que dan alta resolución en sus mediciones y pueden medir algunos contaminantes (PM_{10} , $PM_{2.5}$, PM_1 , CO, SO₂, O₃) hasta los tóxicos en el aire como los orgánicos volátiles. Existen equipos disponibles como analizadores automáticos y monitoreo de partículas, el último es el que nos compete trabajar en nuestro proyecto.

El monitoreo de la calidad del aire, junto con los modelos de predicción de emisiones forman parte de la gestión de la calidad del aire, por ello son necesarios para evaluar los niveles de contaminación en el aire, la idea de monitorear el proporcionar la información pertinente a científicos, legisladores, dueños de empresas, etc. Para que ellos tomen las acciones adecuadas para mejorar el ambiente.

Se realizó un estudio de correlación entre el material particulado $PM_{2.5}$ y las variables meteorológicas en Estados Unidos, con implicaciones de la sensibilidad del $PM_{2.5}$ hacia el cambio climático, mediante la aplicación de un modelo de regresión lineal múltiple (MLR) para estudiar las correlaciones de $PM_{2.5}$ total y sus componentes con variables meteorológicas utilizando un registro de observación de 11 años (1998-2008). Los datos fueron procesados por estaciones para centrarse en las correlaciones de escala sinóptica. Se encontró que la variación diaria en la meteorología descrita por la MLR puede explicar hasta el 50% de la variabilidad de $PM_{2.5}$ con temperatura, humedad relativa (RH), precipitación y circulación, siendo

todos factores predictivos importantes. La precipitación está fuertemente correlacionada negativamente con todos los componentes de PM_{2.5}. Las concentraciones de PM_{2.5} son en promedio 2,6 µg/m³ más altas en los días estancados vs. no estancados. Las correlaciones observadas proporcionan una prueba para los modelos de transporte químico utilizados para simular la sensibilidad de PM_{2.5} al cambio climático. Señalan la importancia de representar adecuadamente la dependencia de la temperatura de las emisiones agrícolas, biogénicas y de incendios forestales en estos modelos. (Tai, A. P., Mickley, L. J., & Jacob, D. J. ,2010).

1.4.11 Conceptos Estadísticos

1.4.11.1 Comparación múltiple

Los procedimientos de comparaciones múltiples permiten identificar si existen relaciones o diferencias significativas al comparar las medias de la variable respuesta con los diferentes tratamientos estadísticos de resultados experimentales. Estas comparaciones pueden ser planeadas o no antes del experimento.

Existen algunas pruebas de comparaciones múltiples, entre las más usadas tenemos:

- 1) Prueba de Diferencia mínima o límite de significación.

Esta prueba realiza comparaciones entre dos medias de variable respuesta con diferentes tratamientos que han sido planeados con anterioridad. La prueba F debe resultar significantes y los ε_{ij} tienen que ser una variable aleatoria independiente, distribuida normalmente con media cero y varianza σ^2 .

$$t_c = \frac{\bar{Y}_i - \bar{Y}_j - k}{S_{\bar{Y}_i - \bar{Y}_j}} \sim t_{(GLE)} \quad H_0 \text{ es verdadera} \quad (1.1)$$

$$S_{\bar{Y}_i - \bar{Y}_{i'}} = \sqrt{CME \left[\frac{1}{n_i} + \frac{1}{n_{i'}} \right]} \quad (1.2)$$

2) Prueba de Tukey.

Esta prueba realiza las comparaciones por pares de todas las medias de la variable respuesta y que están sujetos a tratamientos diferentes, y de la cual se obtienen intervalo de confianza simultáneos $100(1-\alpha)\%$. Se la conoce como método de la diferencia honestamente significativa de Tukey.

Para el procesamiento de la prueba es necesario que los ε_{ij} sea una variable aleatoria independiente distribuida normalmente comedia cero y varianza σ^2 .

$$w = q_\alpha(t, GLE) \frac{1}{\sqrt{2}} S_{\bar{Y}_i - \bar{Y}_{i'}} \quad (1.3)$$

3) Prueba de Bonferroni.

La desigualdad de bonferroni es un método que se da en base a una secuencia de eventos: $\{A_j\}, P\{\cup_j A_j\} \leq \sum P[A_j]$, de manera que uno o más errores de tipo I sean una correlación arbitraria de la prueba. Es decir que permite comparar las medias de los t niveles de un factor después de haber rechazado la hipótesis nula de igualdad de medias.

$$VCB(i, i') = t_{(1 - \frac{\alpha}{2nc}, GLE)} S_{\bar{Y}_i - \bar{Y}_{i'}} \quad (1.4)$$

1.4.11.2 Series de tiempo

Una importante técnica para observar el comportamiento secuencial en el tiempo usada en muchas aéreas como ingenierías, física, meteorología, etc. son las series de tiempo que logran explicar las variaciones observadas históricamente, tratando de determinar si responden a un patrón para poder predecir el comportamiento futuro de las variables de estudio.

De por si el tiempo no es una variable explicativa pero nos sirve de soporte para la realización de la metodología en donde decimos que la serie temporal es una función del tiempo, representado de la siguiente manera:

$$Y_t = f(t) \quad (1.5)$$

Para nuestro proyecto el análisis de serie temporal no es de gran interés por lo que no tenemos datos históricos de años anteriores, sino son específicos del año 2016 y 2017, pero se analizara en esos años para saber el comportamiento.

Así que rápidamente se revisaran los conceptos de Series Temporales.

Tendencia

Esta componente recoge el comportamiento de la serie a largo plazo, para ello es necesario tener un número mayor de observaciones a lo largo de varios años, pueden ser lineal, exponencial, parabólico, logístico, etc.

Variaciones estacionales

Son movimientos periodos de la serie y se basan en tipo climatológico o de orden de tiempo por ejemplo los días de semana u horas. Lo que se requiere analizar es la periodicidad de los datos.

Variaciones cíclicas

Son movimientos a plazo medio, que utiliza periodos de la serie superiores al año y se basan en crear un ciclo completo dependiendo del número de amplitud.

Outliers

Observación de la serie que tiene un comportamiento anormal, sin tener incidencias en el futuro, y que si existe entonces hay que reemplazarlo.

Correlación cruzada:

La correlación cruzada puede ser utilizada para detectar y localizar una señal conocida de referencia inmersa en ruido, así como para correlacionar dos series temporales de un tiempo simultáneo o si están desfasadas en el tiempo.

- Se resaltan asociaciones claras en las series.
- La correlación cruzada se realiza con dos series estacionarias, es decir que su media y varianza no varíen estadísticamente en el tiempo.
- la correlación cruzada se da entonces, entre una serie X_t en un tiempo dado, t , con otra serie en un tiempo posterior Y_{t+k} .
- Los datos que se desplazan en la variable X_t o Y_t determinan el signo positivo o negativo del retardo de los coeficientes
- La correlación cruzada, varía de -1 a +1 para indicar máxima correlación negativa o positiva, respectivamente.

La correlación cruzada $r_{XY(K)}$, entre dos series estacionarias para un retardo dado es la media de los productos de las series X_t y Y_t normalizadas, desfasadas en 1, 2,...(N-1) retardos, como lo indica las siguientes formulas.

$$r_{XY(K)} = \frac{\sum[(X_t - \mu_x) * (Y_{t+k} - \mu_y)]}{N S_x S_y} = \frac{1}{N} \sum \left[\left(\frac{X_t - \mu_x}{S_x} \right) * \left(\frac{Y_{t+k} - \mu_y}{S_y} \right) \right] \text{ en lag } + \quad (1.6)$$

$$r_{YX(-K)} = \frac{\sum[(Y_t - \mu_y) * (X_{t+k} - \mu_x)]}{N S_x S_y} = \frac{1}{N} \sum \left[\left(\frac{Y_t - \mu_y}{S_y} \right) * \left(\frac{X_{t+k} - \mu_x}{S_x} \right) \right] \text{ en lag } - \quad (1.7)$$

Cuando el coeficiente de cros correlación tiene retardo cero, se dice que tiene el mismo valor que el coeficiente de correlación lineal de Pearson, por las series que ocurren simultáneamente.

Si el coeficiente de correlación cruzada es estadísticamente significativo, ósea estadísticamente diferente de cero con un 95% de confianza, quiere decir que cumple con la siguiente característica:

$$r_{xy}(k) > \frac{1.96}{\sqrt{(N-k)}} \quad (1.8)$$

Donde:

- k es el valor absoluto del retardo dado,
- N es el número de datos, y
- $1/\sqrt{(N-k)}$ el error estándar de $r_{xy}(k)$.

Cuando las variables sean estacionarias, estas pueden ser utilizadas en correlación, cros correlación y regresión, solamente si están “cointegradas”, esto quiere decir, que entre ellas existe una combinación lineal que sí es estacionaria, y esa combinación lineal entre ellas es la serie de los residuales de la regresión, (Engle y Granger, 1987).

1.4.11.3 Regresión y Correlación

La regresión consiste en verificar la relación que existe entre una variable dependiente (Y) y una o más variables independientes (x_1, x_2, x_3, \dots). Para poder realizar esta investigación, se debe postular una relación funcional entre las variables. Tenemos algunos tipos de regresión, la más utilizada es la relación lineal, debido a su simplicidad analítica. Esto quiere decir que se reduce a una línea recta que se expresa como:

$$\hat{Y} = \beta_1 X + \beta_0 \quad (1.9)$$

donde los coeficientes β_0 y β_1 son los valores que definen la inclinación de la recta y su posición. El parámetro β_0 , nos indica cuánto es Y cuando $X = 0$. El parámetro β_1 , que nos da la dirección de la recta, nos indica cuánto aumenta Y por cada aumento de una unidad en X. Lo que se quiere obtener son las estimaciones de estos coeficientes a partir de una muestra de observaciones sobre las variables Y y X. En el análisis de regresión, estas estimaciones se obtienen por medio del método de mínimos cuadrados. Para poder visualizar el grado de relación que existe entre las variables, como primer paso se recomienda elaborar un diagrama de dispersión, que consiste en una representación del sistema de coordenadas cartesianas que incluyen los datos numéricos observados. En el eje X se miden las variables independientes, y en Y se mide la variable dependiente o a explicar.

Mediante los diagramas de dispersión observaremos el tipo de relación que tendrán las variables Y y X. Así mismo también apreciamos que no toda la variación en Y puede ser explicada por la variable X.

Existe el caso de que entre las variables exista una relación lineal perfecta, donde los puntos caen a lo largo de la recta de regresión, pero también observamos en la mayoría de los casos que los puntos no caen exactamente en la recta, sino que se encuentran dispersos en torno a ella o dispersos sobre todo el diagrama, este último indica que no existe ninguna relación lineal entre las variables.

Para poder identificar el grado de la posible relación entre las variables, se aplica un estadístico denominado coeficiente de correlación lineal que se calcula de la siguiente manera:

$$r = \frac{\frac{1}{n} * \sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\left(\frac{1}{n} * \sum (x_i - \bar{x})^2\right) * \left(\frac{1}{n} * \sum (y_i - \bar{y})^2\right)}} \quad (1.10)$$

Los valores que puede tomar el coeficiente de correlación "r" son: $-1 < r < 1$

- Si " r " > 0 , la correlación lineal es positiva, y si se aproxima a 1 la correlación es más fuerte.
- Si " r " < 0 , la correlación lineal es negativa
- . Si " r " $= 0$, no existe correlación lineal entre las variables.

El coeficiente de correlación posee las siguientes características:

- El valor del coeficiente de correlación es independiente de cualquier unidad usada para medir las variables.
- El valor del coeficiente de correlación se altera de forma importante ante la presencia de un valor extremo, como sucede con la desviación típica.
- El coeficiente de correlación mide solo la relación con una línea recta. Dos variables pueden tener una relación curvilínea fuerte, a pesar de que su correlación sea pequeña.
- El coeficiente de correlación no se debe extrapolar más allá del rango de valores observado de las variables a estudio ya que la relación existente entre X e Y puede cambiar fuera de dicho rango.
- La correlación no implica causalidad.

El coeficiente de correlación Pearson puede calcularse para cualquier grupo de datos, pero para probar validez de la hipótesis sobre la correlación entre las variables se requiere:

- a) que las dos variables procedan de una muestra aleatoria de individuos.
- b) que al menos una de las variables tenga una distribución normal en la población de la cual la muestra procede.

Para que el cálculo de un intervalo de confianza del coeficiente de correlación de r sea válido, las 2 variables X y Y deben tener una distribución normal. Cuando los datos presentan valores externos, o presentan distribuciones no normales se calcularía un coeficiente de correlación no paramétrico (coeficiente de correlación

de Spearman), el cual se calcula de la misma manera que el coeficiente de correlación de Pearson utilizando el rango de las observaciones.

Para verificar el porcentaje de variación total de la variable independiente que puede ser explicada por la variación de la variable dependiente, se utiliza el estadístico coeficiente de determinación R^2 y se lo denota:

$$R^2 = 1 - \frac{\sum e^2}{\sum (Y - \hat{y})^2} \quad (1.11)$$

Con todo lo anterior, podemos decir que la regresión es una correlación que existe entre dos variables, el cual mide la relación de intensidad a través del estadístico r y que se puede cuantificar el porcentaje de datos pronosticados más cercanos a la relación por el estadístico R^2 .

En el caso de que las variables no presenten normalidad en sus datos o no sean homogéneas, se utilizan las pruebas no paramétricas que trabaja con variables categóricas ordenadas y que se pueden investigar bajo los coeficientes de correlación como tau_b Kendall.

1.4.11.4 Regresión no paramétrica

Lo que desarrolla la regresión no paramétrica, es un modelo libre que predice una respuesta sobre un rango de valores de datos. Se dice que se constituye por métodos que proporcionan una estimación suavizada de la relación para un conjunto de valores de la variable explicativa.

Lo que se pondera en esta regresión no paramétrica es que las variables más cercanas tengan mayor peso que las más alejadas en un conjunto de datos.

Así mismo las pruebas no paramétricas no necesitan establecer supuestos exigentes sobre poblaciones de donde se extraen las muestras y no necesitan trabajar con datos obtenidos con una escala de medida de intervalo o razón.

Por lo que se concluye que se puede utilizar las pruebas no paramétricas para:

- 1) Todos aquellos contrastes que no se ajustan a cualquiera de las tres características de las pruebas paramétricas
- 2) Englobar a ese término genéricos a los contrastes de distribución libre

La regresión no paramétrica simple es importante por dos motivos:

- Ajustar una curva suavizada para los datos preliminares, para confirmar una función respuesta en particular que haya sido ajustada a los datos, para obtener estimaciones de la respuesta media sin especificar la forma de la función respuesta, para estudiar el cumplimiento de supuestos, etc.
- Forma la base a partir de la cual se extienden los conceptos para regresión no paramétrica múltiple.

La filosofía de estos métodos se basan en la idea de dejar que los datos hablen por sí mismos.

El principal objetivo es estimar la función de regresión $E(y_i) = g(x_{i1} + x_{i2} + x_{i3} + \dots + x_{ip})$ en base a una muestra.

El análisis de regresión no paramétrico en lugar de estimar los parámetros tiene por objeto una estimación directa de $g(x_i)$ y sustituye el supuesto de linealidad por uno más débil que consiste en que el valor promedio $g(x_i)$ que es una función de regresión suave.

En regresión múltiple con más de una variable predictiva, aparecen nuevos problemas como correlación entre predictores, interacciones, etc.)

Los Modelos de Regresión Aditivos son una estrategia popular en los cuales se especifica la media de la variable respuesta como la suma de funciones suaves de los regresores como indica la ecuación (1.12):

$$E\left(\frac{y}{x_1, \dots, x_k}\right) = \alpha + f_1(x_1) + \dots + f_k(x_k) \quad (1.12)$$

Cuando la variable respuesta no es Normal, se generaliza todo lo anterior a Modelos de Regresión (Aditivos) Generalizados.

También aparecen el problema de selección de variables y comparación de modelos.

1.4.11.5 Métodos de regresión no paramétricas:

1.4.11.5.1 Estimadores tipo núcleo de regresión.

Los estimadores tipos núcleo son medias ponderadas de las observaciones, los mismos se definen como aquellos estimadores $\hat{m}_h(x)$ de m en el punto x que responden a una expresión del tipo:

$$\hat{m}_h(x) = \sum_{i=1}^n w_{hi}(x)y_i \tag{1.13}$$

donde w_{hi} denota una sucesión de pesos que pueden depender del vector completo x_i

h es el llamado parámetro de suavizado.

Existen diferentes estimadores tales como:

- Estimador núcleo de Nadaraya-Watson, definido para diseños aleatorios.
- Estimador nucleo de regresión para diseños fijo Gasser Muller.

1.4.11.5.2 Splines de suavizamiento.

Los splines de suavizamiento son aquellos estimadores de la función de regresión que hacen mínima la función:

$$S_h(m) = \sum_{i=1}^n (y_i - m(x_i))^2 + h \int_0^1 m''(x)^2 dx, \tag{1.14}$$

Donde h denota el ancho de banda, si $h \rightarrow \infty$, se aproxima a un ajuste lineal por mínimos cuadrados.

1.4.11.5.3 Regresión polinómica local

Esta regresión no paramétrica se realiza en base a ajustar un polinomio por mínimos cuadrados a los puntos en una ventana local, ponderados por la distancia.

El estimador polinómico está dado por la siguiente ecuación:

$$\hat{\beta} = (X'WX)^{-1}X'W_y \quad (1.15)$$

Donde:

$$X = \begin{pmatrix} 1 & x_1 - x_0 & \dots & x_1 - x_0^P \\ \vdots & \vdots & & \vdots \\ 1 & x_n - x_0 & \dots & x_n - x_0^P \end{pmatrix} = [1 \ x_j - x_i \ \dots \ (x_j - x_i)^P]_{j \in U} \quad (1.16)$$

Y la matriz cuadrada de pesos W está dada por:

$$W = \text{diag} \{K_h(x_i - x_0)\} \quad (1.17)$$

CAPÍTULO 2

2. METODOLOGÍA

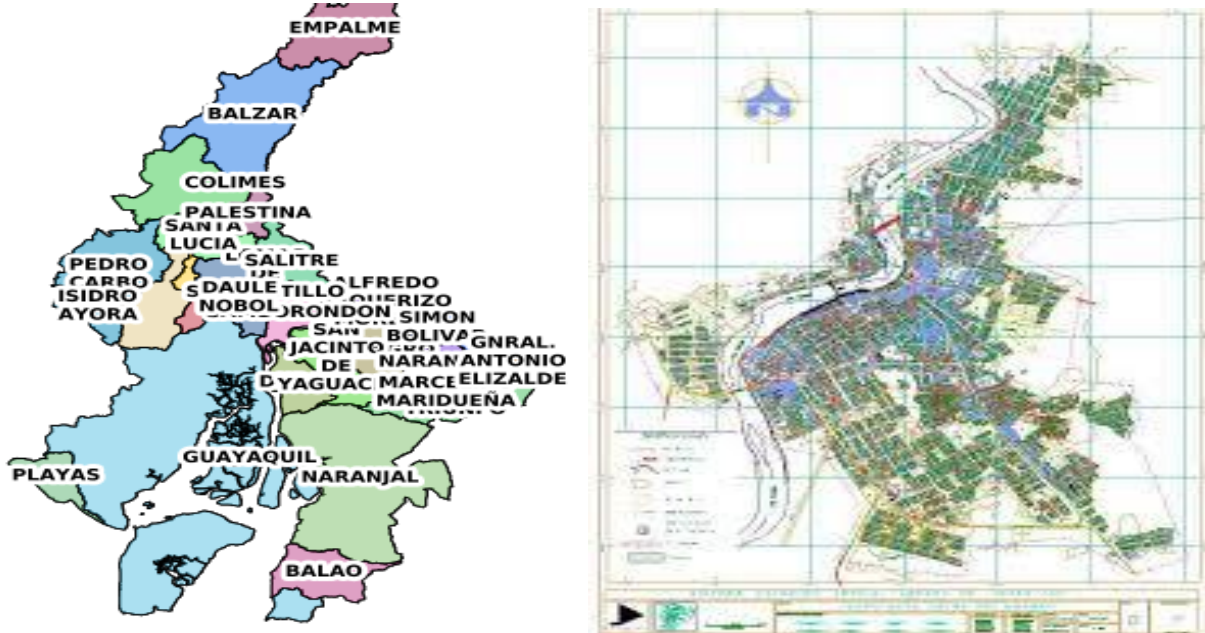
2.1 Área de estudio

El área de estudio corresponde al cantón Guayaquil, provincia de Guayas, ubicada en la región costa del Ecuador, a 4,02 msnm (Figura 2.1); cuenta con un clima cálido característico de las zonas costeras.

Las características del área son las siguientes:

- ❖ La temperatura ambiente media registrada durante todo el año, es de 25.7°C.
- ❖ La humedad relativa en la zona, es de 83%.
- ❖ La velocidad del viento, es de 13 km/h.
- ❖ Precipitación atmosférica de 791 mm.
- ❖ Presión atmosférica de 1012 hPa.
- ❖ Geográficamente se encuentra en las coordenadas: (Latitud: 3°10'S, Longitud: 79°54'O)

Figura 2.1: Cartografía digital de la ciudad de Guayaquil



Fuente: Ecuador en cifras (2010)

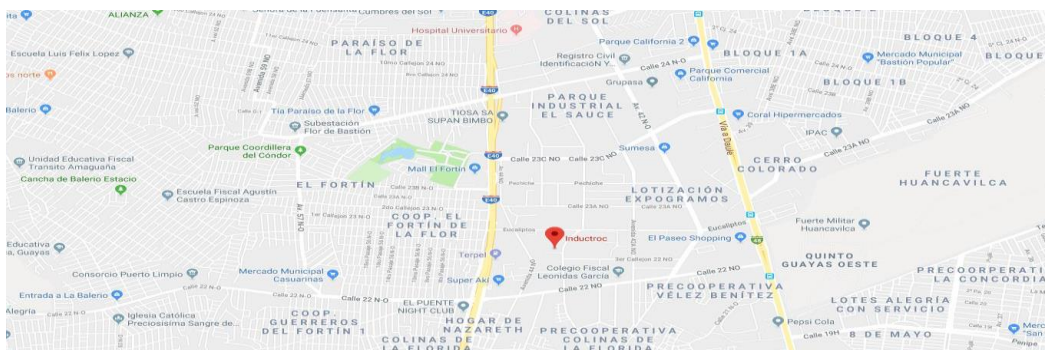
La base de datos recogida de concentraciones de material particulado (PM_{2.5} y PM₁), se la obtuvo de algunos puntos estratégicos de la ciudad de Guayaquil, tenemos a los siguientes sectores: industrial, centro, cementeras y residenciales.

El sector industrial tiene una alta actividad comercial e industrial y está ubicado en las autopistas principales de la ciudad. El sector del centro tiene en cambio alto tráfico vehicular y así mismo tiene alta actividad comercial. Al noroeste se encuentran las cementeras que están ubicadas en la vía a la costa, también tenemos el sector residencial que se caracteriza por encontrarse en un lugar montañoso denominado Bosque seco protector El Paraíso, en este sector se encuentra alta densidad poblacional y cuenta con actividad comercial, institutos, demanda vehicular, etc.

2.1.1 Sector Industrial:

El sector industrial donde se realizó el muestreo del material particulado fue entre la vía Perimetral y la vía a Daule, donde queda INMACONSA. En este sector hay mucha actividad industrial y comercial, así mismo, aquí encontramos el Parque California conocido como PECA, que consta de algunos locales comerciales, también encontramos estaciones de servicios, bodegas, empresas pequeñas, medianas y grandes, también existe alta afluencia vehicular. Específicamente la muestra fue tomada en INDUCTROC que es una empresa dedicada a la producción de mezcla asfáltica elaborada en caliente.

Figura 2.2: Zona Industrial Inmaconsa-Inductroc



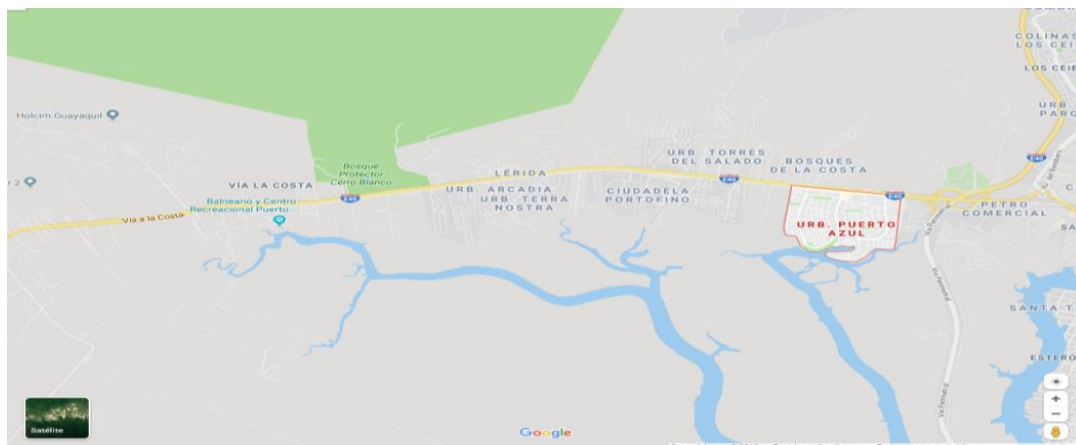
Fuente: Google maps

2.1.2 Sector Cementera:

En este sector, que se encuentra ubicado en la vía a la Costa, se realizó el muestreo de la estación localizada en Puerto Azul, situado estratégicamente, ya que por los alrededores quedan dos importantes cementeras que son Chimborazo y Holcim, estas tienen capacidades de operación altas. Existen también cementeras pequeñas y canteras a lo largo de ese sector, así como también unidades educativas, residencias, y actividades comerciales.

También cabe destacar que a 2 km de distancia de la estación de muestreo está ubicada la Planta Eléctrica de Guayaquil-Salitral.

Figura 2.3: Zona Cementera Puerto Azul

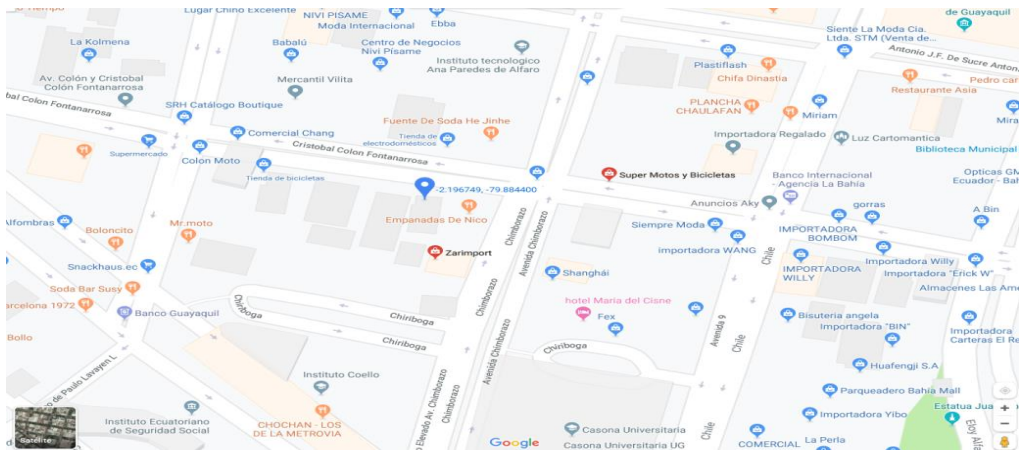


Fuente: Google maps

2.1.3 Sector Centro

El lugar de muestreo que seleccionó en la zona centro, estuvo ubicado en un edificio de las calles Colón y Chimborazo, casco importante de la ciudad ya que es el centro en donde se ubica la mayor actividad comercial, tráfico vehicular, viviendas, hoteles, restaurantes, bodegas o depósitos, etc. También se desarrolla el turismo, ya que muy cerca se encuentra el Malecón 2000. Cabe recalcar, que el tráfico se da en las horas pico como las 07h00 y 18h00 de lunes a viernes.

Figura 2.4: Zona Centro (Colón y Chimborazo)

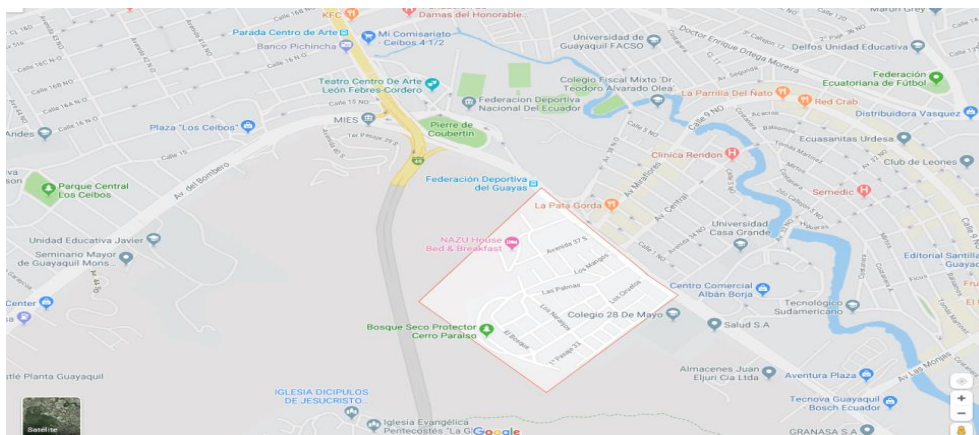


Fuente: Google maps

2.1.4 Sector Residencial:

El muestreo en este sector se realizó en la ciudadela El Paraíso, ubicado en la avenida Carlos Julio Arosemena. Aquí se encuentran los lugares residenciales que tienen urbanizaciones cerradas, libres del tráfico vehicular pero que se encuentran colindando con vías principales o autopistas. También se encuentra cerca de la estación de muestreo el Bosque Seco Protector Cerro Paraíso, centros comerciales grandes y restaurantes, actividad comercial y nocturna, universidades, instituciones educativas, etc.

Figura 2.5: Zona Residencial (Ciudadela el Paraíso)



Fuente: Google maps

2.2 Variables meteorológicas

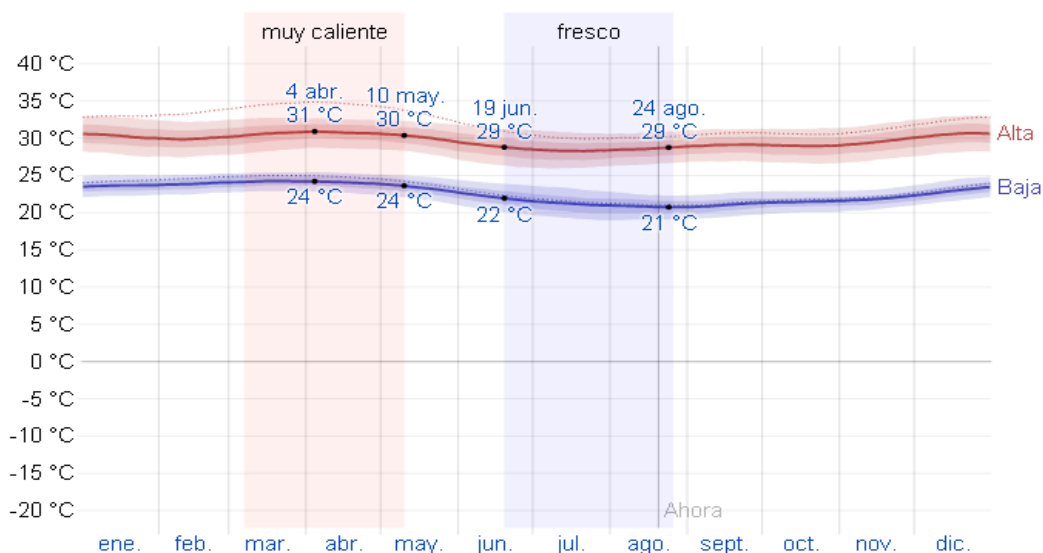
Se analiza una breve descripción del comportamiento (espacial, estacional y horario) de la ciudad de Guayaquil, de las variables meteorológicas medidas en las estaciones de los cuatro sectores ya nombrados anteriormente, y que son motivo de análisis en este documento, siendo estas: Temperatura, Precipitación, Velocidad del Viento, Dirección del viento y Humedad.

2.2.1 Temperatura

En Guayaquil, la temporada calurosa más o menos dura 2 meses, que está entre marzo y mayo, como lo indica la figura 13, alcanzando una temperatura máxima promedio diaria de 30 °C y una temperatura mínima promedio de 24 °C.

En cambio la temperatura fresca dura 2 meses y medio más o menos que comprenden los meses de junio, julio y agosto y corresponde a 29 °C.

Figura 2.6: Temperatura promedio anual



La temperatura máxima (línea roja) y la temperatura mínima (línea azul) promedio diaria con las bandas de los percentiles 25° a 75°, y 10° a 90°. Las líneas delgadas punteadas son las temperaturas promedio percibidas correspondientes.

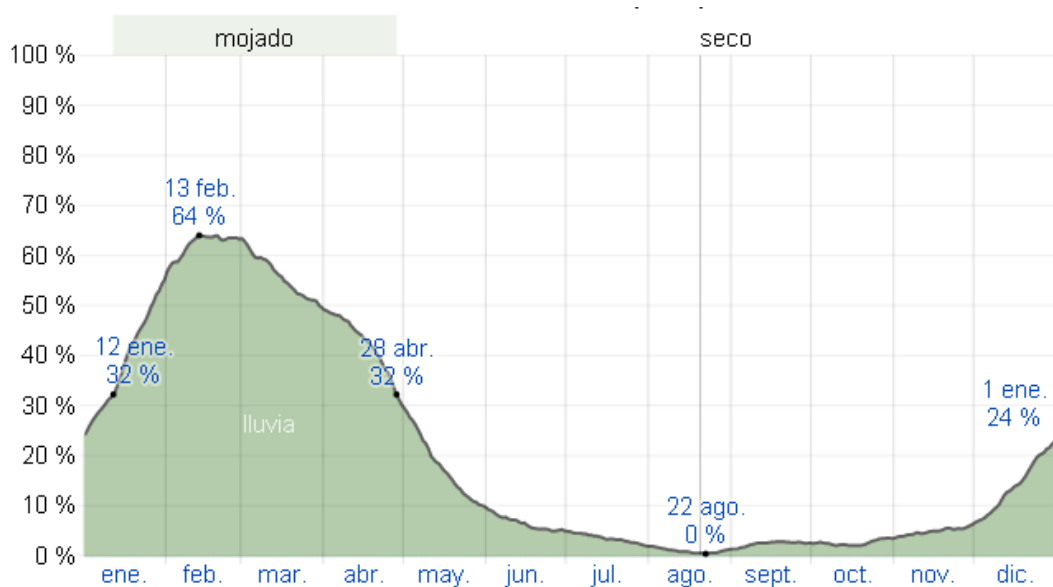
Fuente: Gráfico tomado de es.weatherspark.com

2.2.2 Precipitación

La precipitación de Guayaquil varía según el día y si es muy húmeda o seca, la temporada que tiene más humedad dura aproximadamente de 3 a 4 meses, y la temporada más seca dura ocho meses aproximadamente que comprenden los meses de mayo a diciembre.

EL mes más seco es agosto y la mayor precipitación alcanzada es en el mes de marzo con 199 mm.

Figura 2.7: Precipitación promedio anual



El porcentaje de días en los que se observan diferentes tipos de precipitación, excluidas las cantidades ínfimas: solo lluvia, solo nieve, mezcla (llovió y nevó el mismo día).

Fuente: Gráfico tomado de es.weatherspark.com

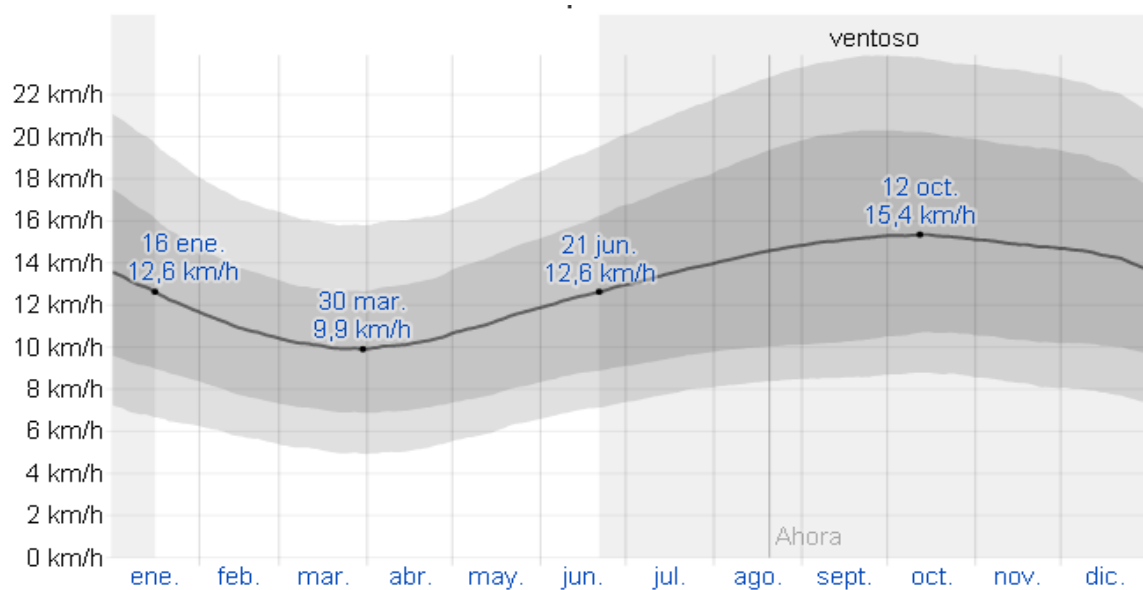
2.2.3 Viento

El viento en la ciudad de Guayaquil es comprendido por la velocidad y dirección del mismo a 10 metros sobre el suelo. Estos factores varían principalmente por la

topografía de la ciudad y otras características que en gran medida se dan según los promedios por hora.

La velocidad promedio del viento por hora en Guayaquil tiene variaciones según la estación en el que se encuentre a lo largo del año, donde mayor viento corre es aproximadamente entre 6 y 7 meses con velocidades promedio de más de 12,6 Kilómetros por hora, como lo indica la figura 2.8.

Donde menos viento corre en el año es en los meses de enero a junio, 6 meses aproximadamente, ya que es muy calmado con una velocidad promedio de 9,9 kilómetros por hora.



El promedio de la velocidad media del viento por hora (línea gris oscura), con las bandas de percentil 25° a 75° y 10° a 90°.

Figura 2.8: Viento promedio anual

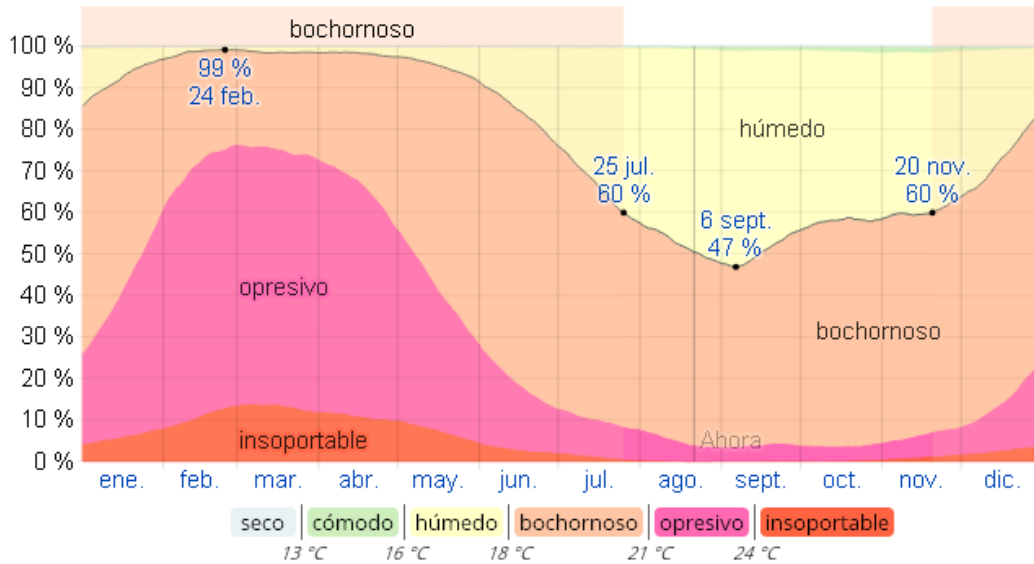
Fuente: Gráfico tomado de es.weatherspark.com

2.2.4 Humedad

El nivel de humedad se lo mide en base al punto de rocío, que determina si el sudor se evapora de la piel, Basamos el nivel de comodidad de la humedad en el punto de rocío, ya que éste determina si el sudor se evaporará de la piel enfriando así el cuerpo.

En la ciudad de Guayaquil, el periodo más húmedo dura 8 meses aproximadamente entre los meses de noviembre a julio, la humedad más alta varía entre un promedio del 99% del tiempo, el mes que presenta menos humedad es septiembre con un promedio de 47%

Figura 2.9: Humedad anual de Guayaquil



El porcentaje de tiempo pasado en varios niveles de comodidad de humedad, categorizado por el punto de rocío.

Fuente: Gráfico tomado de es.weatherspark.com

2.3 Protocolo de muestreo

Para trabajar con la metodología correspondiente se requiere indicar como se dio el proceso de la toma de muestra; esto se realizó para cada tamaño de partícula por separado: $PM_{2.5}$ y PM_1 , en las cuatro estaciones ya nombradas anteriormente, el monitoreo duro 7 días continuos para cada tamaño de partícula, y se recogió muestras en las dos épocas marcadas en la ciudad de Guayaquil que son la de lluvia y la de sequía. Registrándose los valores de concentración de cada 30 minutos durante 24 horas al día, hasta completar la semana. Cabe recalcar que a la vez se midieron las variables meteorológicas: temperatura, dirección y velocidad del viento, precipitación que solo es recolectada durante la época de lluvia. La recolección de los datos se la realizó por sector, primero midiendo $PM_{2.5}$ y continuando con PM_1 de las dos épocas correspondientes, como lo indica la siguiente tabla:

Tabla 2.1: Cronograma de Monitoreo de Material Particulado

Lugar de muestreo	Época seca				Época lluvia			
	PM_1		$PM_{2.5}$		PM_1		$PM_{2.5}$	
	Desde	Hasta	Desde	Hasta	Desde	Hasta	Desde	Hasta
Sector industrial	17-oct-16	24-oct-16	24-oct-16	30-oct-16	13-feb-17	19-feb-17	20-feb-17	26-feb-17
Sector residencial	31-oct-16	07-nov-16	07-nov-16	14-nov-16	07-ene-17	13-ene-17	27-feb-17	05-mar-17
Sector centro	16-nov-16	23-nov-16	23-nov-16	30-nov-16	28-ene-17	03-feb-17	04-feb-17	10-feb-17
Sector Cementera	03-dic-16	10-dic-16	10-dic-16	17-dic-16	14-ene-17	20-ene-17	21-ene-17	27-ene-17
							07-mar-17	13-mar-17

Fuente: Tesis Facultad de Ingeniería Marítima

2.4 Métodos estadísticos

Lo que se requiere con este proyecto es centrar el análisis en buscar una asociación entre las variables meteorológicas (precipitación, temperatura, dirección del viento,

velocidad del viento y humedad), y las concentraciones de Material Particulado ($PM_{2.5}$ y PM_1), así como determinar la influencia que pudiera tener; los sectores en donde se analizó, los días de semana y si hubo o no lluvia, de tal manera que se describe a continuación la metodología que se llevó a cabo para llegar a los resultados:

2.4.1 Validación de la base de datos

La validación de la base de datos recolectados en los sitios de monitoreo, se realizó en base a análisis anteriores revisados de series y comportamientos del clima en otras ciudades y países, entre ellos tenemos:

- ✚ Eliminación de datos nulos, ya que la base de PM_1 y $PM_{2.5}$ presentaban registros de 0 en la concentración, por lo cual ese tipo de datos fueron anulados de la serie, también se encontraron registros nulos de Temperatura, Velocidad y Dirección del viento.
- ✚ Así mismo se identificó algunos valores que presentaron un comportamiento fuera del patrón que PM_1 y $PM_{2.5}$ siguen en un periodo de tiempo. En la base de datos se observó mediciones de $PM_{2.5}$ de 800, 900 ó 1000 $\mu g/m^3$, el procedimiento fue eliminar esos valores aberrantes para tener óptimos resultados, recalcando que no representaban ni el 1% de los datos.

2.4.2 Estadística Descriptiva

Para la verificación de los datos después de tener nuestra base en Excel ya validada, el procesamiento se lo realiza en el software libre estadístico RStudio Versión 1.2.1335, lo que se requirió verificar fue el comportamiento de los indicadores, el máximo nos señala los eventos extremos y si es sensible a cualquier suceso, (en el caso de estudio a las variables meteorológicas); el cuartil, que caracteriza el comportamiento de registros cotidianos del contaminante; el promedio, que permite evaluar el cumplimiento de normas de protección a la salud;

el comportamiento típico diario que permite asociar con la intensidad de las actividades antropogénicas, etc.

Mediante graficas de barras, histogramas y series analizamos la información obtenida de la base de datos recolectada en las estaciones de monitoreo.

2.4.3 Análisis Exploratorio de la serie de datos

El análisis exploratorio de la base de datos recolectada en las 4 estaciones de monitoreo (Cementeras, Centro, Industrial y Residencial), de la ciudad de Guayaquil, se hizo bajo el programa estadístico RStudio de acuerdo a las necesidades de los análisis Este consistió en realizar diagramas de dispersión para cada una de las variables, correlacionar las variables a través de la matriz de Pearson, para las variables cuantitativas y spearman para las cualitativas, series de vientos, precipitación, temperatura y concentración $PM_{2.5}$ y PM_1 para observar su comportamiento a través del tiempo, realizar un proceso de diferenciación en la base de datos para mejorar el comportamiento de la serie.

2.4.4 Comparaciones múltiples

Se realizaron las comparaciones múltiples considerando el ajuste de Bonferroni para controlar el nivel de confianza simultáneo del conjunto de datos, considerando que la probabilidad de al menos uno de los intervalos de confianza no contenga el parámetro de población.

2.4.5 Correlación Cruzada:

Para la determinación de la relación entre las series de Pm_1 y $PM_{2.5}$ se aplicó el método de correlación cruzada, debido a que es uno de los métodos más adecuados al momento de correlacionar series tomadas en distintos tiempos, así lo reflejan también estudios realizados en climatología.

La mayoría de las veces, lo que se utiliza para esta clase de investigaciones es el coeficiente de correlación lineal Pearson, que son para datos cuantitativos y normales, o podríamos utilizar los no paramétricos como Spearman o Kendall.

En este proyecto utilizamos la correlación cruzada entre las dos series estacionarias X_t , en un tiempo dado t , y otra serie Y_{t+k} , ya que los datos recolectados no presentan un comportamiento normal y lo que se requiere es relacionar las variables independientes, con la variable dependiente.

Para este procedimiento se calcularon las correlaciones cruzadas en diferentes retardos entre la concentración de partículas PM_1 y $PM_{2.5}$ mensual en base a las variables meteorológicas de estudio.

Por lo tanto, lo primero que se realizó fue comprobar que ambas series sean estacionarias, ósea que sus medias y varianzas, sean independientes del tiempo.

Luego se correlacionan las variables meteorológicas con $PM_{2.5}$ y PM_1 , buscando sus grados de asociación, y también correlacionarse entre ella, por último, graficar los autocorrelogramas con el fin de observar con precisión su estacionalidad o estacionariedad.

2.4.6 Diseño del modelo

Al elegir diseñar el modelo, se recalca que nos enfocaremos en el modelo de regresión no paramétrica para estimar el valor promedio de la variable respuesta (material particulado PM_1 y $PM_{2.5}$), en función de las variables meteorológica. Cabe recalcar que utilizamos este método porque el contaminante en estudio no se comporta de una forma lineal y se encuentra influenciado por los factores meteorológicos, antropogénicos, etc.

Al igual que el modelo de regresión clásica, una de las formas de expresar el modelo de regresión no paramétrica

Debido a los datos disponibles y la finalidad del modelo, se ha elegido un modelo tipo no paramétrico en base a estimadores llamados suavizadores, cuyo parámetro de suavización es λ , estos modelos surgen cuando la relación entre la predictora y

la respuesta (en el caso de variables numéricas) no se puede escribir de forma lineal, sino más bien a través de una función desconocida.

2.4.6.1 Funciones de suavizado

En este proyecto se utilizarán las funciones de suavizado que permiten capturar el comportamiento de las concentraciones de PM_1 y $PM_{2.5}$ con las variables meteorológicas.

Para ello es necesario utilizar las funciones de predicción y así obtener el modelo resultante.

El modelo de suavizado que utilizaré para una variable predictora y una respuesta viene dado por:

$$Y = s(X) + \epsilon$$

En el caso de estudio las variables predictoras son, X_1 , X_3 , X_4 , X_5 y X_6 de tipo numérico que serían: temperatura, velocidad del viento, dirección del viento, humedad, precipitación y eventos antropogénicos para ellos planteo los modelos saturados siguientes:

$$M0: Y \sim X_1 + X_2 + X_3 + X_4 + X_5 + X_6$$

$$M1: Y \sim s(X_1) + X_2 + X_3 + X_4 + X_5 + X_6$$

$$M2: Y \sim s(X_1) + s(X_2) + X_3 + X_4 + X_5 + X_6$$

$$M3: Y \sim s(X_1) + s(X_2) + s(X_3) + X_4 + X_5 + X_6$$

$$M4: Y \sim s(X_1) + s(X_2) + s(X_3) + s(X_4) + X_5 + X_6$$

$$M5: Y \sim s(X_1) + s(X_2) + s(X_3) + s(X_4) + s(X_5) + X_6$$

$$M6: Y \sim s(X_1) + s(X_2) + s(X_3) + s(X_4) + s(X_5) + s(X_6)$$

Este modelo lo trabajaremos en el Programa Rstudio con la librería *mgcv* que utiliza las funciones de suavizado.

La función de suavizado que se utilizará son los denominados splines penalizados o p-splines, que en R es la función *gam*, que tiene la siguiente estructura:

$$s(\text{variable}, k=, m=, bs="ps", by=\text{factor}),$$

donde k es el tamaño de la base, en el caso de estudio sería 20 ya que la muestra es grande, m toma el valor 2 y como base de splines los penalizados que proporcionan mejor ajuste.

2.4.6.2 Bondad del ajuste

Para la bondad de ajuste del modelo se basa en el estadístico AIC y GCV, que se utilizan para estimar la curva de suavizado asociada a la variable predictora, si el valor resultante de los estadísticos es menor mejor será el modelo obtenido.

Una vez ajustado el modelo, el estudio de los residuos proporciona el análisis diagnóstico del modelo bajo los mismos supuestos de independencia, normalidad y homogeneidad de varianzas.

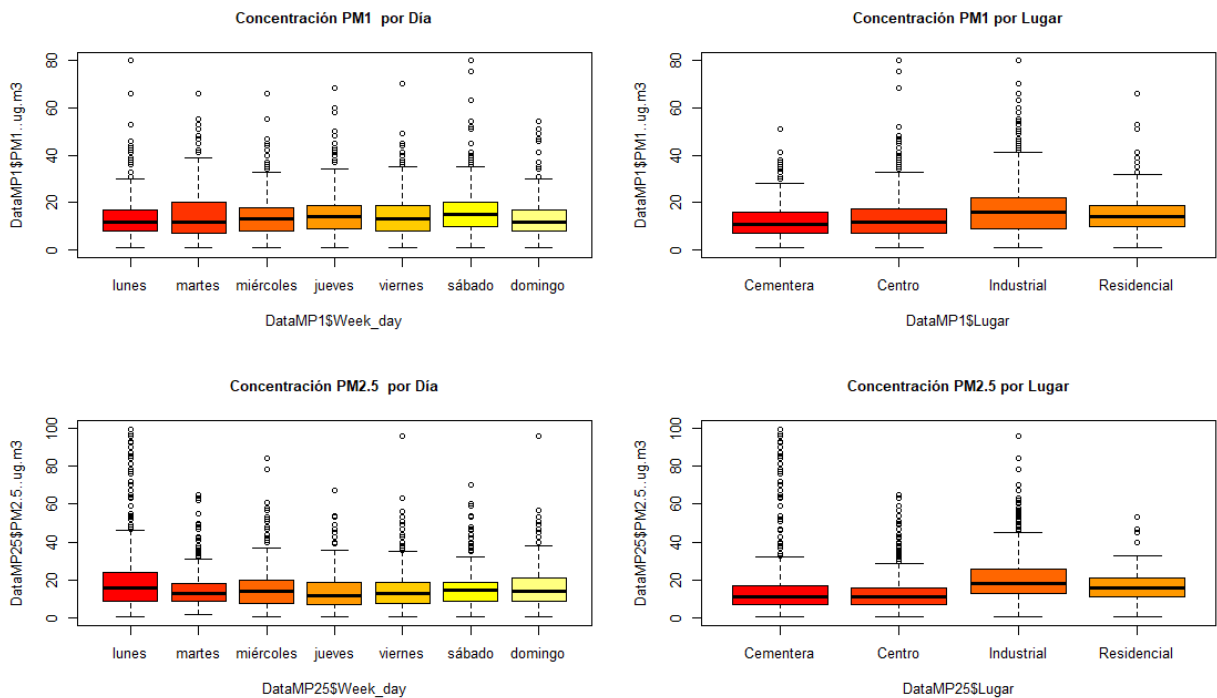
CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

3.1 Descriptiva

Primero analizamos nuestros datos generales de PM1 y PM2.5, en el figura 3.1, denotamos la concentración de PM1 por días de la semana y por lugar o sector donde fueron tomadas las muestras, así mismo la concentración de PM2.5.

Figura 3.1 Concentración de PM1 y PM2.5 por día y lugar



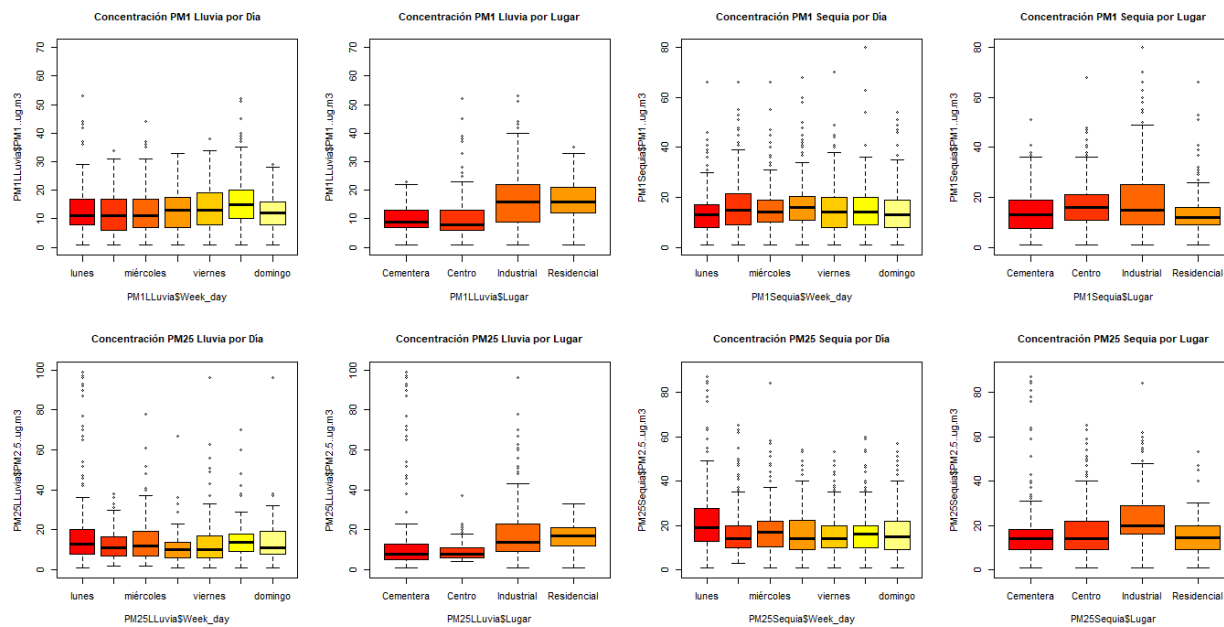
Fuente propia: Gráfica elaborada en Rstudio

Del gráfico se rescata que los datos de la concentración en términos generales, la estación que presenta la mayor dispersión de datos es la del Centro para PM1, y Industrial para PM2.5. Esta dispersión se relaciona con las condiciones atmosféricas y la dinámica de las fuentes aportantes de material particulado en el sector. Adicionalmente, se registraron valores atípicos en todas las estaciones. En cambio

el día que presenta más dispersión en los datos es, para PM1 es el día martes y para PM2.5 los domingos, lunes y jueves. Esto se observó sin tomar en cuenta si es época de lluvia o de sequía.

Para las épocas de lluvia y sequía, se realizaron los diagramas de caja como se observa en la figura 3.2, para material particulado PM1 en lluvia por día y lugar, así como también para sequia por día y lugar, lo mismo para el material particulado PM2.5.

Figura 3.2 Concentración de PM1 y PM2.5 en época lluvia y sequía por día y lugar



Fuente propia: Gráfica elaborada en Rstudio

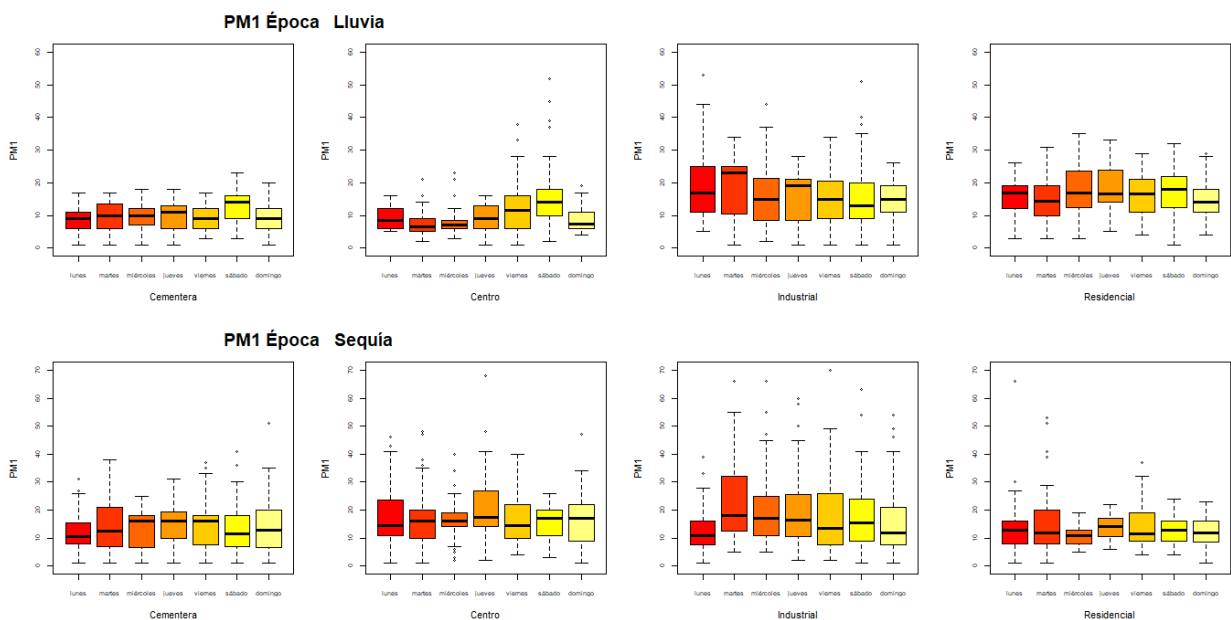
Del gráfico se rescató que los datos de la concentración para material particulado PM1 en la época de lluvia, presento mayor dispersión los lunes en los sectores del centro y cementera; y para la época de sequía la mayor dispersión se presentó el día miércoles en el sector industrial.

En cambio, para el material particulado PM2.5, en época de lluvia, la mayor dispersión se presentó los días domingos en los sectores centro, cementera e

industrial; para el material particulado PM2.5 en la época de sequía, se presentó la mayor dispersión los días jueves y viernes en el sector Industrial. Se registraron valores atípicos en todas las estaciones.

Se representó también los sectores: cementera, centro, industrial y residencial, de donde se recogieron las muestras mediante gráficos representativos tanto para PM1 (figura3.3) y PM2.5 (figura3.4),

Figura 3.3 Concentración de PM1 en época lluvia y sequía por sectores



Fuente propia: Gráfica elaborada en Rstudio

Los resultados para PM1 en la época de lluvia tienen un comportamiento parecido en los sectores de cementera y centro, lo mismo ocurre entre residencial e industrial, el día que más concentración presenta es el domingo en el sector centro, los días viernes en el sector industrial y los días jueves en el sector residencial. En cambio para la época de sequía el comportamiento parecido se da en los sectores de cementera e industrial, y el día que presenta mayor concentración es el día lunes

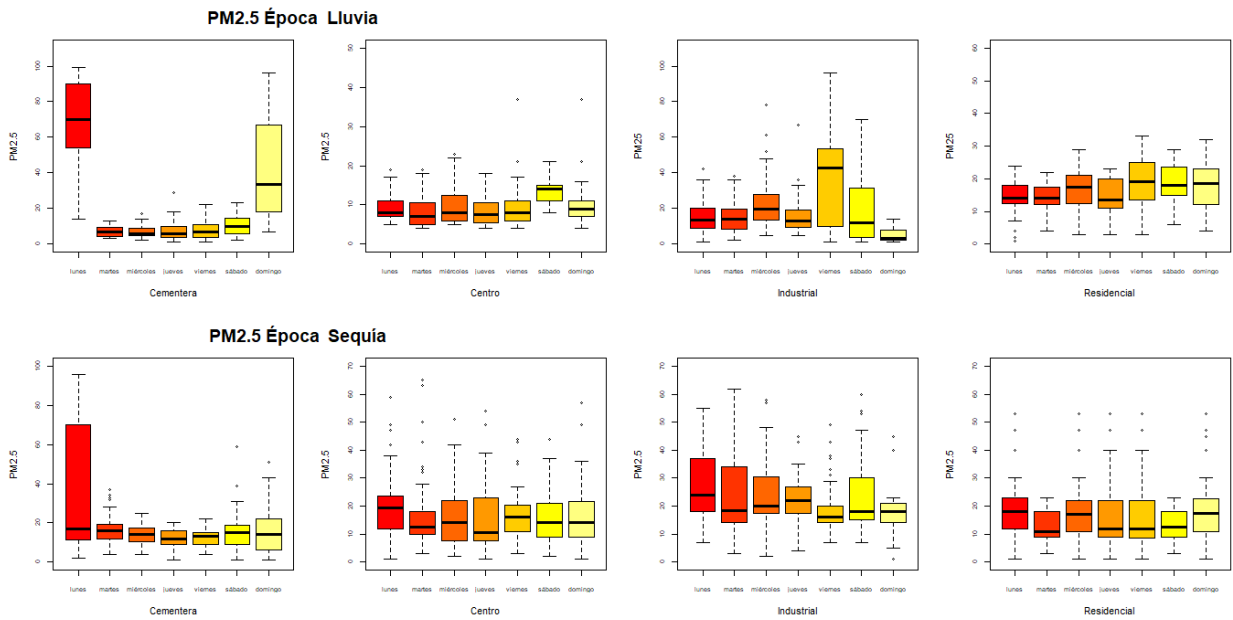
en el sector cementera, el día jueves en el centro, el día domingo en industrial y los viernes en el sector residencial. Esto se da básicamente por las actividades comerciales e industriales que se dan en los diferentes sectores de la ciudad en los cuales se tomó la muestra.

Se realizó el mismo gráfico para el material particulado $PM_{2.5}$ obteniendo el siguiente resultado:

Para la época de lluvia el comportamiento que más se asemeja es entre los sectores centro y residencial, teniendo un diferente comportamiento en los otros dos sectores, con algunos valores atípicos, así mismo los días que más concentración de datos presenta se da los días lunes en los sectores cementera y centro, para el sector industrial los días de mayor concentración son los sábados y domingos y en el sector residencial los días jueves.

En el caso de la época seca, en el sector cementera el día que mayor concentración presenta es el lunes, en el centro los jueves, en el sector industrial los martes y en el residencial se verificó que los días jueves y viernes presentaron la misma concentración.

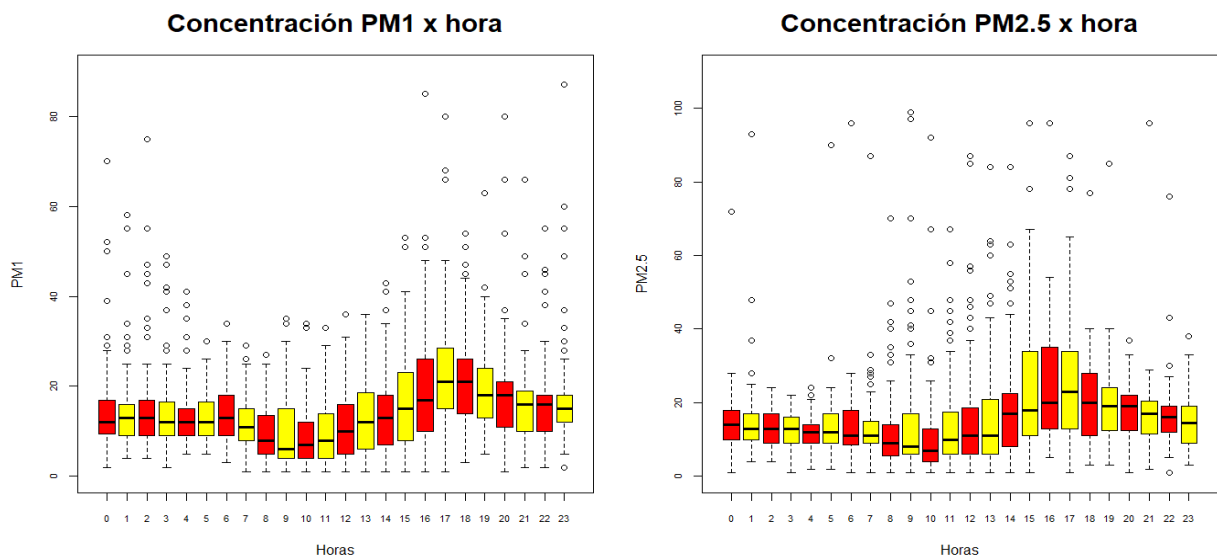
Figura 3.4 Concentración de PM2.5 en época lluvia y sequía por sectores



Fuente propia: Gráfica elaborada en Rstudio

La figura 3.6 ilustra el diagrama de cajas y bigotes de las concentraciones de PM₁ y PM_{2.5} por hora, en general de las 4 estaciones:

Figura 3.6 Concentración de Material particulado PM1 y PM2.5 por hora

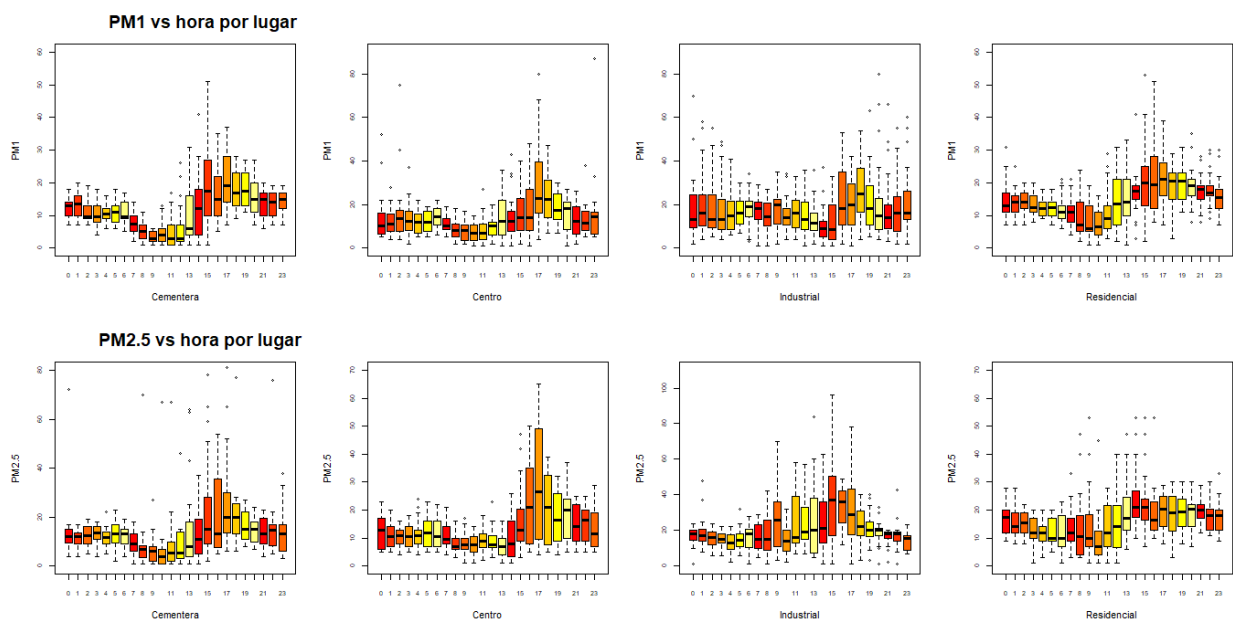


Fuente propia: Gráfica elaborada en Rstudio

Se observó que el comportamiento de cada material particulado es no lineal durante las 24 horas, esto puede ser por el efecto causado por los factores antropogénicos en este horario. Además se denota una heterocedasticidad de las concentraciones de PM_1 y $PM_{2.5}$, cuyas varianzas tienden a cambiar en diferentes horas del día.

Además, se verificó el comportamiento horario de cada material particulado por cada estación, donde se recogieron los datos entre octubre del 2016 y marzo del 2017.

Figura 3.7 Concentración de Material particulado PM_1 y $PM_{2.5}$ por hora



Fuente propia: Gráfica elaborada en Rstudio

En ellos se puede observar que, así como en el general, el comportamiento no es lineal durante las 24 horas del día, siendo las horas de mayor concentración para PM_1 en el sector cementeras entre las dos y tres de la madrugada, a las seis de la mañana y a las trece horas, esto puede deberse a la actividad industrial que realizan las cementeras ubicadas en esa zona. Comportamiento parecido al de $PM_{2.5}$ sumado las seis de la tarde que pueden ser por el cambio de jornada en las fábricas.

En el sector centro las horas de mayor concentración son a las siete de la mañana para PM_1 y entre siete, ocho y doce de la mañana para $PM_{2.5}$.

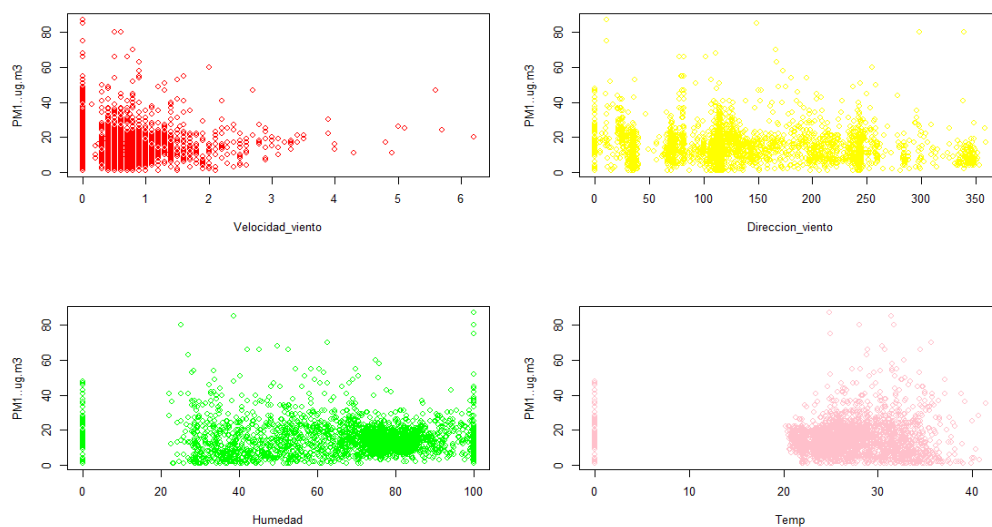
En el sector industrial se denota la mayor concentración en horas de la tarde entre las cuatro y seis para PM_1 y desde las doce hasta las 6 para $PM_{2.5}$ comportamiento parecido en el sector residencial en donde el flujo de material particulado es mayor por las tardes, tal vez debido al tráfico vehicular en esa zona.

3.2 Análisis exploratorio de la serie de datos

3.2.1 Diagramas de dispersión

Como una forma de comenzar a observar la información realizamos los diagramas de dispersión, teniendo en cuenta las inferencias que se hacen por el comportamiento de los datos recalcamos que son superficiales en este ámbito, ya que se necesitaría analizar a fondo los resultados.

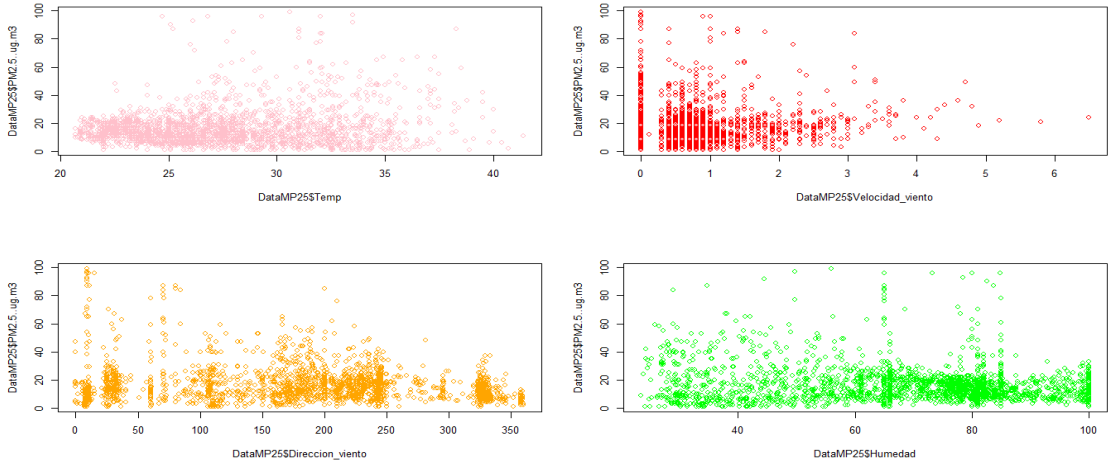
Figura 3.8 Diagramas de Dispersión de la concentración PM_1 vs las variables meteorológicas.



Fuente propia: Gráfica elaborada en el programa Rstudio

En las gráficas se denota que no se existe una asociación entre el Material Particulado PM_1 y las variables meteorológicas recogidas en las estaciones de la ciudad de Guayaquil en general, lo mismo sucede con el material particulado $PM_{2.5}$.

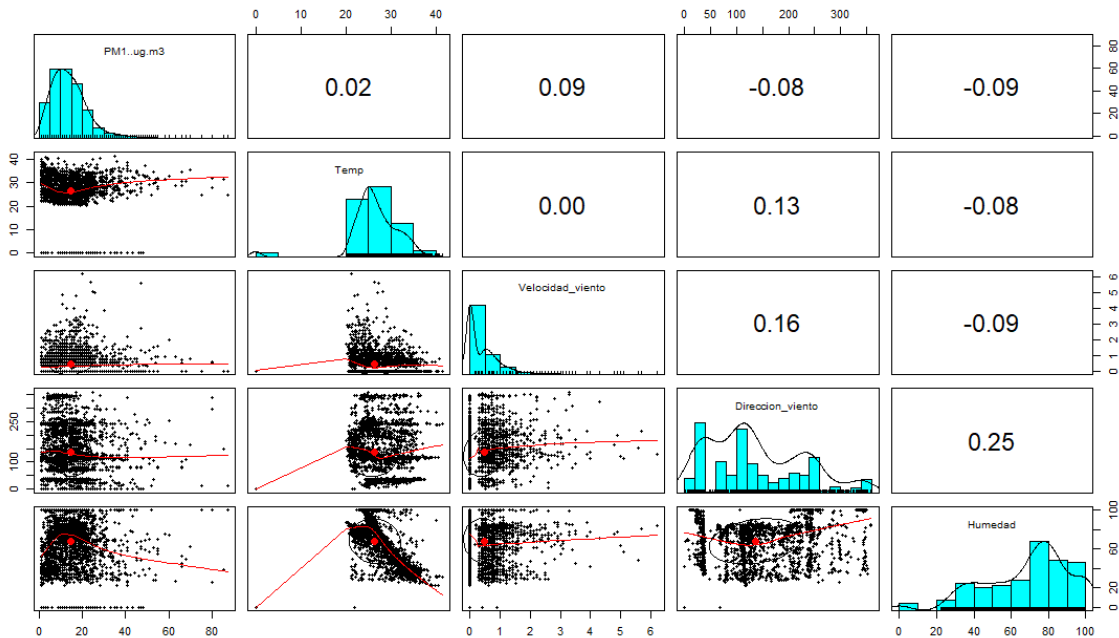
Figura 3.9 Diagramas de Dispersión de la concentración $PM_{2.5}$ vs las variables meteorológicas.



Fuente propia: Gráfica elaborada en el programa Rstudio

En la figura 3.10 se observan las gráficas de dispersión, donde se relacionan las variables meteorológicas con la concentración del Material Particulado entre sí, con el fin de verificar si existe una relación entre ellas, de esta forma se puede establecer que algunas variables se comporten igual ante el PM_1 o $PM_{2.5}$.

Figura 3.10 Panel de diagramas de dispersión de variables meteorológicas y concentración PM1



Gráfica elaborada en el programa Rstudio

Con esta gráfica se puede indicar que de acuerdo a lo observado, no existe una relación marcada entre las variables temperatura, velocidad del viento, dirección del viento y humedad.

Series

Como siguiente paso del análisis exploratorio, se graficó cada una de las variables a evaluar, con el fin de ver su comportamiento a través del tiempo. Mediante estas gráficas se puede observar si la serie tiene tendencia, variaciones estacionales o cíclicas. De acuerdo a lo observado en las figuras 3.11 y 3.12, se puede contemplar lo siguiente:

Figura 3. 11 Gráfica de la serie de tiempo de PM₁ y PM_{2.5}

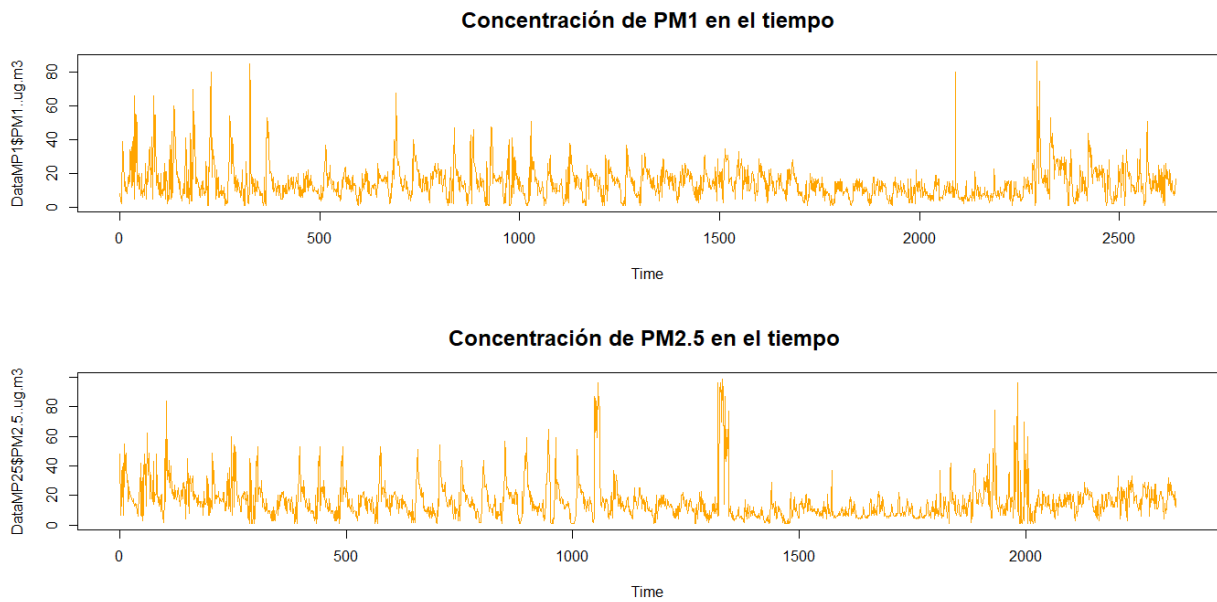
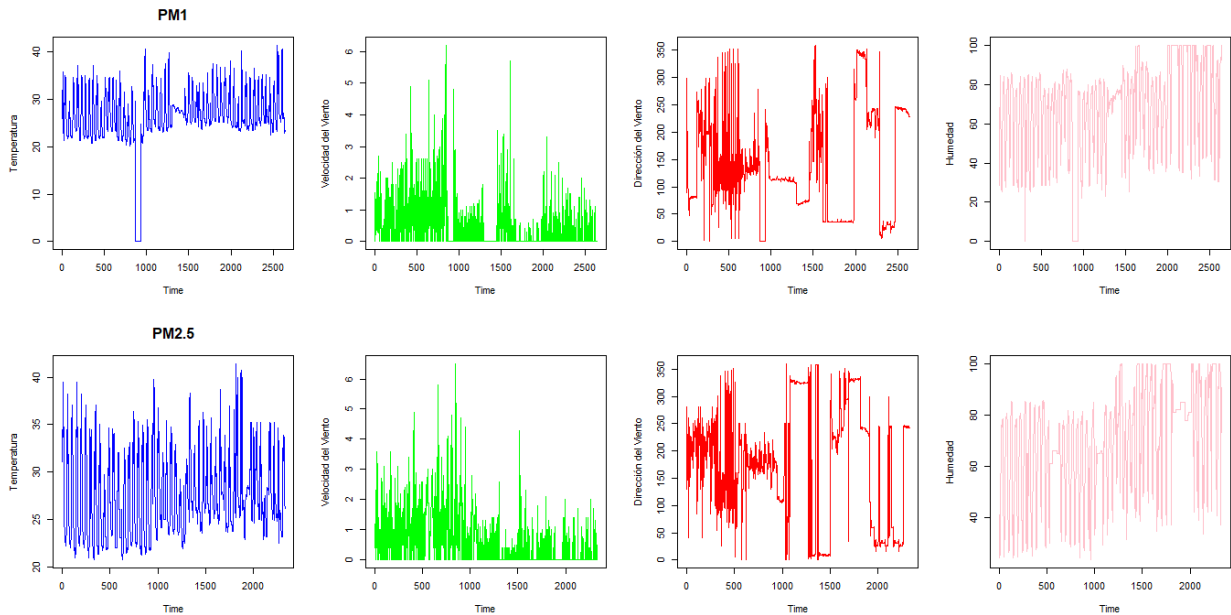


Figura 3. 12 Gráfica de la serie de tiempo de temperatura, velocidad, dirección del viento y humedad tanto para PM₁ como para PM_{2.5}



Las variables de las concentraciones de PM₁ y PM_{2.5}, tienen un comportamiento muy similar en su serie, su variabilidad y media son constantes, por lo cual se puede

inferir que estas variables son estacionarias. La serie de la variable temperatura en la concentración de PM₁ muestra un valor atípico que pudo haber sido causado por un error en la entrada de datos, la serie de humedad en cambio presenta una tendencia ascendente en las dos concentraciones de PM₁ y PM_{2.5}.

En la serie de dirección del viento, se puede rescatar que los datos muestran una variación aleatoria, es decir no hay patrones ni ciclos.

Y por último analizando la serie de velocidad del viento, se observó que tiene un comportamiento parecido tanto en PM₁ como en PM_{2.5}

3.3 Comparaciones múltiples

Para la realización de comparaciones múltiples se utilizó el método de bonferroni, el cual fija un nivel de significación α , que se calcula dividiendo el error global de tipo 1 entre el número de pruebas a realizar.

En las tablas 3.1, 3.2 y 3.3, se observan los resultados de la relación lineal entre la concentración PM₁ y Época, Lugar, Día.

Tabla 3.1: Resumen Modelo Lineal PM₁~Época

Concentración PM₁ vs Época				
Coefficients	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	135.235	0.2620	51.610	< 2e-16
Época Sequía	22.965	0.3735	6.149	8.98e-10

Fuente propia

Tabla 3.2: Resumen Modelo Lineal PM₁~Época

Concentración PM₁ vs Lugar				
Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	119.232	0.3778	31.562	< 2e-16
Lugar Centro	19.729	0.5263	3.748	0.000182
Lugar Industrial	58.299	0.5311	10.977	< 2e-16
Lugar Residencial	30.487	0.5177	5.889	4.37e-09

Fuente propia

Tabla 3.3: Resumen Modelo Lineal PM1~día

Concentración PM1 vs Día				
Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	141.163	0.5075	27.816	<2e-16
Día martes	0.6945	0.7133	0.974	0.3304
Día miércoles	0.1062	0.7078	0.150	0.8808
Día jueves	10.508	0.7073	1.486	0.1375
Día viernes	0.9304	0.7162	1.299	0.1941
Día sábado	17.593	0.7060	2.492	0.0128
Día domingo	-0.7543	0.7021	-1.074	0.2827

Fuente propia

Una vez se tienen los modelos, se identifica si la concentración del material particulado PM₁ varía según época, lugar y día, en los resultados obtenemos que las tres variables tienen efecto en la concentración estudiada como lo indica en la tabla 3.4. que contiene los resultados del ANOVA, el cual rechaza la hipótesis nula. Después se verifica el modelo mediante el LSD.test con el método de Bonferroni,

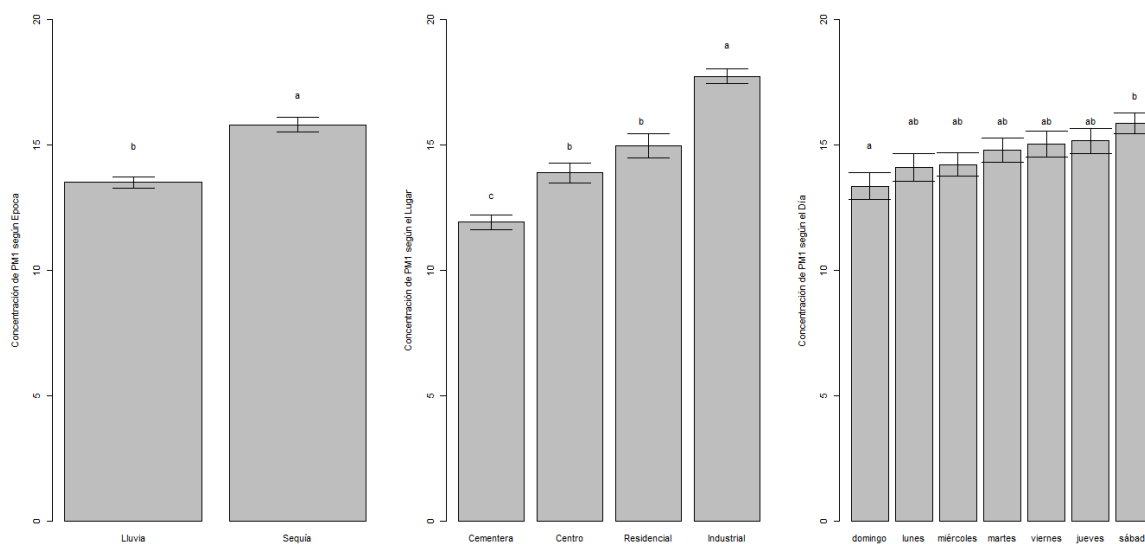
Tabla 3.4: Tabla de análisis de varianza

ANOVA PM1						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significancia
Época	1	3481	3481.3	37.809	8,98E-07	***
Residuals	2639	242986	92.1			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Lugar	3	11261	3753.7	42.084	< 2.2e-16	***
Residuals	2637	235207	89.2			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Día	6	1577	262.825	28.269	0.009582	**
Residuals	2634	244891	92.973			

Fuente propia

Una vez realizado el test de bonferroni, se observa mediante la figura 3.13, tomando en cuenta que se trabajó con este test por el número pequeño de comparaciones que se realizó.

Figura 3. 13 Gráfica test de Bonferroni para la concentración de PM1



Según el gráfico se observó que la concentración PM₁ en la época lluvia y sequía difieren significativamente, en el caso del lugar industrial y cementera difieren significativamente, y respecto a los días los que difieren significativamente es el sábado y domingo.

Para la concentración PM_{2.5} se realizó el mismo procedimiento obteniendo los siguientes resultados:

Tabla 3.5: Resumen Modelo Lineal PM2.5~Época

Concentración PM2.5 vs Época				
Coefficients	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	144.170	0.3900	36.97	< 2e-16
Época Sequía	35.232	0.5282	6.67	3.18e-11

Tabla 3.6: Resumen Modelo Lineal PM2.5~Lugar

Concentración PM2.5 vs Lugar				
Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	157.200	0.5342	29.426	< 2e-16
Lugar Centro	-24.533	0.7272	-3.374	0.000753
Lugar Industrial	54.011	0.7796	6.928	5.49e-12
Lugar Residencial	0.6051	0.7261	0.833	0.404738

Tabla 3.7: Resumen Modelo Lineal PM2.5~Días

Concentración PM2.5 vs Día				
Coefficients:	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	232.448	0.7420	31.326	< 2e-16
Día martes	-80.495	10.082	-7.984	2.21e-15
Día miércoles	-70.620	0.9807	-7.201	8.03e-13
Día jueves	-91.982	0.9910	-9.282	< 2e-16
Día viernes	-82.852	10.028	-8.262	2.38e-16
Día sábado	-71.597	10.062	-7.116	1.48e-12
Día domingo	-73.845	10.628	-6.948	4.79e-12

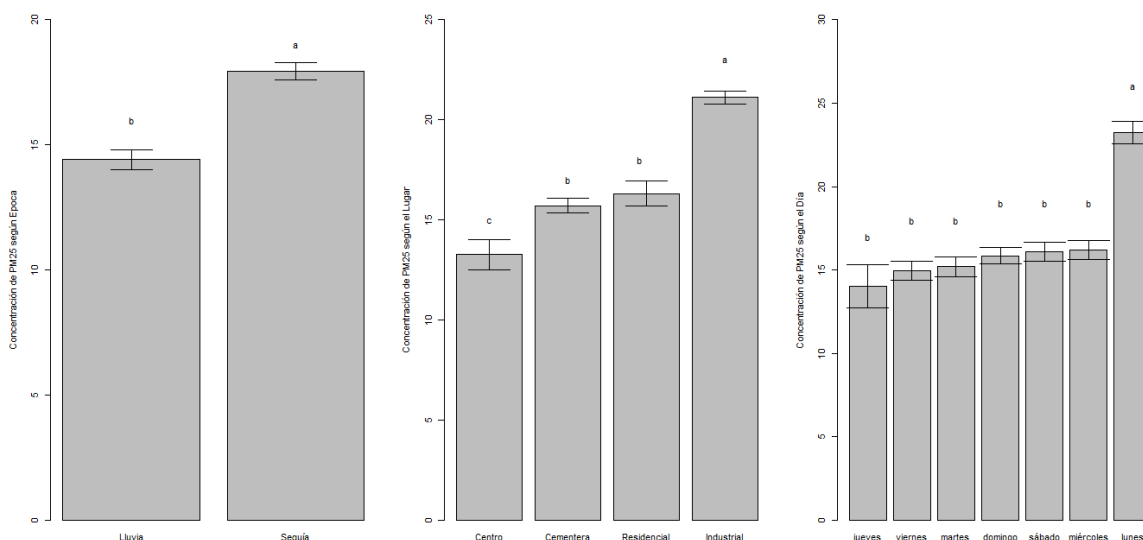
Habiéndose revisado los modelos lineales de la concentración del material particulado PM_{2.5}, se identifica que las tres variables tienen efecto en la concentración estudiada como lo indica en la tabla 3.8. donde las existen diferencias significativas en relación a la época, lugar y día de la concentración PM_{2.5}.

Tabla 3.8: Tabla de análisis de varianza

ANOVA PM2.5						
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significancia
Época	1	7174	7174.5	44.495	3,18E-08	***
Residuals	2329	375533	161.2			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significancia
Lugar	3	17436	5812.1	37.027	< 2.2e-16	***
Residuals	2327	365271	157.0			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significancia
Día	6	16752	2791.93	17.73	< 2.2e-16	***
Residuals	2324	365956	157.47			

Una vez realizado el test de bonferroni, se observa mediante la figura 3.14, los grupos que difieren entre sí.

Figura 3. 14 Gráfica test de Bonferroni para la concentración de PM_{2.5}



Así como el comportamiento que tuvo PM₁ la concentración PM_{2.5} en la época lluvia y sequía difieren significativamente, en el caso del lugar industrial y cementera difieren significativamente, centro e industrial también difieren significativamente y respecto a los días el que difiere significativamente es el lunes, podría explicarse en este caso porque se da inicio a las actividades tanto comerciales como productivas. Cabe recalcar que para un nivel de significancia del 95% el valor crítico de t para cada variable es 1.96 para época, 2.64 para lugar y 3.04 en el día respectivamente.

3.4 Correlaciones

Las correlaciones de las variables meteorológicas, tomando en cuenta que son numéricas, se la realizó en base al estadístico de Pearson, que nos dio como

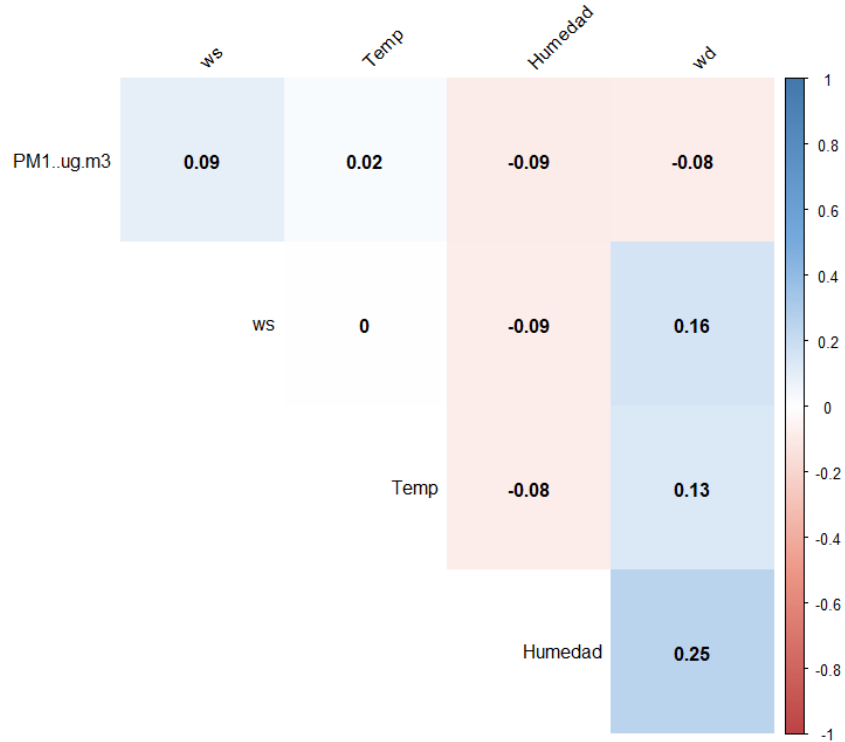
resultado el índice de relación entre las variables cuantitativas, mismos que se pueden observar en la tabla 3.9.

Tabla3.9 Tabla de valores de coeficiente de correlación para PM1.

PM1	t	df	p-value	Interval	cor
Temp	1,11040	2369	0.2669	-0.01654485; 0.05970296	0.02161048
Velocidad del Viento	4,60720	2369	4,27E-03	0.0513600; 0.1270356	0.08932671
Dirección del viento	-4,29470	2369	1,81E-02	-0.12106771; -0.04531302	-0.08331073
Humedad	-4,54600	2369	5,71E-03	-0.12586810; -0.05017661	-0.0881496

Con estos resultados en los coeficientes de correlación se comprueba que no existe una relación lineal entre las variables meteorológicas y el material particulado PM₁. Se puede corroborar los datos de la tabla con la figura 3.15, donde se pueden observar claramente las relaciones entre las variables.

Figura 3. 15 Gráfica Matriz de Correlación PM₁



Esta gráfica denota claramente que no existe correlación entre las variables, sin embargo, el comportamiento de la velocidad del viento denota que se acerca un poco a la relación lineal con PM₁.

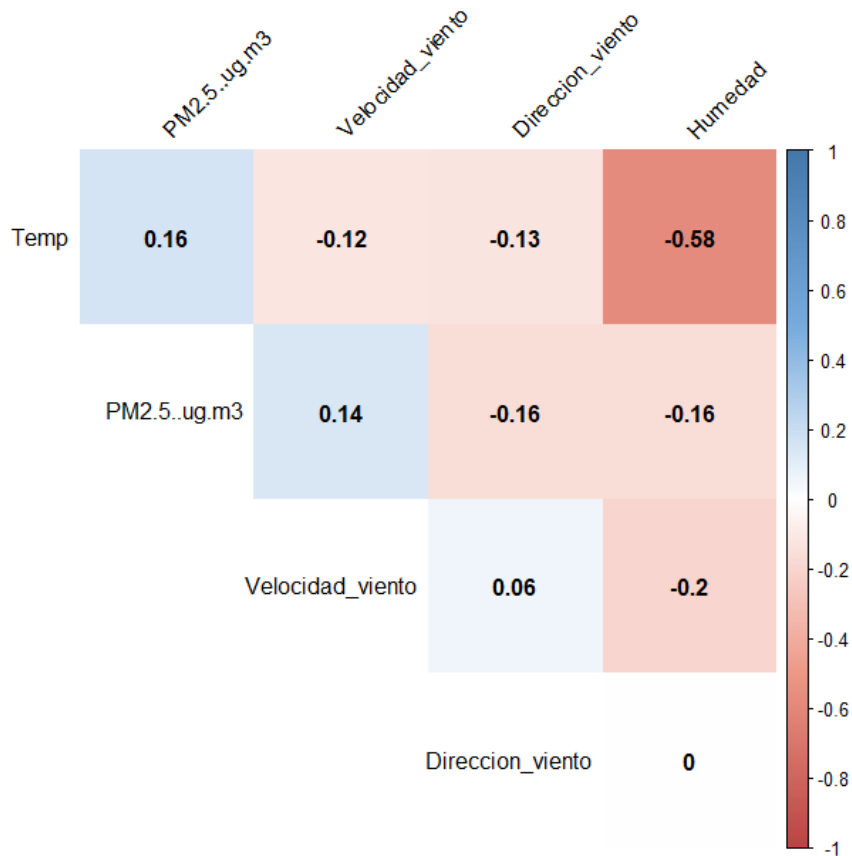
Para PM_{2.5} se realizó el mismo procedimiento obteniendo los siguientes resultados:

Tabla3.10 Tabla de valores de coeficiente de correlación para PM2.5.

PM2.5	t	df	p-value	Interval	cor
Temp	7,69480	2329	2,08E-11	0.1176095; 0.1967981	0.1574569
Velocidad del Viento	6,80750	2329	1,26E-08	0.0996426; 0.1792595	0.1396768
Dirección del viento	-7,84630	2329	6,48E-12	-0.1997747; -0.1206640	-0.1604771
Humedad	-7,83740	2329	6,94E-12	-0.1996001 -0.1204847	-0.1602998

Según la tabla de resultados descrita se comprueba que no existe una relación lineal entre las variables meteorológicas y el material particulado PM₁, al tener un p-value menor al 0.05 que nos hace rechazar la hipótesis nula.

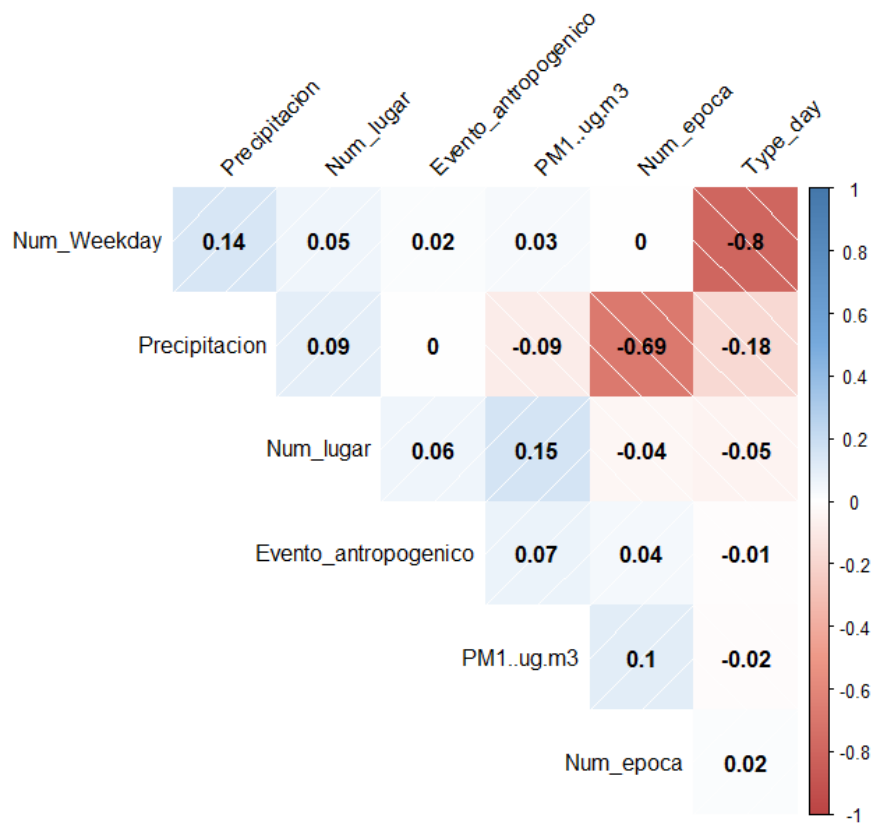
Figura 3. 16 Gráfica Matriz de Correlación PM_{2.5}



Así como se muestra en la figura 3.16 existe una mínima relación entre las variables de la concentración PM_{2.5} y la temperatura y velocidad del viento.

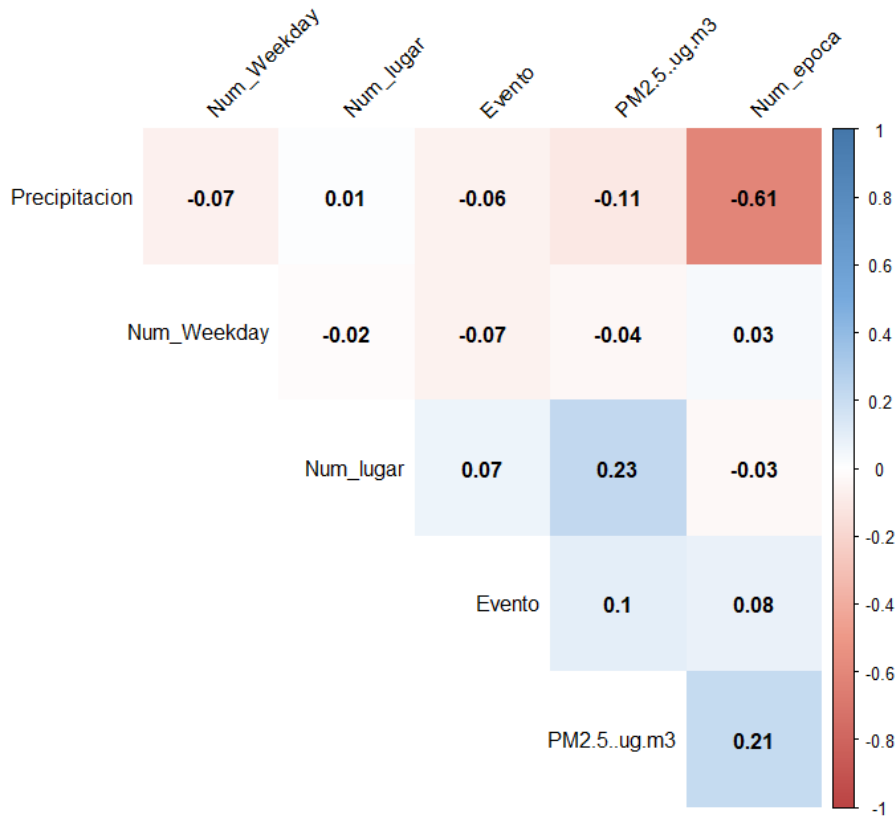
Para las variables continuas u ordinales como lugar, época, tipo de día, evento antropogénicos y si llovió o no, se utilizó el coeficiente de correlación Spearman, examinando la relación entre las variables.

Figura 3. 17 Gráfica Matriz de Correlación PM₁



La gráfica 3.17 nos explica, mediante el coeficiente de correlación Spearman, la relación más fuerte entre las variables, obteniendo como resultado una mínima relación entre la concentración PM₁ y el lugar donde se encontró la estación de muestreo.

Figura 3. 18 Gráfica Matriz de Correlación PM_{2.5}



Para la concentración de PM_{2.5} denotamos la relación existente entre el material particulado el lugar con una coeficiente de 0.23, así como de la época con un coeficiente de correlación Spearman de 0.21. Pero también se analiza la relación negativa de las variables de concentración PM_{2.5} con la precipitación de -0.11, lo que indica que a medida que la precipitación aumente la concentración del material particulado disminuye, esto quiere decir que cuando hay lluvia la concentración se dispersa.

3.5 Correlación cruzada

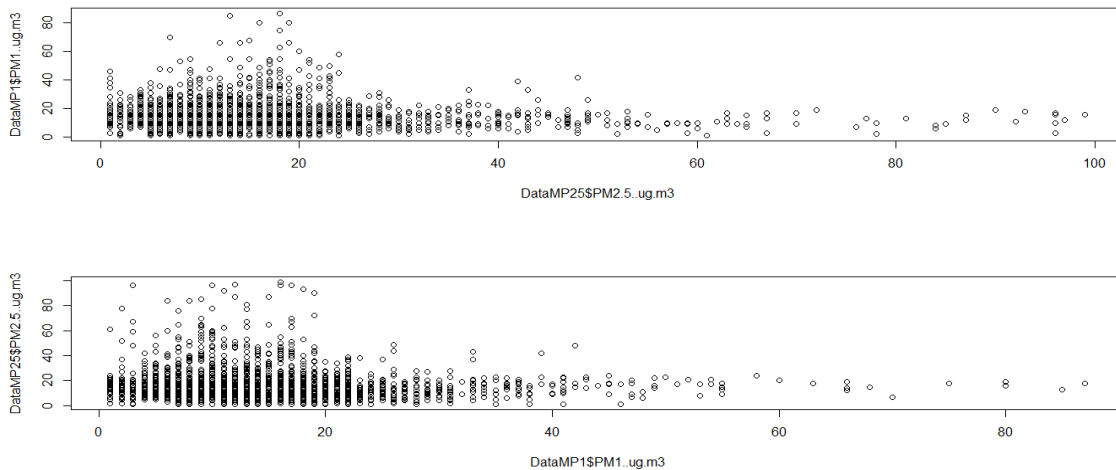
En este proyecto lo que se requiere observar es, la correlación entre dos series temporales tanto en el tiempo simultaneo o desfasadas en el tiempo, así como cuantificar la relación o asociación entre las series o entre una de ella y otras

variables. Para ello el primer paso es asegurarse de que las dos series sean estacionarias, en el apartado 3.2 de este documento se verifico la estacionariedad de las series mediante las gráficas 3.19 y 3.20.

En los resultados, analizados anteriormente, se destaca la estacionariedad debido a que sus medias y varianzas son independiente del tiempo, tanto de la concentración PM_1 como la de $PM_{2.5}$. El test de Dickey-Fuller con un estadístico de (-12.369) para PM_1 y (-9.889) para $PM_{2.5}$, con un 95% de confiabilidad y con un valor $p < 0.05$ se rechaza la hipótesis nula, aceptando la hipótesis alternativa de estacionariedad.

El segundo paso para la correlación de variables es verificar si los residuales presentan un comportamiento normal, como se puede observar en la figura 3.17, donde se realizó el diagrama de dispersión entre las dos variables para verificar linealidad o normalidad.

Figura 3. 19 Diagramas de dispersión entre las variables PM_1 y $PM_{2.5}$



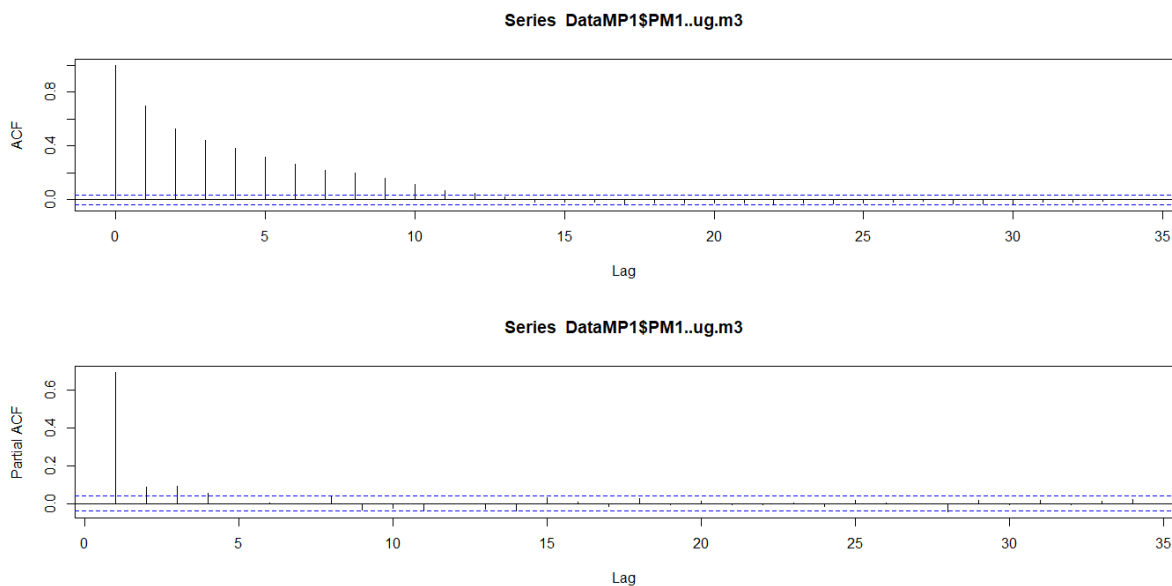
Este gráfico nos indica que no existe una linealidad entre las series de PM_1 y de $PM_{2.5}$, siendo este, un requisito esencial para la regresión.

Además que $PM_{2.5}$ es función de PM_1 , puesto que PM_1 lidera, ya que al aumentar PM_1 aumenta $PM_{2.5}$

Las variables meteorológicas que se analizaron como, temperatura, velocidad del viento, dirección del viento, humedad y precipitación, también se obtuvo estacionariedad en cada una de las series por lo que se dice que son significativamente independientes del tiempo.

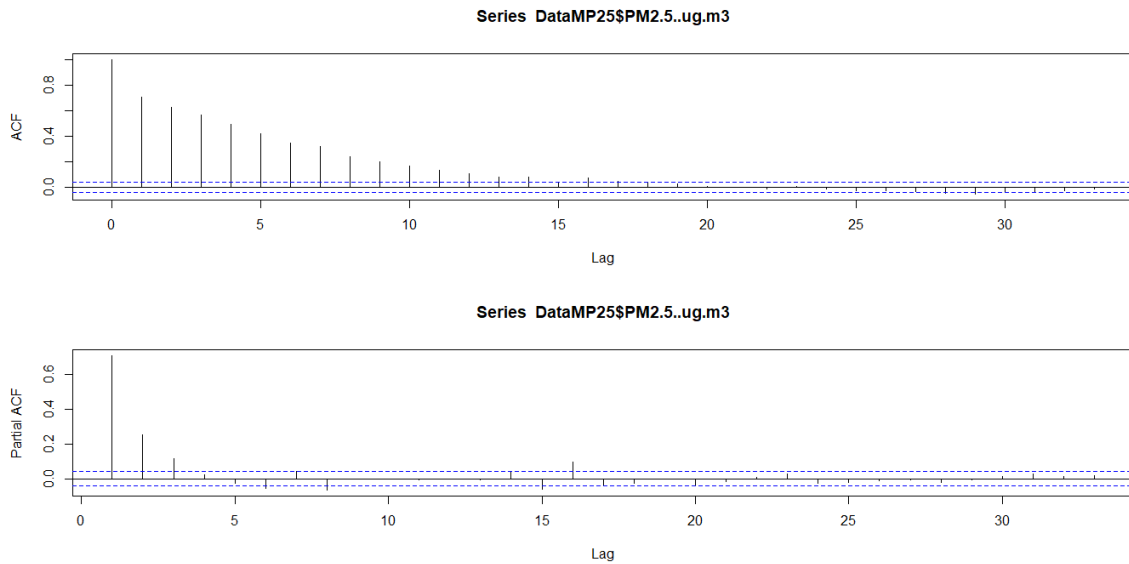
Como siguiente paso se realizó los autocorrelogramas de las series, que indicaron los coeficientes de correlación serial y los lags que son más significativos, así como lo indican las figuras 3.20 para PM_1 3.15 para $PM_{2.5}$.

Figura 3. 20 Autocorrelograma de la serie PM_1



Estos gráficos de Autocorrelogramas ACF y Partial ACF, sirven para comprobar la correlación que sigue la serie a través del tiempo, en el caso de PM_1 no tiene valores significativos, sin embargo algún componente significativo permanece en el retardo 1. Para verificar la estacionariedad se realiza el test de Dickey Fuller aumentada (ADF) y Phillips-Perron (PP), el cual nos da como resultado -12.369 un valor mayor al valor crítico del 5%, esto nos indica que la serie si es estacionaria

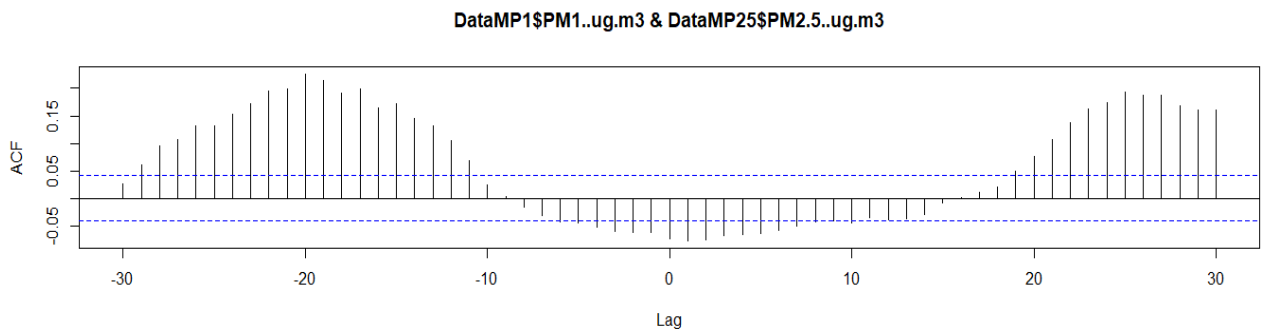
Figura 3. 21 Autocorrelograma de la serie PM_{2.5}



Para el caso de la variable PM_{2.5}, muy parecido al caso de PM₁ con la diferencia que este tiene más tiempo de retardo, sin embargo en las pruebas de Dick Fuller y PP, dio como resultado con un p-value =0.01, -9.889 que afirma que si es estacionaria la serie.

Como se verificó que ambas series son estacionarias, se aplicó la correlación cruzada calculando los coeficientes de correlación con retardos desde el lag -30 hasta el lag 30 (tabla 3.9 y Croscorrelograma respectivo, figura 3.22), según el programa Rstudio.

Figura 3. 22 Autocorrelograma de la serie PM₁ & PM_{2.5}



En esta figura, que muestra el correlograma de la correlación cruzada o también llamado croscorelograma, se aprecia la influencia del desfase temporal en las asociaciones de las dos variables, el lag 0 y los lags o retardos positivos del 1 al 30 en días, indican que la variable independiente PM_1 lidera el proceso y hacia la izquierda del lag 0, los retardos negativos, indican que la variable independiente $PM_{2.5}$, lidera el proceso.

Tabla 3.9 Autocorrelaciones PM_1 & $PM_{2.5}$

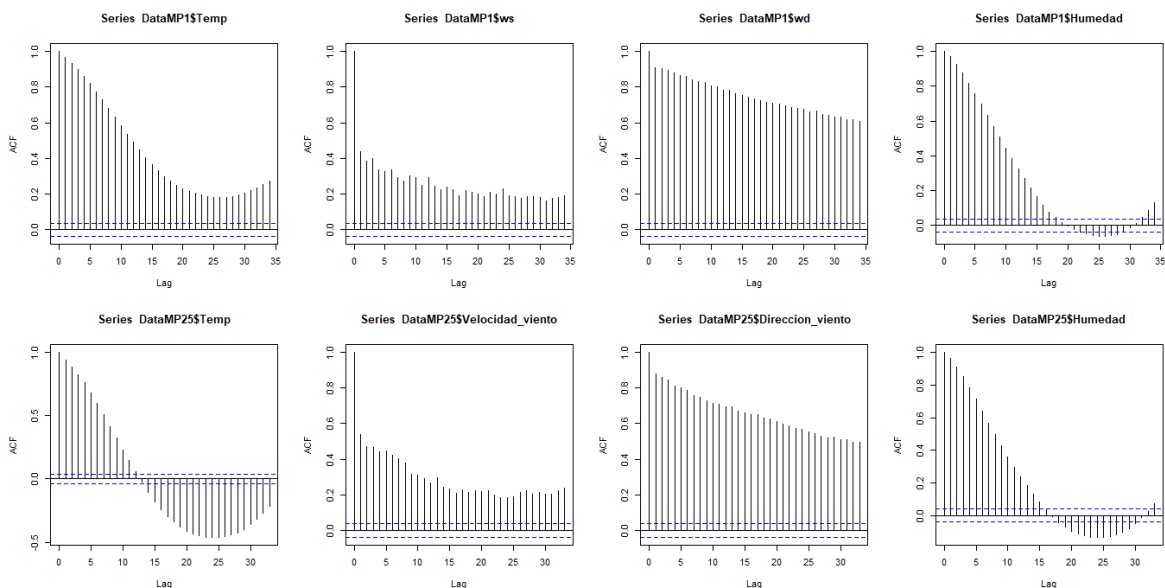
PM1 & PM2.5			
ACF	LAG	ACF	LAG
0.0266506316	-30	-0.0742954281	0
0.0599050963	-29	-0.0779973547	1
0.0952593270	-28	-0.0761896923	2
0.1067800162	-27	-0.0683953574	3
0.1321832115	-26	-0.0667950924	4
0.1327257379	-25	-0.0648914238	5
0.1533200311	-24	-0.0585202679	6
0.1722033185	-23	-0.0511617547	7
0.1945117311	-22	-0.0430831699	8
0.1990467641	-21	-0.0417659289	9
0.2267468806	-20	-0.0451984546	10
0.2144169422	-19	-0.0355167623	11
0.1915432461	-18	-0.0388956423	12
0.1985705492	-17	-0.0375903908	13
0.1638718559	-16	-0.0288043755	14
0.1719557775	-15	-0.0081969164	15
0.1455319700	-14	0.0002744766	16
0.1324848992	-13	0.0100466176	17
0.1044250092	-12	0.0202670882	18
0.0682515174	-11	0.0501665592	19
0.0239383903	-10	0.0770489356	20
0.0040261050	-9	0.1076906301	21
-0.0159675254	-8	0.1371592093	22
-0.0312636982	-7	0.1631068520	23
-0.0425691409	-6	0.1743977176	24
-0.0451799418	-5	0.1941670887	25
-0.0517304310	-4	0.1878882845	26

-0.0595169368	-3	0.1866181870	27
-0.0616975756	-2	0.1676818669	28
-0.0612712316	-1	0.1605977324	29
		0.1614221772	30

Según estos resultados se observa que si la concentración PM_1 se da el día lunes, la de $PM_{2.5}$ será el martes, ósea un día después por el retardo, o viceversa, si PM_1 se da el lunes, $PM_{2.5}$ se da el domingo. Esto quiere decir que $PM_{2.5}$, depende de PM_1 .

Para las variables meteorológicas se realizó los autocorrelogramas de cada una, por cada Data, tanto para PM_1 y para $PM_{2.5}$.

Figura 3. 23 Autocorrelogramas de variables meteorológicas de PM_1 & $PM_{2.5}$



Los correlogramas muestran que las funciones de autocorrelación para las series temporales de: Temperatura, velocidad del viento y dirección del viento de la

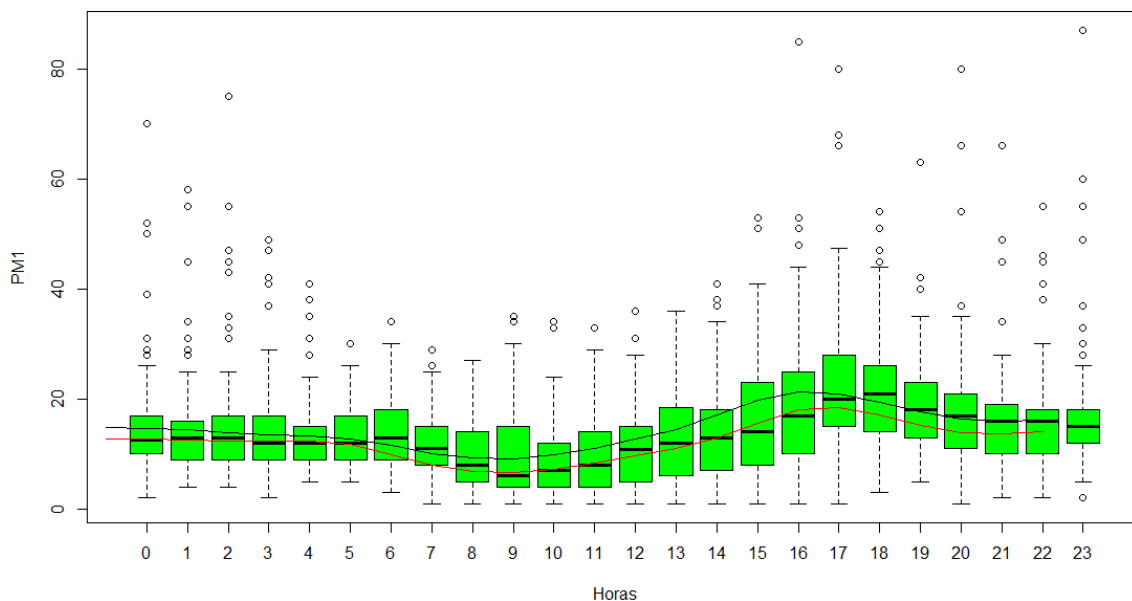
concentración PM_1 son significativas durante todo el tiempo, en cambio la humedad es significativa los primeros 15 días. Con esto se puede concluir que la serie de tiempo dada no es estacionaria.

Para la concentración de $PM_{2.5}$ en cambio, la figura 3.23, nos indica que las funciones de autocorrelación para las series temporales de temperatura y humedad son significativas los primeros 15 días.

3.6 Diseño del modelo

Mediante la figura 3.24 y 3.25 se revisó el comportamiento horario de la concentración del Material Particulado para PM_1 y para $PM_{2.5}$.

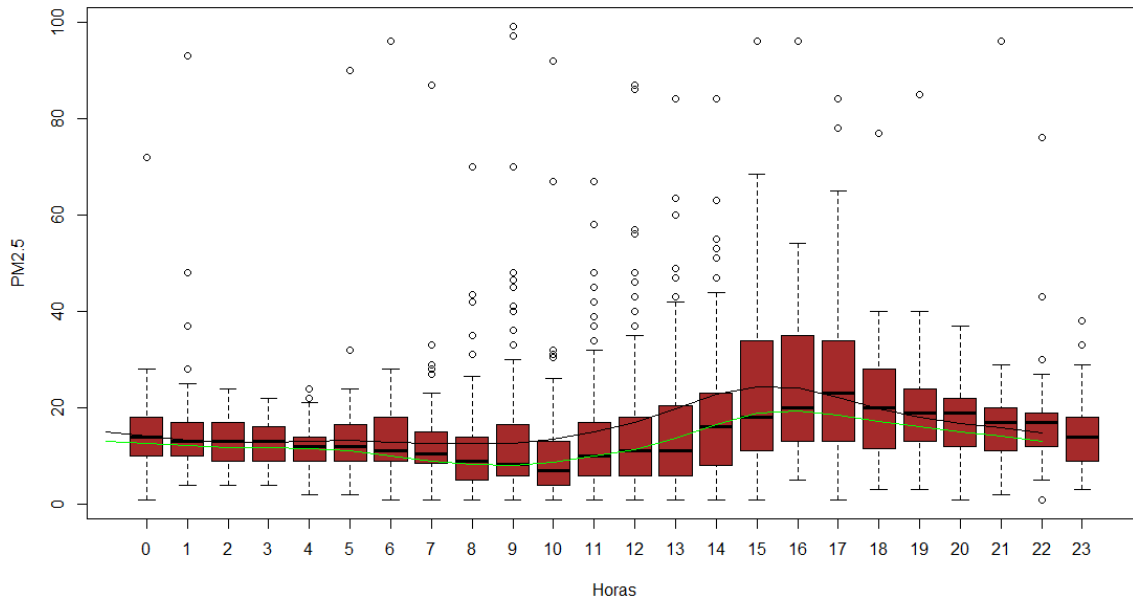
Figura 3. 24 Comportamiento horario de la concentración de PM_1



Para realizar este gráfico se tomó en cuenta el promedio de cada hora, ya que las mediciones recogidas se tomaron cada media hora, lo que se requirió es identificar el comportamiento que sigue la concentración del material particulado a lo largo del día. Partiendo del análisis exploratorio de los datos, el gráfico explica que la mayor concentración horaria del material particulado se da en las horas de la tarde a partir de las 15h00 hasta las 20h00 aproximadamente, esto se podría explicar debido al

tráfico vehicular que existe en las horas picos, por las salidas de los trabajos o el cambio de horario en las actividades industriales.

Figura 3. 25 Comportamiento horario de la concentración de PM_{2.5}



Para el material particulado PM_{2.5} la mayor concentración también se da en las horas de la tarde a partir de las 15h00 hasta las 18h00 aproximadamente.

Cabe recalcar que no se observa un comportamiento lineal durante las 24 horas en las dos medidas del material particulado estudiado, muy probablemente por el efecto causado por las actividades antropogénicas en este horario. Se puede decir, además, que existe una clara indicación de heterocedasticidad de las concentraciones de PM₁ y PM_{2.5} cuyas varianzas tienden a cambiar en diferentes horas del día.

Para el diseño del modelo no paramétrico multivariado, se utilizó métodos de suavización, ya que estos resultan ser los más adecuados cuando la distribución de una variable no sigue una tendencia lineal.

Para el diseño del modelo múltiple, lo primero que se realizó es representar la relación entre las variables respuestas PM1 (figura 3.26), PM2.5 (figura 3.27) y las variables predictoras de tipo numérico (temperatura, velocidad del viento, dirección del viento, humedad, precipitación y factores antropogénicos)

Figura 3. 26 Relaciones entre PM₁ y variables meteorológicas

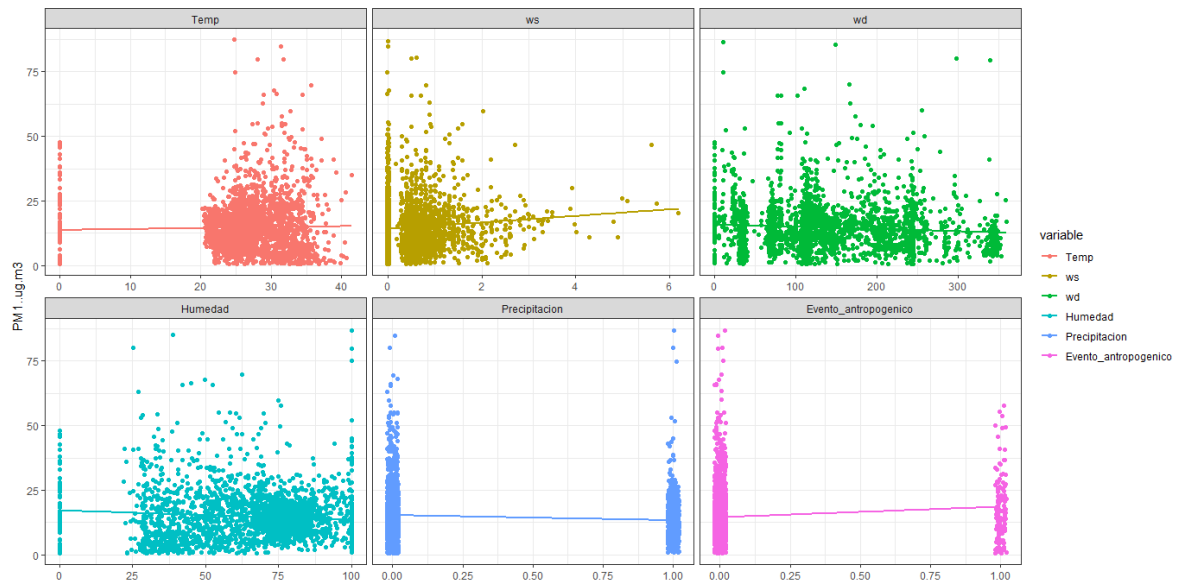
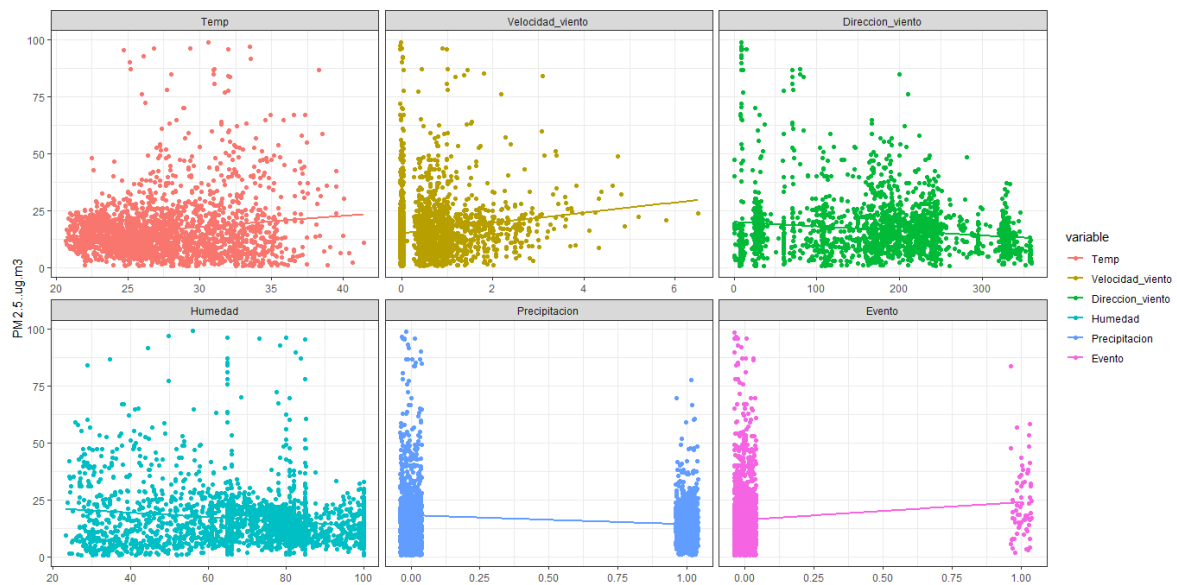


Figura 3. 27 Relaciones entre PM_{2.5} y las variables meteorológicas



Una vez que se establece mediante las gráficas que no existe la relación lineal, se representó el modelo suavizado o también denominado splines penalizados o p-splines, que se da sobre base en polinomios, donde lo que se necesita es el tamaño de la base de polinomios, el orden, tipo de spline, y el factor para las curvas de suavizado.

Como la muestra que tenemos en nuestra base es grande, se escogió el tamaño de la base de polinomios 20 de orden 2 y base de splines penalizados, ya que proporcionan el mayor ajuste.

3.6.1 Construcción del modelo

El análisis consistió en plantear un modelo aditivo, para verificar si es posible eliminar algunos de los efectos. Como ajuste de bondad utilizamos el estadístico GCV y escogemos al más pequeño.

Para la construcción del modelo nos quedamos con las variables más influyentes según los gráficos analizados que son: temperatura, humedad, velocidad y dirección del viento.

El primer modelo resultante es el completo sin suavizar:

```
M1 <- gam(PM1..ug.m3 ~ Temp + ws + wd + Humedad, data = DataMP1)
```

El segundo modelo resultante es el que se le aplica suavizado a la variable temperatura:

```
M2 <- gam(PM1..ug.m3 ~ s(Temp, k = 20, m = 2, bs = "ps") + ws + wd + Humedad + Precipitacion+Evento_antropogenico, data = DataMP1)
```

El tercer modelo resultante es el que se le aplica suavizado a la variable velocidad del viento:

```
M3 <- gam(PM1..ug.m3 ~ Temp + s(ws, k = 20, m = 2, bs = "ps")
          + wd + Humedad, data = DataMP1)
```

El cuarto modelo resultante es el que se le aplica suavizado a la variable dirección del viento:

```
M4 <- gam(PM1..ug.m3 ~ Temp + ws + s(wd, k = 20, m = 2, bs = "ps")
          + Humedad, data = DataMP1)
```

El quinto modelo resultante es el que se le aplica suavizado a la variable humedad:

```
M5 <- gam(PM1..ug.m3 ~ Temp + ws + wd + s(Humedad, k = 20, m = 2, bs = "ps"),
          data = DataMP1)
```

El sexto modelo resultante es el que se le aplica suavizado a las variables temperatura y velocidad del viento:

```
M6 <- gam(PM1..ug.m3 ~ s(Temp, k = 20, m = 2, bs = "ps")
          + s(ws, k = 20, m = 2, bs = "ps") + wd + Humedad, data = DataMP1)
```

El séptimo modelo resultante es el que se le aplica suavizado a las variables temperatura, velocidad del viento y dirección del viento:

```
M7 <- gam(PM1..ug.m3 ~ s(Temp, k = 20, m = 2, bs = "ps")
          + s(ws, k = 20, m = 2, bs = "ps") + s(wd, k = 20, m = 2, bs = "ps")+
          Humedad, data = DataMP1)
```

El octavo modelo resultante es el que se le aplica suavizado a las variables temperatura, velocidad del viento, dirección del viento y humedad:

```
M8 <- gam(PM1..ug.m3 ~ s(Temp, k = 20, m = 2, bs = "ps")
          + s(ws, k = 20, m = 2, bs = "ps") + s(wd, k = 20, m = 2, bs = "ps")+
          Humedad, data = DataMP1)
```


s(Humedad, k = 20, m = 2, bs = "ps"), data = DataMP1)

Los GCV del modelo se establecen en la tabla 3.10, en donde se explica el estadístico de cada modelo construido:

Tabla 3.10 GCV de los modelos para la concentración PM₁

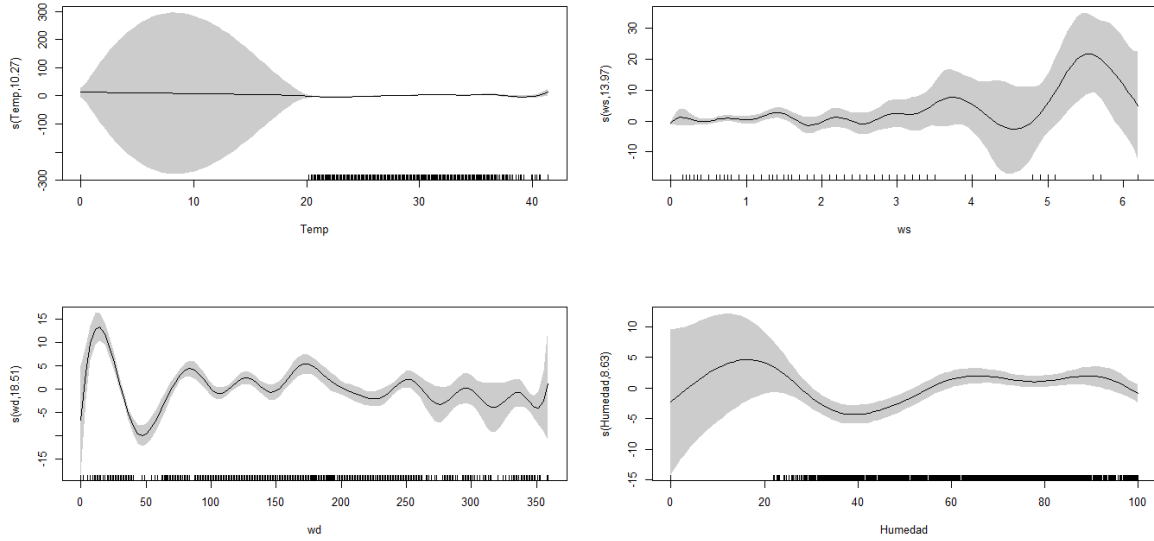
Estadístico	M1	M2	M3	M4	M5	M6	M7	M8
GCV	91,64841	89,1271	91,5862	83,6995	89,3478	88,9293	81,72194	80,24504

Como se dijo anteriormente el modelo que explica mejor es el que tiene el estadístico de menor valor, por lo tanto el modelo resultante es el 8. Sin embargo en estas situaciones es recomendable ajustar un suavizado conjunto para ambas variables.

Al aplicar el suavizado conjunto el mejor modelo sigue siendo el de suavizado individual para todas las variables, es decir el modelo completo con un GCV de 80,24504.

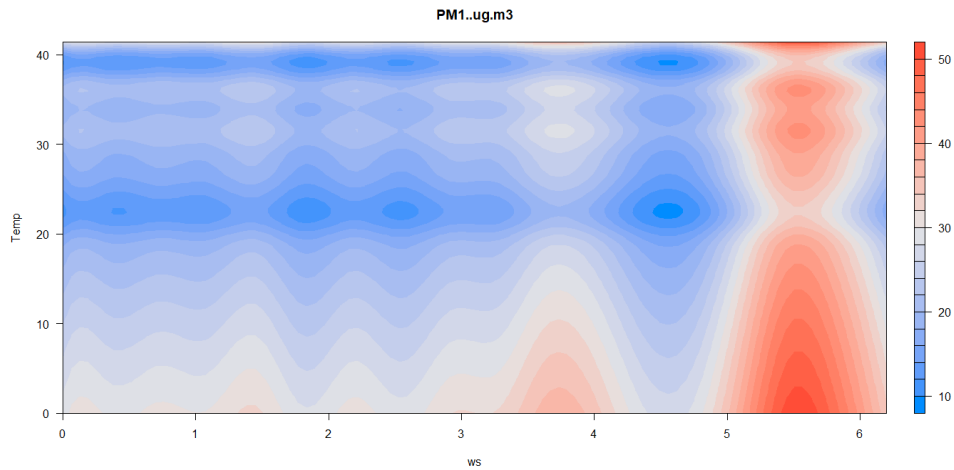
Al hacer la predicción de modelo obtenemos las curvas típicas de las concentraciones de PM₁ y bandas de variabilidad respecto a las variables meteorológicas estudiadas según los datos recogidos en las cuatro estaciones de muestro en época de lluvia y sequía. Mediante la figura 3.28 se puede observar el efecto generado por cada una de las variables meteorológicas, notando que la variable que explica mejor el modelo es la de velocidad del viento, ya que tiende a la linealidad.

Figura 3. 28 Gráfica de modelo ajustado



En la figura 3.29 se visualiza mejor la relación de suavizado de la variable velocidad del viento con Temperatura.

Figura 3. 29 Gráfica de pares



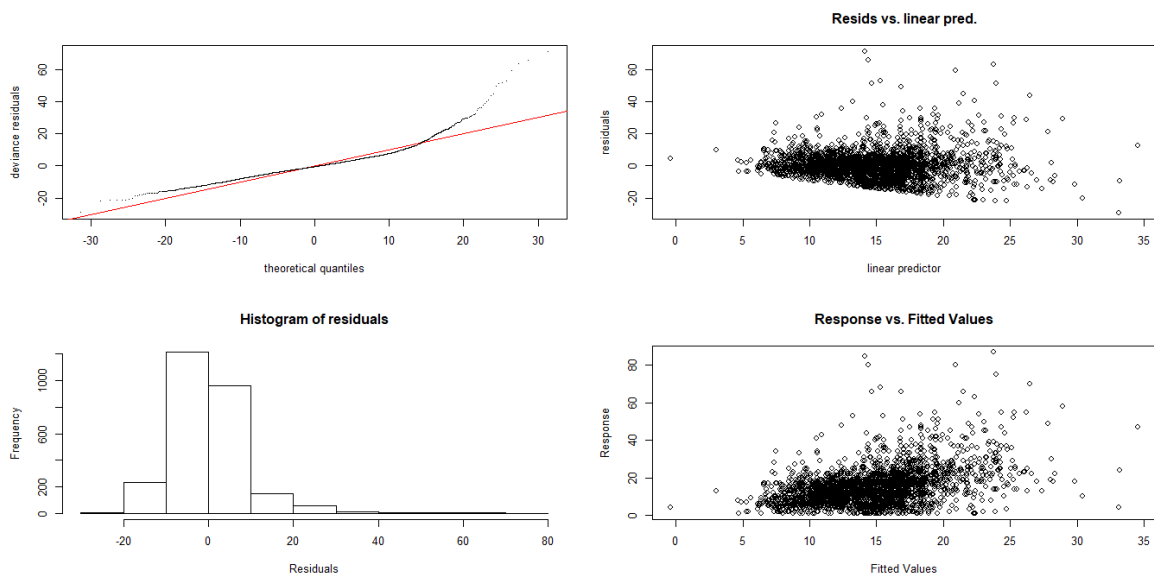
3.6.2 Diagnóstico del modelo

Para la hipótesis del modelo se utiliza las mismas del modelo de regresión, es decir se verifica normalidad y varianza constante de los residuos del modelo.

Mediante la función `gam.check` de Rstudio podemos obtener los gráficos (figura3.30), que explican el modelo de regresión.

Para el diagnostico se utilizó el modelo ajustado:

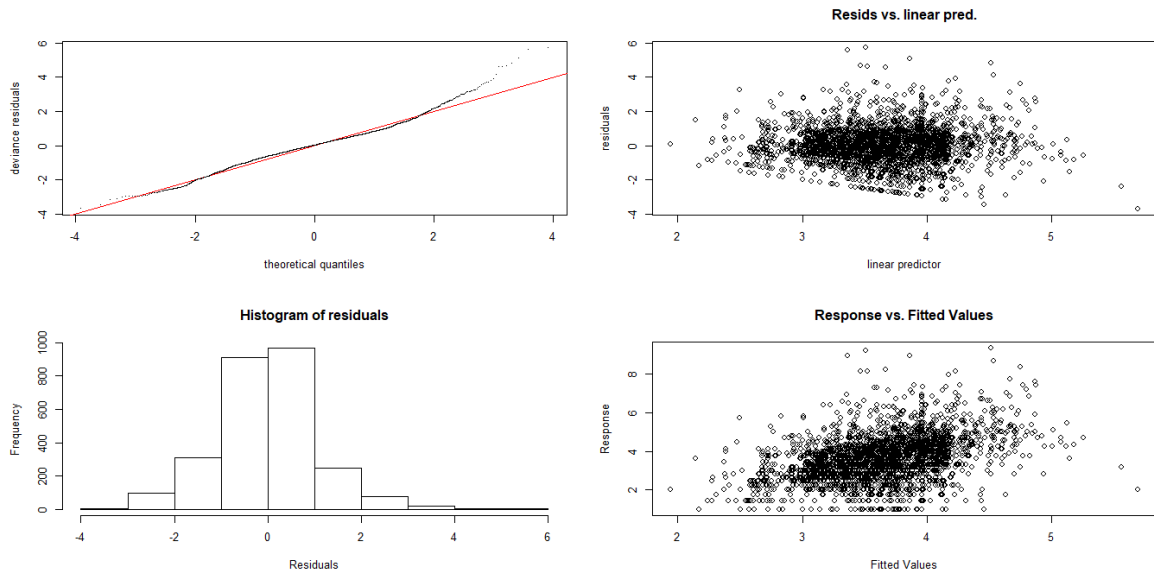
Figura 3. 30 Gráficos de diagnóstico del modelo



El modelo incumple la hipótesis de varianza constante ya que la variabilidad de los residuos aumente conforme al predictor lineal, por ello se aplicó la transformación de raíz cuadrada al modelo, y se ajusta un nuevo modelo.

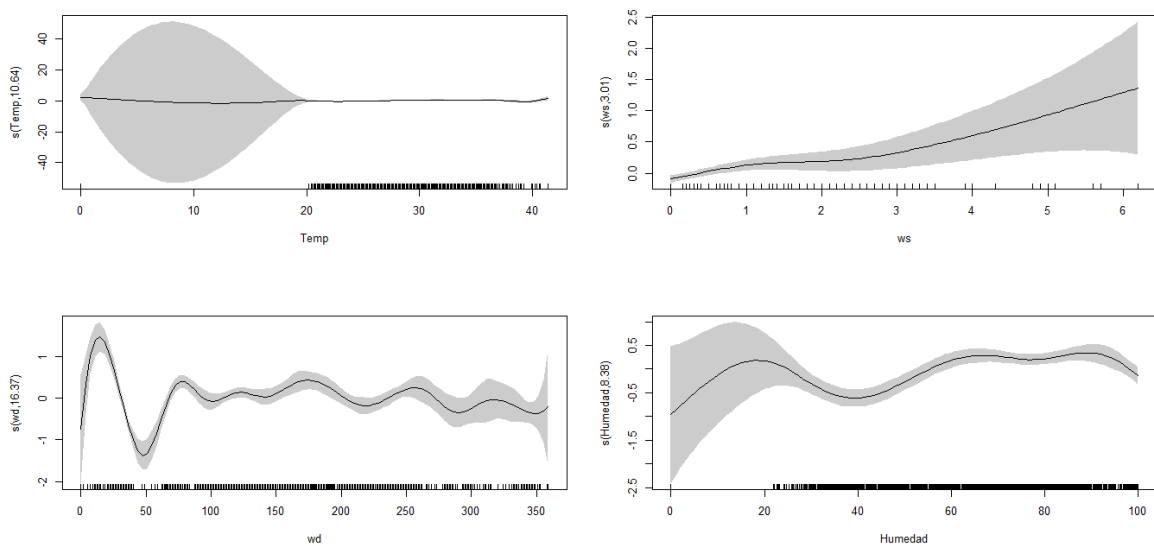
Se realiza el diagnostico nuevamente obteniendo los siguientes resultados:

Figura 3. 31 Gráficos de diagnóstico del modelo ajustado



Con este nuevo modelo se verifican todas las hipótesis, por lo que se resuelve que este modelo es el resultante para la concentración de PM_1 .

Figura 3. 32 Modelo Resultante de concentración PM_1



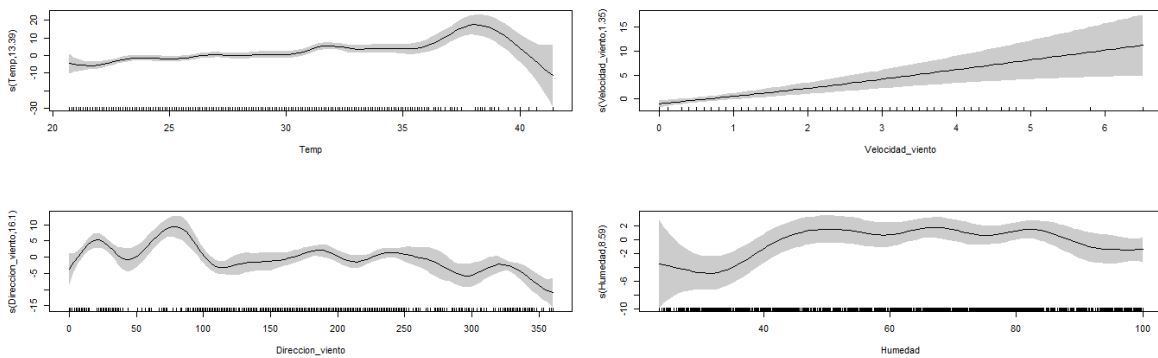
Para la concentración de PM2.5 se llevó a cabo el mismo proceso teniendo los siguientes resultados:

Modelo escogido el suavizado completo:

```
M8 <- gam(PM2.5..ug.m3 ~ s(Temp, k = 20, m = 2, bs = "ps")  
          + s(Velocidad_viento, k = 20, m = 2, bs = "ps") + s(Direccion_viento, k = 20,  
m = 2, bs = "ps")+ s(Humedad, k = 20, m = 2, bs = "ps"), data = DataMP25)
```

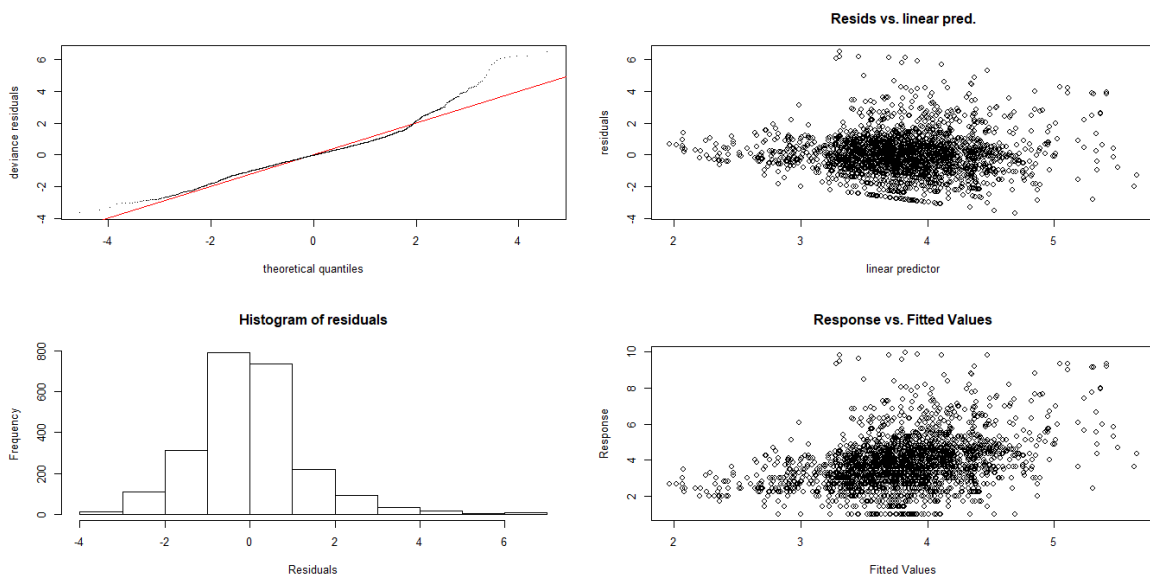
El estadístico GCV es 143.44, que fue el menor escogido.

Figura 3. 33 Gráfica del modelo ajustado PM_{2.5}



Diagnóstico del modelo:

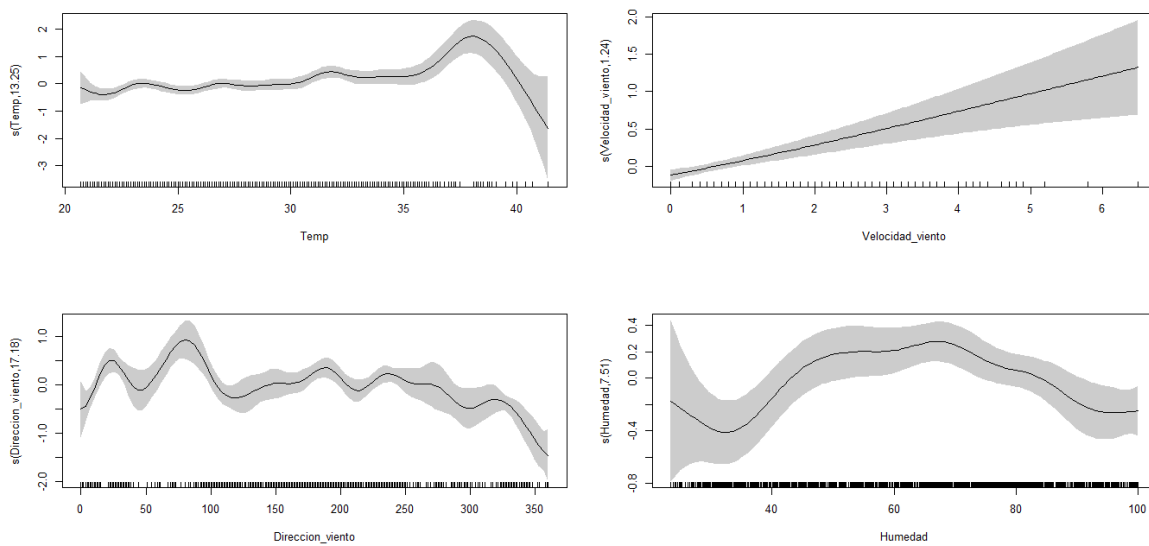
Figura 3. 34 Diagnóstico del modelo ajustado PM_{2.5}



Se verificaron las hipótesis resultantes, y se resuelve que es el modelo ajustado para la concentración de $PM_{2.5}$

Haciendo los ajustes necesarios obtenemos nuestro modelo resultante para la concentración $PM_{2.5}$ que indica que la variable que más explica el modelo es la de velocidad del viento así como el de PM_1 .

Figura 3. 35 Modelo final ajustado para $PM_{2.5}$



3.7 Implementación del modelo

Para poder implementar el modelo es necesario seguir un procedimiento que implique el correcto manejo de los datos, como:

- ✚ Tener bases de datos con información referenciada o georeferenciada, que mantenga un inventario de las emisiones generadas en la ciudad.
- ✚ Para los datos meteorológicos se debe tener un sistema de almacenados de datos con una resolución mínima de 1 hora y que se descarguen en una base de datos de fácil acceso.
- ✚ Los datos históricos de las emisiones se deben actualizar periódicamente.

El modelo diseñado en este trabajo de investigación permite recopilar y analizar los datos metodológicamente para establecer un pronóstico. Quienes lleguen a implementarlo deberán especificar características necesarias para el uso correcto del modelo.

3.8 Análisis costo beneficio

El análisis de costo del proyecto se da en base a la implementación del modelo no paramétrico múltiple realizado para la ciudad de Guayaquil.

Actualmente existen redes de monitoreo que están instalándose en lugares estratégicos de la ciudad de Guayaquil, por lo que lo único que se necesita es tener acceso a los datos que se recojan en esas estaciones de monitoreo.

Es importante que las mediciones se realicen durante todo el año, por lo que se requerirá hacer un mantenimiento a los equipos de medición, así como la calibración respectiva.

Para la utilización del modelo, es necesario que accedan a un programa o paquete estadístico denominado RStudio, la inversión en el sistema tampoco tiene costo ya que tienes descargables libres en la web.

Sin embargo en la tabla 3.11 se detallan los costos de funcionamiento y mantenimiento del modelo estadístico.

Nombre	Actividad	Costo Anual
Recursos humanos	Personal que maneja el modelo y da mantenimiento	\$8.000
Tecnológicos	Equipos, mantenimiento de equipos, actualizaciones	\$2.000
Inversión Total Anual		\$10.000

Por lo que la inversión inicial sería de \$1000 teniendo un costo anual de \$10000.

CAPÍTULO 4

CONCLUSIONES Y RECOMENDACIONES

Las técnicas de suavización no paramétricas inciden en la creación de modelos múltiples que caracterizan el comportamiento diario de los contaminantes que afectan a la atmósfera, frente a las variables meteorológicas o antropogénicas que pueden ocasionar de una u otra formas cambios importantes en el diseño.

El análisis estadístico realizado, mostró las posibles relaciones existentes entre el material particulado en estudio (PM_1 y $PM_{2.5}$) y las variables meteorológicas presentes en la ciudad de Guayaquil, del periodo 2016 y 2017, sin embargo, se puede inferir que las relaciones dadas no tengan un nivel de confiabilidad óptimo, debido a la falta de datos o el haber recogido muy poca información de las estaciones de muestreo en esas fechas.

El modelo propuesto muestra, que la única asociación existente se dio entre la concentración del material particulado y la variable meteorológica velocidad del viento, cabe resaltar que la dirección del viento también juega un rol importante en estos estudios, debido a que es o bien tarde en la noche, o en la mañana y cuando es época de sequía en donde se observa la dispersión del material particulado suspendido.

Cabe resaltar que los modelos splines o suavizados resultan ser adecuados para modelar el comportamiento diario que tiene el material particulado, el mismo que se puede aplicar a cualquier contaminante atmosférico, no solo al material particulado

y pueda generar una herramienta estadística para que sea utilizada para estimar concentraciones a cualquier hora del día.

Mediante la construcción del modelo se describió la variabilidad de la concentración del material particulado PM_1 y $PM_{2.5}$ a partir de una variable respuesta o explicativa como la velocidad del viento, la cual presentó un comportamiento lineal, después de haberse ajustado los datos mediante un suavizado no paramétrico.

Al validar el modelo, resultó confiabilidad al 95%, tomando en cuenta el estadístico de validación cruzada generalizada que para PM_1 fue de 80 y para $PM_{2.5}$ de 143, resultando ser los mejores modelos para explicar la variable. Después de ajustar, realizar la predicción y comprobar supuestos de regresión, normalidad, homocedasticidad e independencia, se obtuvo el modelo final para cada medida del material particulado, que explica el 17 y 16% de los casos aproximadamente.

Este proyecto otorga grandes beneficios, tanto para la población en general, así como las industrias y el medio ambiente, el trabajo en conjunto debe ser de todos desde casa, instituciones y ministerios para poder lograr tener una calidad de vida con un ambiente limpio y lleno de aire puro.

Recomendaciones

Siendo que este estudio se realizó con datos de años anteriores, es recomendable realizar el modelo con datos actuales, recogidos de las estaciones de muestreo que están en funcionamiento, con el fin de verificar el comportamiento de estas variables que afectan significativamente la calidad de aire en la ciudad de Guayaquil

Cabe recalcar también que hay que tener en cuenta al momento de realizar los muestreos se tomen en cuenta todos los datos y no existan faltantes o por error de digitación no existan datos aberrantes que podrían desequilibrar el modelo diseñado, en el caso de estudio, tuvimos inconvenientes por faltante de datos o datos aberrantes, que se los pudo corregir en el tratamiento de los datos.

Se recomienda, que se realice una investigación más a fondo para poder identificar, no solo las variables meteorológicas que influyan en la concentración del material particulado, sino también la existencia de otros factores tanto antropogénicos o atmosféricos.

BIBLIOGRAFÍA

BASANTES, C. A. (2018). *DETERMINACIÓN DE CONCENTRACIONES DE MATERIAL VOLATIL*. RIOBAMBA: DSPACE.

Cedar Lake Ventures, Inc. (25 de mayo de 2018). *weatherspark*. Obtenido de WeatherSpark.com: <https://es.weatherspark.com/y/19346/Clima-promedio-en-Guayaquil-Ecuador-durante-todo-el-a%C3%B1o#Sections-Summary>

Díaz, J. M. (2014). *Uso correcto de la correlación cruzada en Climatología*. Venezuela: Terra nueva etapa.

Ltd., W. (2019). *woespana*. Obtenido de weather online: <https://www.woespana.es/weather/maps/city>

masteres. (2017). *masteres*. Obtenido de masteres.ugr.es: [http://masteres.ugr.es/moea/pages/tfm0910/metodosderegresionnparametricaenmuestreoenpoblacionesfinitas/!](http://masteres.ugr.es/moea/pages/tfm0910/metodosderegresionnparametricaenmuestreoenpoblacionesfinitas/)

Rstudio. (2018). */rstudio*. Obtenido de https://rstudio-pubs-static.s3.amazonaws.com/365376_df3b77f32af8496aa879023dc270575c.html#funciones_de_suavizado

SALLE, M. M.-U. (2016). *ASOCIACIÓN EXISTENTE ENTRE LAS VARIABLES*. Bogota, Bogota, Colombia.