

**METODOLOGÍA PARA ESTIMAR MATERIAL PARTICULADO PM2.5 INTEGRANDO
DATOS DE ESTACIONES DE MONITOREO Y SENSORES REMOTOS EN EL ÁREA
URBANA DE LA CIUDAD DE MEDELLÍN**

IVÁN ANDRÉS CULMA ORTIZ



**FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA CIVIL Y GEOMÁTICA
SANTIAGO DE CALI, VALLE DEL CAUCA
2021**

**METODOLOGÍA PARA ESTIMAR MATERIAL PARTICULADO PM2.5 INTEGRANDO
DATOS DE ESTACIONES DE MONITOREO Y SENSORES REMOTOS EN EL ÁREA
URBANA DE LA CIUDAD DE MEDELLÍN**

Iván Andrés Culma Ortiz

Trabajo de grado para optar al título de:
Ingeniero Topográfico

Directores:

Francisco Luis Hernández M.Sc.

Johann Ospina Galíndez M.Sc.

**Universidad del Valle
Facultad de Ingeniería
Escuela de Ingeniería Civil y Geomática
Ingeniería Topográfica
Santiago de Cali
2021**

Nota de Aceptación

Director de Trabajo de Grado

Codirector de Trabajo de Grado

Jurado 1

Jurado 2

Santiago de Cali, octubre de 2021

A mi madre Luz Mary, por todo el amor, dedicación y apoyo incondicional brindados en todo momento.

AGRADECIMIENTOS

Agradezco a la Universidad del Valle, por darme el apoyo y la oportunidad de una formación académica profesional. Al profesor y director de tesis Francisco Luis Hernández, por sus enseñanzas y dirección en este trabajo de grado. Al profesor y codirector de tesis Johann Alexis Ospina, por su apoyo y disposición frente a inquietudes durante el desarrollo del trabajo de grado.

Al Sistema de Alerta Temprana de Medellín y el Valle de Aburrá – SIATA, por proveer la información empleada en este trabajo de grado, por parte de la red de estaciones de monitoreo de la calidad del aire en el área urbana de la ciudad de Medellín.

TABLA DE CONTENIDO

1. INTRODUCCIÓN.....	1
2. PLANTEAMIENTO DEL PROBLEMA	3
3. OBJETIVOS.....	5
3.1. Objetivo general.....	5
3.2. Objetivos específicos	5
4. JUSTIFICACIÓN.....	6
5. MARCOS DE REFERENCIA	8
5.1. MARCO CONCEPTUAL	8
5.1.1. Contaminación Atmosférica	8
5.1.1.1. Material Particulado (PM2.5)	9
5.1.1.2. Espesor óptico de aerosoles (AOD)	10
5.1.2. Percepción Remota	10
5.1.2.1. Espectro Electromagnético.....	11
5.1.2.2. Resoluciones.....	12
5.1.2.3. Espectroradiómetro de Imágenes de Resolución Moderada (MODIS).....	13
5.1.3. Sistemas de Vigilancia de la Calidad del Aire (SVCA)	16
5.1.4. Dato Funcional.....	16
5.2. MARCO TEÓRICO	17
5.2.1. Reflectancia tope de la atmósfera.....	17
5.2.2. Reflectancia superficial	19
5.2.3. Profundidad óptica de aerosoles (AOD).....	20
5.2.3.1. Reflectancia de la trayectoria de Raleigh	21
5.2.3.2. Ángulo de fase de dispersión	22
5.2.3.3. Función fase del aerosol	22

5.2.3.4.	Transmisiones atmosféricas	22
5.2.3.5.	Retrodispersión superficie – atmósfera	23
5.2.3.6.	Algoritmo simplificado de recuperación de aerosoles (SARA)	23
5.2.4.	Red Robótica de Aerosoles (AERONET)	24
5.2.5.	Análisis de datos funcionales	26
5.2.5.1.	Espacio funcional	26
5.2.5.2.	Base de funciones.....	27
5.2.5.3.	Elección del número de funciones base	28
5.2.5.4.	Construcción de datos funcionales.....	28
5.2.5.5.	Elección del número de funciones de la base (K).....	30
5.2.5.6.	Estadística descriptiva funcional.....	31
5.2.6.	Modelo de regresión funcional concurrente	32
5.2.6.1.	Intervalos de confianza en el modelo de regresión funcional.....	33
5.2.6.2.	Bondad de ajuste del modelo	33
5.3.	MARCO REFERENCIAL.....	35
6.	METODOLOGÍA.....	39
6.1.	ÁREA DE ESTUDIO	40
6.2.	FUENTES DE INFORMACIÓN	41
6.2.1.	Datos MODIS.....	41
6.2.2.	Datos de estaciones de monitoreo.....	44
6.2.2.1.	Depuración de datos in situ	45
6.2.3.	Datos Red Robótica de Aerosoles (AERONET).....	45
6.3.	MÉTODOS.....	46
6.3.1.	Georreferenciación	46
6.3.2.	Procesamiento de información espacial.....	47
6.3.3.	Imputación de datos faltantes	50

7. RESULTADOS Y DISCUSIÓN	51
7.1. IMPUTACIÓN Y ANÁLISIS DESCRIPTIVO	51
7.2. ESPESOR ÓPTICO DE AEROSOL (AOD)	54
7.3. ANÁLISIS DE DATOS FUNCIONALES (ADF)	58
7.3.1. Construcción y análisis de datos funcionales	58
7.3.2. Modelo de regresión funcional PM2.5	61
7.4. MAPAS DE ESTIMACIÓN DE CONCENTRACIONES PM2.5	64
8. CONCLUSIONES Y RECOMENDACIONES	67
8.1. CONCLUSIONES	67
8.2. RECOMENDACIONES	68
9. BIBLIOGRAFÍA	70
10. ANEXOS	75

LISTA DE FIGURAS

Figura 1 Elementos fundamentales de la percepción remota.....	11
Figura 2 El espectro electromagnético y la región del espectro visible.	12
Figura 3 Relación entre los valores de enteros escalados (<i>SI</i>) y valores corregidos instrumentalmente (<i>dn **</i>).....	18
Figura 4 Características principales del sistema de medición AERONET.....	25
Figura 5 Esquema metodológico general propuesto	39
Figura 6 Área de estudio y ubicación de estaciones PM2.5 y AERONET.....	40
Figura 7 Esquema metodológico detallado para recuperar la profundidad óptica de aerosoles.....	49
Figura 8 Frecuencia de los registros de datos faltantes PM2.5 y AOD	51
Figura 9 Diagrama de cajas de las concentraciones de PM2.5 in situ horario por estación, para las 26 fechas de estudio comprendidas entre los años 2019 y 2020.	52
Figura 10 Comportamiento de las concentraciones de PM2.5 in situ horario para las 26 fechas de estudio comprendidas entre los años 2019 y 2020.....	53
Figura 11 Comportamiento del Espesor óptico de aerosoles (AOD) para las 26 fechas de estudio comprendidas entre los años 2019 y 2020	56
Figura 12 Datos funcionales de las concentraciones de PM2.5 y su media funcional (Línea roja) y desviación estándar funcional (Línea azul).	59
Figura 13 Datos funcionales del espesor óptico de aerosoles (AOD) y su media funcional (Línea roja) y desviación estándar funcional (Línea azul).	59
Figura 14 Comportamiento del intercepto funcional (a) y pendiente funcional (b) del modelo de regresión funcional propuesto.	61

Figura 15 Datos funcionales de PM2.5 ajustados por el modelo de regresión funcional (Línea roja) a partir de datos funcionales y datos de PM2.5 medidos en tierra (Línea negra) para las 9 estaciones dentro del área de estudio. Eje x: Corresponde a los 26 días de evaluación durante los años 2019 y 2020. Eje y: Equivale a concentraciones de PM2.5..... 62

Figura 16 (a): Representa el coeficiente de determinación funcional del modelo ($R^2(t)$). (b): Equivale a los errores residuales funcionales del modelo. 63

Figura 17 Distribución espacial de concentraciones de PM2.5 en la ciudad de Medellín para el día primero de enero del 2019. 65

Figura 18 Distribución espacial de concentraciones de PM2.5 en la ciudad de Medellín para el día 10 de octubre del 2020..... 66

LISTA DE TABLAS

Tabla 1 Niveles máximos permisibles de contaminantes atmosféricos en Colombia.	9
Tabla 2 Características de las bandas espectrales del sensor MODIS	14
Tabla 3 Productos MOD02HKM para los años 2019 - 2020	42
Tabla 4 Productos MOD09GA para los años 2019 - 2020.....	43
Tabla 5 Estaciones de monitoreo SIATA para la ciudad de Medellín.	44
Tabla 6 Significado de valores Flag para la red de estaciones de la calidad del aire ...	45
Tabla 7 Regresiones lineales entre el AOD y el PM2.5 para las 26 fechas de estudio durante los años 2019 – 2020.....	55

LISTA DE ANEXOS

Anexo A Datos de PM2.5 terrestre y del espesor óptico de aerosoles (AOD) obtenido mediante el algoritmo simplificado de recuperación de aerosoles (SARA) para las 9 estaciones de monitoreo y las 26 fechas de estudio comprendidas entre los años 2019 y 2020.	75
Anexo B Gráficos de dispersión y regresiones lineales para las 26 fechas de estudio comprendidas entre los años 2019 y 2020 entre las concentraciones de PM2.5 medido en tierra y el AOD obtenido mediante el modelo SARA.	81
Anexo C Mapas de concentraciones de PM2.5 en la ciudad de Medellín, para las fechas de estudio.....	91

GLOSARIO

PM2.5: Contaminante atmosférico que se compone de partículas menores a 2.5 micras.

AOD (aerosol optical depth): El espesor óptico de aerosoles cuantifica la cantidad de energía absorbida, reflejada o dispersada por las partículas en la atmosfera.

Sensor remoto: Instrumento que mide la cantidad de energía reflejada por la Tierra para representar las características superficiales y/o climáticas de la misma, sin estar directamente en contacto con la Tierra.

MODIS (Moderate-Resolution Imaging Spectroradiometer): Es un sensor lanzado por la NASA en el programa de satélites Terra y Aqua, para el estudio de la tierra y variaciones climáticas.

Contaminación atmosférica: Proceso donde la atmósfera se ve afectada cambiando su composición debido a presencia de diferentes agentes externos.

Banda espectral: Rango de ondas electromagnéticas en función de la longitud o frecuencia, que permite la cuantificación de una parte específica del espectro electromagnético.

Sensor multiespectral: Instrumento que censa radiación electromagnética dentro y fuera del rango visible que puede captar el ojo humano.

Sensor hiperespectral: Instrumento que capta la mayoría del espectro electromagnético, conformado aproximadamente por más de 20 bandas espectrales.

RESUMEN

La contaminación atmosférica es una problemática que enfrentan la mayoría de las ciudades del mundo, puesto que puede causar enfermedades respiratorias, daños en cultivos agrícolas, entre otros. La presente investigación estima concentraciones de PM2.5 en el área urbana de la ciudad de Medellín para 26 fechas disponibles libres de nubosidad para datos del Espectroradiómetro de imágenes de resolución moderada (MODIS) a una resolución de 500 metros, durante los años 2019 y 2020; empleando técnicas de percepción remota mediante un algoritmo simplificado de recuperación de aerosoles (SARA) aplicado a los datos de entrada se recuperó el espesor óptico de aerosoles (AOD). Dado que distintos estudios han demostrado la relación positiva entre AOD y PM2.5, se generaron modelos de regresión lineal obteniendo un coeficiente de determinación $R^2 = 0.886$ para el mejor de los casos y un promedio de $R^2 = 0.66$ y desviación estándar de $4.85 \mu g/m^3$ para todas las fechas evaluadas; a partir de las regresiones lineales se estimó las concentraciones de PM2.5 dentro de toda el área de estudio, obteniendo mediciones que se encuentran apartadas de la red de estaciones de monitoreo tradicional. Se utilizó el análisis de datos funcionales (ADF) para modelar el comportamiento del PM2.5 durante las fechas de estudio para los años 2019 y 2020. Durante las fechas del año 2019 se evidenciaron días que superaron los $25 \mu g/m^3$; mientras que para las fechas del año 2020 se presentaron valores inferiores promedios de $10 \mu g/m^3$, reflejando probablemente el impacto positivo en las concentraciones de PM2.5 durante la pandemia, aumento de precipitaciones y/o velocidades del viento, entre otras variaciones climáticas. El uso de metodologías basadas en sensores remotos como alternativa para complementar y/o sustituir los métodos tradicionales, presenta gran importancia en temas de costos y distribución espacial de los datos.

Palabras clave: PM2.5, AOD, datos funcionales, sensores remotos, MODIS, contaminación atmosférica, ADF.

1. INTRODUCCIÓN

La contaminación atmosférica es la alteración de la composición natural de la atmósfera debido a factores externos generados de forma natural o producidos por actividades humanas. Existen contaminantes atmosféricos primarios y secundarios. Los primarios son aquellos que se generan en la atmósfera directamente de una fuente y conservan una misma estructura química. Los secundarios se definen como cualquier contaminante que se origina por reacciones químicas atmosféricas primarias o precedentes. También se define como contaminantes criterio a aquellos que puedan producir efectos negativos en la salud, infraestructura o el medio ambiente (Landrigan et al., 2017).

Para monitorear los efectos negativos de la contaminación atmosférica hacia el medio ambiente se implementan normas que regulan, estas se basan en fundamentos científicos para determinar los niveles máximos permisibles dentro de una zona de estudio (DNP, 2018). En Colombia, las principales ciudades como Bogotá, Medellín y Cali emplean un Sistema de Vigilancia de la Calidad del Aire (SVCA); este se define como un grupo de equipos de monitoreo de los principales contaminantes atmosféricos, que se encuentran ubicados en sitios donde previamente se realizaron estudios técnicos que determinaron una ubicación geográfica estratégica, tipos de contaminantes principales a monitorear y tipo de SVCA a implementar (Minambiente, 2010).

El material particulado (PM) es un conjunto de partículas líquidas y sólidas, de sustancias orgánicas e inorgánicas que se encuentran suspendidas en el aire. Algunos de sus componentes primarios son el amoníaco, carbón, cloruro de sodio, agua, entre otros (EPA, 2016). Estas se clasifican según su diámetro, donde se encuentran las de menor a 10 micrómetros (PM10) y las de menor a 2.5 micrómetros (PM2.5). Estas dos clases se utilizan como indicadores para determinar la calidad del aire en una zona estudiada. Las fuentes de emisiones contaminantes globales están representadas en un 70% para fuentes móviles y un 30% para fuentes fijas (DNP, 2018). En el mundo, las poblaciones que se encuentran en zonas con niveles mayores a la exposición normal de PM2.5 tienden a ser cada vez mayores, lo cual resulta alarmante y perjudicial para la salud (Luo et al., 2015; OMS, 2005).

Los sensores remotos son alternativas que se emplean para estimar mediante diferentes longitudes de onda el material particulado presente en la atmósfera (Abad & Mejía, 2017; Alvarez & Padilla, 2016; Cespedes et al., 2015). El uso de imágenes satelitales tiene un gran potencial en su resolución espacial, puesto que en cada píxel de la imagen se registra la información estimada de material particulado, teniendo así múltiples datos espaciales en sola captura de información. Algunas imágenes satelitales se pueden obtener sin costo alguno dependiendo de su resolución espacial y temporal (Zapata & Anaya, 2011). Con el desarrollo de nuevos sensores e implementación en nuevas misiones, se han desarrollado técnicas eficientes que permiten detectar la contaminación atmosférica, empleando datos multiespectrales mediante sensores remotos (IGAC, 2007; Luo et al., 2015; Mei et al., 2013; Othman et al., 2010).

La metodología empleada en este trabajo se basa en estimar el material particulado PM_{2.5} integrando dos fuentes de información; una fuente obtenida de la red de estaciones de monitoreo del Sistema de Alerta Temprana de Medellín y el Valle de Aburra (SIATA), y otra fuente a través de sensores remotos recuperando el espesor óptico de aerosoles (AOD) mediante un algoritmo simplificado de recuperación de aerosoles (SARA) (Bilal et al., 2013), utilizando información del sensor TERRA MODIS de La Administración Nacional de la Aeronáutica y del Espacio (NASA), en el área urbana de la ciudad de Medellín, Colombia. Esta región se encuentra entre las dos ciudades que registran mayor cantidad de contaminación atmosférica en Colombia (DNP, 2018). Se pretende integrar la información mediante un método estadístico de regresión que permita relacionar la funcionalidad de los datos puntuales de la red de estaciones y los de la imagen satelital (Santiago & Chicangana, 2017).

2. PLANTEAMIENTO DEL PROBLEMA

Según el Departamento Nacional de Planeación (2018), en Colombia, el 51% de la población considera que la contaminación del aire es el principal problema ambiental de la sociedad. El 76% de los 78 municipios que miden la calidad del aire, registran niveles de contaminación que podrían causar enfermedades en la población. El material particulado que presenta los mayores índices de afectación en Colombia es el PM2.5, que es generado principalmente por automotores que emplean combustible diésel (DNP, 2018). De acuerdo con el Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM (2017), la evaluación de la calidad del aire determinó que las ciudades de Medellín, Bogotá y Cali obtuvieron las mayores concentraciones de contaminantes en el aire respectivamente.

Teniendo en cuenta esta problemática de calidad del aire, en Colombia, se ha venido implementando sucesivamente desde 1982 políticas de prevención y control de la contaminación del aire. La primera norma de emisiones atmosféricas se cristalizó mediante decreto presidencial 02, en 2005, en la que se establecieron los lineamientos para la formulación de la política de prevención y control de la contaminación del aire a través del documento CONPES 3344; en 2008 se establece un mejoramiento del combustible disminuyendo gradualmente el contenido de azufre del combustible diésel de 3500 a 50 ppm; en 2010 se establece formalmente la política de prevención y control de la contaminación del aire planteando los primeros planes de reducción de contaminación del aire en Bogotá y Medellín; finalmente en 2017 se implementó la evaluación de las políticas planteadas en el 2011 mediante la nueva norma calidad del aire Resolución 2254 del 2017 (DNP, 2018).

Los sistemas que vigilan la calidad del aire (SVCA) empleados en Colombia están controlados por las Corporaciones Autónomas Regionales y las autoridades ambientales de los grandes centros urbanos, con el objetivo de verificar el cumplimiento de las normas en la calidad del aire. Los contaminantes de mayor interés se establecen según el tipo área y estación a implementar; para un área de tipo urbana y estación de tipo tráfico y punto crítico, se miden los contaminantes gaseosos como los óxidos de nitrógeno (NOx),

monóxido de carbono (CO), compuestos orgánicos volátiles (COV) y ozono troposférico (O_3); para el caso de las partículas, se miden el PM10 y PM2.5. Para el caso del PM2.5, actualmente los niveles máximos permisibles para un tiempo de exposición anual y de 24 horas son $25 \mu g/m^3$ y $37 \mu g/m^3$ respectivamente (MAVDT, 2010).

Estos equipos requieren de una gran logística e inversión por sistema de monitoreo para su funcionamiento, limitando el número de equipos y municipios que puedan llevar a cabo su implementación. Esto se evidencia en que muchos municipios no se evaluó la calidad del aire o se cuenta con muy pocos equipos, lo que conduce a una problemática espacial debido a los pocos equipos y calidad de los datos por la carencia de los mismo, teniendo en cuenta que se mide una variable continua de forma discreta, lo que demanda un monitoreo más constante y preciso (IDEAM, 2017).

La ciudad de Medellín y Valle de Aburra cuentan con 16 estaciones que monitorean la calidad del aire, 9 de estas miden el material particulado PM2.5. La información puntal de estas estaciones se emplea para estimar mediante técnicas de interpolación el PM2.5 en toda su área metropolitana de aproximadamente $382 km^2$. Esta técnica presenta una principal desventaja, que entre más distante este la información obtenida a través de la interpolación del SVCA, mayor será el error asociado en los datos. Esto evidencia una desventaja en la resolución espacial en la toma de datos y por consiguiente en la calidad de estos. Teniendo en cuenta esto, se evidencia una problemática de características espaciales en la adquisición de la información. Dado esto, se plantea la siguiente pregunta de investigación, ¿Cómo implementar una metodología que permita estimar el material particulado PM2.5 integrando datos de una red de estaciones de monitoreo y las técnicas de percepción remota?

3. OBJETIVOS

3.1. Objetivo general

Implementar una metodología para estimar concentraciones de PM2.5 en la ciudad de Medellín integrando datos de estaciones de monitoreo y sensores remotos.

3.2. Objetivos específicos

- Estimar concentraciones de PM2.5 mediante técnicas de percepción remota.
- Determinar las concentraciones de PM2.5 en sitios o zonas sin cobertura de la red de la red de estaciones de monitoreo.
- Comparar la distribución espacial del material particulado PM2.5 entre los datos de la red de estaciones de monitoreo y sensores remotos.

4. JUSTIFICACIÓN

El monitoreo de la calidad del aire es un proceso que cada día se vuelve más indispensable, puesto que principalmente en las grandes ciudades el uso de automotores y funcionamiento de grandes y medianas industrias generan cantidades considerables de polución. La necesidad de implementar equipos y metodologías que permitan optimizar la relación espacial y temporal integrando diferentes métodos, son procedimientos que tienden a ser cada vez más importantes para obtener un monitoreo más eficaz en una zona determinada. La contaminación del aire produce efectos negativos en la salud, ambiente y la infraestructura. Estas afectaciones están relacionadas con muertes humanas por causas respiratorias, daños en cultivos alimenticios y deterioro en materiales de edificaciones (DNP, 2018).

La percepción remota para el monitoreo de la calidad del aire ha tenido una importante aceptación en la comunidad científica, desde el uso de productos derivados de imágenes satelitales directamente, hasta la implementación de metodologías que logren recuperar la profundidad óptica de aerosoles (AOD) en una imagen satelital compuesta por bandas del espectro visible (RGB). Una escena aporta una visión completa de la zona de estudio, detallando las características espaciales de la variable estudiada. El uso de estas herramientas presenta limitantes con respecto a la presencia de nubosidad y la resolución temporal de los satélites (Zapata & Anaya, 2011); estas limitaciones justifican la necesidad de llevar a cabo estudios enfocados en la teledetección que permitan desarrollar ampliamente estas herramientas. Dado esto, los sensores remotos aportan características que complementan y/o sustituyen los métodos tradicionales que monitorean la calidad del aire (Abad & Mejía, 2017).

Mediante la integración de información espacial obtenida por una red de estaciones de monitoreo e imágenes de satélite, se espera obtener un monitoreo de PM_{2.5} de mayor alcance en zonas donde la cobertura de la red de estaciones de monitoreo es limitada. Este tipo de metodologías han demostrado tener resultados positivos al abordar la problemática de espacialidad en una red de monitoreo de calidad del aire (Cespedes et al., 2015). La implementación de una metodología que integre datos de satélite y

estaciones de monitoreo para estimar un contaminante criterio sería de gran interés para entidades encargadas del funcionamiento de las redes de estaciones de monitoreo en los principales cascos urbanos del país. Este tipo de metodologías presentan un gran potencial para las entidades regionales, dado que su implementación se puede llevar a cabo para diferentes tipos de variables a estudiar.

5. MARCOS DE REFERENCIA

5.1. MARCO CONCEPTUAL

5.1.1. Contaminación Atmosférica

La contaminación atmosférica o contaminación del aire se define como la presencia en la atmósfera de sustancias o partículas que involucren un grado de peligro considerable durante su exposición, generando molestias graves en los seres humanos o deterioro en objetos de cualquier naturaleza (Ataz & de Mera Morales, 2004). Los principales tipos de contaminantes atmosféricos son de naturaleza gaseosos y aerosoles, también conocidos como partículas en suspensión o material particulado. Los contaminantes atmosféricos se clasifican principalmente en primarios, secundarios y contaminantes criterio. Entre los contaminantes criterio de destacan el material particulado (PM), ozono troposférico (O_3), monóxido de carbono (CO), óxidos de Azufre (SOx) y el óxido de Nitrógeno (NOx). La tabla 1 presenta los niveles máximos permisibles de estos contaminantes para Colombia. La contaminación por material particulado (PM) esta categorizada en material particulado menor a 10 (PM10) y 2.5 (PM2.5) micrómetros de diámetro (EPA, 2016).

La resolución 2254 del 1 de noviembre del 2017 establece la norma de calidad de aire para la gestión del aire en el territorio nacional, donde establece los niveles máximos permisibles de exposición a contaminantes atmosféricos para garantizar un ambiente sano y reducir el riesgo de afectaciones sobre la salud humana.

Tabla 1 Niveles máximos permisibles de contaminantes atmosféricos en Colombia.

Contaminante	Nivel máximo permisible ($\mu\text{g}/\text{m}^3$)	Tiempo de exposición
PM10	50	Anual
	75	24 horas
PM2.5	25	Anual
	37	24 horas
SO_2	50	Anual
	100	1 hora
NO_2	60	Anual
	200	1 hora
O_3	100	8 horas
	5000	8 horas
CO	35000	1 hora

Fuente: Minambiente, 2017

5.1.1.1. Material Particulado (PM2.5)

El material particulado PM2.5 son partículas finas inhalables que tienen un diámetro de 2.5 micrómetros o menor, que es igual a 1/30 parte de un cabello humano promedio. Estas son una mezcla de partículas sólidas y pequeñas gotas líquidas que se encuentran suspendidas en el aire, a diferencia de partículas como el polvo, el humo o el hollín que pueden ser vistas a simple vista, el PM se tiene que detectar empleando un microscopio (EPA, 2016).

Estas partículas pueden ser de origen primario, siendo emitidas directamente por la fuente, como por ejemplo chimeneas, incendios o sitios de construcción. Las de origen secundario, son aquellas resultantes de reacciones complejas de productos químicos como el óxido de nitrógeno y dióxido de azufre, que son principalmente emitidas por automotores, centrales eléctricas e industrias (EPA, 2016). La forma tradicional de estimar el PM es mediante una red de estaciones de monitoreo de la calidad del aire y empleado técnicas y herramientas sensoramiento remoto.

5.1.1.2. Espesor óptico de aerosoles (AOD)

El espesor óptico de aerosoles es la cuantificación de la absorción y dispersión de luz visible provocada por las partículas que se encuentran en la columna vertical de la atmósfera (Kaufman et al., 1997). Esta variable es ampliamente usada para estudiar la calidad del aire, puesto que es un valor proporcional a las concentraciones de material particulado (Bilal et al., 2013). Sus principales limitaciones son la dependencia del espectro visible para su estimación, lo que traduce que para escenarios con nubosidad o no diurnos, no se puede calcular el AOD (Alegría, 2015); igualmente para el caso de superficies muy brillantes como desiertos y planicies de hielo.

5.1.2. Percepción Remota

La percepción remota se define como la toma y procesamiento de información que se genera en la superficie de la Tierra sin estar en contacto con ella, mediante sensores incorporados en satélites y sondas artificiales. Esta información se produce a través de la interacción electromagnética entre la superficie terrestre y el sensor (SEGEMAR, 2018). Según la figura 1, el proceso para llevar a cabo la percepción remota o teledetección cuenta con siete pasos o elementos fundamentales, que se organizan secuencialmente como: 1: Fuente de energía o radiación electromagnética, 2: Medio de transmisión o atmósfera, 3: Interacción energía–objeto, 4: Percepción de la energía por el sensor, 5: Transmisión - recepción, 6: Procesamiento de imágenes satelitales y 7: El usuario final de la información que la interpretara e implementara una solución (Lillesand et al., 2015).

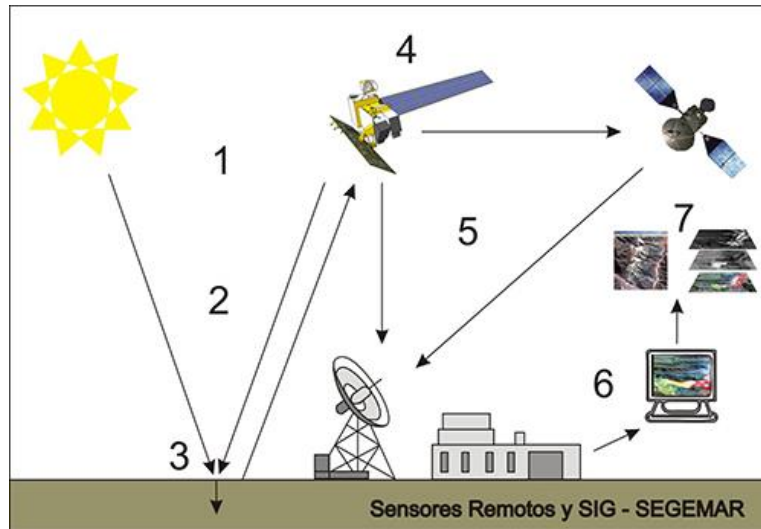


Figura 1 Elementos fundamentales de la percepción remota

Fuente: Servicio Geológico Minero Argentino, 2018

5.1.2.1. Espectro Electromagnético

El espectro electromagnético o espectro de la radiación electromagnética es la distribución del conjunto de frecuencias a las que se puede producir radiación electromagnética. Este conjunto se ordena en función de la frecuencia o longitud de onda. Dentro de este se encuentra el espectro visible como se indica en la figura 2, que es el conjunto de ondas que son por el ojo humano (Ordóñez, 2012). En la percepción remota, existen sensores que pueden capturar información de múltiples regiones del espectro electromagnético, generando imágenes multiespectrales e hiperespectrales.

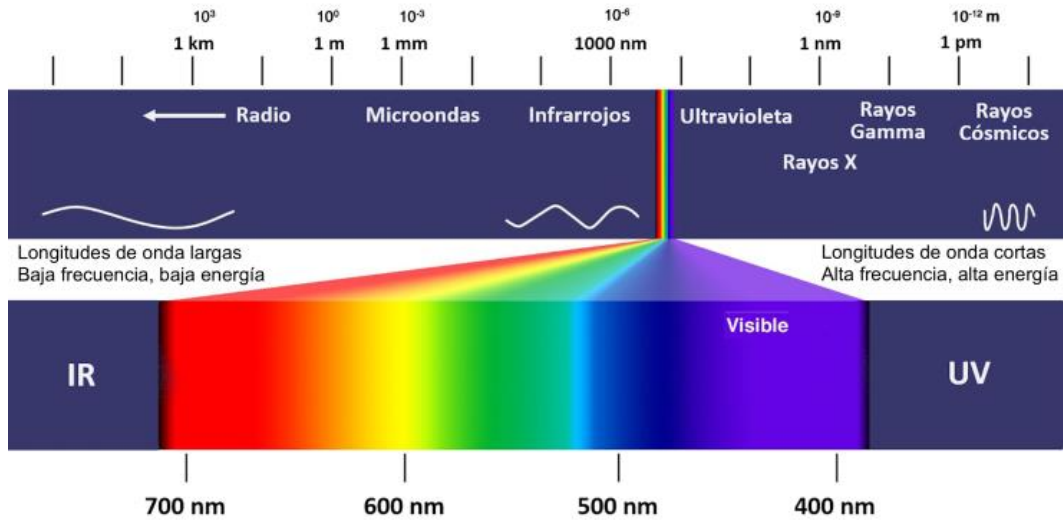


Figura 2 El espectro electromagn3tico y la regi3n del espectro visible.

Fuente: Adaptado de (NASA, 2007)

5.1.2.2. Resoluciones

Los sensores remotos tienen caracter3sticas propias dependiendo del prop3sito de la misi3n o informaci3n a sensar. Unas de sus principales caracter3sticas es la capacidad de registrar, detallar y discriminar informaci3n a trav3s de sus resoluciones (Chuvi3co, 1995), que se detallan a continuaci3n:

- **Resoluci3n espacial:** Hace referencia al tama1o de la superficie terrestre que se representa con el p3xel de una imagen satelital, esta se relaciona con el nivel de detalle que un sensor puede obtener, si se tiene una mayor resoluci3n espacial, el sensor podr3 obtener informaci3n espacial m3s detallada de una zona de estudio.
- **Resoluci3n temporal:** Es el tiempo o periodo en el cual un sensor es capaz de adquirir la misma informaci3n espacial de una zona determinada. Esta se deriva de las caracter3sticas orbitales del sensor, como la altura, la velocidad e inclinaci3n.

- **Resolución espectral:** Menciona la capacidad que tiene el sensor en discriminar un número y anchura de regiones del espectro electromagnético o bandas espectrales. Según la capacidad del sensor, se puede obtener información multiespectral (decenas de bandas espectrales) o hiperespectral (centenas de bandas espectrales).
- **Resolución radiométrica:** Indica la sensibilidad del sensor, es decir, la capacidad de percibir variaciones energéticas que recibe de la radiancia espectral en una zona determinada. Estas variaciones se cuantifican en formato binario (bits) que dan lugar al número de niveles digitales empleados por el sensor.

5.1.2.3. Espectroradiómetro de Imágenes de Resolución Moderada (MODIS)

El sensor MODIS es un instrumento a bordo de la plataforma Terra lanzado en el año 1999, que observa la superficie completa de la tierra cada 1 o 2 días. Su órbita se encuentra a una altura de 705 km con un tiempo de cruce ecuatorial a las 10:30 am de nodo descendente. Cuenta con una resolución radiométrica de 12 bits y una resolución espectral de 36 bandas que van en longitudes de onda de $0,4 \mu m$ a $14,4 \mu m$, detalladas en la tabla 2. Los productos de datos MODIS tienen una resolución espacial de 250m, 500m y 1km para las bandas 1-2, 3-7 y 8-36 respectivamente. Su principal misión es recolectar información de los procesos y dinámicas globales de la atmósfera, los océanos y la superficie terrestre (NASA, 2019).

Tabla 2 Características de las bandas espectrales del sensor MODIS

Principal Uso	Banda	Longitud de Onda (A)	Radiancia Espectral (B)	SNR o NE [delta]T(K) requerido
Límites de tierra / nube / aerosoles	1	620 - 670	21,8	128
	2	841 - 876	24,7	201
	3	459 - 479	35,3	243
Propiedades de tierra / nube / aerosoles	4	545 - 565	29,0	228
	5	1230-1250	5,4	74
	6	1628-1652	7,3	274
	7	2105 - 2155	1	110
Color del océano / Fitoplancton / Biogeoquímica	8	405 - 420	44,9	880
	9	438 - 448	41,9	838
	10	483 - 493	32,1	802
	11	526 - 536	27,9	754
	12	546 - 556	21	750
	13	662 - 672	9,5	910
	14	673 - 683	8,7	1087
	15	743 - 753	10,2	586
Vapor de agua atmosférico	16	862 - 877	6,2	516
	17	890 - 920	10,0	167
	18	931 - 941	3,6	57
Temperatura de la superficie / nube	19	915 - 965	15,0	250
	20	3.660 - 3.840	0,45 (300 K)	0,05
	21	3.929 - 3.989	2,38 (335 K)	0,2
	22	3.929 - 3.989	0,67 (300 K)	0,07
Temperatura Atmosférica	23	4.020 - 4.080	0,79 (300 K)	0,07
	24	4.433 - 4.498	0,17 (250 K)	0,25
Vapor de agua de nubes cirros	25	4.482 - 4.549	0,59 (275.000)	0,25
	26	1.360 - 1.390	6	150 (SNR)
	27	6.535 - 6.895	1,16 (240 K)	0,25
Propiedades de la nube	28	7,175 - 7,475	2,18 (250 K)	0,25
	29	8.400 - 8.700	9,58 (300 K)	0,05
Ozono	30	9.580 - 9.880	3,69 (250 K)	0,25
Temperatura de la superficie / nube	31	10,780 - 11,280	9,55 (300 K)	0,05
	32	11.770 - 12.270	8,94 (300 K)	0,05
Altitud de la cima de la nube	33	13.185 - 13.485	4,52 (260 K)	0,25
	34	13.485 - 13.785	3,76 (250 K)	0,25
	35	13.785 - 14.085	3,11 (240 K)	0,25
	36	14.085 - 14.385	2,08 (220 K)	0,35

Fuente: NASA, 2019

A: Bandas de 1 a 7 en unidades de nm. Bandas de 20 a 36 en μm .
B: Valores de Radiancia espectral en $\text{W}/\text{m}^2/\mu\text{m}/\text{sr}$.
SNR: Radio de Señal-Ruido.

Los datos del sensor MODIS se estructuran en colecciones. Cada colección de productos se caracteriza por ser generados mediante la misma rutina, empleando en algunos casos diferentes versiones de algoritmos (Remerl et al., 2009). Según Mas (2011), el nivel de procesamiento de los productos MODIS se clasifica en 5 niveles, detallados a continuación:

- **Nivel 0:** Corresponde a los datos sin ningún tipo de procesamiento.
- **Nivel L1:** Es la información de geolocalización (MOD03) en coordenadas geodésicas, máscara de tierra/agua, la altura del terreno, ángulo de elevación, cenit y azimut de iluminación y observación.
- **Nivel 1A:** Corresponde a los productos de calibración, geolocalización y procesamiento. Contiene los niveles de radiancia bruta captada por el sensor.
- **Nivel 1B:** Son los productos que contienen las radiancias calibradas y geolocalizadas para las bandas del sensor MODIS.
- **Nivel L2G:** Corresponde a las variables geofísicas proyectadas en una malla homogénea. Estos datos se agrupan en una cuadrícula llamada tile, la cual está proyectada sinusoidalmente en un área equivalente.
- **Nivel L2:** Este nivel contiene las variables geofísicas generadas a partir del nivel 1B, aplicando diferentes algoritmos durante su procesamiento. Estos son almacenados en gránulos equivalentes a 5 minutos de datos censados por MODIS.
- **Nivel 3:** Son los productos generados por las variables geofísicas mapeadas, donde se pueden procesar a una menor resolución espacial o como un intervalo de tiempo de los mejores resultados.
- **Nivel 4:** El último nivel corresponde a los datos que serán agregados a modelos o tablas de búsqueda para calcular las variables geofísicas. Por lo general, estos productos se estiman empleando datos de nivel L2 e información complementaria.

5.1.3. Sistemas de Vigilancia de la Calidad del Aire (SVCA)

En Colombia los SVCA emplean estaciones fijas de monitoreo que usan diferentes técnicas de medición, estas técnicas se clasifican en activas (manuales o semiautomáticos), pasivas y automáticas (Minambiente, 2010). Las técnicas manuales toman muestras con métodos físicos y químicos in situ, que posteriormente requieren ser analizadas en laboratorio para estimar la contaminación atmosférica. Por parte de los semiautomáticos, estos realizan un muestro similar a los manuales, diferenciándolos en que estos integran sistemas electrónicos que permiten disminuir la incertidumbre de la información. Para los que emplean tecnología automática, se tiene que son capaces de tomar, analizar y elaborar una estimación en tiempo real de la contaminación atmosférica con una posterior validación (Minambiente, 2010).

5.1.4. Dato Funcional

Para estudiar una variable continua como el material particulado en un periodo de tiempo específico, se cuenta tradicionalmente con 24 mediciones durante un día, es decir, que cada medición corresponde a la evaluación de la calidad del aire durante cada hora de dicho día. Si se tiene en cuenta lo anterior, el estudio detallado de la variable perdería información en los rangos de tiempo en que se toman las muestras, lo que llevaría a una limitación de la metodología para el estudio de la variable. El proceso continuo o función que describe la variable en cualquier instante de tiempo a partir de las observaciones discretas, se le conoce como dato funcional (Santiago & Chicangana, 2017).

5.2. MARCO TEÓRICO

En esta sección se especifican las técnicas y modelos matemáticos empleados para el desarrollo de la metodología propuesta, con base en la revisión bibliográfica consultada que se adaptan a la estimación de material particulado empleando sensores remotos y datos de fuentes de información puntuales.

5.2.1. Reflectancia tope de la atmósfera

La reflectancia en el tope de la atmósfera (TOA) es la composición de la reflectancia de la superficie y la reflectancia atmosférica, la cual es generada a partir del flujo de energía irradiante o incidente con respecto a una unidad de área en una superficie, también conocido como radiancia. La reflectancia cuantifica el ratio de la energía reflejada con respecto a la cantidad total de energía incidente (Ihlen & USGS, 2019). Según Moran et al. (1999), la reflectancia al tope de la atmósfera o aparente se define como:

$$\rho_A = \frac{\pi L_\lambda}{ESUN_\lambda \cos \theta_s} \quad (1)$$

Donde L_λ es la radiancia espectral en el sensor, $ESUN_\lambda$ significa la irradiancia solar en la banda de interés y θ_s es el ángulo cenital solar en grados. Para estudiar a profundidad la reflectancia a nivel de TOA, esta se puede descomponer en función de la reflectancia del aerosol ρ_{AER} , la reflectancia de Raleigh ρ_{Ray} , la reflectancia superficial ρ_S , la relación de retrodispersión atmosférica S , la transmitancia sol – tierra T_s y sensor – tierra T_v mediante la siguiente ecuación (Vermote et al., 1997):

$$\rho_{TOA} = \rho_{AER} + \rho_{Ray} - \frac{T_s T_v \rho_S}{1 - \rho_S S} \quad (2)$$

Estas variables son fundamentales para estudiar la profundidad óptica de aerosoles y se detallan más adelante.

Los productos de nivel 1B generados por el sensor MODIS para las bandas solares reflectantes (RSB) son la reflectancia TOA y radiación TOA provenientes de la superficie terrestre. Estos se corrigen solamente a nivel instrumental para efectos como compensaciones electrónicas, variaciones angulares de la reflectancia del espejo de escaneo, no linealidades de los convertidores de analógico al digital y variaciones de ganancia causadas por las temperaturas del instrumento y el plano focal. Estos valores corregidos dn^{**} se encuentran en un rango dinámico $[D_{min}, D_{max}]$ y se escalan a una variable entera, denominado entero escalado (SI) que toma valores radiométricos en el rango $[0, 32767]$ (Toller et al., 2017) que corresponde a la mitad de los valores de 2^{16} bits. La figura 3 muestra la relación que existe entre los valores corregidos y los valores escalados.

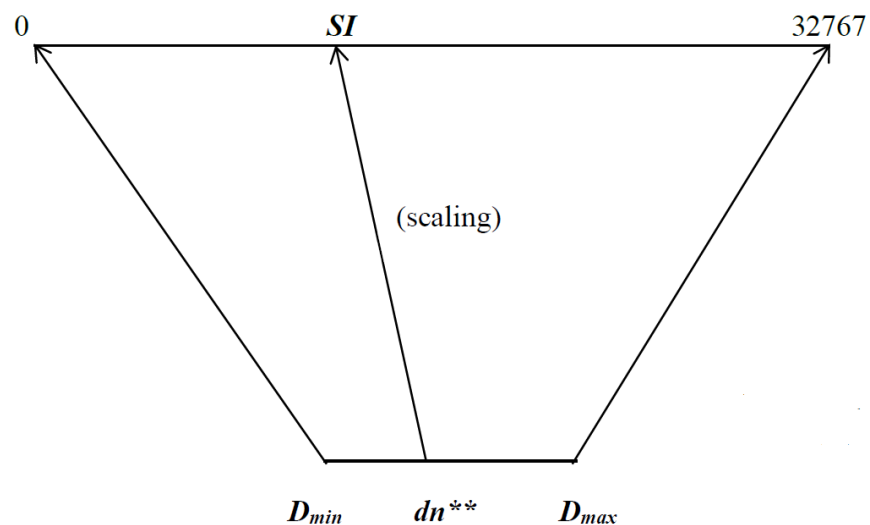


Figura 3 Relación entre los valores de enteros escalados (SI) y valores corregidos instrumentalmente (dn^{**})

Fuente: Toller et al., 2017

La reflectancia a nivel de TOA (ρ) se calcula mediante el algoritmo de calibración lineal desarrollado por el equipo de soporte y caracterización de MODIS (MCST) para cualquier píxel, dado por:

$$\rho = m_0 + m_1 dES^2 dn^* \quad (3)$$

Donde m_0 y m_1 son representados por parámetros de calibración obtenidos por lecturas del difusor solar iluminado por el sol y se guardan como tablas de consultas; dES es equivalente a la distancia Tierra – Sol en unidades astronómicas (AU) y dn^* son los valores digitales corregidos instrumentalmente. Los productos de nivel 1B almacenan estos parámetros de calibración en un par de parámetros llamados *reflectance_scales* y *reflectance_offsets*; estos parámetros varían y dependen únicamente de la banda reflectante que se quiera escalar. Dado esto, el producto de reflectancia a nivel de TOA (ρ) viene dado por:

$$\rho \cos \theta_s = reflectance_scales_B (SI - reflectance_offsets_B) \quad (4)$$

Donde *reflectance_scales_B* y *reflectance_offsets_B* son los parámetros de calibración para una banda específica, θ_s es el ángulo cenital solar y *SI* corresponde a los enteros escalados o pixeles del producto de nivel 1B. Puesto que la reflectancia a nivel de TOA viene expresada como un producto del coseno del ángulo cenital solar, solo basta con despejar ρ para estimarla (Roger et al., 2011).

5.2.2. Reflectancia superficial

Las condiciones atmosféricas causan perturbaciones en la información reflectante captada por los sensores, si se eliminan estos efectos se podrá cuantificar únicamente la cantidad de luz reflejada por la superficie de la tierra, y por lo tanto, estaría libre de distorsiones producidas por los efectos atmosféricos. A estos datos corregidos se le conoce como reflectancia superficial (Moran et al., 1999). La reflectancia superficial está dada por:

$$\rho_s = \frac{\pi(L_\lambda - L_p)}{T_v((ESUN_\lambda \cos \theta_s T_z) + E_{down})} \quad (5)$$

Donde L_p son los efectos provocados por la bruma, T_v es la transmitancia de la atmósfera en la dirección de visión, T_z es la transmitancia atmosférica en la dirección de iluminación, y E_{down} es la irradiancia difusa descendiente. La reflectancia no tiene unidades, y por lo general se escala entre valores de 0 y 1.

Los productos de nivel 2 generados por el sensor MODIS se calculan a partir de los productos de nivel 1B, ha estos datos se les aplica el algoritmo MOD_PR09.exe para corregir efectos de gases, aerosoles atmosféricos y la recuperación de la profundidad óptica de aerosoles, que posteriormente se aplica en tablas de consultas en una segunda pasada para la corrección atmosférica. Un producto derivado de este algoritmo es la reflectancia superficial terrestre. Los valores de este producto se encuentran en un rango valido de $[-100, 16000]$, dado que los valores de reflectancia superficial se escalan por 10000 y luego se convierten en valores enteros de 16 bits. Para estimar la reflectancia superficial ρ_s , se calcula el producto entre el valor de reflectancia escalada ρ_{Esc} y el factor de escala F_{Esc} almacenado en el campo *scale_factor* contenido en los metadatos del archivo (Roger et al., 2011).

$$\rho_s = \rho_{Esc} * F_{Esc} \quad (6)$$

5.2.3. Profundidad óptica de aerosoles (AOD)

Para cuantificar el AOD se debe tener en cuenta las propiedades de los aerosoles y su comportamiento al interactuar con las ondas electromagnéticas. La luz sufre variaciones durante su trayectoria desde la fuente, hasta el sensor o receptor que la capte, estas perturbaciones generadas por las partículas producen dispersión o absorción del haz de luz. La recuperación de estas variaciones en un parámetro, se denomina espesor óptico de aerosoles (Cespedes et al., 2015).

La dispersión atmosférica se puede modelar empleando la dispersión de Mie y Raleigh. La dispersión de Mie modela la dispersión de las ondas electromagnéticas para partículas esféricas, cuyo diámetro es igual a la longitud de onda. Por otra parte, la dispersión de Raleigh modela la dispersión de partículas que tienen un diámetro menor al tamaño de la longitud de onda incidente (Vermote et al., 1997). A continuación, se describen las ecuaciones y parámetros empleados para recuperar la profundidad óptica de aerosoles.

5.2.3.1. Reflectancia de la trayectoria de Raleigh

La reflectancia de la trayectoria de Raleigh cuantifica la capacidad de las partículas de dispersar la luz en función de su longitud de onda. Esta alteración se puede calcular mediante la siguiente ecuación (Mei et al., 2013) :

$$R_{Ray} = \frac{\tau_{Ray} P_{Ray}}{4 \cos \theta_s \cos \theta_v} \quad (7)$$

Donde τ_{Ray} es el espesor óptico de Raleigh, P_{Ray} es la función fase de Raleigh, θ_s es el ángulo cenital solar y θ_v es el ángulo cenital de observación. Estas variables se pueden calcular empleando las siguientes ecuaciones (Liang, 2005):

$$\tau_{Ray} = \frac{P_z}{P_o} (0.00864 + 6.5 * 10^{-6} * z) \lambda^{-(3.96 + 0.074\lambda + \frac{0.05}{\lambda})} \quad (8)$$

$$P_{Ray} = \frac{3}{4} (1 + \cos^2 \theta) \quad (9)$$

Donde P_o es la presión a nivel del mar con un valor aproximado de 1013.25 mbar, P_z es la presión atmosférica de la zona de estudio en unidades de milibar (mbar), z es la elevación sobre el nivel del mar en kilómetros, λ es el ancho de longitud de onda en micrómetros (mm) y θ es el ángulo de fase de dispersión.

5.2.3.2. Ángulo de fase de dispersión

El ángulo de fase de dispersión Θ se relaciona con la variación que sufre la luz causada por la interacción con las partículas, este se calcula utilizando la siguiente ecuación (Levy et al., 2007):

$$\Theta = \cos^{-1}(-\cos \theta_s \cos \theta_v + \sin \theta_s \sin \theta_v \cos \phi) \quad (10)$$

Donde θ_s es el ángulo cenital solar, θ_v es el ángulo cenital de observación y ϕ es el azimut relativo. Mediante el ángulo de fase de dispersión se puede estimar la función fase de dispersión, que emplea en modelos de transferencia radiativa para calcular la cantidad de radiación difusa, esta función es derivada de las ecuaciones propuestas en la teoría de Mie.

5.2.3.3. Función fase del aerosol

Teniendo el ángulo de fase de dispersión se calcula la función fase del aerosol, que integra las diferentes direcciones de la luz, describiendo la distribución angular de la dispersión para cada una de estas. La función fase del aerosol está dada por (Rahman et al., 1993):

$$P_a = \frac{1 - g^2}{[1 + g^2 - 2g \cos(\pi - \Theta)]^{\frac{3}{2}}} \quad (11)$$

Donde g es el factor de asimetría y Θ es el ángulo de fase de dispersión.

5.2.3.4. Transmisiones atmosféricas

La interacción de la radiación electromagnética con la atmósfera produce efectos de transmisión atmosférica en la trayectoria sensor – superficie T_s y sol – superficie T_v , estos

efectos son estimados en función de variables anteriormente detalladas empleando las siguientes ecuaciones (Bilal et al., 2013).

$$T_s = e^{-\left(\frac{\tau_{Ray} + \tau_a}{\cos \theta_s}\right)} \quad (12)$$

$$T_v = e^{-\left(\frac{\tau_{Ray} + \tau_a}{\cos \theta_v}\right)} \quad (13)$$

5.2.3.5. Retrodispersión superficie – atmósfera

Este tipo de retrodispersión describe la cantidad de luz que es reflejada por la atmósfera en la misma dirección de en qué se reflejó por la superficie, es decir, aquellas ondas electromagnéticas que vuelven en la dirección de incidencia. La relación de retrodispersión atmosférica S se emplea para explicar la reflexión múltiple existente entre la atmósfera y la superficie en función del espesor óptico de Raleigh τ_{Ray} , la profundidad óptica de aerosoles τ_a y el factor de asimetría g , mediante la siguiente ecuación (Liang, 2005):

$$S = (0.92\tau_{Ray} + (1 - g)\tau_a) e^{-(\tau_{Ray} + \tau_a)} \quad (14)$$

5.2.3.6. Algoritmo simplificado de recuperación de aerosoles (SARA)

La profundidad óptica de aerosoles (AOD) se estimó empleando un algoritmo simplificado de recuperación de aerosoles (SARA) propuesto por Bilal et al. (2013), que se destaca por aplicar las ecuaciones de transferencia radiativa (RTM) directamente al conjunto de datos de entrada, evitando la tradicional técnica de crear tablas de búsqueda (LUT). Este modelo se ejecuta bajo tres supuestos, como primera media se supone que la superficie del área de estudio sea de tipo lambertiana, es decir que refleja la energía incidente en todas las direcciones de forma homogénea. Como segundo supuesto se estima la aproximación de una dispersión simple y por último supone que el albedo de dispersión simple y el factor asimétrico no presentan cambios importantes

específicamente en el área y periodo de estudio. La ecuación del SARA para una longitud de onda específica es (Bilal & Nichol, 2015; Bilal et al., 2013, 2014; Luo et al., 2015) :

$$\tau = \frac{4\mu_s\mu_v}{\omega_o P_a} \left[\rho_{TOA} - \rho_{Ray} - \frac{e^{-\left(\frac{\tau_{Ray} + \tau_a}{\mu_s}\right)} e^{-\left(\frac{\tau_{Ray} + \tau_a}{\mu_v}\right)} \rho_S}{1 - \rho_S (0.92\tau_{Ray} + (1 - g)\tau_a) e^{-(\tau_{Ray} + \tau_a)}} \right] \quad (15)$$

Donde ρ_{TOA} es la reflectancia en el tope de la atmósfera (TOA), ρ_S representa la reflectancia superficial, ρ_{Ray} es equivalente a la reflectancia de Raleigh, τ_{Ray} es el espesor óptico de Raleigh, τ_a significa el espesor óptico de aerosoles, g es el factor de asimetría, P_a es la función fase del aerosol, ω_o equivale al albedo de dispersión simple, μ_s y μ_v corresponden al coseno de los ángulos cenit solar y de observación respectivamente.

5.2.4. Red Robótica de Aerosoles (AERONET)

AERONET es una red de fotómetros solares terrestres administrados por la NASA y el Photométrie pour le Traitement Opérationnel de Normalisation Satellitaire (PHOTONS). Estos fotómetros miden cantidades y propiedades de los aerosoles atmosféricos; entre estos tenemos la profundidad óptica de aerosoles, el albedo de dispersión simple y el factor de asimetría del aerosol, entre otros. Actualmente los algoritmos de procesamiento son de versión 3, que cuenta con 3 niveles de calidad de los datos. El nivel 1.0 corresponde a los datos sin procesamiento posterior a su observación, a los datos de nivel 1.5 se les aplica un control de calidad y mascara de nubosidad, y los datos pertenecientes al nivel 2.0 se garantizan que estén libres de distorsiones o cualquier error asociado durante la observación (AERONET, 2021b).

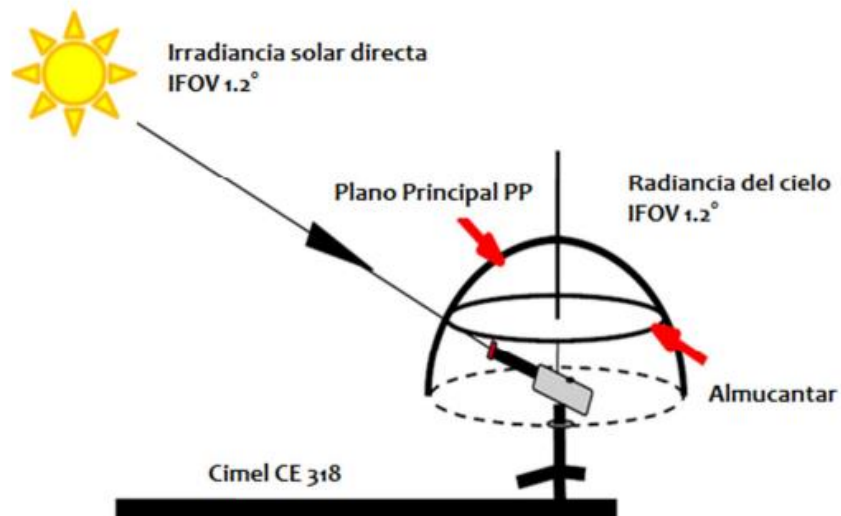


Figura 4 Características principales del sistema de medición AERONET.

Fuente: Alegría, 2015.

La figura 4 muestra de manera general las características del sistema de medición de los fotómetros solares terrestres. Los datos de AERONET se obtuvieron de manera gratuita mediante su página oficial: https://aeronet.gsfc.nasa.gov/new_web/index.html. Estos datos corresponden a la profundidad óptica de aerosoles (AOD), el albedo de dispersión simple (SSA) y el factor de asimetría del aerosol (AST) a una longitud de onda de 440 y 675 nm para los años 2019 y 2020. Se descargaron datos de nivel 1.5 y 2.0, empleando datos de nivel 1.5 solamente si no había registro de datos de nivel 2.0. En los casos donde no se encontraron registros de ningún nivel se tomó el dato del día siguiente, dado que las partículas en suspensión no se sedimentan en periodos cortos, sino que permanecen suspendidas en el aire por su tamaño y densidad (MAVDT, 2010).

5.2.5. Análisis de datos funcionales

El Análisis de Datos Funcionales (ADF) es una rama de la estadística que estudia y analiza la información contenida en curvas, superficies o cualquier elemento que varía sobre un proceso continuo, principalmente en función del tiempo (Ramsay & Silverman, 2005). Las características de las funciones hacen que surja todo un conjunto de herramientas para estudiarlas correctamente (Manteiga & Vieu, 2007). En general, cualquier observación que varíe de forma continua se puede representar por un dato funcional, desde la precipitación de una región hasta la contaminación atmosférica. En la práctica, la información es evaluada por sensores que toman las mediciones de una determinada variable en diferentes momentos del tiempo en un rango específico. De esta forma, una observación puede expresarse mediante la familia $\{X(t_j)\}_{j=1,\dots,J}$. Actualmente, los avances tecnológicos en los sistemas de información hacen posible disminuir la periodicidad en la que es registrado el dato, este es uno de los motivos importantes del uso del ADF, de modo que se podría considerar estas muestras como observaciones de un espacio continuo $\chi = \{X(t); t \in (t_{min}, t_{max})\}$. Dado esto, una variable funcional es aquella variable aleatoria que toma valores en un espacio funcional, es decir, de dimensión infinita, donde una observación de X se denomina un dato funcional (Ferraty & Vieu, 2006).

5.2.5.1. Espacio funcional

Un espacio funcional F es un espacio de funciones. Cada función en el espacio puede ser pensado como un punto (Ramsay & Silverman, 2005), por ejemplo:

- $C[a, b]$ es el conjunto de todas las funciones continuas con valores reales en el intervalo $[a, b]$.
- $L_1[a, b]$ es el conjunto de todas las funciones reales cuyos valores absolutos es integrable en el intervalo $[a, b]$.
- $L_2[a, b]$ es el conjunto de todas las funciones reales cuyos valores absolutos es cuadrado integrable en el intervalo $[a, b]$.

5.2.5.2. Base de funciones

En la práctica, los datos funcionales son obtenidos de forma discreta, por lo que es necesario convertirlos a funciones suaves sobre las cuales se pueda trabajar, para esto pueden emplearse diferentes tipos de ajustes. Teniendo en cuenta lo anterior, la representación de estos datos como datos funcionales a partir de una base ortonormal brinda una ventaja teórica y práctica, que se utilizan para transformar datos discretos en datos funcionales.

A continuación, se presenta una revisión teórica del concepto de bases de funciones y las bases más utilizadas para la construcción de datos funcionales (Ramsay & Silverman, 2005).

- **Bases:** Es un conjunto de funciones conocidas $\{\phi_k\}_{k \in \mathbb{N}}$ tales que cualquier función puede ser aproximada, tan bien como se requiera, mediante una combinación lineal de K funciones.
- **Bases de Fourier:** Estas bases están formadas por senos y cosenos, y es factible conseguir la expresión derivada en cualquier orden, dada sus características de funciones periódicas, los datos que mejor se ajustan a este tipo de bases son los que presentan un cierto tipo de ciclo repetitivo.
- **Bases B-Splines:** Estas bases combinan la eficiencia computacional de los polinomios con una mayor flexibilidad, que permite elegir un número menor de funciones (K) para obtener buenos resultados. Estas bases se emplean comúnmente para datos no periódicos.

Las bases de B-Splines permiten aproximar todas estas funciones facilitando así su manejo. Sus características se basan en las siguientes propiedades:

- Cada elemento de la base $\phi_k(t)$ será una función Spline de orden m y partición τ .
- Cualquier combinación lineal de funciones Splines es una función Spline.
- Cualquier función Spline de orden m sobre la partición τ se puede expresar como combinación lineal de las funciones de la base.

Sea $B_k(t, \tau)$ la representación del valor del k –ésimo elemento de la base sobre la partición τ en el instante t , la función Spline $S(t)$ se representa como:

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau) \quad (16)$$

5.2.5.3. Elección del número de funciones base

El número de funciones k determina que tan suave o detalladas son las funciones que representan el comportamiento de los datos. Un número alto de funciones incluiría la variabilidad en el comportamiento de las funciones y un número bajo podría no representar ciertas características que pueden brindar gran información en la reconstrucción de los datos funcionales. Dado esto, no se puede determinar un número k estándar, sino un parámetro que se calcula dependiendo de las características de los datos (Ramsay & Silverman, 2005).

Para elegir el número óptimo de funciones base se debe encontrar un equilibrio entre valores muy altos y valores muy bajos. Según Ramsay & Silverman (2005), empleando el Error Cuadrático Medio (ECM) se puede encontrar el número óptimo k , puesto que, se puede expresar como la suma de dos componentes, el primero que cuantifica la varianza y el segundo cuantifica el sesgo en la estimación.

5.2.5.4. Construcción de datos funcionales

Para explicar la metodología de construcción de datos funcionales se debe definir el concepto de variable aleatoria funcional. Según Ferraty & Vieu (2006), este concepto se define como: “Una variable aleatoria X es llamada variable funcional (f.v) si toma valores en un espacio dimensional infinito (o espacio funcional). Una observación x de X se denomina un dato funcional”. Teniendo en cuenta esto, a partir de observaciones discretas se busca reconstruir su observación funcional teniendo presente su posible error de medición.

Un dato funcional se puede representar como: sea $x = \{x(t); t \in (t_{min}, t_{max})\}$ un dato funcional y sean y_1, y_2, \dots, y_n las n observaciones de la función x para los valores $t = 1, t = 2, \dots, t = n$ donde t por lo general corresponde al tiempo. Suponiendo de que x es una función suave, se emplea la teoría de regresión en el contexto funcional, al representar a $x(t)$ como una combinación lineal de funciones de una base, de la siguiente manera:

$$x(t) = \sum_{k=1}^K c_k \phi_k(t) \quad (17)$$

Donde $\phi_k = (\phi_1(t), \dots, \phi_k(t))$ representa el número de K funciones conocidas de una base de funciones definida y los c_k son los coeficientes desconocidos en la combinación lineal. Estos coeficientes se pueden estimar mediante la suma de cuadrados del error definida en forma matricial como:

$$SMSSE(y|c) = (y - \Phi c)'(y - \Phi c) \quad (18)$$

Donde Φ es una matriz $n \times K$ compuesta por los valores de las bases ϕ_k determinadas en los tiempos del intervalo (t_{min}, t_{max}) . Optimizando la expresión (18) se pueden obtener los coeficientes estimados \hat{c} que minimizan la suma de cuadrados del error, de la siguiente manera:

$$\hat{c} = (\Phi' \Phi)^{-1} \Phi' y \quad (19)$$

Teniendo en cuenta los coeficientes calculados, los datos funcionales o curvas deseadas se expresan como:

$$\Phi \hat{c} = \hat{y} = \Phi (\Phi' \Phi)^{-1} \Phi' y \quad (20)$$

Para la creación de los datos funcionales de PM2.5 y AOD, se agruparon las mediciones y datos por día, teniendo un total de 234 datos para la variable estudiada en las 9 estaciones. Para crear los datos funcionales por día, se procede a calcular el número óptimo de funciones base k empleando la función *optim.basis* de la librería *fda.usc* en Rstudio y el parámetro de suavización λ (Ramsay et al., 2009); Estos valores se entienden por óptimos a los números que minimizan el criterio de validación cruzada generalizada (GCV). Esta función calcula la suma de cuadrados por GCV en interacciones secuenciales de valores de ambos parámetros y calcula la mejor combinación resultante de la suma de cuadrados más baja.

5.2.5.5. Elección del número de funciones de la base (K)

El número de funciones de la base indica en cierta medida el grado en que los datos son suavizados mediante la curva construida. Ramsay & Silverman (2005), proponen que para seleccionar el número K se debe tomar en cuenta el sesgo y la varianza de la estimación, representados por:

$$Sesgo[\hat{x}(t)] = x(t) - E[\hat{x}(t)] \quad (21)$$

$$Var[\hat{x}(t)] = E\{[\hat{x}(t) - E[\hat{x}(t)]]^2\} \quad (22)$$

Teniendo en cuenta esto, un indicador que permite controlar tanto el sesgo como la varianza es el Error Cuadrático Medio (ECM), dado por:

$$ECM[\hat{x}(t)] = Sesgo^2[\hat{x}(t)] + Var[\hat{x}(t)] \quad (23)$$

La relación entre el sesgo y la varianza indican que se puede tolerar un poco de sesgo, siempre y cuando la varianza presente una gran reducción. Adicional a esto, Ramsay & Silverman (2005), plantean emplear la validación cruzada generalizada (VCG). Para explicar la VCG se debe conceptualizar sobre la Validación Cruzada (VC). La VC es una metodología usada para la elección de parámetros que controlen el equilibrio entre la

bondad del ajuste y la capacidad predictiva. El método se basa en extraer del conjunto de muestra sucesivamente cada una de las observaciones x_i , estimar el modelo con los datos restantes y predecir el dato ausente para finalmente comparar la predicción con el dato real. La validación cruzada generalizada es una modificación de la VC, donde se basa en reemplazar los valores de la diagonal principal de la matriz S por su valor promedio, expresada de una forma simplificada como (Delicado, 2008):

$$ECM_{VCG}(K) = \frac{n \hat{\sigma}_\varepsilon^2}{n - \nu} \quad (24)$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n - \nu} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (25)$$

Donde $\nu = \text{Traza}(S)$ y $\hat{\sigma}_\varepsilon^2$ es una estimación de la varianza de los valores residuales del modelo y K_{VCG} es el parámetro determinado por el criterio de validación cruzada generalizada (Delicado, 2008).

5.2.5.6. Estadística descriptiva funcional

Los análisis estadísticos para un conjunto de datos deben ser de tipo descriptivo y exploratorio. Teniendo en cuenta esto, Ramsay & Silverman (2005), proponen las siguientes técnicas descriptivas y exploratorias para un conjunto de datos funcionales: Sea $x_i(t)$ $i = 1, 2, \dots, N$ una muestra de curvas o datos funcionales, la media, varianza, covarianza y correlación funcional son iguales a:

- Media funcional:

$$\bar{x}(t) = N^{-1} \sum_{i=1}^N x_i(t) \quad (26)$$

- Varianza funcional:

$$Var_x(x(t)) = (N - 1)^{-1} \sum_{i=1}^N (x_i(t) - \bar{x}(t))^2 \quad (27)$$

- Covarianza funcional:

$$Cov_x(t_1, t_2) = (N - 1)^{-1} \sum_{i=1}^N (x_i(t_1) - \bar{x}(t_1))(x_i(t_2) - \bar{x}(t_2)) \quad (28)$$

- Correlación funcional:

$$Cov_x(t_1, t_2) = (N - 1)^{-1} \sum_{i=1}^N (x_i(t_1) - \bar{x}(t_1))(x_i(t_2) - \bar{x}(t_2)) \quad (29)$$

5.2.6. Modelo de regresión funcional concurrente

Este modelo se caracteriza por tener una respuesta funcional a partir de variables regresoras funcionales, es decir, que acepta e integra parámetros funcionales. Este modelo está representado en bases restringidas tanto en la respuesta como en la variable regresora (Ramsay & Silverman, 2005). El modelo se define como

$$Y_i(t) = \beta_0(t) + \beta_1(t)X_{i1}(t) + \beta_2(t)X_{i2}(t) + \dots + \beta_q(t)X_{iq}(t) + \epsilon_i(t) \quad (30)$$

En notación matricial se denota como:

$$Y(t) = X(t)\beta(t) + \epsilon(t) \quad (31)$$

Para estimar las concentraciones de PM2.5 empleando las curvas o datos funcionales de las variables anteriormente construidas, se emplea un modelo de regresión funcional como se explica en la sección 5.2.5. A continuación, se expresa el modelo previamente descrito:

$$PM2.5_i(t) = \beta_0(t) + \beta_1(t)AOD_i(t) + \epsilon_i(t) \quad (32)$$

Donde, $PM2.5_i$ equivale al valor de PM2.5 calculado por el modelo de regresión funcional; β_0 es el intercepto funcional, β_1 corresponde a la pendiente funcional y ϵ_i es el componente aleatorio del modelo correspondiente al error funcional, $i = 1, 2, \dots, 9$ correspondiente a las 9 estaciones de monitoreo y (t) representa el tiempo en días.

5.2.6.1. Intervalos de confianza en el modelo de regresión funcional

Según las expresiones propuestas por Ramsay & Silverman (2005), primero se definen los residuales del modelo de regresión funcional concurrente mediante la siguiente forma:

$$r(t) = T(t) - X(t) \Theta(t) \hat{b} \quad (33)$$

La matriz de varianzas y covarianzas de los residuales funcionales queda definida como:

$$\sum_e^* = \frac{1}{N} r(t) r^T(t) \quad (34)$$

Donde r es la matriz de residuales.

5.2.6.2. Bondad de ajuste del modelo

Empleando la metodología de validación cruzada se evalúa el ajuste del modelo. Este se basa en predecir la curva de material particulado excluyendo una ubicación donde se conozca el dato funcional y así predecir la curva; seguido a esto, se compara con el dato

funcional conocido correspondiente a la ubicación anteriormente excluida y así determinar el poder predictivo. Otros métodos de evaluación son el coeficiente de determinación funcional $R^2(t)$, propuesto por Ramsay & Silverman (2005).

$$R^2(t) = 1 - \int \{\hat{y}_i(t) - y_i(t)\}^2 dt / \int \{y_i(t) - \bar{y}_i(t)\}^2 dt \quad (35)$$

5.3. MARCO REFERENCIAL

Las investigaciones a nivel internacional del uso de los sensores remotos como alternativa para la evaluación de la calidad del aire toman gran interés, a medida que se lanzan diferentes tipos de sensores con múltiples aplicaciones; como por ejemplo la investigación realizada por Gupta et al. (2006), que evaluaron la calidad del aire con respecto al material particulado PM_{2.5} en áreas urbanas ubicadas en Sídney, Delhi, Hong Kong, Nueva York y países como Suiza. Emplearon datos recolectados de 1 año del espesor óptico de aerosoles (AOD) del Espectroradiómetro de imágenes de media resolución (MODIS) e información tomada en tierra por estaciones de monitoreo PM_{2.5}. Estos obtuvieron una correlación lineal de 0.96 entre los dos métodos de toma de datos. También determinaron que la relación entre métodos depende fundamentalmente de la humedad relativa (HR), la nubosidad, concentración de aerosoles y la capa de mezcla atmosférica. Ellos demuestran que el AOD obtenido a través de técnicas de percepción remota funciona como sustituto o complemento para el estudio de la calidad del aire sobre la tierra.

Por otra parte, en Australia, Woldemeskel et al. (2013), desarrollaron una metodología para determinar el beneficio de combinar datos de red de estaciones y datos de imágenes de satélite para cuantificar la precipitación espacial. Estos emplearon datos de precipitaciones observadas en un periodo de 10 años (enero de 1998 – diciembre de 2007) en 230 pluviómetros que garantizaron obtener mediciones de lluvia mensual de alta calidad, según la investigación de Lavery et al. (1997). Para calcular la cuadrícula de precipitación pluviométrica utilizaron el método de ponderación linealizado y los pesos en cada cuadrícula obtenidos en función de las variaciones de error en cada grupo de datos. Para evaluar los resultados, los autores emplean la validación cruzada (CV) en ubicaciones de las estaciones de monitoreo. De acuerdo con los resultados de la validación, concluyeron que la combinación de estas diferentes fuentes de información optimizó la estimación de precipitación espacial, aun cuando la red de estaciones es limitada.

En Hong Kong, China, los autores Bilal et al. (2013), desarrollaron un algoritmo simplificado de recuperación de aerosoles (SARA) que no emplea los métodos comunes de construir tablas de búsquedas (LUT) mediante el modelo de transferencia radiativa (RTM). Este algoritmo fue probado utilizando productos MODIS a una resolución espacial de 500 metros y mediciones de la red robótica de aerosoles (AERONET). El algoritmo fue validado empleando datos del fotómetro solar Microtops II ubicado en el aeropuerto internacional de Hong Kong y datos de una estación AERONET localizada en la zona urbana de Hong Kong. El espesor óptico de aerosoles recuperado por SARA evidenció una alta correlación con datos medidos en tierra, tanto en zonas urbanas, como rurales, obteniendo un coeficiente de correlación promedio de $R = 0.964$. Los resultados de validación demuestran la capacidad óptima que tiene el SARA de estimar las concentraciones de aerosoles en una zona de estudio.

Otros autores como Bilal et al. (2014), Bilal & Nichol (2015) y Luo et al. (2015) emplean el SARA como base teórica para estimar las concentraciones de aerosoles atmosféricos. Estos últimos, estiman la profundidad óptica de aerosoles en Beijing, China utilizando imágenes satelitales del programa Landsat, con una resolución espacial de 30 metros e integrando parámetros meteorológicos. Validaron los resultados obtenidos con mediciones AERONET en tierra, obteniendo un coeficiente de correlación promedio de $R = 0.741$. Los autores concuerdan con la simplicidad y precisión del algoritmo para estudiar la profundidad óptica de aerosoles a una escala más detallada.

En Latinoamérica, la percepción remota como alternativa para la evaluación de la calidad del aire también ha tenido importantes aportes, como es el caso del estudio llevado a cabo por Hinojosa-Baliño et al. (2019) en ciudad de México, Alvarez & Padilla (2016) y Abad & Mejía (2017) en Ecuador; estos últimos emplearon una ecuación empírica para determinar las concentraciones de PM₁₀ en zonas donde no se tiene cobertura por los métodos de medición tradicionales. Se correlacionaron datos de 3 estaciones de monitoreo y valores obtenidos de imágenes satelitales Landsat 8 para los años 2013 y 2014, mediante un modelo de regresión lineal múltiple entre la reflectancia tope de la atmósfera (TOA) y los niveles digitales (ND); obtuvieron como resultado un coeficiente de determinación múltiple $R^2 = 0.55$ y un valor- p de 0.002. Para validar el modelo

empleado, realizaron una validación cruzada empleando el método que deja un dato por fuera. También concluyeron que el uso de bandas térmicas en los modelos optimiza considerablemente el coeficiente de determinación.

A nivel nacional, en Colombia, la percepción remota ha tenido aportes considerables, como por ejemplo el estudio realizado en el Valle del Cauca por Rojas et al. (2017), donde presentaron una metodología para estimar el material particulado PM10 usando datos del Espectrorradiómetro de imágenes de media resolución (MODIS), específicamente el producto MOD02 y una red de 7 estaciones de monitoreo, procesando la información a través de Scripts en el lenguaje de programación Python y empleando modelos gaussianos para determinar la distribución espacial del PM10 para los años 2011 y 2012. Los resultados presentados demuestran una relación directamente proporcional entre el espesor óptico de aerosoles (AOD) y el PM10, lo que destaca el método de teledetección como uno confiable y complementario a otros tipos de métodos tradicionales para la evaluación de la calidad del aire.

Otro aporte nacional es la investigación realizada por Castro (2019), donde plantea una metodología para estimar concentraciones de PM10 en la capa de la troposfera mediante sensores remotos. El autor empleó 4 imágenes satelitales del sensor LANDSAT 8 para el año 2016 y una red de 6 estaciones de monitoreo ubicadas en el Valle de Aburrá, que contiene 10 municipios del departamento de Antioquia. El modelo que mejor se ajustó obtuvo un coeficiente de determinación de $R^2 = 0.963$, el modelo se validó usando una quinta imagen satelital de la zona de estudio. Los resultados evidenciaron la relación proporcional existente entre el AOD y el PM10, obteniendo resultados positivos de los métodos que emplean sensores remotos para el monitoreo de la calidad del aire.

Las investigaciones que modelan variables continuas utilizando análisis de datos funcionales, han incrementado en la última década, destacando la publicación realizada por Ramsay & Silverman (2005), donde se detalla la definición, construcción y múltiples aplicaciones que tienen este tipo de métodos. Entre estas, se destaca el trabajo realizado por Suhaila et al. (2011), que a partir de datos de lluvia observados diariamente de forma discreta para 50 estaciones, modelaron la precipitación diaria en la península de

Malasya. Como resultado obtuvieron las curvas o funciones que describen el proceso de precipitación de una forma continua en cada región del país empleando el análisis de datos funcionales (ADF).

En Colombia, se destaca el aporte realizado por Ospína et al. (2019), que presentan un modelo de regresión funcional concurrente para integrar datos de estaciones climáticas e imágenes de satélite para modelar el comportamiento funcional de la precipitación pentadal (acumulada de 5 días) en el Valle del Cauca, Colombia. Los resultados presentados muestran una relación directa entre ambas fuentes de información, presentando una mayor correlación en los tiempos donde los registros de precipitación terrestre fueron mayores, principalmente en los fenómenos del NIÑO y NIÑA. También concluyen en la importancia de la información de las imágenes de satélite como complemento para estimar el comportamiento funcional de la precipitación en base a información de mediciones en tierra.

6. METODOLOGÍA

La metodología propuesta se basa en la adquisición de productos MODIS, datos de estaciones de monitoreo, parámetros de calibración MODIS, datos AERONET de las propiedades físicas de la atmósfera, procesamiento de datos y estimación de concentraciones de PM2.5 en el área de estudio. A continuación, en la figura 5 se detalla la metodología propuesta.

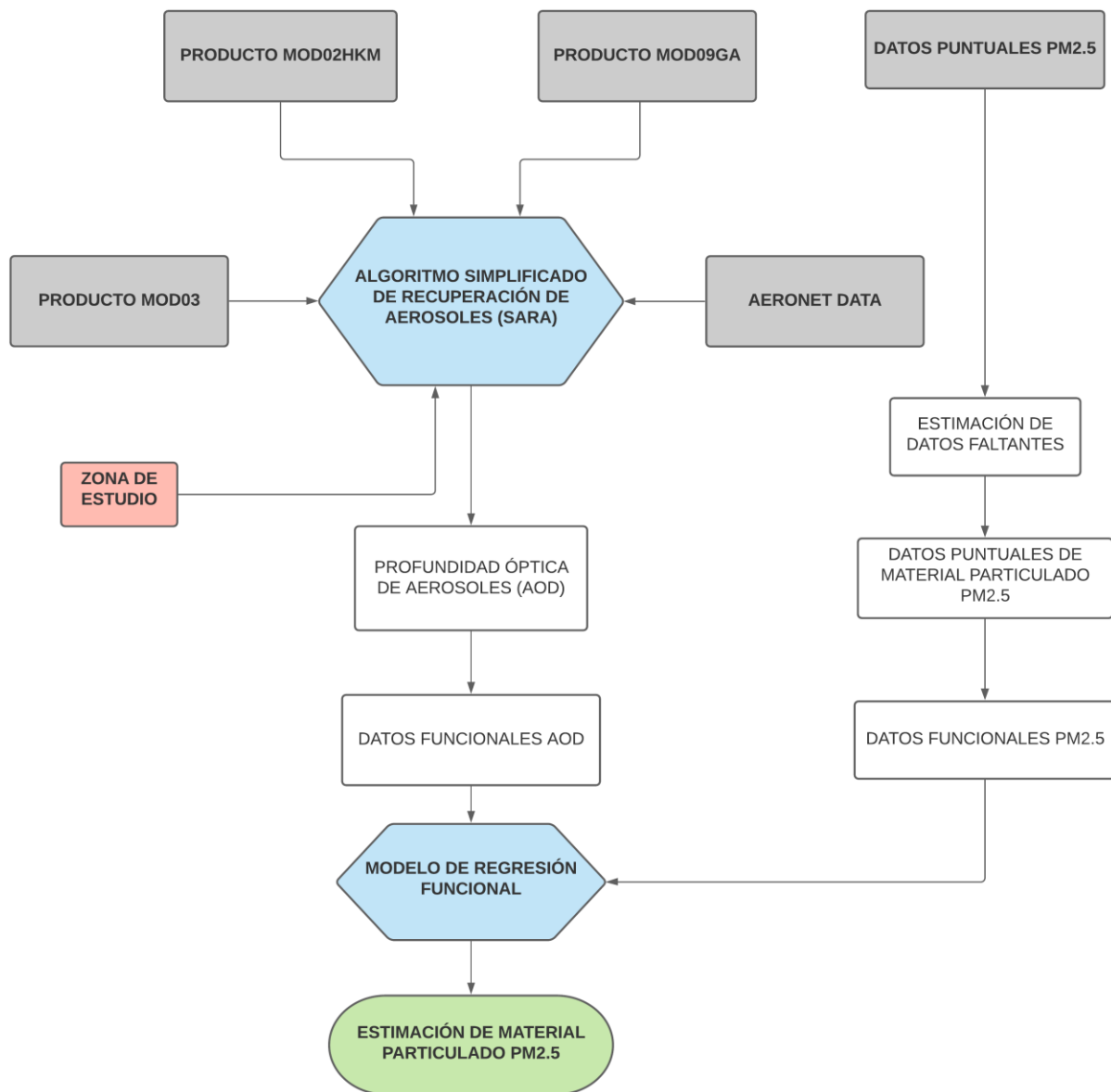


Figura 5 Esquema metodológico general propuesto

6.1. ÁREA DE ESTUDIO

La ciudad de Medellín es la capital del departamento de Antioquia, Colombia. Su área metropolitana se estima en 382 km^2 aproximadamente y hace parte de la región natural llamada el Valle de Aburra, la cual está compuesta por los 10 principales municipios del departamento. Su precipitación promedio anual es de 1.571 mm , su temperatura media anual de 24°C y su altitud es de 1495 msnm aproximadamente (Alcaldía de Medellín, 2006). A continuación, la figura 6 detalla la ubicación espacial del área de estudio.

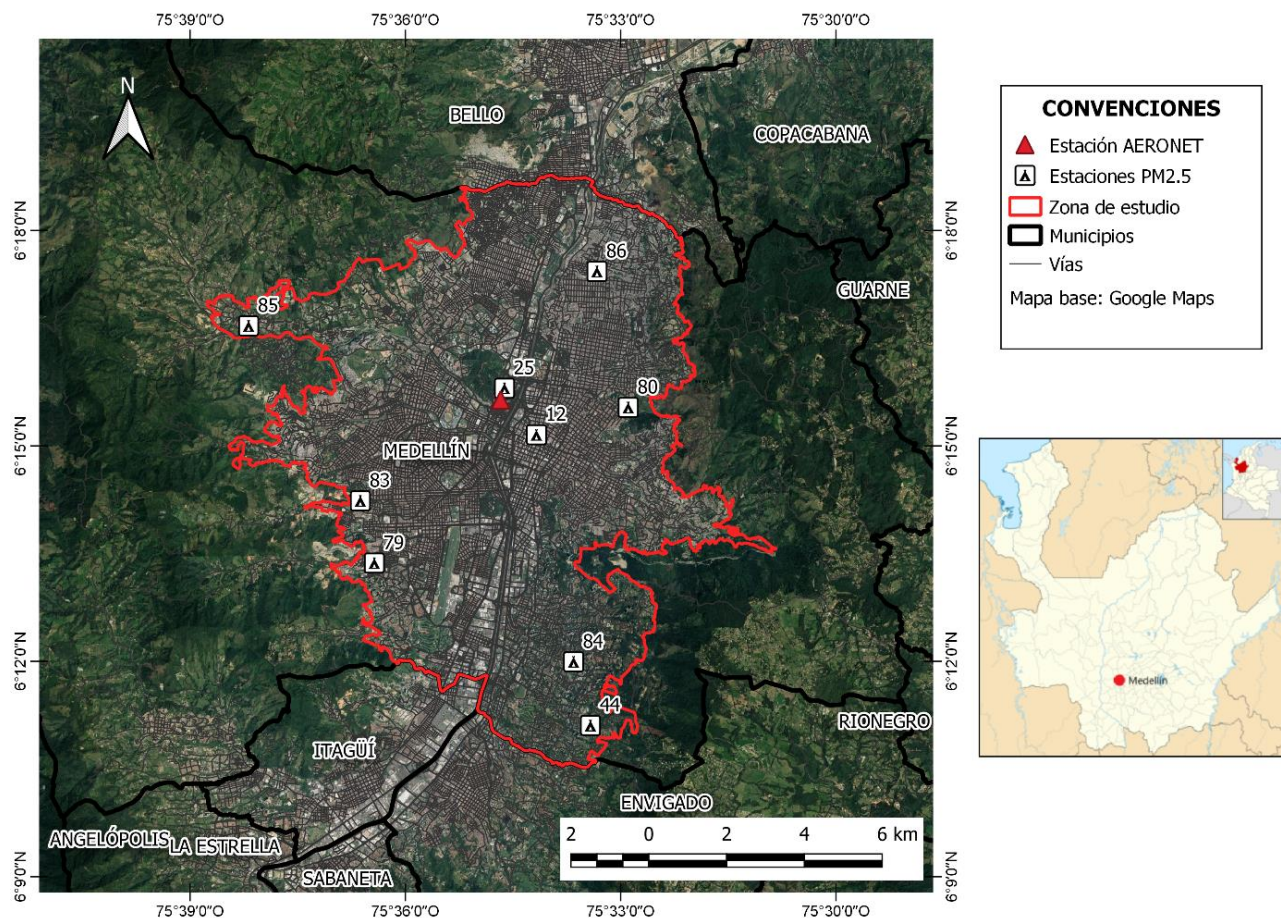


Figura 6 Área de estudio y ubicación de estaciones PM2.5 y AERONET.

6.2. FUENTES DE INFORMACIÓN

Las principales fuentes de información para llevar a cabo la metodología corresponden a las imágenes satelitales, productos MODIS, datos de estaciones de monitoreo de la calidad de aire y datos de la Red Robótica de Aerosoles (AERONET). Estos datos se obtuvieron para los años 2019 – 2020 y 9 estaciones disponibles en la zona de estudio. A continuación, se detallan estos procesos de adquisición.

6.2.1. Datos MODIS

Para el desarrollo de la metodología se emplearon tres conjuntos de imágenes satelitales o productos derivados del sensor MODIS Terra. Los datos de estos productos se almacenan en formato de datos jerárquico (HDF-EOS), este formato permite estandarizar la distribución en múltiples plataformas del conjunto de datos científicos (SDS), entre esta información se destacan los datos espectrales y sus características (Metadatos).

El primer producto corresponde al MOD02HKM el cual es de nivel 1B (Ver tabla 3) que tiene una resolución espacial de 500 metros por pixel, conteniendo las radiancias de apertura calibradas y geolocalizadas para 7 bandas, cubre la región de 0.45 a 2.20 μm del espectro electromagnético (MODIS Characterization Support Team (MCST), 2017a). Estas imágenes se obtuvieron de forma gratuita mediante el sitio web de la NASA LAADS DAAC: <https://ladsweb.modaps.eosdis.nasa.gov/search/> . La banda 4 centrada en la longitud de onda 0.55 μm , correspondiente al verde, fue empleada para esta investigación, puesto que esta es requerida por el SARA en la metodología propuesta (Bilal et al., 2013). Los criterios de selección de imágenes se basaron en la disponibilidad y nubosidad dentro de la zona de estudio, se obtuvieron un total de 26 imágenes satelitales para los años 2019 y 2020.

Tabla 3 Productos MOD02HKM para los años 2019 - 2020

Fecha (DD-MM-AAA)	Día Juliano	Hora	Escena MOD02HKM
1/01/2019	001	16:05	MOD02HKM.A2019001.1605.061.2019002013822.hdf
2/01/2019	002	15:10	MOD02HKM.A2019002.1510.061.2019003012022.hdf
3/01/2019	003	15:50	MOD02HKM.A2019003.1550.061.2019004013351.hdf
4/01/2019	004	14:55	MOD02HKM.A2019004.1455.061.2019005012616.hdf
30/01/2019	030	15:35	MOD02HKM.A2019030.1535.061.2019030212728.hdf
15/02/2019	046	15:35	MOD02HKM.A2019046.1535.061.2019047011837.hdf
22/02/2019	053	15:40	MOD02HKM.A2019053.1540.061.2019054012839.hdf
1/05/2019	121	15:15	MOD02HKM.A2019121.1515.061.2019122042637.hdf
24/05/2019	144	15:20	MOD02HKM.A2019144.1520.061.2019145013401.hdf
26/06/2019	177	16:05	MOD02HKM.A2019177.1605.061.2019178010953.hdf
6/07/2019	187	15:05	MOD02HKM.A2019187.1505.061.2019188011728.hdf
5/08/2019	217	15:15	MOD02HKM.A2019217.1515.061.2019218011114.hdf
9/12/2019	343	15:25	MOD02HKM.A2019343.1525.061.2019344012020.hdf
6/01/2020	006	15:50	MOD02HKM.A2020006.1550.061.2020007011107.hdf
10/01/2020	010	15:25	MOD02HKM.A2020010.1525.061.2020011170710.hdf
11/01/2020	011	16:10	MOD02HKM.A2020011.1610.061.2020012010929.hdf
12/01/2020	012	15:15	MOD02HKM.A2020012.1515.061.2020013012053.hdf
18/02/2020	049	15:35	MOD02HKM.A2020049.1535.061.2020050020107.hdf
20/05/2020	141	16:00	MOD02HKM.A2020141.1600.061.2020142011927.hdf
4/06/2020	156	15:15	MOD02HKM.A2020156.1515.061.2020160202544.hdf
13/07/2020	195	15:20	MOD02HKM.A2020195.1520.061.2020196011341.hdf
31/07/2020	213	15:10	MOD02HKM.A2020213.1510.061.2020214010726.hdf
2/08/2020	215	14:55	MOD02HKM.A2020215.1455.061.2020216012100.hdf
10/10/2020	284	15:15	MOD02HKM.A2020284.1515.061.2020285012756.hdf
2/11/2020	307	15:20	MOD02HKM.A2020307.1520.061.2020308012111.hdf
7/11/2020	312	15:40	MOD02HKM.A2020312.1540.061.2020315034233.hdf

El segundo producto empleado fue el MOD09GA versión 6 (Ver tabla 4) que proporciona la reflectancia superficial terrestre estimada para el rango de bandas 1 – 7 del sensor MODIS Terra y tiene una resolución espacial de 500 metros. Este producto viene corregido para efectos atmosféricos como gases, aerosoles, dispersión de Rayleigh y efectos instrumentales (Vermote & Wolfe, 2015), puesto que este producto se genera a partir del producto de radiancias calibradas MOD02 de nivel 1B anteriormente detallado, el número de imágenes empleadas es igual a las utilizadas en el primer producto.

Tabla 4 Productos MOD09GA para los años 2019 - 2020

Fecha (DD-MM- AAA)	Día Juliano	Hora	Escenas MOD09GA
1/01/2019	001	16:05	MOD09GA.A2019001.h10v08.061.2020283143124.hdf
2/01/2019	002	15:10	MOD09GA.A2019002.h10v08.061.2020283154103.hdf
3/01/2019	003	15:50	MOD09GA.A2019003.h10v08.061.2020283164831.hdf
4/01/2019	004	14:55	MOD09GA.A2019004.h10v08.061.2020283175519.hdf
30/01/2019	030	15:35	MOD09GA.A2019030.h10v08.061.2020285021735.hdf
15/02/2019	046	15:35	MOD09GA.A2019046.h10v08.061.2020287151157.hdf
22/02/2019	053	15:40	MOD09GA.A2019053.h10v08.061.2020288095216.hdf
1/05/2019	121	15:15	MOD09GA.A2019121.h10v08.061.2020293052016.hdf
24/05/2019	144	15:20	MOD09GA.A2019144.h10v08.061.2020294155322.hdf
26/06/2019	177	16:05	MOD09GA.A2019177.h10v08.061.2020298232306.hdf
6/07/2019	187	15:05	MOD09GA.A2019187.h10v08.061.2020302113656.hdf
5/08/2019	217	15:15	MOD09GA.A2019217.h10v08.061.2020304104623.hdf
9/12/2019	343	15:25	MOD09GA.A2019343.h10v08.061.2020319123233.hdf
6/01/2020	006	15:50	MOD09GA.A2020006.h10v08.061.2020324044826.hdf
10/01/2020	010	15:25	MOD09GA.A2020010.h10v08.061.2020324235625.hdf
11/01/2020	011	16:10	MOD09GA.A2020011.h10v08.061.2020325021114.hdf
12/01/2020	012	15:15	MOD09GA.A2020012.h10v08.061.2020325074507.hdf
18/02/2020	049	15:35	MOD09GA.A2020049.h10v08.061.2020329111527.hdf
20/05/2020	141	16:00	MOD09GA.A2020141.h10v08.061.2020334202036.hdf
4/06/2020	156	15:15	MOD09GA.A2020156.h10v08.061.2020335125745.hdf
13/07/2020	195	15:20	MOD09GA.A2020195.h10v08.061.2020338095032.hdf
31/07/2020	213	15:10	MOD09GA.A2020213.h10v08.061.2020340223241.hdf
2/08/2020	215	14:55	MOD09GA.A2020215.h10v08.061.2020341015026.hdf
10/10/2020	284	15:15	MOD09GA.A2020284.h10v08.061.2020348005841.hdf
2/11/2020	307	15:20	MOD09GA.A2020307.h10v08.061.2020353021535.hdf
7/11/2020	312	15:40	MOD09GA.A2020312.h10v08.061.2020353060846.hdf

El tercer producto auxiliar corresponde al MOD03 que contiene la información geolocalizada del campo visual instantáneo (IFOV) MODIS a una resolución de 1 kilómetro. En cada IFOV se destaca información complementaria como la altura del sensor sobre la superficie del geoide, ángulos azimutales y cenitales de observación e iluminación. Este producto se destaca por ser empleado en otros productos MODIS posteriores (MODIS Characterization Support Team (MCST), 2017b).

6.2.2. Datos de estaciones de monitoreo

Los datos puntuales de estaciones de monitoreo empleados corresponden al Sistema de Alerta Temprana de Medellín y el Valle de Aburrá (SIATA), que es un proyecto administrado y gestionado por el Departamento Administrativo de Gestión del Riesgo de Desastres (DAGR) y la Alcaldía de Medellín (SIATA, 2019). Se utilizaron datos de 9 estaciones de monitoreo que estiman PM2.5 y datos de 1 estación meteorológica que mide la presión atmosférica, estas toman muestras horarias recolectadas durante las 24 horas del día. Se emplearon los datos horarios de PM2.5 que correspondían a la fecha y hora de adquisición de las imágenes satelitales.

Tabla 5 Estaciones de monitoreo SIATA para la ciudad de Medellín.

Código	Estación	Medición	Coordenadas	
			Latitud	Longitud
12	Estación Tráfico Centro	PM 2.5	6° 15' 09" N	75° 34' 10" W
25	Centro occidente - Universidad Nacional, sede El Volador	PM 2.5	6° 15' 48" N	75° 34' 37" W
44	El Poblado - Tanques La Ye EPM	PM 2.5	6° 10' 57" N	75° 33' 02" W
79	Altavista - I.E. Pedro Octavio Amado	PM 2.5	6° 13' 19" N	75° 36' 38" W
80	Villahermosa - Planta de producción de agua potable EPM	PM 2.5	6° 15' 32" N	75° 32' 54" W
83	Belén - I.E Pedro Justo Berrio	PM 2.5	6° 14' 14" N	75° 36' 38" W
84	El Poblado - I.E INEM sede Santa Catalina	PM 2.5	6° 12' 00" N	75° 33' 39" W
85	San Cristóbal - Parque Biblioteca Fernando Botero	PM 2.5	6° 16' 40" N	75° 38' 11" W
86	Aranjuez - I.E. Ciro Mendía	PM 2.5	6° 17' 26" N	75° 33' 20" W
201	Torre SIATA	PA	6° 15' 33" N	75° 35' 19" W

Fuente: (SIATA, 2019)

PM2.5: Material particulado de diámetro menor a 2.5 micrómetros

PA: Presión atmosférica

6.2.2.1. Depuración de datos in situ

Para evaluar y depurar los datos de estaciones de la calidad del aire suministrados por el SIATA, se debe tener en cuenta los datos faltantes y los datos “bandera” o Flag, que indican la calidad o confianza de la medición reportada. Este valor es validado según el reporte del operario de la estación fija. Para las estaciones de la calidad del aire, este valor Flag oscila entre 1 y 4.9 en su columna *calidad_pm25* para el PM2.5 (SIATA, 2019). Para la metodología propuesta se emplearon solamente datos con un valor Flag igual a 1, que indica que la medición reportada es válida. La tabla 6 muestra la relación entre el valor Flag y la calidad del dato asociada.

Tabla 6 Significado de valores Flag para la red de estaciones de la calidad del aire

Valor Flag	Calidad del Dato
1	Dato válido
-1	Dato válido por el operado anterior
1.8 - 2.5	Dato dudoso
2.6 - 3.9	Dato malo
>= 4.0	Dato faltante
-9999	Equipo fuera de operación

Fuente: SIATA, 2019

6.2.3. Datos Red Robótica de Aerosoles (AERONET)

Estos datos pertenecen a 1 de los 6 fotómetros instalados en Colombia, que está ubicado en la ciudad de Medellín con una latitud: 06°15'39" N, Longitud: 75°34'40" W y con una elevación de 1471 msnm (AERONET, 2021a). Los datos AOD de AERONET no se estiman para la longitud de onda de 550 nm debido a su compleja caracterización, estos datos de AOD se calcularon empleando las longitudes de ondas adyacentes (440 y 675 nm) mediante el exponente de Ångström (Ångström, 1964; Wang et al., 2013):

$$\tau(\lambda) = \beta\lambda^{-\alpha} \quad (36)$$

Donde β es la turbidez atmosférica, λ es la longitud de onda y α corresponde al exponente de Ångström que viene dado en función de la de la tendencia espectral y su valor de AOD τ_a correspondiente, expresado mediante la Ecuación (36):

$$\alpha_{\lambda_2/\lambda_1} = - \frac{\ln(\tau_a(\lambda_2)/\tau_a(\lambda_1))}{\ln(\lambda_2/\lambda_1)} \quad (37)$$

6.3. MÉTODOS

6.3.1. Georreferenciación

Dado que los productos adquiridos se encuentran inicialmente en el formato HDF-EOS (Hierarchical Data Format - Earth Observing System) que almacenan los datos espectrales, metadatos y parámetros de georreferenciación para el estudio científico de la Tierra empleando tecnologías de detección remota en las misiones del sistema de observación de la Tierra (EOS), es necesario utilizar la herramienta de conversión HDF-EOS a GEOTIFF (HEG) creada por la NASA.

Mediante el software de conversión HDF-EOS a GeoTIFF (HEG) se ingresan los siguientes parámetros para la georreferenciación de cada una de las bandas de interés:

- Ruta de archivo
- Campo del metadato
- Banda del campo del metadato
- Coordenadas del área de procesamiento
- Tipo de remuestreo
- Proyección cartográfica
- Ruta del archivo de salida
- Tipo de archivo de salida

Ingresando los parámetros anteriormente descritos, se generó un archivo en formato GeoTIFF georreferenciado para la banda seleccionada (Taaheri & Murthy, 2019). Los parámetros de entrada en el campo del metadato fueron *EV_250_Aggr500_RefSB* para la banda 1 y *EV_500_RefSB* para las bandas 3 y 4 para el producto MOD02HKM. Para el producto MOD09GA se emplearon los campos *sur_refl_b01_1*, *sur_refl_b03_1* y *sur_refl_b04_1* para las bandas 1, 3 y 4 respectivamente. Estos parámetros de entrada corresponden a los atributos de los Metadatos contenidos en los productos MODIS. Ambos productos se georreferenciaron con los parámetros de tipo de muestreo Nearest Neighbor (Vecino más cercano) que es el único disponible, presentando una limitación en la herramienta, proyección Geográfica y tipo de archivo de salida GeoTIFF. Este procedimiento se llevó a cabo empleando la opción de procesamiento por lotes del programa, teniendo que realizar 3 procesos por año de estudio, puesto que las bandas de interés corresponden a la B1, B3 y B4.

6.3.2. Procesamiento de información espacial

El procesamiento de la información espacial se llevó a cabo empleado el lenguaje de programación R y su entorno de desarrollo integrado Rstudio. R es un proyecto de computación estadística que se caracteriza por implementar una variedad de métodos estadísticos avanzados y técnicas gráficas de gran calidad configuradas por defecto. R es de software libre y código abierto licenciado bajo la General Public License (GNU) (R-Project, 2021).

Georreferenciados y transformados a GeoTIFF los tres productos MODIS se obtuvieron 78 bandas para cada uno de los productos MOD02, MOD09 y MOD03, teniendo un total de 234 bandas a procesar. Primero se procesó el producto MOD03 del que se extrajo el valor promedio de píxeles dentro de la zona de estudio para las variables de altura, ángulo cenital solar, ángulo cenital de observación, ángulo azimutal solar y ángulo azimutal de observación, almacenados en los atributos *Height*, *SolarZenith*, *SensorZenith*, *SolarAzimuth* y *SensorAzimuth* del producto MOD03 respectivamente.

Estos parámetros se obtuvieron para cada día de estudio, adquiriendo un total de 26 parámetros para cada una de las 5 variables anteriormente descritas.

Para calcular la reflectancia a nivel de TOA se eliminaron los píxeles que se encuentran fuera del rango válido $[0, 32767]$ de los niveles digitales de la imagen, luego se extrajeron los valores de escalas dentro de los metadatos del producto MOD02HKM que corresponden a los atributos *reflectance_scales* y *reflectance_offsets*. Teniendo los valores de escala y los ángulos cenitales solares se aplicó la ecuación (4) para estimar la reflectancia en el tope de la atmósfera con un rango de valores de 0 a 1 en unidades adimensionales.

La reflectancia superficial se estimó aplicando el factor de escala almacenado en el atributo *scale_factor* del producto MOD09GA, que tiene un valor de 10^{-4} y se aplica en toda la imagen satelital para los píxeles del rango válido $[-100, 16000]$ mediante la ecuación (6), obteniendo la reflectancia superficial corregida por efectos atmosféricos.

Después de computadas las ecuaciones de la sección 5.2.3 para ejecutar el algoritmo simplificado de recuperación de aerosoles (SARA), se creó un archivo de parámetros que contiene las variables de entrada para el modelo, que son la altura en kilómetros, los ángulos azimutales y cenitales en dirección sol – tierra y sensor – tierra, la presión atmosférica en milibares, la profundidad óptica de aerosoles puntual a 550nm, el albedo de dispersión simple a 550 nm, el factor de asimetría del aerosol a 550 nm y por último las reflectancias TOA y superficial para las banda del verde, puesto que según Bilal et al. (2013), el algoritmo emplea la longitud de onda perteneciente al verde ($0.55 \mu m$). Teniendo las entradas del modelo se empleó la ecuación (15) para estimar la profundidad óptica de aerosoles (AOD) en toda la zona de estudio. La figura 7 detalla las variables empleadas y su respectiva fuente de información.

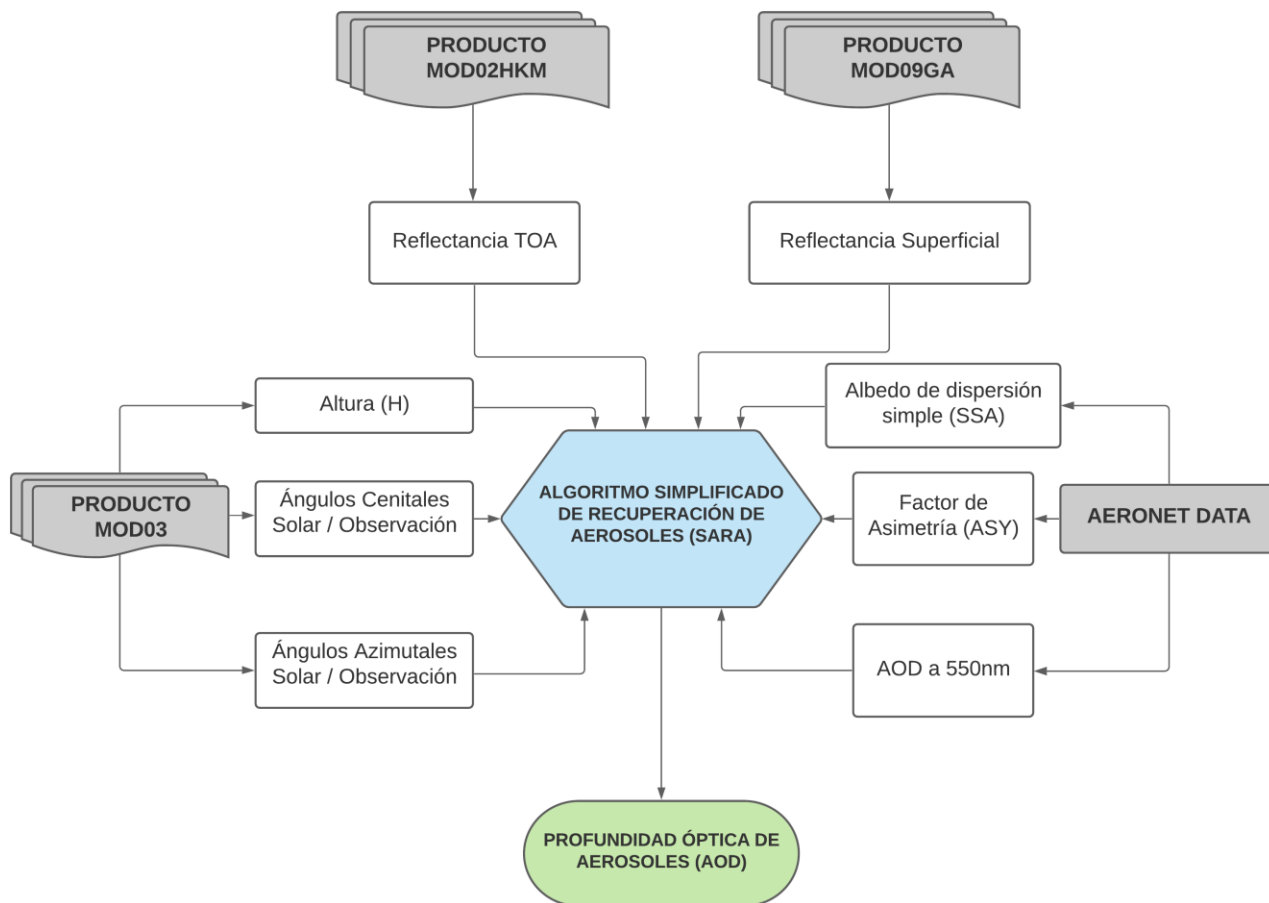


Figura 7 Esquema metodológico detallado para recuperar la profundidad óptica de aerosoles.

Los resultados de AOD para píxeles con presencia de nubosidad o sombras fueron descartados, puesto que estos arrojan valores atípicos que afectan los resultados del algoritmo. Para corroborar el desempeño y relación directa proporcional entre el AOD obtenido a partir de datos MODIS y mediciones de PM_{2.5} in situ (Wang et al., 2013), se estimó su correlación mediante una regresión lineal, tomando el mejor píxel (Sin nubosidad o Bruma) de una cuadrilla de píxeles 3x3 centrados en la ubicación de la estación in situ PM_{2.5}, dado a la sensibilidad del modelo a este tipo de alteraciones (Bilal et al., 2013). Los resultados del desempeño del modelo serán analizados y discutidos más adelante.

6.3.3. Imputación de datos faltantes

Para no perder información importante se imputaron los datos faltantes de PM2.5 y AOD empleando imputación múltiple Fully Conditionally Specification (FCS) implementado por el algoritmo MICE (Multiple Imputation by Chained Equations), que se caracteriza por emplear un modelo propio de imputación para cada variable. Para este caso, se utilizaron modelos de imputación para datos continuos (Predictive mean matching - pmm) (Castro, 2014). Mediante este método, se obtiene la totalidad de los datos faltantes, para tener el 100% de los datos observados.

7. RESULTADOS Y DISCUSIÓN

7.1. IMPUTACIÓN Y ANÁLISIS DESCRIPTIVO

Con los registros de datos de las estaciones de monitoreo de PM_{2.5} depurados como se explicó en la sección 6.2.2.1 y los datos del espesor óptico de aerosoles depurados para nubosidad y pixeles distorsionados para los años 2019 y 2020, se obtuvo 12% de datos faltantes de los 234 registrados para PM_{2.5} y AOD mediante las técnicas estadísticas detalladas en la sección 6.3.3. Estos corresponden a datos horarios de las 9 estaciones de monitoreo de la calidad del aire evaluadas y datos horarios de AOD para la fecha de adquisición, presentados en el Anexo A. La figura 8 muestra la cantidad y distribución de datos faltantes en el conjunto de datos.

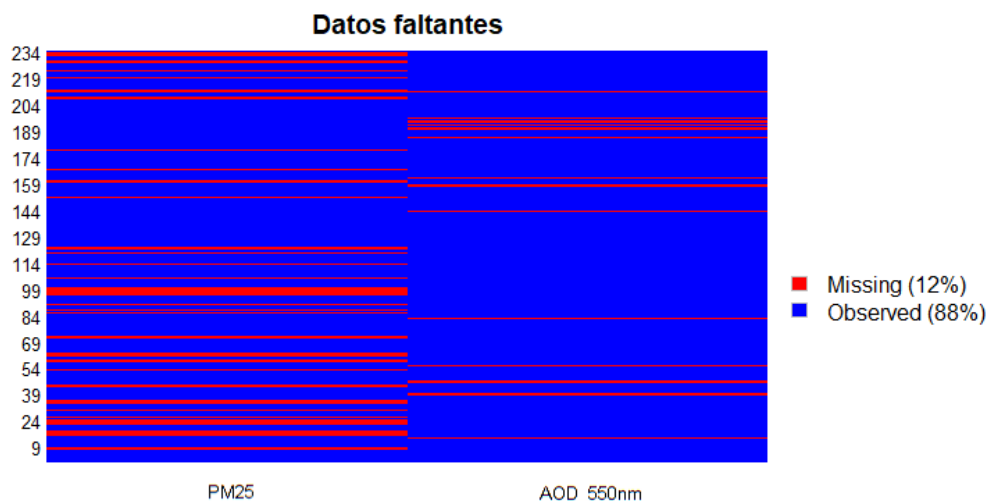


Figura 8 Frecuencia de los registros de datos faltantes PM_{2.5} y AOD

Con los datos de observación completos después del proceso de imputación, se analiza mediante las figuras 9 y 10 la dispersión, distribución y patrones de comportamiento de los datos puntuales de concentraciones de PM 2.5 para las 26 fechas de estudio.

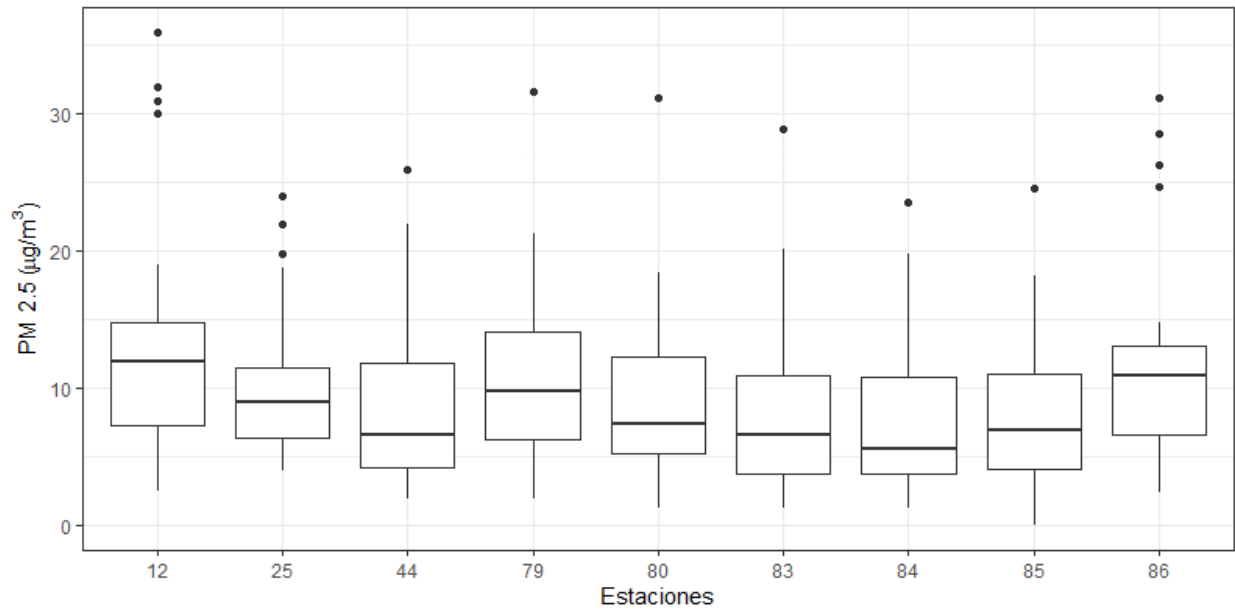


Figura 9 Diagrama de cajas de las concentraciones de PM2.5 in situ horario por estación, para las 26 fechas de estudio comprendidas entre los años 2019 y 2020.

En la figura 9 se observan los diagramas de cajas correspondientes al comportamiento de los datos de la concentración de PM2.5 de las estaciones SIATA para las fechas y horas en que se obtuvieron los datos satelitales. Según el comportamiento de los datos, se evidencia la existencia de datos atípicos registrados en las estaciones 12, 25 y 86 mayoritariamente, con valores máximos entre $25 \mu g/m^3$ y $36 \mu g/m^3$, teniendo en cuenta que se optaron por mantener todos los datos que se encontraban en un rango menor a $40 \mu g/m^3$ registrados y validados por cada estación como un dato válido. La mediana presenta un comportamiento de fluctuaciones alrededor de $10 \mu g/m^3$ en casi todas las estaciones evaluadas.

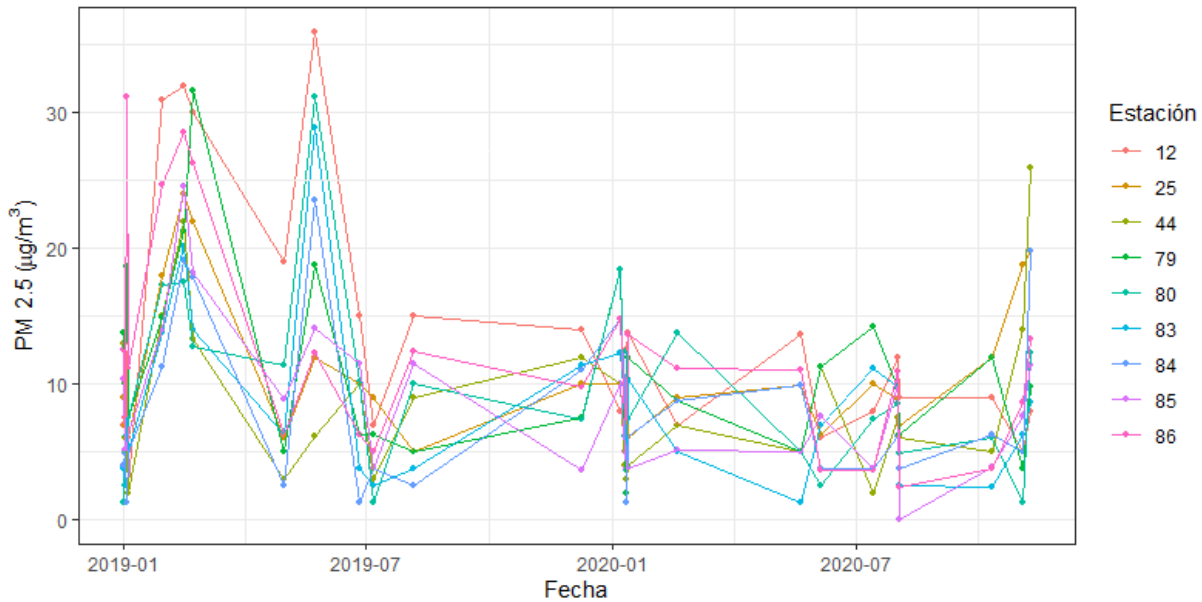


Figura 10 Comportamiento de las concentraciones de PM_{2.5} in situ horario para las 26 fechas de estudio comprendidas entre los años 2019 y 2020.

Según la figura 10 se puede observar que para las fechas del año 2019 la mayoría de los datos se concentraron en registros mayores a $10 \mu\text{g}/\text{m}^3$, destacando algunos meses donde se registraron valores máximos entre $25 \mu\text{g}/\text{m}^3$ y $36 \mu\text{g}/\text{m}^3$ aproximadamente. Para las fechas del año 2020 se evidencia una reducción considerable con registros promedio menores a $10 \mu\text{g}/\text{m}^3$, teniendo en cuenta que se evaluaron diferentes fechas para el periodo estudiado; para el final de las fechas durante el año 2020, se presenta un incremento en las concentraciones de PM_{2.5} alcanzando valores máximos de $25 \mu\text{g}/\text{m}^3$ aproximadamente. Teniendo en cuenta que no se tiene un conjunto completo de datos de PM_{2.5}, se pueden presentar variaciones temporales debido a factores climatológicos y posiblemente una reducción en las concentraciones de PM_{2.5} asociado a la pandemia del SARS-CoV-2 (COVID-19); esta puede generar disminución en las concentraciones de PM_{2.5} por el cierre y cese de actividades humanas no esenciales (Berman & Ebisu, 2020).

7.2. ESPESOR ÓPTICO DE AEROSOLES (AOD)

Las concentraciones de AOD se recuperaron como se detalla en la sección 6.3.2, los resultados de las regresiones lineales para evaluar el desempeño del algoritmo SARA para estimar AOD, se llevaron a cabo empleando datos depurados de 9 estaciones de monitoreo de la calidad del aire SIATA y estimaciones de AOD, los datos y gráficos de dispersión entre PM_{2.5} y AOD se encuentran en los anexos A y B respectivamente. La tabla 7 muestra los resultados que se obtuvieron de las regresiones para cada fecha de datos MODIS disponibles, estos resultados arrojaron valores de coeficiente de determinación $R^2 = 0.886$ y $R^2 = 0.855$ para el mejor de los casos correspondiente a 3 de enero del año 2019 y 31 de julio del año 2020 respectivamente. Los resultados de los datos agrupados por año mostraron una correlación mayor a $R^2 = 0.68$ para las fechas del año 2019 y $R^2 = 0.47$ para las fechas del año 2020. Para el total de las fechas comprendidas entre los años 2019 y 2020 se obtuvo en promedio un nivel de similitud del 57% entre los datos del AOD y PM_{2.5}. Estos resultados confirman la relación directamente proporcional entre el espesor óptico de aerosoles y el material particulado expuesta por diferentes autores consultados (Céspedes et al., 2017; Othman et al., 2010; Wang et al., 2013).

Tabla 7 Regresiones lineales entre el AOD y el PM2.5 para las 26 fechas de estudio durante los años 2019 – 2020.

Fecha (DD-MM-AAA)	Día Juliano	Hora	Ecuación Lineal	R ²
1/01/2019	001	16:05	$y = 199.68x - 2.8692$	0,8536
2/01/2019	002	15:10	$y = 339.43x - 5.4531$	0,5632
3/01/2019	003	15:50	$y = 74.046x - 4.846$	0,8863
4/01/2019	004	14:55	$y = 425.85x - 6.171$	0,6823
30/01/2019	030	15:35	$y = 63.734x - 7.7974$	0,7272
15/02/2019	046	15:35	$y = 28.08x + 12.585$	0,4636
22/02/2019	053	15:40	$y = 24.552x + 5.6578$	0,848
1/05/2019	121	15:15	$y = 196.74x - 12.416$	0,5501
24/05/2019	144	15:20	$y = 72.586x + 5.2213$	0,5478
26/06/2019	177	16:05	$y = 209.97x - 2.4701$	0,6945
6/07/2019	187	15:05	$y = 197.3x - 2.795$	0,5447
5/08/2019	217	15:15	$y = 150.81x - 5.5154$	0,6322
9/12/2019	343	15:25	$y = 39.545x + 5.575$	0,5611
6/01/2020	006	15:50	$y = 44.228x - 6.3373$	0,6585
10/01/2020	010	15:25	$y = 31.705x - 1.3357$	0,7169
11/01/2020	011	16:10	$y = 332.66x - 24.602$	0,7576
12/01/2020	012	15:15	$y = 78.142x + 0.062$	0,6026
18/02/2020	049	15:35	$y = 13.519x + 2.4842$	0,6834
20/05/2020	141	16:00	$y = 43.541x - 1.6232$	0,5492
4/06/2020	156	15:15	$y = 267.44x - 19.723$	0,5893
13/07/2020	195	15:20	$y = 61.294x - 1.1556$	0,4987
31/07/2020	213	15:10	$y = 65.555x + 6.3215$	0,8547
2/08/2020	215	14:55	$y = 502.1x - 13.326$	0,6623
10/10/2020	284	15:15	$y = 97.029x - 1.5729$	0,845
2/11/2020	307	15:20	$y = 48.132x + 0.9095$	0,5406
7/11/2020	312	15:40	$y = 26.483x - 6.1757$	0,7363

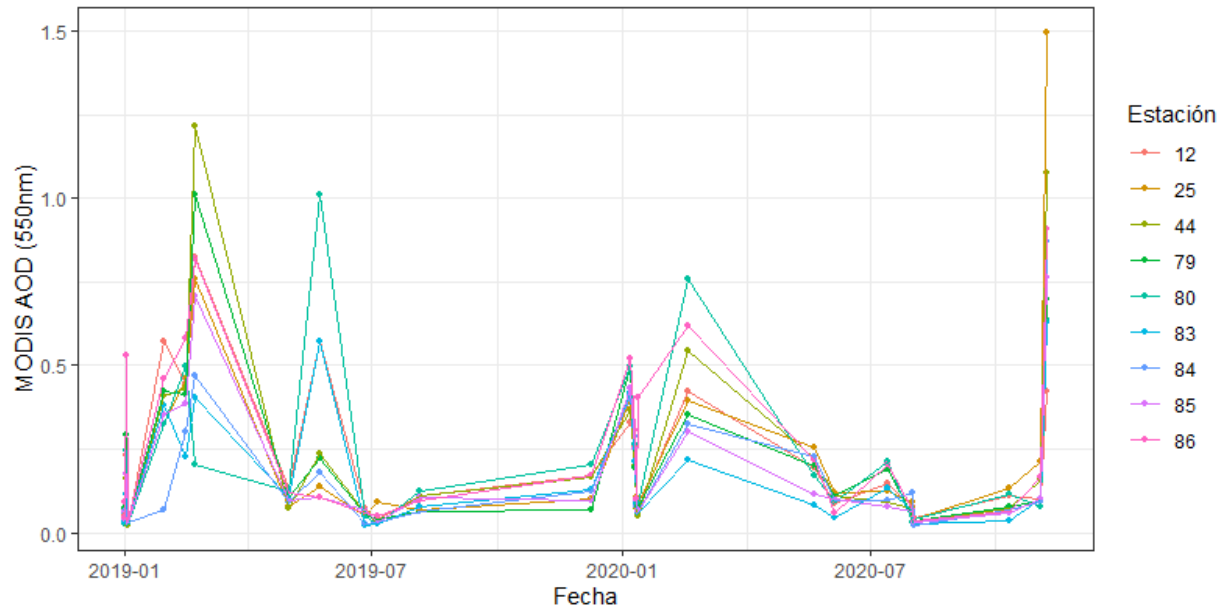


Figura 11 Comportamiento del Espesor óptico de aerosoles (AOD) para las 26 fechas de estudio comprendidas entre los años 2019 y 2020

La figura 11 muestra los resultados obtenidos de AOD en función de tiempo, a partir de los resultados del modelo SARA, para las fechas y horas de disponibilidad de datos MODIS en el periodo 2019 – 2020. Teniendo en cuenta los resultados de las regresiones lineales entre AOD y PM_{2.5}, se puede evidenciar un comportamiento similar a los resultados de las concentraciones de PM_{2.5} representados en la figura 10; para las fechas que corresponden al año 2019 se presenta la mayor cantidad de valores de AOD cercanos y mayores a 1, que indica la presencia de mayor carga de partículas de contaminación atmosférica dentro del área de estudio. Para el año 2020, se evidencia una menor carga de partículas en la zona de estudio, para las fechas estudiadas.

En Beijing, Wang et al., (2013), empleando un algoritmo mejorado similar al SARA y el producto MODIS de AOD (MOD04), obtuvieron coeficientes de correlación de las regresiones lineales entre el AOD y el PM_{2.5} correspondientes a $R^2 = 0.818$ y $R^2 = 0.750$ para los mejores casos, resultados similares a los obtenidos de $R^2 = 0.886$ y $R^2 = 0.855$ en la investigación presente. La metodología de Wang et al., (2013) emplea un mayor número de estaciones en tierra y un producto adicional de AOD medido en tierra, lo que

indica su importancia y validez de implementación en metodologías para estimar AOD satelital, según los resultados positivos obtenidos.

Bilal et al., (2013), en Hong Kong, plantean y evalúan el desempeño del algoritmo SARA, a partir del que obtienen un coeficiente de correlación de $R = 0.964$ entre los resultados del SARA y el PM_{2.5} medido en tierra. En comparación con la investigación presente, Bilal et al., (2013) emplearon un mayor número de estaciones en tierra y el doble de estaciones AERONET con una mayor disponibilidad y calidad de los datos. Estos factores tienen una gran influencia en el algoritmo SARA, debido a su sensibilidad por no emplear tablas de búsqueda (LUT) y aplicar los modelos de transferencia radiativa (RTM) directamente en los datos MODIS.

Céspedes et al. (2017), presentaron los resultados de una metodología para medir las concentraciones de aerosoles atmosféricos empleando técnicas de teledetección. Estos estiman el material particulado PM₁₀ usando un algoritmo multiespectral, donde emplean todas las bandas del espectro visible del sensor MODIS. Obtuvieron resultados destacables para varios días del año, con coeficientes de determinación cercanos a 1, comparado con el $R^2 = 0.818$ para el mejor de los casos en el presente estudio. Discutiendo los resultados obtenidos, las metodologías difieren en el uso de datos auxiliares de estaciones AERONET, caso contrario de la metodología propuesta que si los emplea. Céspedes et al., (2017) usan todas las bandas del espectro visible, diferente del SARA, que solo emplea la banda visible correspondiente al verde ($0.55 \mu m$). A pesar de que las dos zonas de estudio son topográficamente diferentes, los resultados obtenidos en ambos casos son positivos, lo que puede llegar a indicar que las características geográficas de la región tienen poca influencia en los resultados obtenidos.

La estimación del espesor óptico de aerosoles se llevó a cabo empleando datos que no presentaban una variación anual homogénea en la distribución de las fechas, pudiendo generar influencia en los resultados y evidenciando una de las principales limitaciones de la metodología propuesta. La estimación de datos faltantes empleando técnicas estadísticas permitió tener una base de datos completa para generar el modelo de

regresión funcional; estas se estimaron para mediciones de PM2.5 in situ donde no se encontraban disponibles.

7.3. ANÁLISIS DE DATOS FUNCIONALES (ADF)

7.3.1. Construcción y análisis de datos funcionales

De acuerdo con los resultados obtenidos el número de bases óptimo que minimiza el criterio de GCV es $k = 13$, en base a lo expuesto en la sección 5.2.4.3; este valor se puede asociar a la variabilidad de los datos y su grado de dificultad en que el número de bases pueda representar detalladamente el comportamiento de estos. En cuanto al parámetro de suavización, se obtuvo que el valor que minimiza el criterio de GCV es $\lambda = 0$, por lo tanto, las curvas ajustadas se obtuvieron empleando el método de ajuste usando la función *smooth.basis* que usa el criterio de mínimos cuadrados; esta suavización permite evaluar la variable en cualquier momento, caracterizar sus tasas de cambio y reducir el ruido en los datos.

La base de funciones elegida para la creación de los datos funcionales fueron las bases de Fourier. En un principio se planeaba emplear bases B-spline, puesto que la teoría del comportamiento de la contaminación atmosférica es no periódica; pero los resultados obtenidos no se ajustaron correctamente, por lo que se decidió usar bases de Fourier obteniendo mejores ajustes en las curvas de los resultados. A continuación, se presentan los resultados de las curvas ajustadas o datos funcionales de las variables estudiadas.

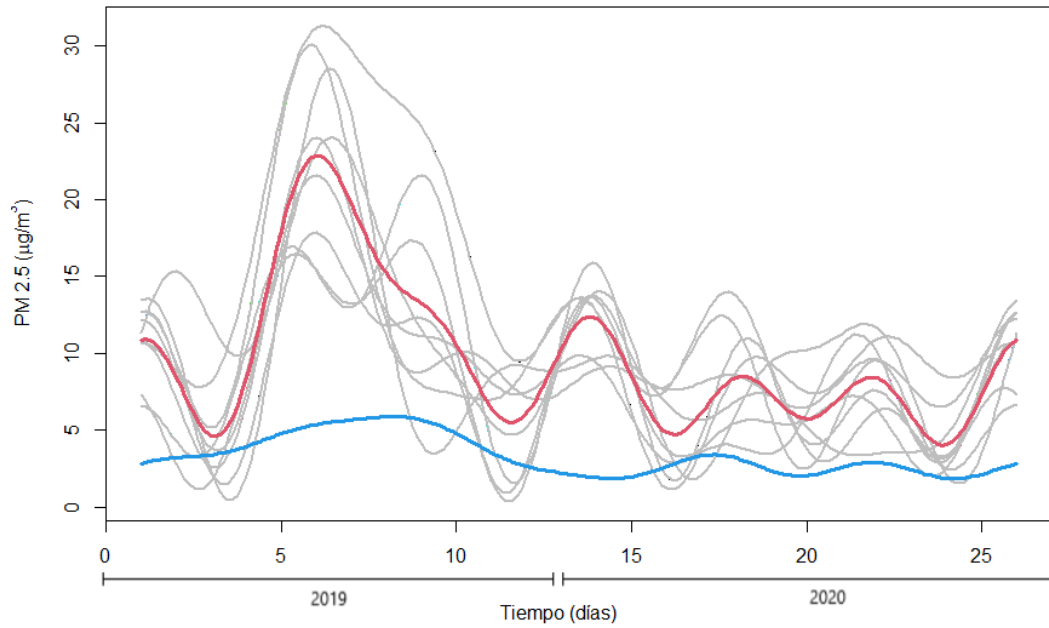


Figura 12 Datos funcionales de las concentraciones de PM_{2.5} y su media funcional (Línea roja) y desviación estándar funcional (Línea azul).

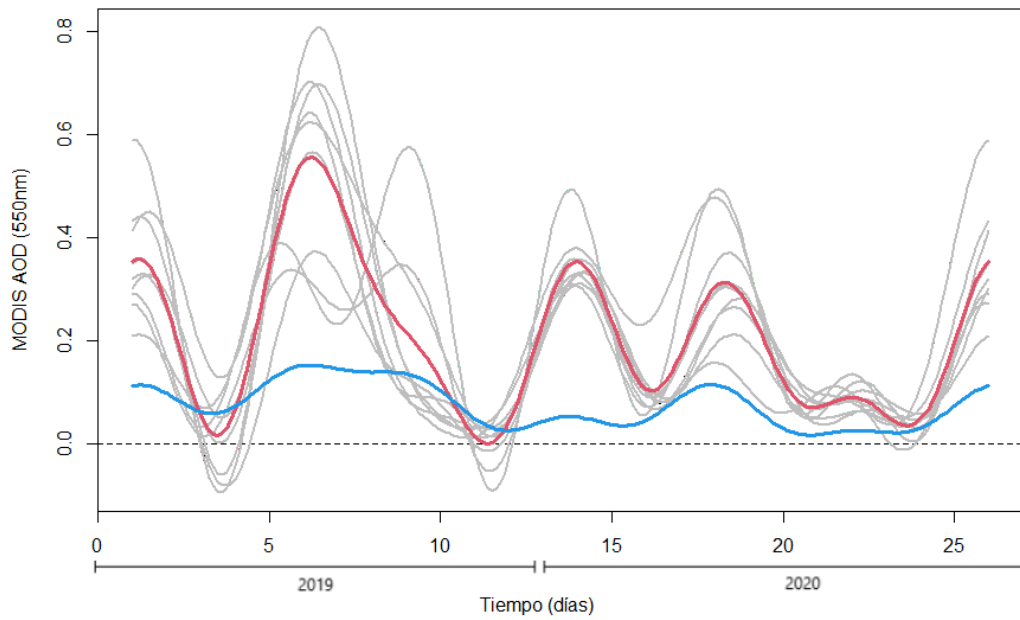


Figura 13 Datos funcionales del espesor óptico de aerosoles (AOD) y su media funcional (Línea roja) y desviación estándar funcional (Línea azul).

Las figuras 12 y 13 representan el comportamiento de los datos funcionales de PM2.5 y AOD respectivamente. En ambos casos se puede evidenciar que para la mitad de los días evaluados el comportamiento de las variables toma valores altos según su respectiva escala, se tiene este comportamiento dado que los primeros 13 días corresponden a registros del año 2019, época en la que los niveles de contaminación atmosférica fueron mayores con respecto a registros del año 2020 como se evidencia en las figuras 12 y 13. Para los datos de PM2.5 correspondientes al año 2019, se evidencia picos máximos en las funciones que superan los $25 \mu g/m^3$; por parte del AOD, su comportamiento presenta picos máximos en los primeros días evaluados, reflejando una atmósfera densa o con presencia considerable de contaminación atmosférica. En ambas gráficas, se evidencia una reducción considerable de picos en las funciones de las variables estudiadas, puesto que corresponden a las fechas del año 2020, donde anteriormente se detalló este comportamiento decreciente en las concentraciones de PM2.5.

El comportamiento promedio de todas las curvas para las variables estudiadas, se evidencia el comportamiento de valores máximos para los primeros 13 días estudiados, donde se alcanza valores promedio de $22 \mu g/m^3$, así como un pico durante el último mes del año situado aproximadamente a los 14 días. Después del día 14, el comportamiento de la curva decrece teniendo variaciones poco considerables, tomando valores promedio dentro de la normativa de $8 \mu g/m^3$ aproximadamente. Finalmente se evidencia un incremento hacia los días 25 y 26 correspondientes a las fechas de los meses de noviembre y diciembre del año 2020 donde se alcanzan valores de hasta $12 \mu g/m^3$. La variabilidad de la desviación estándar refleja el cambio en el comportamiento de los datos durante los meses de diciembre del 2019 y enero del 2020, donde se evidencia un fuerte decrecimiento durante los días 10 y 11 del estudio, momento exacto donde se implementaron las restricciones y medidas de cuarentenas obligatorias dentro del territorio nacional, atribuyendo probablemente como un factor en la cantidad de concentraciones de PM2.5 registradas.

7.3.2. Modelo de regresión funcional PM2.5

A continuación, en la figura 14 se expresa el comportamiento de la pendiente y el intercepto funcional del modelo de regresión funcional concurrente propuesto en la sección 5.2.6.

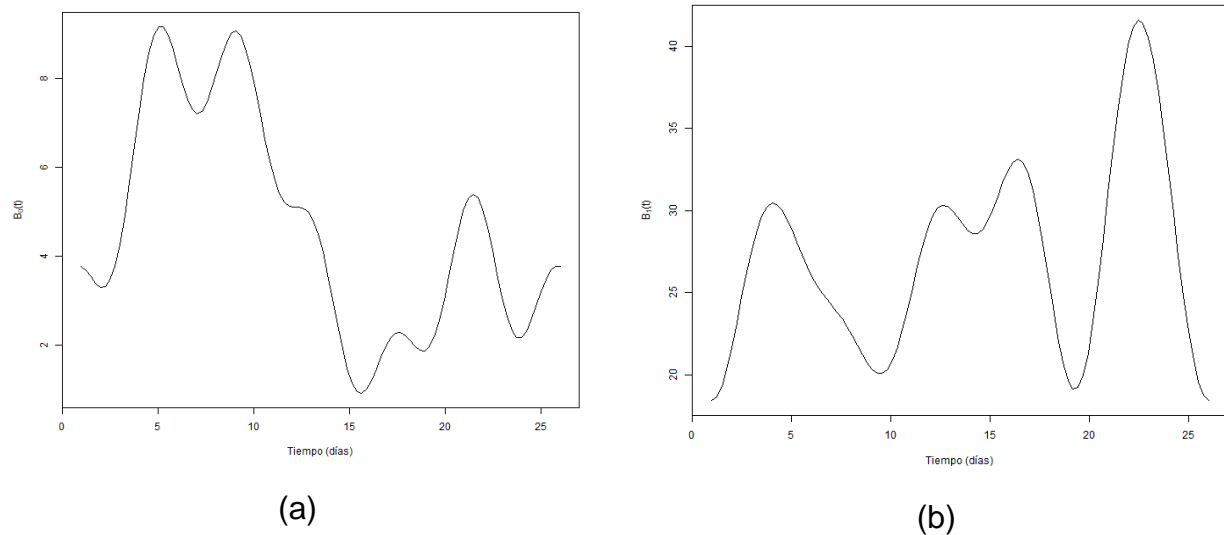


Figura 14 Comportamiento del intercepto funcional (a) y pendiente funcional (b) del modelo de regresión funcional propuesto.

A continuación, en la figura 15 se presentan los resultados obtenidos por el modelo de regresión funcional propuesto a partir de datos de PM2.5 para cada estación y AOD estimado a partir de imágenes satelitales. La línea negra representa los datos de PM2.5 medidos en cada estación y la línea roja representa los valores de PM2.5 ajustados por el modelo de regresión funcional.

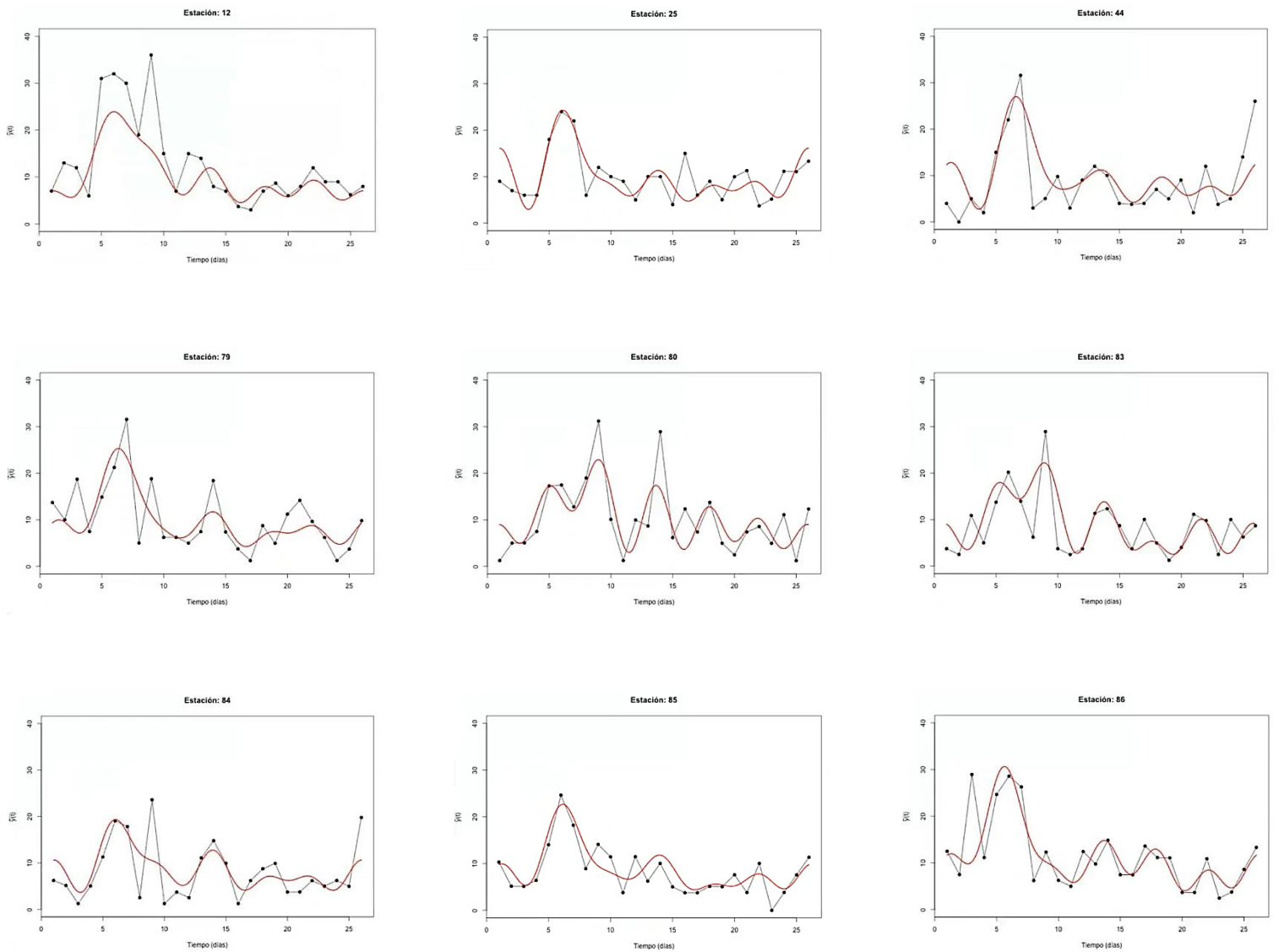
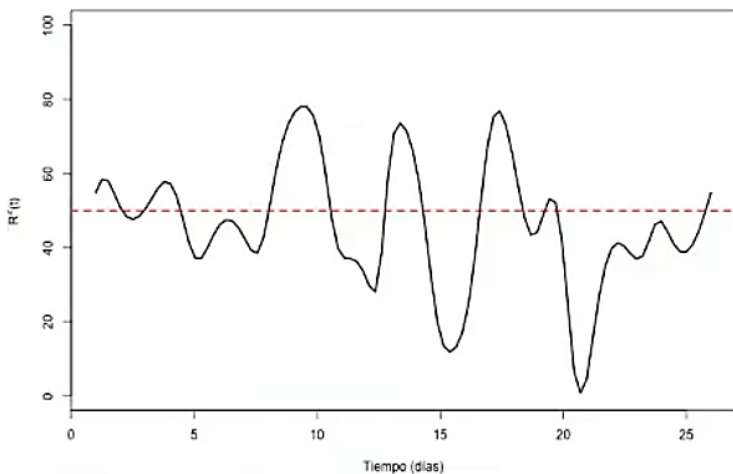


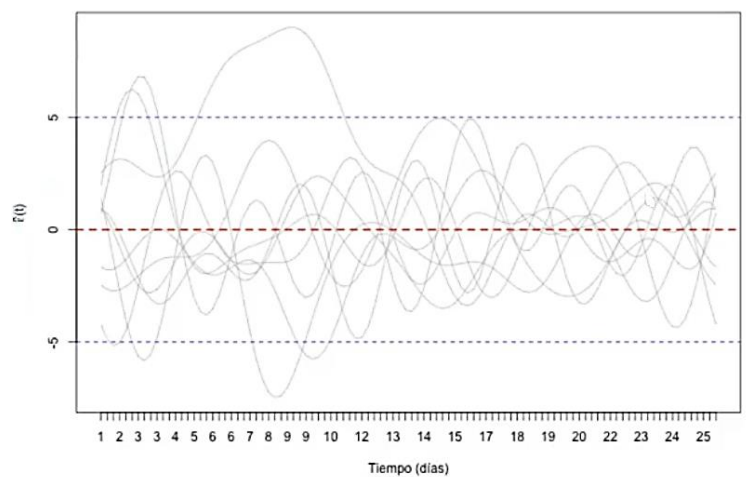
Figura 15 Datos funcionales de PM2.5 ajustados por el modelo de regresión funcional (Línea roja) a partir de datos funcionales y datos de PM2.5 medidos en tierra (Línea negra) para las 9 estaciones dentro del área de estudio. Eje x: Corresponde a los 26 días de evaluación durante los años 2019 y 2020. Eje y: Equivale a concentraciones de PM2.5.

Según la figura 16(a), el coeficiente de determinación funcional obtenido a partir del modelo general de regresión funcional presentó valores en promedio de $R^2 = 0.55$, destacando picos máximos de $R^2 = 0.80$ para los días de estudio 10, 14 y 17 equivalentes a los meses junio del 2019 y enero del 2020; durante estos meses se obtuvieron las mejores respuestas de estimación por parte del modelo. Durante el mes de junio del 2020, día 21 del estudio, se evidencia un pico decreciente que alcanza valores mínimos de R^2 afectando el poder predictivo del modelo; este comportamiento se asocia a valores atípicos registrados en los datos.

La figura 16(b) muestra el comportamiento de los valores residuales del modelo, entre los datos de PM2.5 ajustados y los datos de PM2.5 medidos en tierra (Figura 17). Los residuales muestran un umbral promedio de error de $5 \mu g/m^3$ para valores ajustados por el modelo; en algunos días de estudio el umbral supera el error promedio, atribuyendo esta característica a los datos atípicos registrados por las estaciones. El comportamiento de los residuales se estabiliza para el año 2020, donde presenta umbrales promedio de $3 \mu g/m^3$.



(a)



(b)

Figura 16 (a): Representa el coeficiente de determinación funcional del modelo ($R^2(t)$).
 (b): Equivale a los errores residuales funcionales del modelo.

El modelo de regresión funcional se empleó para comparar el comportamiento de las curvas ajustadas de PM2.5 dependientes, obtenidas a partir de curvas de AOD independientes, con los datos puntuales de PM2.5 para las 9 estaciones de monitoreo. Comparando el comportamiento de los errores residuales funcionales del modelo, con los obtenidos por Ospína et al. (2019), para la evaluación de una variable de estudio diferente, empleando el mismo modelo de regresión funcional, donde los errores residuales funcionales de los autores, son estables cercanos a 0, salvo algunos casos donde al final presenta pequeñas fluctuaciones; estos emplearon información diaria y pentadal para 18 años de estudio, destacando en comparación con el estudio presenten una base de datos más completa. El modelo no se empleó para espacializar la variable de estudio en toda el área de estudio, dada la complejidad de estimar datos funcionales para cada píxel de la imagen satelital durante todas las fechas de estudio; dejando planteado el modelo para futuros estudios y evidenciando una limitación de este.

7.4. MAPAS DE ESTIMACIÓN DE CONCENTRACIONES PM2.5

Con los modelos generados de las regresiones lineales entre el PM2.5 y el espesor óptico de aerosoles (AOD) contenidas en la tabla 7, lo cuales se obtuvieron a partir de estimar la relación entre las variables PM2.5 y AOD como se detalla en sección 6.3.2, se logró espacializar la variable dependiente, para este caso el PM2.5, a partir de la variable independiente, el AOD; confirmando la relación directamente proporcional entre las variables como se explica en la sección 7.2. Se obtuvieron 13 salidas cartográficas correspondientes a los modelos que presentaron un coeficiente de determinación mayor a $R^2 = 0.65$. Estos mapas son de suma importancia, puesto que, en estos se aprecia el comportamiento espacial de las concentraciones de PM2.5 para la ciudad de Medellín durante los 26 días de estudio durante los años 2019 y 2020. Debido al flujo de trabajo automatizado empleando el lenguaje de programación Python, los pixeles válidos y depurados para nubosidad corresponden únicamente a los que se encuentran dentro del área de estudio delimitada en la cartografía generada, cualquier registro de concentraciones de PM2.5 fuera del área de estudio, debe ser descartado.

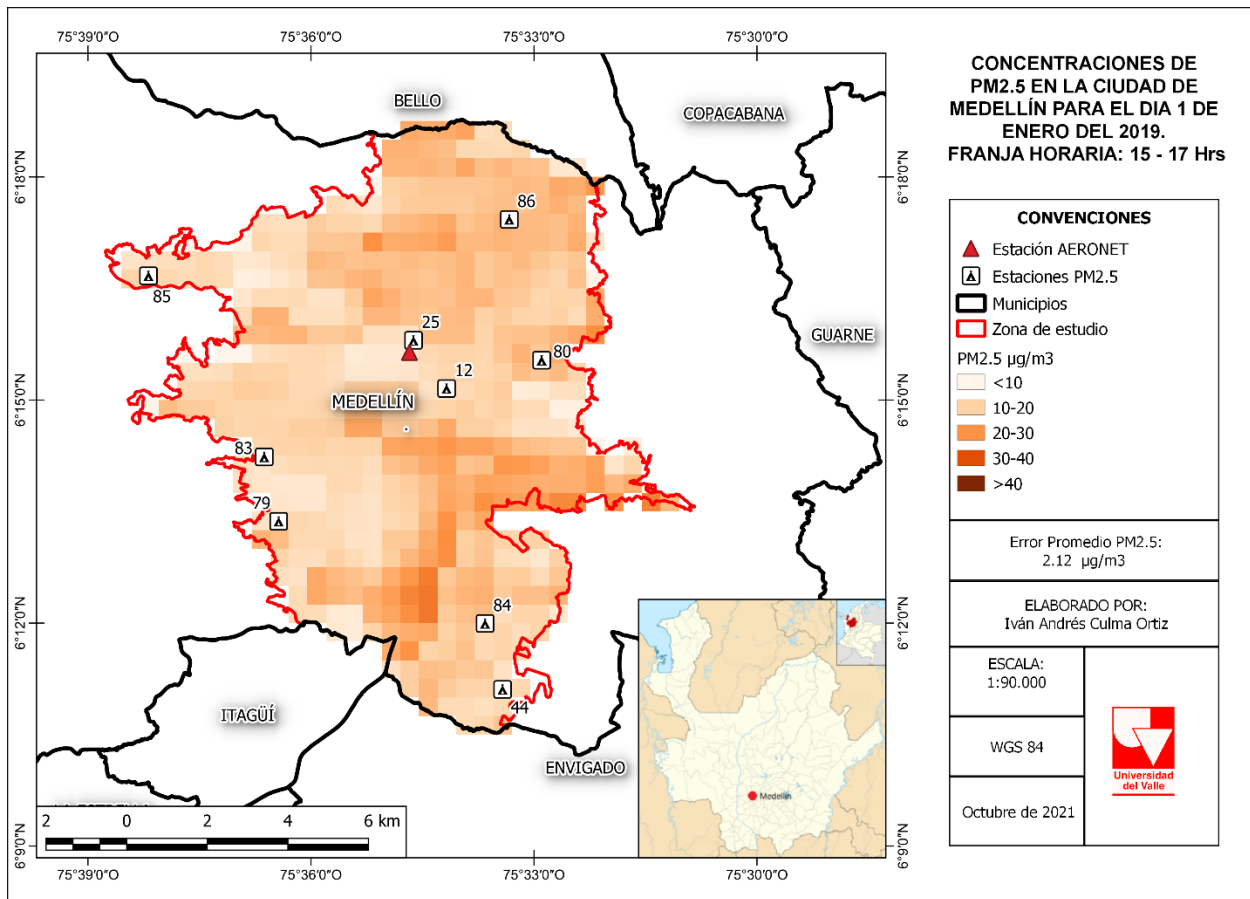


Figura 17 Distribución espacial de concentraciones de PM2.5 en la ciudad de Medellín para el día primero de enero del 2019.

En los mapas generados se obtuvieron resultados estimados de PM2.5 que alcanzaron concentraciones máximas superiores a $40 \mu\text{g}/\text{m}^3$. Analizando la distribución espacial de estos valores máximos, se observó que dentro del área de estudio las zonas de corredores viales primarios como la Autopista Sur y Avenida Regional, coinciden en su localización con los mayores valores estimados de PM2.5, por lo que se constituyen en una importante fuente de emisiones de concentraciones de PM2.5; situación que igualmente se evidencia en el corredor vial principal que conecta a los municipios de Itagüí y Envigado, donde existe un importante flujo vehicular; según el comportamiento espacial de la variable, en las zonas como El Poblado y Terminal Norte, se evidenciaron valores de concentraciones PM2.5 mayores, con respecto a las otras zonas residenciales

de la ciudad. Para las fechas evaluadas del año 2020, se registraron valores promedio de $15 \mu\text{g}/\text{m}^3$ como se presenta en la figura 18.

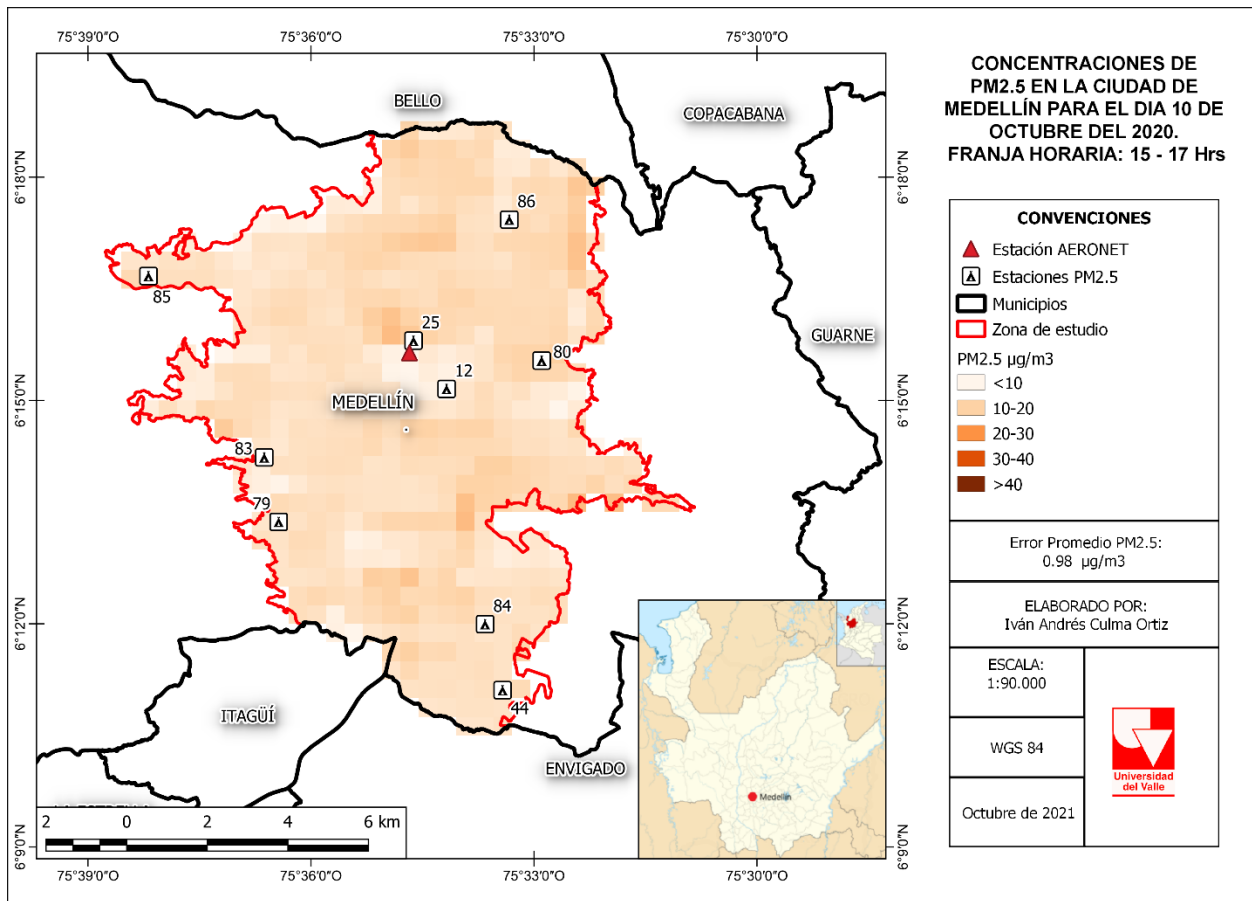


Figura 18 Distribución espacial de concentraciones de PM2.5 en la ciudad de Medellín para el día 10 de octubre del 2020.

Comparando la distribución espacial de las concentraciones de PM2.5 estimadas a partir del AOD y las registradas en tierra, se tiene una ventaja espacial superior, puesto que las estimaciones de PM2.5 en zonas alejadas de la red de estaciones en tierra, se le asocia un error con respecto a la distancia de esta. En la distribución de PM2.5 a partir de AOD, se puede apreciar zonas alejadas de la red de estaciones con mediciones más homogéneas sin tener mucha dependencia espacial del conglomerado de estaciones. Los mapas de estimación de concentraciones de PM2.5 para todas las fechas de estudio se encuentra en la sección de Anexos, C.

8. CONCLUSIONES Y RECOMENDACIONES

8.1. CONCLUSIONES

En esta investigación se realizó la estimación de concentraciones de PM2.5 empleando mediciones en tierra y datos de sensores remotos. Se utilizó una metodología que permitió integrar las diferentes fuentes de información haciendo uso de un algoritmo simplificado de recuperación de aerosoles para extraer la profundidad óptica de aerosoles. El análisis de datos funcionales permitió caracterizar el comportamiento de la variable estudiada a lo largo del periodo evaluado mediante un modelo de regresión funcional que alcanzó coeficientes de determinación $R^2 = 0.80$. Los resultados de AOD demostraron tener un comportamiento similar a los registros de PM2.5 según los coeficientes de correlación lineal promedio de $R^2 = 0.65$. Estos resultados confirman la efectividad e importancia de emplear y desarrollar metodologías que permitan estimar contaminantes atmosféricos empleando sensores remotos; ya sea haciendo uso de la información directamente, o en el mejor de los casos, combinar la información con mediciones en tierra.

Según los resultados obtenidos por el algoritmo SARA, se concluye la importancia de emplear datos AERONET con un nivel de procesamiento superior y de varias estaciones en tierra, dado que estas permiten modelar las características de las partículas dentro de la zona de estudio mejorando el desempeño del algoritmo. La sensibilidad del algoritmo a la presencia de pixeles con nubosidad o bruma fue evidente, obteniendo valores atípicos y concluyendo que el uso del algoritmo es solo para imágenes satelitales con presencia menor al 10% de nubosidad. La imputación de datos basada en el algoritmo MICE logró recuperar el 12% de los datos entre mediciones de concentraciones de PM2.5 y el espesor óptico de aerosoles AOD, ocasionadas por ausencia de registros en el caso de las estaciones, y nubosidad o alteraciones para los datos de satélite. Lo que permitió tener una base de datos completa para la construcción óptima de los datos funcionales, requisito indispensable para alcanzar un mejor desempeño de esta técnica.

El modelo de regresión funcional mostro tener un buen desempeño pese a la limitada cantidad de estaciones y días de estudio según las imágenes satelitales disponibles.

Teniendo en cuenta el coeficiente de determinación y los errores residuales funcionales estables alrededor de cero, se concluye que el uso del análisis de datos funcionales para modelar variables continuas que se registran de forma discreta en el tiempo evidenció tener resultados favorables que permiten describir y estudiar la variable en cualquier instante de tiempo para el periodo estudiado. El modelo de regresión funcional planteado puede ser empleado en futuras investigaciones, mediante un algoritmo que permita espacializar la información a partir de las curvas generadas de la variable de estudio.

La cartografía generada es un suministro de vital importancia, dado que mediante esta se logró representar e interpretar el comportamiento espacial de las concentraciones de PM2.5, evidenciando fuentes o zonas potencialmente generadoras de PM2.5, que con mediciones en tierra difícilmente se puede reflejar; logrando así, tener una herramienta para la toma de decisiones o planes que busquen mitigar este impacto ambiental negativo, actuando directamente en las zonas o fuentes que presenten registros mayores al límite máximo permisible.

8.2. RECOMENDACIONES

Durante la investigación se presentaron diferentes problemas, obtener imágenes satelitales libres de nubosidad lo que limitó la generación de una base de datos robusta que permitiera contar con los datos durante todo el período de estudio. Esto provocó que hubiese pequeños saltos de comportamiento donde no se logró obtener la información. Para investigaciones futuras, se recomienda obtener el mayor número posible de imágenes satelitales libres de nubosidad, así mismo como la implementación de algoritmos de predicción para imágenes que presenten nubosidad baja. Una base de datos más robusta también permitirá realizar un análisis de datos funcionales más ajustado a la realidad, construyendo curvas que se comporten con más afinidad a la variable continua estudiada.

La implementación de sensores remotos y variables meteorológicas han tenido resultados positivos en otras investigaciones. La resolución espacial empleada demostró

tener buenos resultados, aunque perdiendo información importante dada el área censada por cada píxel de las imágenes satelitales empleadas; Según esto, el uso de un sensor complementario con mayor resolución espacial permitiría detallar aún mejor la variable estudiada. El uso de variables meteorológicas en futuros estudios, presentarían una incidencia positiva en los resultados, puesto que los aerosoles atmosféricos varían su dinámica en base a cambios meteorológicos; generando la posibilidad de recuperar las concentraciones de material particulado con mayor precisión y exactitud.

9. BIBLIOGRAFÍA

- Abad, L., & Mejía, D. (2017). Estimación De La Concentración De Material Particulado Menor a 10 Micras a Través De Sensores Remotos En El Área Urbana De La Ciudad De Cuenca. *XVI Conferencia Iberoamericana de Sistemas de Información Geográfica*, 381–390.
- AERONET. (2021a). *AERONET site information Database*.
https://aeronet.gsfc.nasa.gov/new_web/photo_db_v3/Medellin.html
- AERONET. (2021b). *Aerosol Robotic Network - MISSION*.
https://aeronet.gsfc.nasa.gov/new_web/index.html
- Alcaldía de Medellín. (2006). Medellín y su población. *Documento Técnico De Soporte Pot [Acuerdo 46/2006]*, 83–90.
[http://www.medellin.gov.co/irj/go/km/docs/wpccontent/Sites/Subportal del Ciudadano/Plan de Desarrollo/Secciones/Informaci?n General/Documentos/POT/medellinPoblacion.pdf](http://www.medellin.gov.co/irj/go/km/docs/wpccontent/Sites/Subportal%20del%20Ciudadano/Plan%20de%20Desarrollo/Secciones/Informaci%20n%20General/Documentos/POT/medellinPoblacion.pdf)
- Alegría, D. (2015). *Propiedades microfísicas de los aerosoles troposféricos derivadas de medidas de fotómetro solar y Lidar en el Valle de Aburrá*. Facultad de Ciencias, Escuela de Física Universidad Nacional de Colombia.
- Alvarez, C., & Padilla, O. (2016). Estimación de la contaminación del aire por PM10 en Quito a través de índices ambientales con imágenes LANDSAT ETM+. *Revista Cartográfica*, 147, 135–147.
- Ångström, A. (1964). The parameters of atmospheric turbidity. *Tellus*, 16(1), 64–75.
<https://doi.org/10.3402/tellusa.v16i1.8885>
- Ataz, E., & de Mera Morales, Y. (2004). *Contaminación atmosférica*. (Vol. 45). Univ de Castilla La Mancha.
- Berman, J. D., & Ebisu, K. (2020). Changes in U.S. air pollution during the COVID-19 pandemic. *Science of the Total Environment*, 739, 139864.
<https://doi.org/10.1016/j.scitotenv.2020.139864>
- Bilal, M., & Nichol, J. (2015). Evaluation of MODIS aerosol retrieval algorithms over the Beijing-Tianjin-Hebei region during low to very high pollution events. *Journal of Geophysical Research*, 175(4449), 238.
- Bilal, Muhammad, Nichol, J. E., Bleiweiss, M. P., & Dubois, D. (2013). A Simplified high resolution MODIS aerosol retrieval algorithm (SARA) for use over mixed surfaces. *Remote Sensing of Environment*, 136, 135–145.
- Bilal, Muhammad, Nichol, J. E., & Chan, P. W. (2014). Validation and accuracy assessment of a Simplified Aerosol Retrieval Algorithm (SARA) over Beijing under low and high aerosol loadings and dust storms. *Remote Sensing of Environment*,

153, 50–60. <https://doi.org/10.1016/j.rse.2014.07.015>

- Castro, G. (2019). *Metodología para determinar la cantidad de material particulado en la capa troposfera a partir de teledetección . Área de estudio: Valle de Aburrá , Antioquia , para el año 2016 .* UNIVERSIDAD CATÓLICA DE MANIZALES.
- Castro, M. (2014). *Imputación de datos faltantes en un modelo de tiempo de fallo acelerado.* 53.
- Cespedes, J., Francisco, H., & Cuero, J. (2015). Metodología para medir las concentraciones de aerosoles atmosféricos usando técnicas de teledetección. *Puente*, 9(2), 61–77.
- Chuvienco, E. (1995). *Fundamentos de teledetección espacial* (Segunda ed). Ediciones Rialp.
- Delicado, P. (2008). Curso de Modelos no Paramétricos. *Departamento de Estadística Investigación Operativa, Universidad Politécnica de Cataluña.* http://www.eio.upc.es/~delicado/docencia/Apuntes_Models_No_Parametrics.pdf
- DNP. (2018). Política para el Mejoramiento de la Calidad del Aire. *Documento CONPES 3943.*
- EPA. (2016). *Environmental topics: Particulate Matter (pm) Pollution.* <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM>
- Ferraty, F., & Vieu, P. (2006). Nonparametric functional data analysis: theory and practice. *Springer Science & Business Media.*
- Gupta, P., Christopher, S. A., Wang, J., Gehrig, R., Lee, Y., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment*, 40(30), 5880–5892.
- Hinojosa-Baliño, I., Infante-Vázquez, O., & Vallejo, M. (2019). Distribution of PM2.5 air pollution in Mexico City: spatial analysis with land-use regression model. *Applied Sciences*, 9(14), 2936.
- IDEAM. (2017). *Informe del Estado de la Calidad del Aire en Colombia 2016.*
- IGAC. (2007). *Mejora de los Sistemas de Cartografía del Territorio Colombiano*, 104 p. :il.
- Ihlen, V., & USGS. (2019). Landsat 7 (L7) Data Users Handbook. *USGS Landsat User Services*, 7(November), 151.
- Kaufman, Y. J., Tanré, D., Remer, L. A., Vermote, E. F., Chu, A., & Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *Journal of Geophysical Research Atmospheres*, 102(14), 17051–17067.

- Landrigan, P. J., Fuller, R., Acosta, N. J. R., Adeyi, O., Arnold, R., Basu, N. (Nil), Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J. I., Breyse, P. N., Chiles, T., Mahidol, C., Coll-Seck, A. M., Cropper, M. L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., ... Zhong, M. (2017). The Lancet Commission on pollution and health. *The Lancet*, 391(10119), 462–512.
- Lavery, B., Kariko, A., & Nicholls, N. (1997). A historical rainfall data set for Australia. *Australian Meteorological Magazine*, 40(1), 33–39.
- Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., & Kaufman, Y. J. (2007). Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *Journal of Geophysical Research Atmospheres*, 112(13).
- Liang, S. (2005). *Quantitative remote sensing of land surfaces (Vol. 30)*. John Wiley & Sons.
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation (Seventh Ed)*. John Wiley & Sons.
- Luo, N., Wong, M., Zhao, W., Yan, X., & Xiao, F. (2015). Improved aerosol retrieval algorithm using Landsat images and its application for PM10 monitoring over urban areas. *Atmospheric Research*, 153, 264–275.
- Manteiga, W. G., & Vieu, P. (2007). Statistics for Functional Data. *Computational Statistics and Data Analysis*, 51(10), 4788–4792. <https://doi.org/10.1016/j.csda.2006.10.017>
- MAVDT. (2010). *Ministerio de Ambiente Vivienda y Desarrollo Territorial. Resolución número 610*. <https://www.minambiente.gov.co/images/normativa/app/resoluciones/bf-Resolución 610 de 2010 - Calidad del Aire.pdf>
- Mei, L., Xue, Y., & Kokhanovsky, A. A. (2013). Retrieval of aerosol optical depth over land surfaces from AVHRR data. *Atmospheric Measurement Techniques Discussions*, 6(1), 2227–2251. <https://doi.org/10.5194/amtd-6-2227-2013>
- Minambiente. (2010). Manual de diseño de sistemas de vigilancia de la calidad del aire. *Protocolo Para El Monitoreo y Seguimiento de La Calidad Del Aire*, 137.
- Minambiente. (2017). *Resolución No. 2254 “Por el cual se adopta la norma de calidad del aire ambiente y se dictan otras disposiciones”. MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE*.
- MODIS Characterization Support Team (MCST). (2017a). *MODIS 500m Calibrated Radiance Product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center*. <https://doi.org/http://dx.doi.org/10.5067/MODIS/MOD02HKM.061>

- MODIS Characterization Support Team (MCST). (2017b). *MODIS Geolocation Fields Product. NASA MODIS Adaptive Processing System, Goddard Space Flight Center.* <https://doi.org/http://dx.doi.org/10.5067/MODIS/MOD03.061>
- Moran, M. S., Jackson, R. D., Slater, P. N., & Teillet, P. M. (1999). Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sensing of Environment*, 41(2–3), 169–184.
- NASA. (2007). *Electromagnetic Spectrum Diagram.* 2007. <https://mydasdata.larc.nasa.gov/basic-page/electromagnetic-spectrum-diagram>
- NASA. (2019). *Terra & Aqua Moderate Resolution Imaging Spectroradiometer (MODIS).* <https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/modis/>
- OMS. (2005). Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre. *Actualización Mundial 2005*, 5(1), 1–21.
- Ordóñez, J. (2012). Espectro electromagnético y espectro radioeléctrico. *Acta*, 62, 17–31.
- Ospína, J., Giraldo, R., & Andrade, M. (2019). Functional regression concurrent model with spatially correlated errors: application to rainfall ground validation. *Journal of Applied Statistics*, 46(8), 1350–1363.
- Othman, N., Mat Jafri, M. Z., & San, L. H. (2010). Estimating Particulate Matter Concentration over Arid Region Using Satellite Remote Sensing: A Case Study in Makkah, Saudi Arabia. *Modern Applied Science*, 4(11), 131–142.
- R-Project. (2021). *The R Project for Statistical Computing.* <https://www.r-project.org/>
- Rahman, H., Pinty, B., & Verstraete, M. M. (1993). Coupled surface-atmosphere reflectance (CSAR) model 2. Semiempirical surface model usable with NOAA advanced very high resolution radiometer data. *Journal of Geophysical Research*, 98(D11). <https://doi.org/10.1029/93jd02072>
- Ramsay, J., Hooker, G., & Graves, S. (2009). Functional Data Analysis with R and MATLAB. In R. Gentleman, K. Hornik, & G. Parmigiano (Eds.), *Applied Spatial Data Analysis with R*. Springer.
- Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis (Second Ed.)*. Springer.
- Remer, L., Tanré, D., & Kaufman, Y. (2009). *Algorithm for remote sensing of tropospheric aerosol from MODIS: Collection 5.*
- Roger, J., Vermote, E., & Ray, P. (2011). *MODIS Surface Reflectance User 's Guide.* 40.

- Santiago, E., & Chicangana, F. (2017). *Análisis Exploratorio Funcional de las Curvas Diarias de Concentraciones Promedio Horarias de Material Particulado Fino (PM2.5) en el Aire de la Ciudad de Santiago de Cali para el Periodo Diciembre 2014- Noviembre 2015*. Universidad Del Valle.
- Servicio Geológico Minero Argentino. (2018). *Sensores Remotos*. <http://www.segemar.gov.ar/igrm/sensores-remotos/>
- SIATA. (2019a). *Generalidades Información de estaciones*. Sistema de Alerta Temprana de Medellín y El Valle de Aburrá. https://siata.gov.co/descarga_siata/index.php/info/estaciones/
- SIATA. (2019b). *Quienes somos*. Sistema de Alerta Temprana de Medellín y El Valle de Aburrá. https://siata.gov.co/sitio_web/index.php/quienesSomos
- Suhaila, J., Jemain, A. A., Hamdan, M. F., & Wan Zin, W. Z. (2011). Comparing rainfall patterns between regions in Peninsular Malaysia via a functional data analysis technique. *Journal of Hydrology*, 411(3–4), 197–206.
- Taaheri, A., & Murthy, S. (2019). *HDF-EOS to GeoTIFF Conversion Tool (HEG) Stand-alone Users Guide, Version 2.15, Technical Paper*.
- Toller, G., Isaacman, A., & Kuyper, J. (2017). MODIS Level 1B Product Users Guide Operation Operation. *Signature*, 2. <https://mcst.gsfc.nasa.gov/content/l1b-documents>
- Vermote, E., & Wolfe, R. (2015). MOD09GA MODIS / Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V006. *Procesos Terrestres EOSDIS de La NASA DAAC*. <https://doi.org/https://doi.org/10.5067/MODIS/MOD09GA.006>
- Vermote, Tanré, D., Deuzé, J. L., Herman, M., & Morcrette, J. J. (1997). Second simulation of the satellite signal in the solar spectrum, 6s: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3), 675–686.
- Wang, C., Liu, Q., Ying, N., Wang, X., & Ma, J. (2013). Air quality evaluation on an urban scale based on MODIS satellite images. *Atmospheric Research*, 132–133(November 2020), 22–34.
- Woldemeskel, F. M., Sivakumar, B., & Sharma, A. (2013). Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *Journal of Hydrology*, 499, 167–176. <https://doi.org/10.1016/j.jhydrol.2013.06.039>
- Zapata, I. B., & Anaya, J. A. (2011). Los sensores remotos en los proyectos de mitigación de gases de efecto invernadero. *Entorno Geográfico*, 7–8(1692–0074), 67–83.

10. ANEXOS

Anexo A Datos de PM2.5 terrestre y del espesor óptico de aerosoles (AOD) obtenido mediante el algoritmo simplificado de recuperación de aerosoles (SARA) para las 9 estaciones de monitoreo y las 26 fechas de estudio comprendidas entre los años 2019 y 2020.

Código Estación	Fecha	Hora	Dia juliano	AOD_550nm	PM2,5
12	1/01/2019	16:00	1	0,04362	7
25	1/01/2019	16:00	1	0,04216	
44	1/01/2019	16:00	1	0,04495	
79	1/01/2019	16:00	1	0,07018	13,72
80	1/01/2019	16:00	1	0,02841	1,25
83	1/01/2019	16:00	1	0,03650	3,74
84	1/01/2019	16:00	1	0,05404	
85	1/01/2019	16:00	1	0,06126	10,32
86	1/01/2019	16:00	1	0,08938	12,48
12	2/01/2019	15:00	2	0,04781	13
25	2/01/2019	15:00	2	0,03665	7
44	2/01/2019	15:00	2	0,02919	
79	2/01/2019	15:00	2	0,03867	10,02
80	2/01/2019	15:00	2	0,02449	5,01
83	2/01/2019	15:00	2	0,03049	2,51
84	2/01/2019	15:00	2	0,03887	
85	2/01/2019	15:00	2	0,04013	5,16
86	2/01/2019	15:00	2	0,04215	7,49
12	3/01/2019	15:00	3	0,23011	12
25	3/01/2019	15:00	3	0,11454	6
44	3/01/2019	15:00	3	0,16074	5
79	3/01/2019	15:00	3	0,29334	18,71
80	3/01/2019	15:00	3	0,24705	
83	3/01/2019	15:00	3		
84	3/01/2019	15:00	3	0,06991	1,25
85	3/01/2019	15:00	3	0,17368	5,12
86	3/01/2019	15:00	3	0,52760	
12	4/01/2019	14:00	4	0,02566	
25	4/01/2019	14:00	4	0,02767	6
44	4/01/2019	14:00	4	0,02204	2

79	4/01/2019	14:00	4	0,03226	7,48
80	4/01/2019	14:00	4	0,02513	7,50
83	4/01/2019	14:00	4	0,03079	5,01
84	4/01/2019	14:00	4	0,02826	5,04
85	4/01/2019	14:00	4	0,02951	6,41
86	4/01/2019	14:00	4	0,03912	11,14
12	30/01/2019	15:00	30	0,57156	31
25	30/01/2019	15:00	30	0,32355	18
44	30/01/2019	15:00	30		15
79	30/01/2019	15:00	30	0,42292	14,89
80	30/01/2019	15:00	30		17,29
83	30/01/2019	15:00	30	0,37891	13,74
84	30/01/2019	15:00	30		11,29
85	30/01/2019	15:00	30	0,35308	14,02
86	30/01/2019	15:00	30		24,64
12	15/02/2019	15:00	46	0,44507	32
25	15/02/2019	15:00	46	0,46070	24
44	15/02/2019	15:00	46	0,42997	22
79	15/02/2019	15:00	46	0,41253	21,24
80	15/02/2019	15:00	46		17,48
83	15/02/2019	15:00	46	0,22666	20,19
84	15/02/2019	15:00	46	0,30057	19,06
85	15/02/2019	15:00	46	0,38509	24,64
86	15/02/2019	15:00	46	0,58197	28,58
12	22/02/2019	15:00	53	0,82345	30
25	22/02/2019	15:00	53	0,75806	22
44	22/02/2019	15:00	53	1,21489	
79	22/02/2019	15:00	53	1,01002	31,58
80	22/02/2019	15:00	53	0,20344	12,79
83	22/02/2019	15:00	53	0,40188	13,96
84	22/02/2019	15:00	53	0,46714	17,82
85	22/02/2019	15:00	53	0,70503	18,21
86	22/02/2019	15:00	53	0,81995	26,27
12	1/05/2019	15:00	121	0,13216	19
25	1/05/2019	15:00	121	0,07686	6
44	1/05/2019	15:00	121	0,07449	3
79	1/05/2019	15:00	121	0,10377	5,03
80	1/05/2019	15:00	121	0,12294	
83	1/05/2019	15:00	121	0,10487	6,22
84	1/05/2019	15:00	121	0,08872	2,53
85	1/05/2019	15:00	121	0,09616	8,92
86	1/05/2019	15:00	121	0,11726	6,22

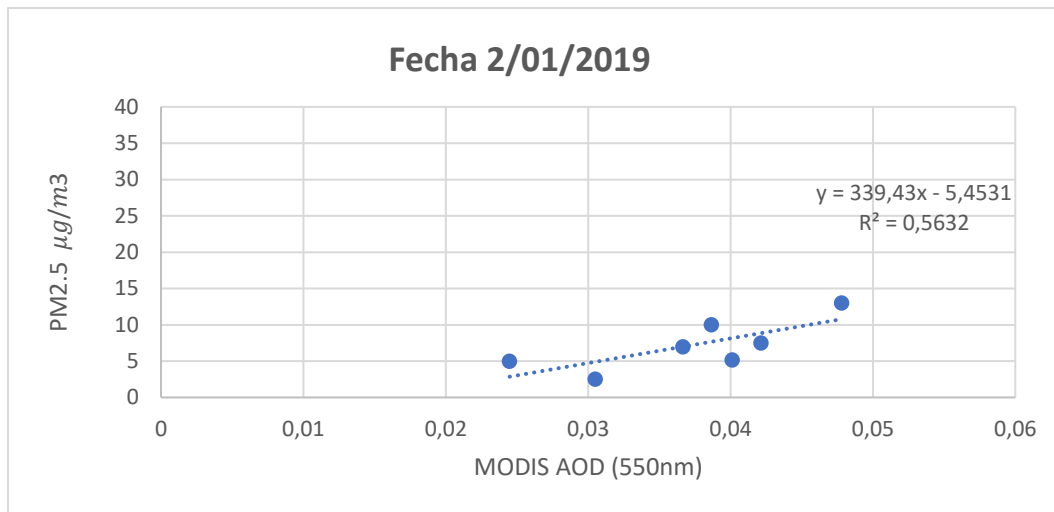
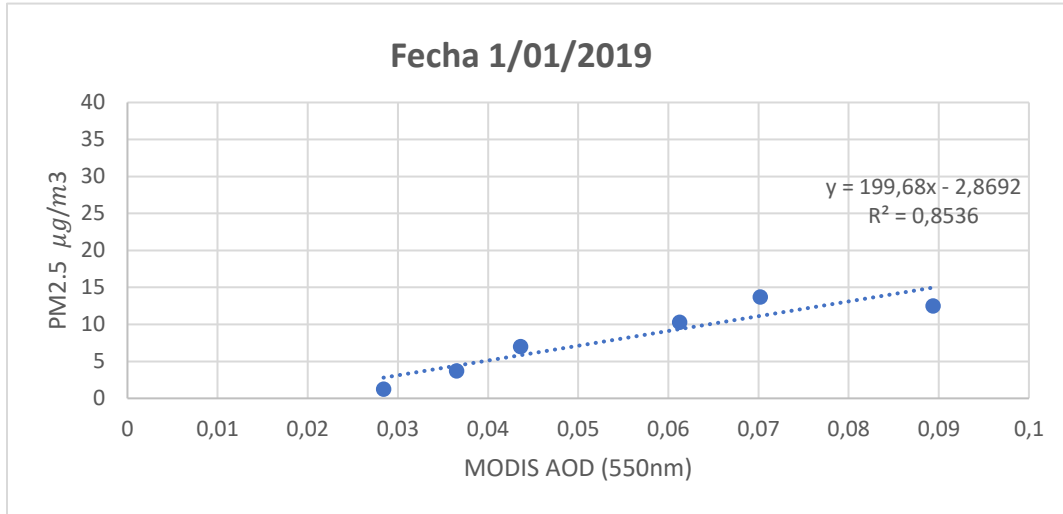
12	24/05/2019	15:00	144		36
25	24/05/2019	15:00	144	0,13542	12
44	24/05/2019	15:00	144	0,23794	
79	24/05/2019	15:00	144	0,22301	18,81
80	24/05/2019	15:00	144		31,21
83	24/05/2019	15:00	144		28,93
84	24/05/2019	15:00	144	0,18170	23,59
85	24/05/2019	15:00	144	0,10692	14,11
86	24/05/2019	15:00	144	0,10692	12,31
12	26/06/2019	16:00	177	0,07037	15
25	26/06/2019	16:00	177	0,04947	10
44	26/06/2019	16:00	177	0,05919	
79	26/06/2019	16:00	177	0,05545	6,23
80	26/06/2019	16:00	177	0,05077	10,09
83	26/06/2019	16:00	177	0,02040	3,74
84	26/06/2019	16:00	177	0,02773	1,25
85	26/06/2019	16:00	177	0,06803	11,44
86	26/06/2019	16:00	177	0,05692	6,26
12	6/07/2019	15:00	187	0,04058	7
25	6/07/2019	15:00	187		9
44	6/07/2019	15:00	187	0,03202	3
79	6/07/2019	15:00	187	0,03748	6,23
80	6/07/2019	15:00	187	0,02793	1,25
83	6/07/2019	15:00	187	0,02721	2,48
84	6/07/2019	15:00	187	0,03176	3,73
85	6/07/2019	15:00	187	0,03229	3,79
86	6/07/2019	15:00	187	0,04871	4,98
12	5/08/2019	15:00	217	0,11173	15
25	5/08/2019	15:00	217	0,06775	5
44	5/08/2019	15:00	217	0,10975	9
79	5/08/2019	15:00	217	0,06450	5,01
80	5/08/2019	15:00	217	0,12474	9,97
83	5/08/2019	15:00	217	0,07935	3,73
84	5/08/2019	15:00	217	0,06450	2,53
85	5/08/2019	15:00	217	0,10496	11,47
86	5/08/2019	15:00	217	0,09360	12,42
12	9/12/2019	15:00	343	0,16939	14
25	9/12/2019	15:00	343	0,10064	10
44	9/12/2019	15:00	343	0,16775	12
79	9/12/2019	15:00	343	0,06997	7,47
80	9/12/2019	15:00	343	0,20223	
83	9/12/2019	15:00	343	0,12881	11,34

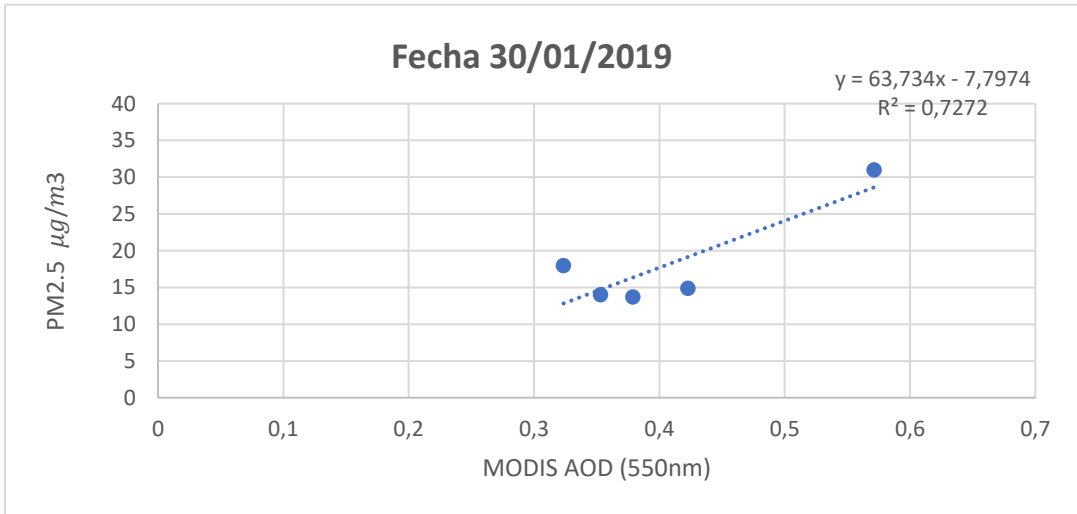
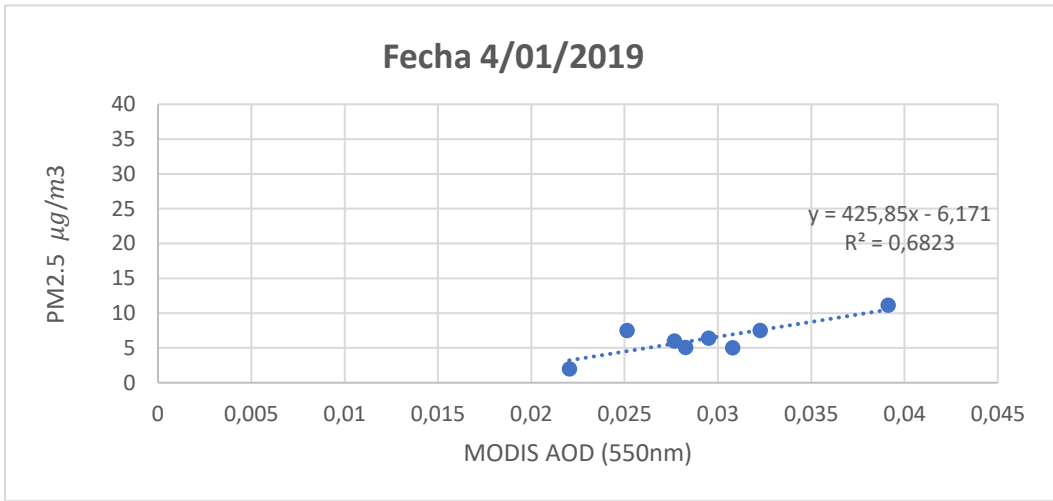
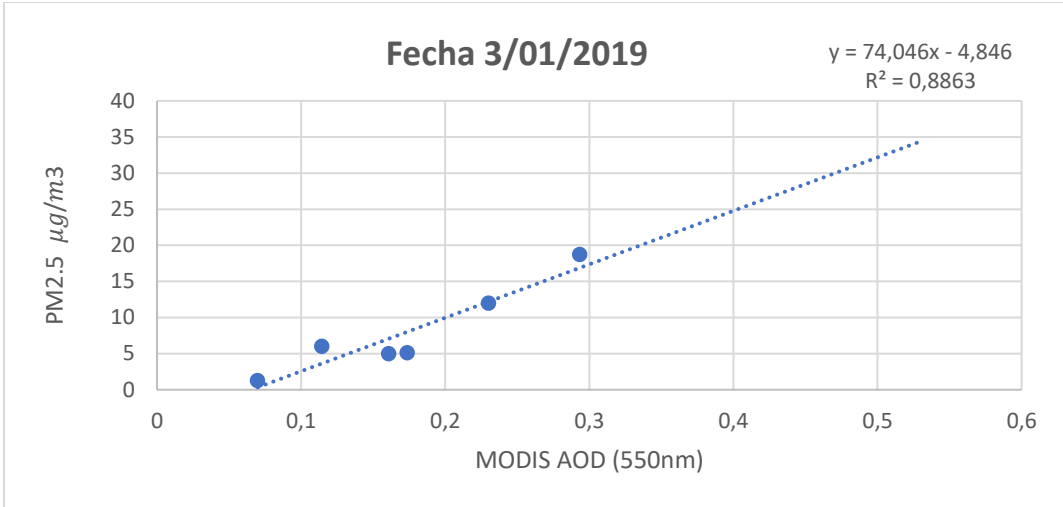
84	9/12/2019	15:00	343	0,12242	11,09
85	9/12/2019	15:00	343	0,09558	
86	9/12/2019	15:00	343	0,16814	9,77
Año 2020					
12	6/01/2020	15:00	6	0,32677	8
25	6/01/2020	15:00	6	0,39738	10
44	6/01/2020	15:00	6	0,36812	10
79	6/01/2020	15:00	6	0,49851	18,43
80	6/01/2020	15:00	6	0,49898	
83	6/01/2020	15:00	6	0,41807	12,34
84	6/01/2020	15:00	6	0,41055	14,80
85	6/01/2020	15:00	6	0,43255	10,04
86	6/01/2020	15:00	6	0,52094	14,85
12	10/01/2020	15:00	10	0,25879	7
25	10/01/2020	15:00	10	0,21312	4
44	10/01/2020	15:00	10	0,19729	4
79	10/01/2020	15:00	10	0,19190	
80	10/01/2020	15:00	10	0,26271	6,18
83	10/01/2020	15:00	10	0,21227	8,69
84	10/01/2020	15:00	10	0,28754	9,93
85	10/01/2020	15:00	10	0,25130	5,04
86	10/01/2020	15:00	10	0,30420	7,46
12	11/01/2020	16:00	11	0,08940	
25	11/01/2020	16:00	11	0,10611	
44	11/01/2020	16:00	11	0,07370	
79	11/01/2020	16:00	11	0,08939	
80	11/01/2020	16:00	11	0,10248	12,35
83	11/01/2020	16:00	11	0,08444	3,70
84	11/01/2020	16:00	11	0,07506	1,23
85	11/01/2020	16:00	11	0,09368	3,76
86	11/01/2020	16:00	11	0,09982	7,45
12	12/01/2020	15:00	12	0,07238	
25	12/01/2020	15:00	12	0,08458	6
44	12/01/2020	15:00	12	0,05000	4
79	12/01/2020	15:00	12	0,07588	
80	12/01/2020	15:00	12	0,08563	7,40
83	12/01/2020	15:00	12	0,05689	
84	12/01/2020	15:00	12	0,06308	6,20
85	12/01/2020	15:00	12	0,06293	3,75
86	12/01/2020	15:00	12		13,61
12	18/02/2020	15:00	49	0,42429	7

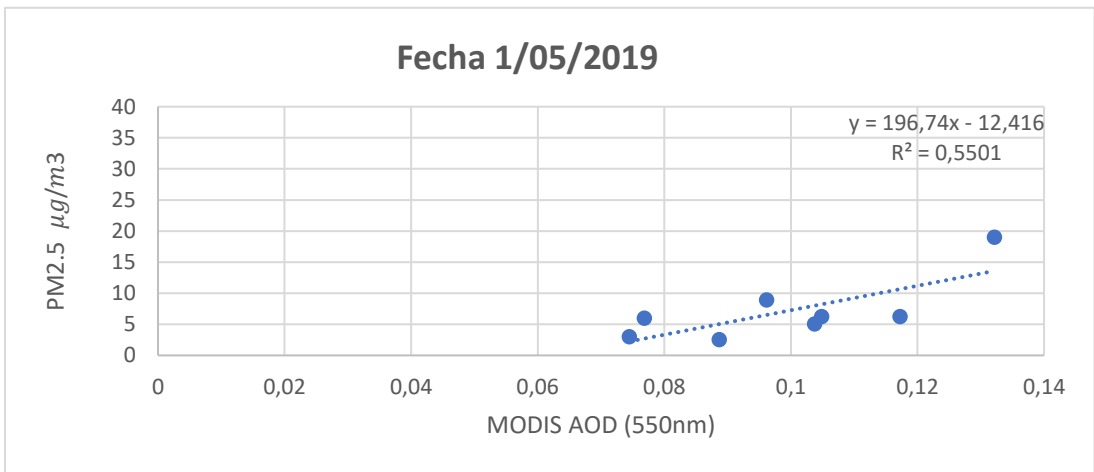
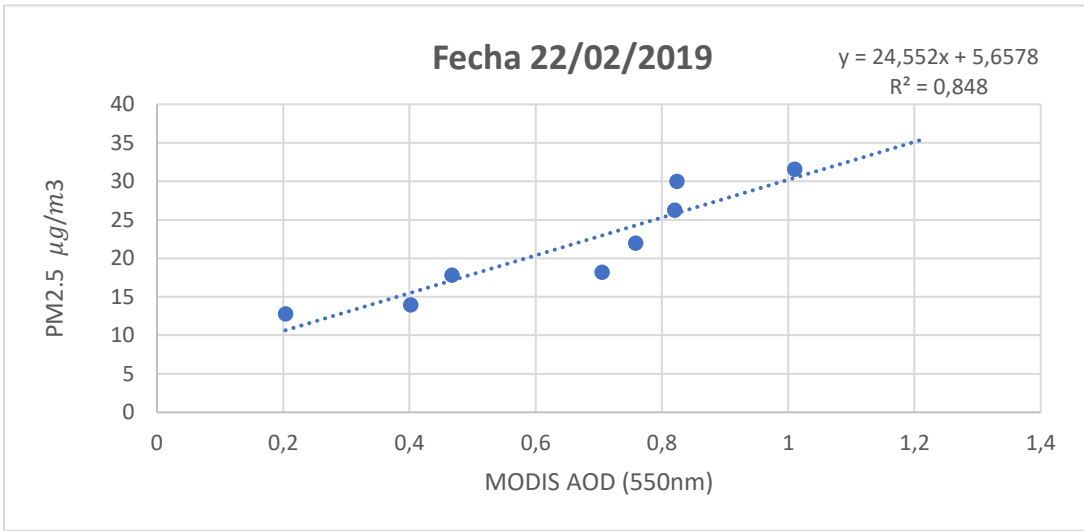
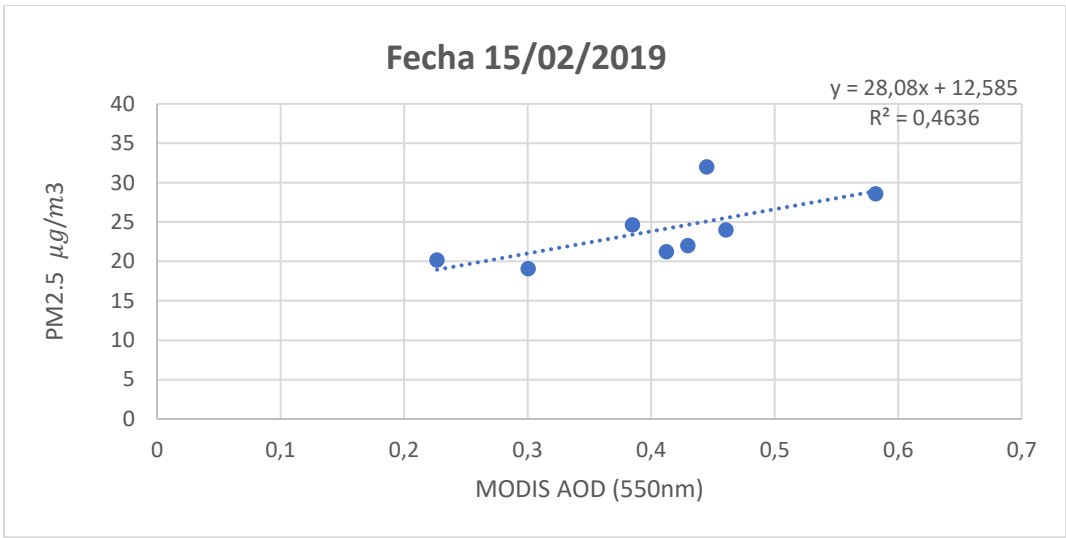
25	18/02/2020	15:00	49	0,39458	9
44	18/02/2020	15:00	49	0,54210	7
79	18/02/2020	15:00	49	0,35426	8,76
80	18/02/2020	15:00	49	0,75875	13,75
83	18/02/2020	15:00	49	0,21633	4,98
84	18/02/2020	15:00	49	0,32405	8,75
85	18/02/2020	15:00	49	0,30250	5,14
86	18/02/2020	15:00	49	0,61729	11,14
12	20/05/2020	16:00	141	0,19228	
25	20/05/2020	16:00	141	0,25430	
44	20/05/2020	16:00	141	0,22239	5
79	20/05/2020	16:00	141	0,19726	4,97
80	20/05/2020	16:00	141	0,17081	4,97
83	20/05/2020	16:00	141	0,08193	1,24
84	20/05/2020	16:00	141	0,22637	9,91
85	20/05/2020	16:00	141	0,11432	5,06
86	20/05/2020	16:00	141	0,21819	11,07
12	4/06/2020	15:00	156	0,08596	6
25	4/06/2020	15:00	156	0,11671	
44	4/06/2020	15:00	156	0,10741	
79	4/06/2020	15:00	156	0,10999	11,22
80	4/06/2020	15:00	156	0,08880	2,48
83	4/06/2020	15:00	156	0,04428	
84	4/06/2020	15:00	156	0,09777	3,75
85	4/06/2020	15:00	156	0,10239	7,60
86	4/06/2020	15:00	156		3,67
12	13/07/2020	15:00	195	0,14514	8
25	13/07/2020	15:00	195	0,12596	
44	13/07/2020	15:00	195	0,09317	2
79	13/07/2020	15:00	195	0,18944	14,19
80	13/07/2020	15:00	195	0,21269	7,40
83	13/07/2020	15:00	195	0,13515	11,13
84	13/07/2020	15:00	195	0,09778	3,76
85	13/07/2020	15:00	195	0,07937	3,81
86	13/07/2020	15:00	195		3,66
12	31/07/2020	15:00	213	0,07815	12
25	31/07/2020	15:00	213	0,09202	
44	31/07/2020	15:00	213	0,07158	
79	31/07/2020	15:00	213	0,05068	9,66
80	31/07/2020	15:00	213	0,02999	8,55
83	31/07/2020	15:00	213	0,06348	9,79
84	31/07/2020	15:00	213		6,18

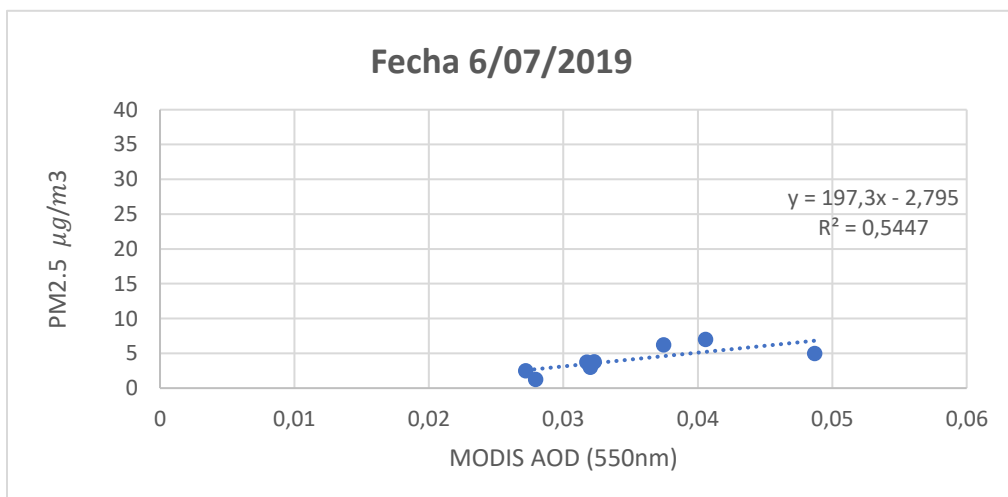
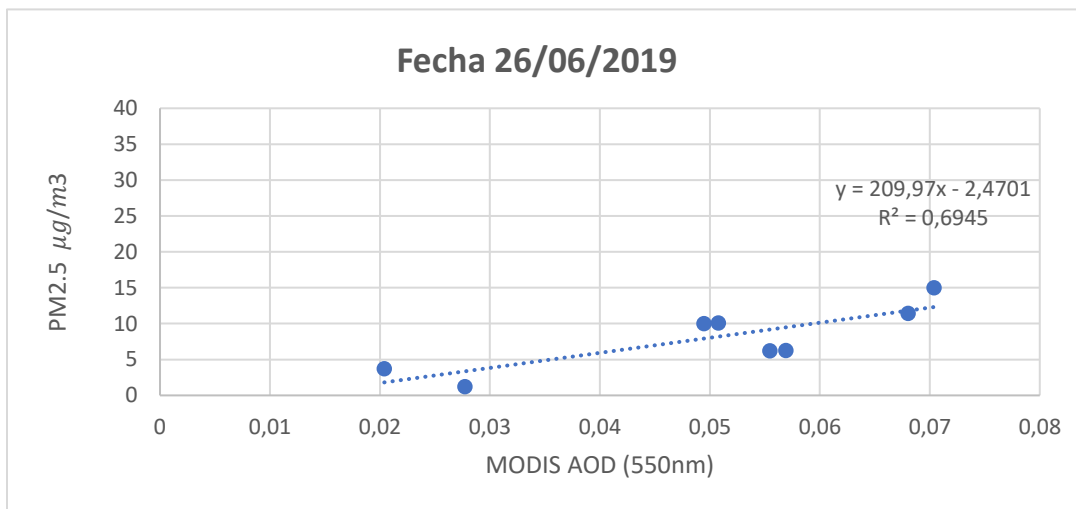
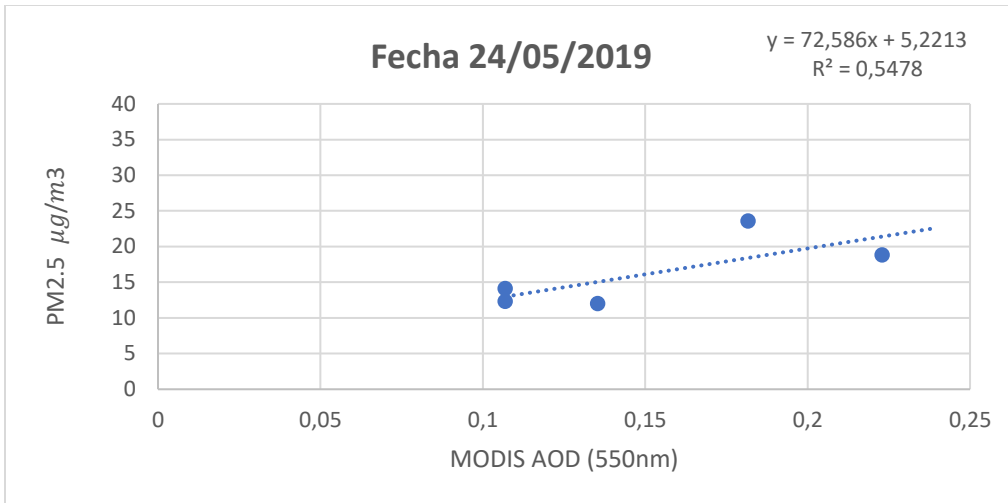
85	31/07/2020	15:00	213	0,06185	10,04
86	31/07/2020	15:00	213	0,06698	10,88
12	2/08/2020	14:00	215	0,04282	9
25	2/08/2020	14:00	215	0,03779	
44	2/08/2020	14:00	215	0,02980	
79	2/08/2020	14:00	215	0,03575	6,21
80	2/08/2020	14:00	215	0,03742	4,94
83	2/08/2020	14:00	215	0,02740	2,50
84	2/08/2020	14:00	215	0,02208	
85	2/08/2020	14:00	215	0,03230	0
86	2/08/2020	14:00	215	0,03357	2,45
12	10/10/2020	15:00	284	0,10872	9
25	10/10/2020	15:00	284	0,13415	
44	10/10/2020	15:00	284	0,07424	5
79	10/10/2020	15:00	284	0,07695	
80	10/10/2020	15:00	284	0,11446	
83	10/10/2020	15:00	284	0,03602	
84	10/10/2020	15:00	284	0,06562	6,24
85	10/10/2020	15:00	284	0,05817	3,83
86	10/10/2020	15:00	284	0,06119	3,75
12	2/11/2020	15:00	307	0,10031	
25	2/11/2020	15:00	307	0,21036	
44	2/11/2020	15:00	307	0,15860	
79	2/11/2020	15:00	307	0,09589	3,71
80	2/11/2020	15:00	307		1,24
83	2/11/2020	15:00	307	0,10106	6,25
84	2/11/2020	15:00	307	0,09160	4,97
85	2/11/2020	15:00	307	0,10034	7,58
86	2/11/2020	15:00	307	0,16358	8,60
12	7/11/2020	15:00	312	0,42415	8
25	7/11/2020	15:00	312	1,49713	
44	7/11/2020	15:00	312	1,07354	26
79	7/11/2020	15:00	312	0,69619	9,83
80	7/11/2020	15:00	312	0,63533	12,33
83	7/11/2020	15:00	312	0,62529	8,68
84	7/11/2020	15:00	312	0,86894	19,79
85	7/11/2020	15:00	312	0,76145	11,35
86	7/11/2020	15:00	312	0,90852	13,31

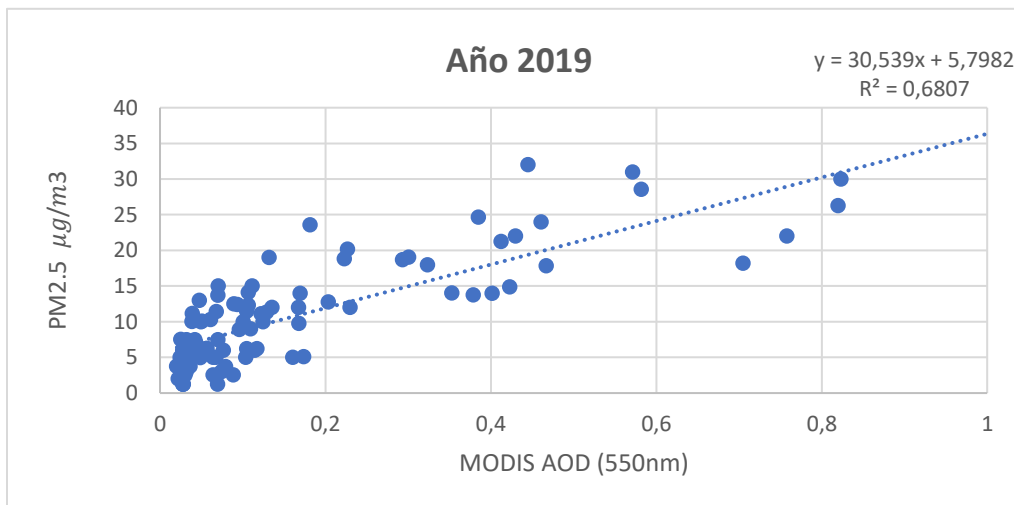
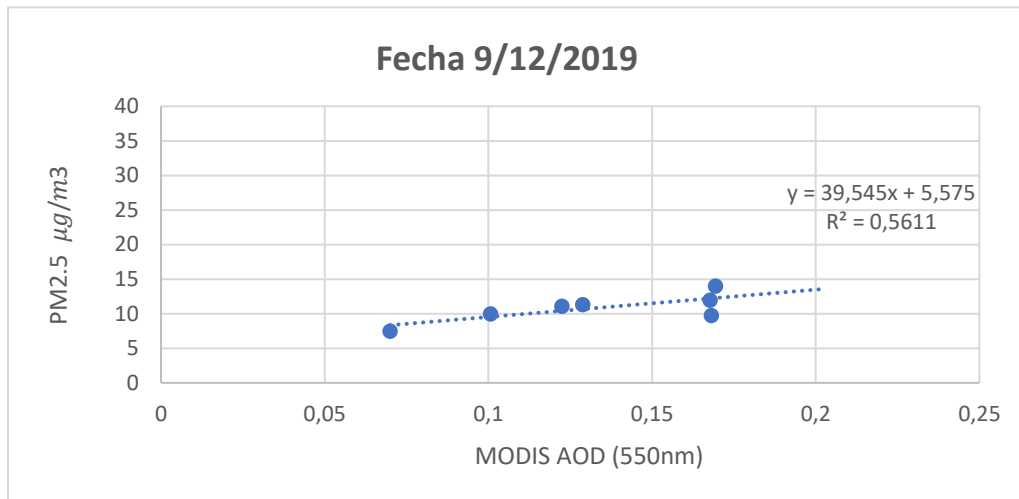
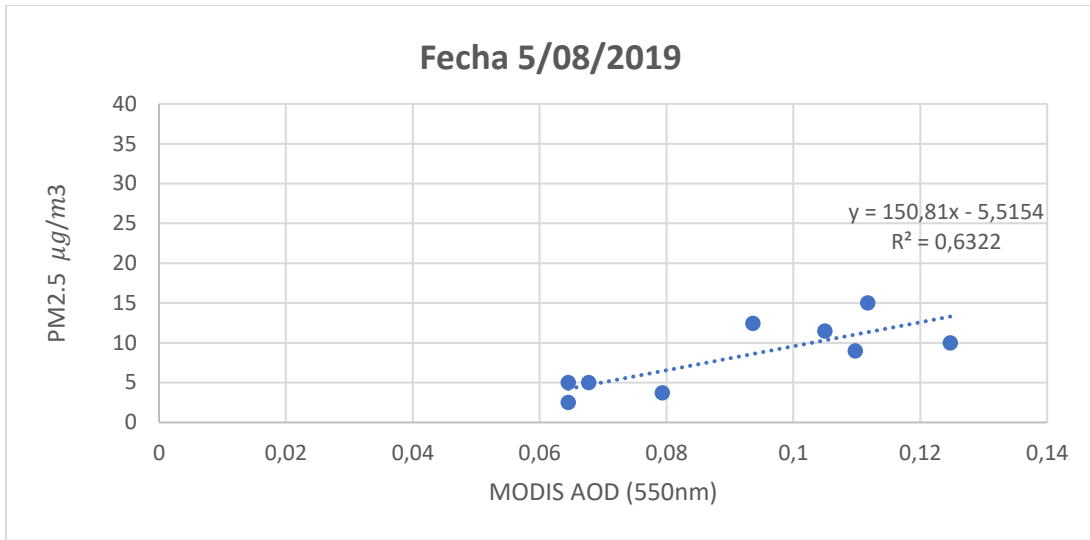
Anexo B Gráficos de dispersión y regresiones lineales para las 26 fechas de estudio comprendidas entre los años 2019 y 2020 entre las concentraciones de PM2.5 medido en tierra y el AOD obtenido mediante el modelo SARA.

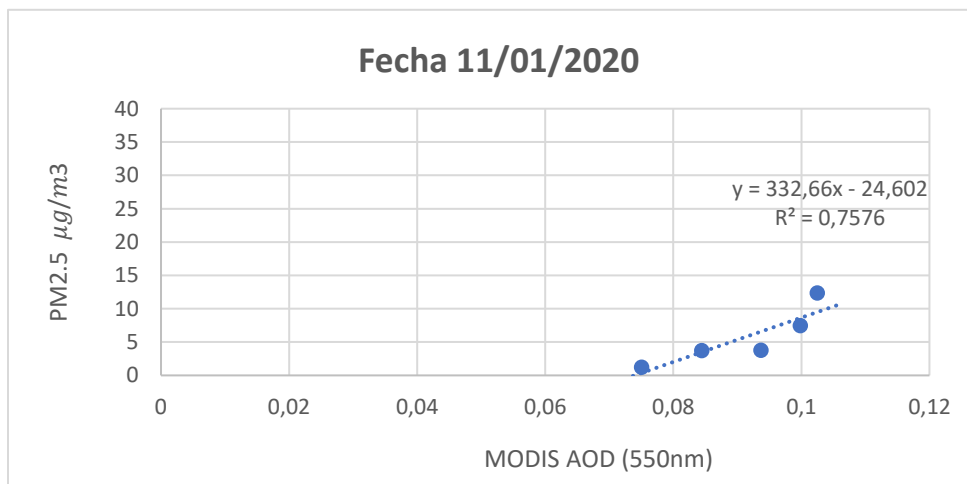
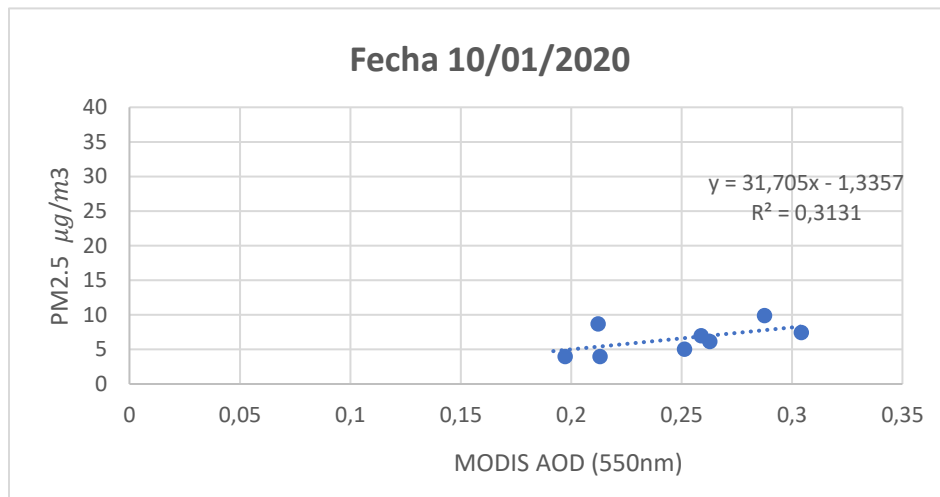
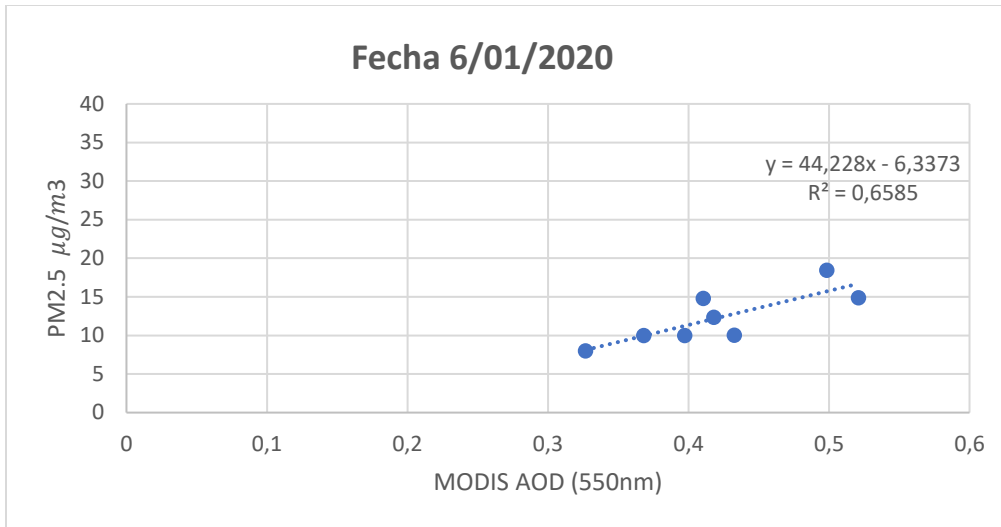


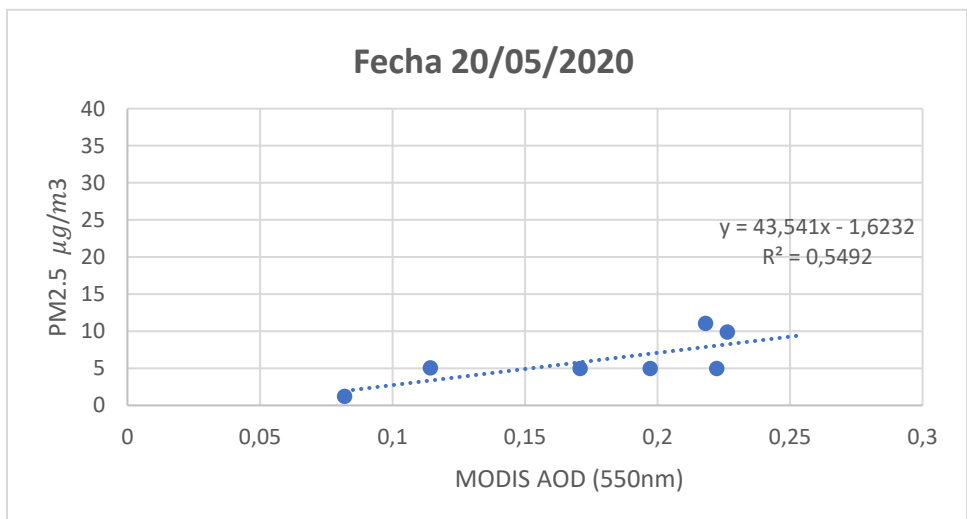
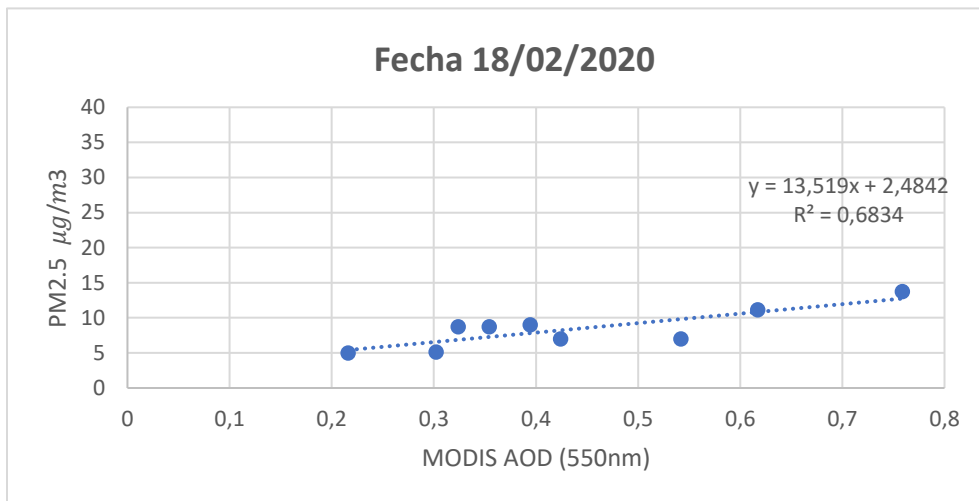
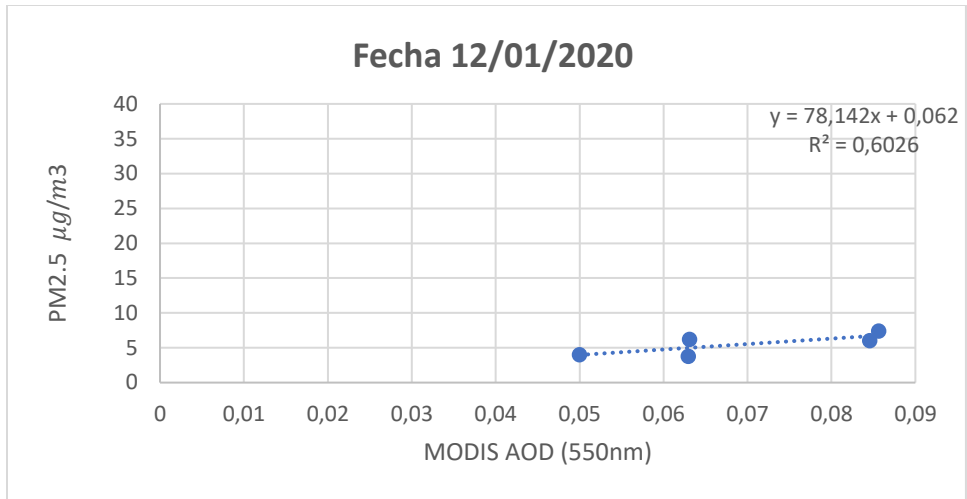


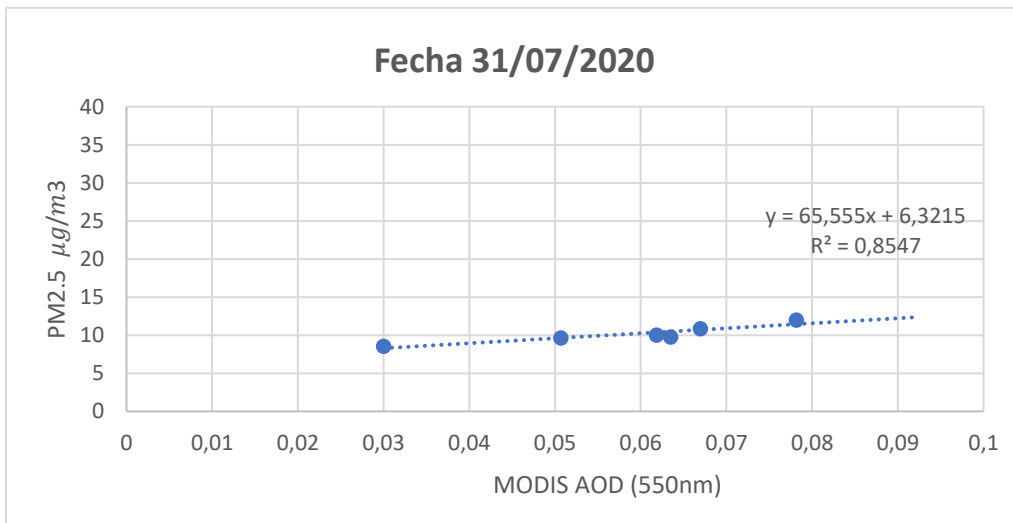
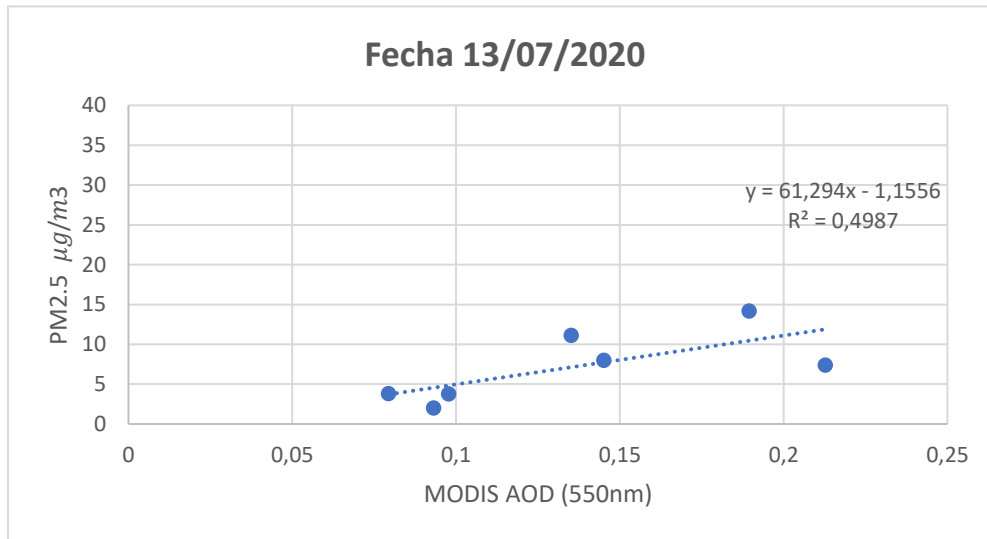
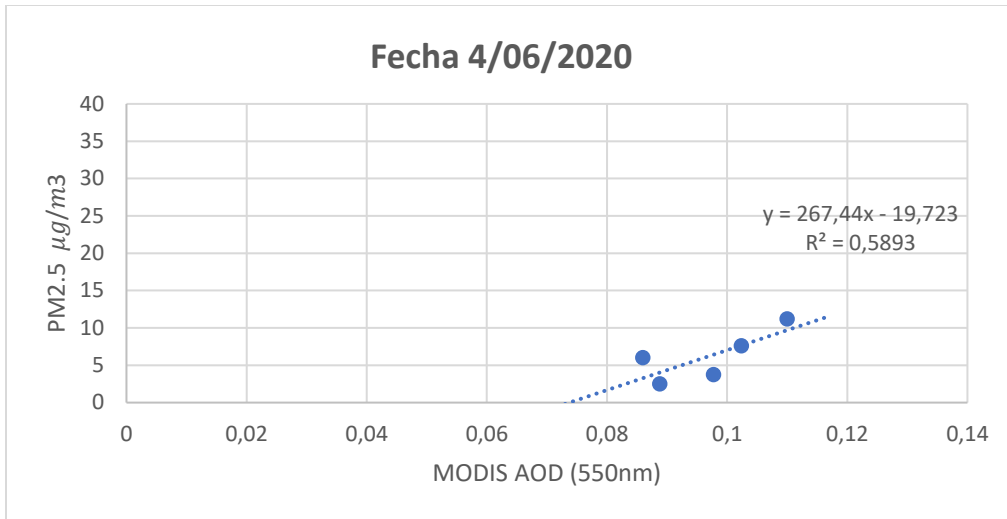


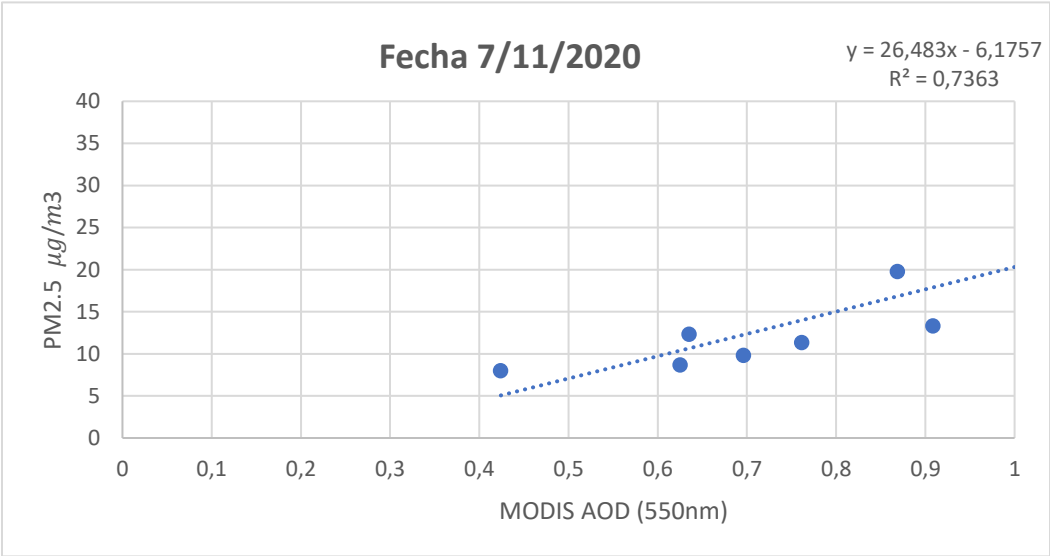
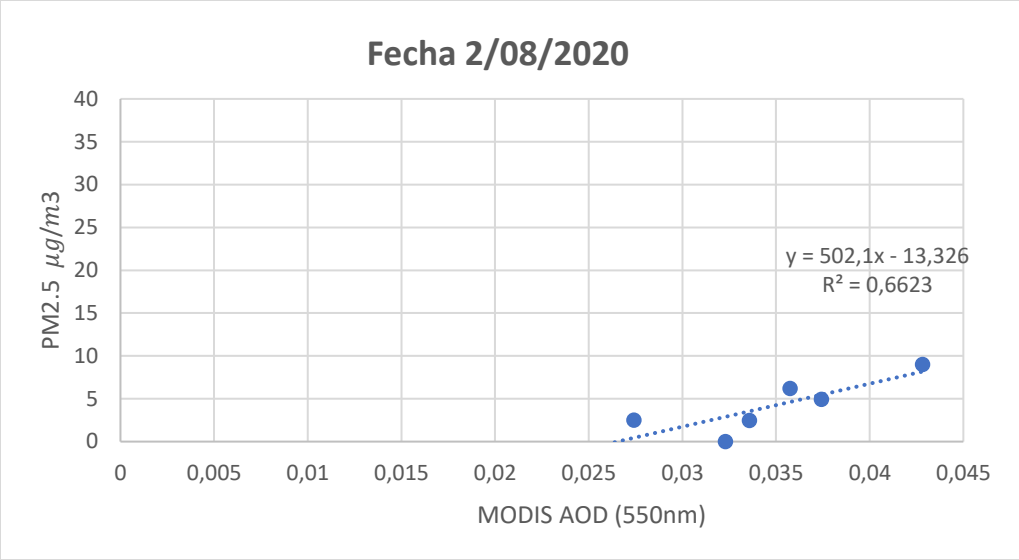


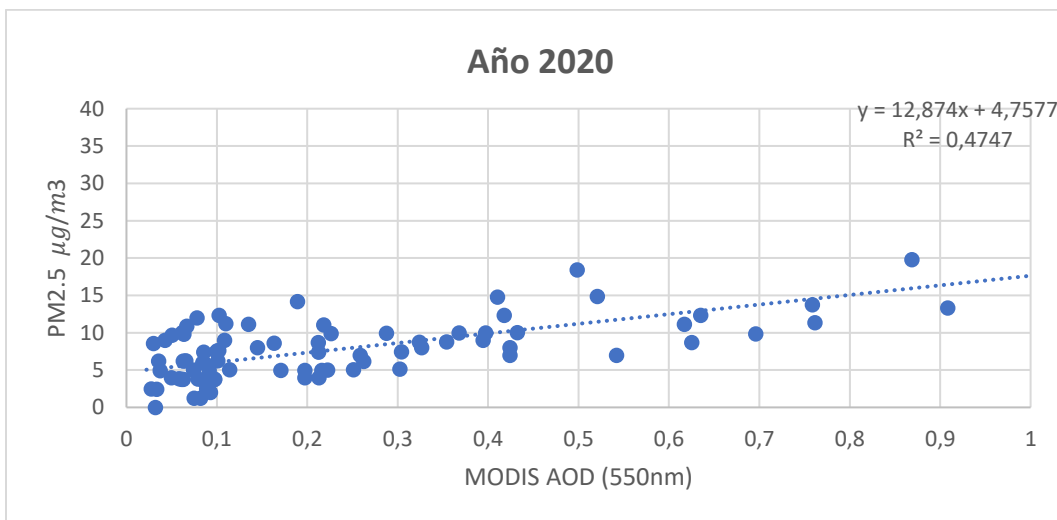
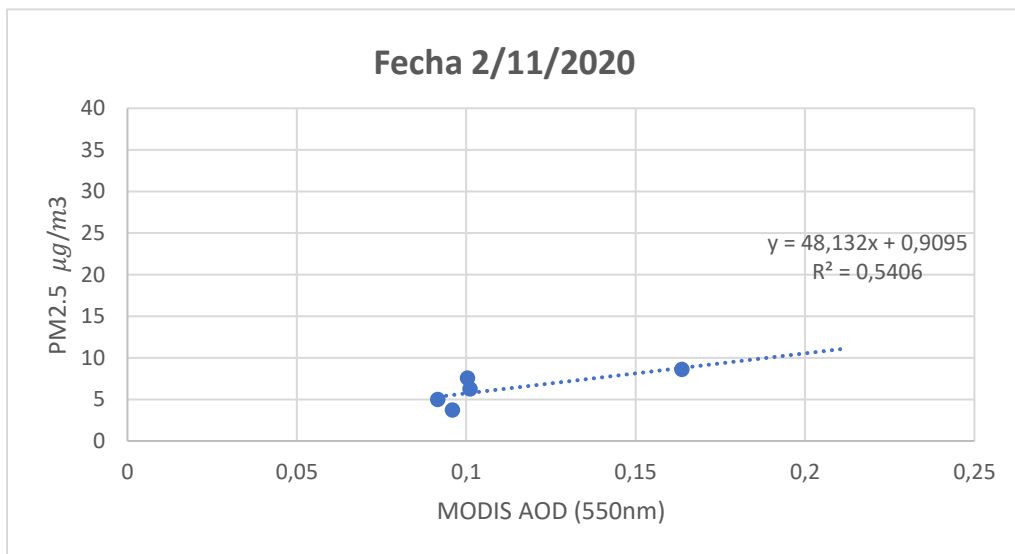
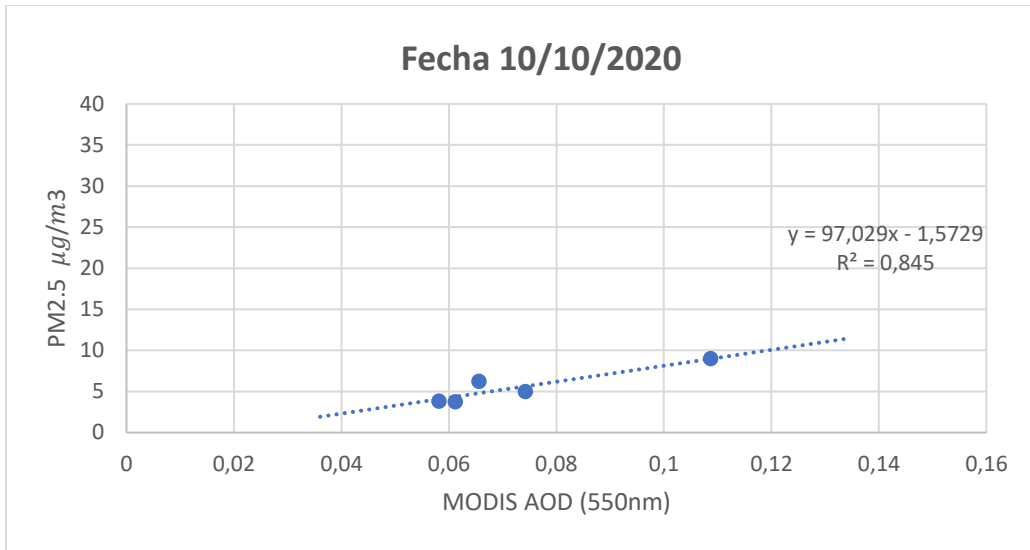












Anexo C Mapas de concentraciones de PM2.5 en la ciudad de Medellín, para las fechas de estudio.

