



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA AMBIENTAL**

**Evaluación de la contaminación y calidad del aire por material
particulado vinculado a las variables meteorológicas en Lima
Metropolitana entre 2017-2020**

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniera Ambiental

AUTORAS:

Mendez Reyes, Rubi Katerine (orcid.org/0000-0002-3205-5386)
Urquia Ochavano, Anayca Nayra (orcid.org/0000-0002-1804-4912)

ASESOR:

Dr. Munive Cerron, Ruben Víctor (orcid.org/0000-0001-8951-2499)

LÍNEA DE INVESTIGACIÓN:

Calidad y Gestión de los Recursos Naturales

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo sostenible y adaptación al cambio climático

LIMA – PERÚ

2023

Dedicatoria

A mi amada madre Milagros Emperatriz Reyes Diaz, quien ha sido siempre la inspiración para alcanzar mis metas y ser una profesional.

A mi querida hermanita Saori Michelle Flores Reyes, que siempre la llevo en mi corazón.

A mi amada abuela Nelly Emperatriz Diaz Maturrano, a quien agradezco todo su apoyo en lograr ser una profesional.

A mis tíos Fernando, Ricardo, Verónica y Sandy por sus consejos y constante aliento para conseguir mis metas.

A mis primos Alejandro, Antonio, Isaias, Karumi, July, Echani, para quienes deseo ser ejemplo de éxito y superación.

A mi hija Bibi-Ying, que estuvo incondicionalmente conmigo brindándome su personalidad única.

A mi amado Oppa Cesar Antonio Guerrero Barrios, quien me apoyo en los momentos más difíciles de mi vida, brindándome su lealtad, paciencia, cariño, respeto, valores y más. A pesar de todo, 사랑해요.

Rubi Katerine Mendez Reyes

A Dios por todo, por la vida y la salud que día a día nos brinda, por darme la fuerza y sabiduría para salir adelante.

A mi padre y a mi madre por todo el apoyo y siempre estar conmigo en las buenas y malas pase lo que pase, por ser mi soporte de afrontar muchos retos que se me presenten en el camino.

A mis hermanos(as) que me incentivan para poder lograr mis objetivos. A mi novio que siempre me apoya y me aconseja a nunca rendirme.

A mi amiga Rubí Mendez por aconsejarme a que si puedo lograr mis metas.

Anayca Nayra Urquia Ochavano

Agradecimiento

A nuestro asesor el Dr. Rubén Munive Cerrón por dedicación, motivación, paciencia y conocimientos aportados en la corrección y mejora en la redacción de la tesis.

A nuestro coasesor el Ing. Oswaldo Miguel Flores Huaman por darnos los conocimientos necesarios en el procesamiento de datos, programas informáticos y métodos estadísticos aplicados en el análisis de los datos ambientales.

A los investigadores del Instituto del Mar del Perú: Ing. Marceliano Segura Zamudio y Msc. Manuel Guzmán Roca, por el aporte de conocimiento en relación al monitoreo, evaluación y recomendaciones en el desarrollo de esta tesis.

A los docentes de la Universidad Nacional José Faustino Sánchez Carrión: Dr. Oswaldo Francisco Flores Saldaña, MSC Carlos Maldonado Vázquez y MSC María Hurtado Zamora, por sus sugerencias para el presente trabajo de investigación.



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA AMBIENTAL**

Declaratoria de Autenticidad del Asesor

Yo, RUBEN VICTOR MUNIVE CERRON, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA AMBIENTAL de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis titulada: "Evaluación de la contaminación y calidad del aire por material particulado vinculado a las variables meteorológicas en Lima Metropolitana entre 2017-2020", cuyos autores son MENDEZ REYES RUBI KATERINE, URQUIA OCHAVANO ANAYCA NAYRA, constato que la investigación tiene un índice de similitud de 17.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 15 de Diciembre del 2023

Apellidos y Nombres del Asesor:	Firma
RUBEN VICTOR MUNIVE CERRON DNI: 19889810 ORCID: 0000-0001-8951-2499	Firmado electrónicamente por: RMUNIVEC el 22-12- 2023 16:27:41

Código documento Trilce: TRI - 0698403





Declaratoria de Originalidad de los Autores

Nosotros, MENDEZ REYES RUBI KATERINE, URQUIA OCHAVANO ANAYCA NAYRA estudiantes de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA AMBIENTAL de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, declaramos bajo juramento que todos los datos e información que acompañan la Tesis titulada: "Evaluación de la contaminación y calidad del aire por material particulado vinculado a las variables meteorológicas en Lima Metropolitana entre 2017-2020", es de nuestra autoría, por lo tanto, declaramos que la Tesis:

1. No ha sido plagiada ni total, ni parcialmente.
2. Hemos mencionado todas las fuentes empleadas, identificando correctamente toda cita textual o de paráfrasis proveniente de otras fuentes.
3. No ha sido publicada, ni presentada anteriormente para la obtención de otro grado académico o título profesional.
4. Los datos presentados en los resultados no han sido falseados, ni duplicados, ni copiados.

En tal sentido asumimos la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de la información aportada, por lo cual nos sometemos a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

Nombres y Apellidos	Firma
RUBI KATERINE MENDEZ REYES DNI: 73698673 ORCID: 0000-0002-3205-5386	Firmado electrónicamente por: RMENDEZRE el 15-12-2023 20:00:46
ANAYCA NAYRA URQUIA OCHAVANO DNI: 77228488 ORCID: 0000-0002-1804-4912	Firmado electrónicamente por: AURQUIAO el 15-12-2023 10:59:42

Código documento Trilce: TRI - 0698404

ÍNDICE DE CONTENIDOS

Carátula	i
Dedicatoria.....	ii
Agradecimiento	iii
Declaratoria de Autenticidad del asesor	iv
Declaratoria de Originalidad de los Autores	v
Índice de contenidos.....	vi
Índice de tablas.....	vii
Índice de figuras	viii
Resumen	ix
Abstract.....	x
I. INTRODUCCIÓN	1
II. MARCO TEÓRICO	5
III. METODOLOGÍA.....	16
3.1 Tipo y diseño de investigación	16
3.2 Variables y operacionalización	16
3.3 Población, muestra y muestreo	19
3.4 Técnicas e instrumentos de recolección de datos	20
3.5 Procedimientos	21
3.6 Métodos de análisis de datos.....	22
3.7 Aspectos éticos	34
IV. RESULTADOS	35
4.1. Evaluación de la concentración de partículas.....	35
4.2. Distribución espacial de la concentración de partículas	37
4.3. Evolución horaria de la calidad del aire según el INCA	40
4.4. Material particulado relacionado a los parámetros meteorológicos.....	43
V. DISCUSIÓN.....	51
VI. CONCLUSIONES.....	55
VII. RECOMENDACIONES	56
REFERENCIAS.....	57
ANEXOS.....	68

Índice de tablas

Tabla 1. Estaciones de monitoreo del SENAMHI en Lima Metropolitana.	20
Tabla 2. Test de Kolmogorov-Smirnov para material particulado y los parámetros meteorológicos. ...	23
Tabla 3. Asimetría, curtosis y test de Jarque-Bera (JB) para material particulado y parámetros meteorológicos	24
Tabla 4. Modelos Aditivos Generalizados para PM _{2.5} y parámetros meteorológicos	27
Tabla 5. AIC de los modelos para PM _{2.5} y los parámetros meteorológicos.....	27
Tabla 6. Resumen del modelo de la familia gaussiana con enlace identidad para PM _{2.5} y parámetros meteorológicos	28
Tabla 7. Modelos Aditivos Generalizados para PM ₁₀ y parámetros meteorológicos	29
Tabla 8. AIC de los modelos para PM ₁₀ y parámetros meteorológicos	29
Tabla 9. Resumen del modelo de la familia gaussiana con enlace identidad para PM ₁₀ y parámetros meteorológicos	30
Tabla 10. Resumen de los modelos de predicción para la distribución espacial del PM _{2.5} y PM ₁₀	33
Tabla 11. Desviación estándar, error estándar y coeficiente de determinación (R ²) de los modelos de predicción para la distribución espacial del PM _{2.5} y PM ₁₀	34
Tabla 12. Número de veces excedido el ECA para la concentración de PM _{2.5} en un periodo de 24 horas, en Lima Metropolitana 2017-2020.....	35
Tabla 13. Concentración promedio anual para PM _{2.5} , en Lima Metropolitana entre 2017-2020.	35
Tabla 14. Número de veces excedido el ECA para la concentración de PM ₁₀ durante un periodo de 24 horas, en Lima Metropolitana 2017-2020.....	36
Tabla 15. Concentración promedio anual para PM ₁₀ , en las estaciones del SENAMHI de Lima Metropolitana entre 2017-2020.	37
Tabla 16. Área donde la concentración promedio de PM ₁₀ fue mayor al ECA anual en Lima Metropolitana.	38
Tabla 17. Área donde la concentración promedio de PM _{2.5} fue mayor al ECA anual en las estaciones del SENAMHI de Lima Metropolitana.	39
Tabla 18. Calidad del aire según el INCA para la concentración de PM _{2.5} , en Lima Metropolitana entre 2017-2020.	41
Tabla 19. Calidad del aire según el INCA para la concentración de PM ₁₀ , en las estaciones del SENAMHI de Lima Metropolitana entre 2017-2020.....	42
Tabla 20. Resumen de la validación cruzada para la poda del árbol de PM _{2.5} y los parámetros meteorológicos	45
Tabla 21. Validación cruzada para la poda del árbol de regresión entre PM _{2.5} y los parámetros meteorológicos según el INCA “buena” y “moderada”	45
Tabla 22. Validación cruzada para la poda del árbol entre PM _{2.5} y los parámetros meteorológicos según el INCA “mala” y “VUEC”	46
Tabla 39. Resumen de la validación cruzada para la poda del árbol de PM ₁₀	49
Tabla 40. Resumen de la validación cruzada para la poda del árbol de PM ₁₀	49

Índice de figuras

Figura 1. Modelos para el ajuste del variograma teórico	14
Figura 2. Estaciones meteorológicas del SENAMHI en Lima Metropolitana.	20
Figura 3. Análisis de normalidad y homocedasticidad de varianzas del modelo lineal: PM_{10} (arriba) y $PM_{2.5}$ (abajo) respecto a las variables meteorológicas.	24
Figura 4. Coeficiente de Pearson entre las partículas y parámetros meteorológicos, de las estaciones del SENAMHI de Lima Metropolitana entre el 2017 y 2020	25
Figura 5. Coeficiente de Spearman entre las partículas y parámetros meteorológicos, de las estaciones del SENAMHI de Lima Metropolitana entre el 2017 y 2020	25
Figura 6. Coeficiente Pearson entre material particulado y los parámetros meteorológicos con observaciones por horas (derecha) y promedios mensuales (izquierda).	26
Figura 7. Gráficos de los cuantiles normalizados (Normal Q-Q), histograma de los residuos y probabilidad normal de los residuales (Fitted values vs. Index)	28
Figura 8. GAM de la concentración de $PM_{2.5}$ en relación a los parámetros meteorológicos en las estaciones de Lima Metropolitana 2017-2020	29
Figura 9. Gráficos de los cuantiles normalizados (Normal Q-Q), histograma de los residuos y probabilidad normal de los residuales (Fitted values vs. Index)	30
Figura 10. GAM de la concentración de PM_{10} en relación a los parámetros meteorológicos en las estaciones de Lima Metropolitana 2017-2020	31
Figura 11. Estructura del árbol de regresión	32
Figura 12. Variograma experimental y teórico de los datos de $PM_{2.5}$ y PM_{10} , para los métodos de interpolación co-kriging ordinario y kriging ordinario	33
Figura 13. Número de veces excedido el ECA para la concentración de $PM_{2.5}$ en 24 horas en Lima Metropolitana 2017-2020	35
Figura 14. Concentración promedio y ECA anual para $PM_{2.5}$, en las estaciones de monitoreo del SENAMHI de Lima Metropolitana entre 2017-2020	36
Figura 15. Número de veces excedido el ECA para la concentración de PM_{10} en 24 horas en Lima Metropolitana 2017-2020	36
Figura 16. Concentración promedio anual y ECA anual para la concentración de PM_{10} , en Lima Metropolitana entre 2017-2020	37
Figura 17. Distribución espacial de la concentración de $PM_{2.5}$ con el método de interpolación kriging ordinario en Lima Metropolitana entre 2017-2020	38
Figura 18. Distribución espacial de la concentración de PM_{10} con el método de interpolación co-kriging ordinario en Lima Metropolitana entre 2017-2020	39
Figura 19. Calidad del aire según el INCA para la concentración de $PM_{2.5}$, las estaciones de SENAMHI en Lima Metropolitana entre 2017-2020	40
Figura 20. Calidad del aire según el INCA para la concentración de PM_{10} , las estaciones de SENAMHI en Lima Metropolitana entre 2017-2020	42
Figura 21. Modelos aditivos generalizados de la relación entre $PM_{2.5}$ y los parámetros meteorológicos en Lima Metropolitana entre el 2017 y 2020	43
Figura 22. Árbol sin podar de la concentración de $PM_{2.5}$ en relación a los parámetros meteorológicos	44
Figura 23. Árboles de regresión entre $PM_{2.5}$ y los parámetros meteorológicos según la calificación del INCA 2017-2020	46
Figura 24. Modelos aditivos generalizados de la relación entre el PM_{10} y los parámetros meteorológicos en Lima Metropolitana entre el 2017 y 2020	47
Figura 25. Árbol sin podar de la concentración de PM_{10} en relación a los parámetros meteorológicos	48
Figura 26. Árbol de regresión entre PM_{10} y los parámetros meteorológicos en las estaciones de Lima Metropolitana	50

Resumen

El objetivo general de la investigación fue evaluar la contaminación, la calidad del aire por material particulado y su relación con los parámetros meteorológicos en Lima Metropolitana entre 2017-2020. La investigación fue aplicada con diseño no experimental y de tipo longitudinal-correlacional. Se colectaron datos de la concentración de material particulado del SENAMHI, siendo analizados a través de series temporales, distribución espacial y regresiones no paramétricas. Los resultados demuestran que la concentración del $PM_{2.5}$ y PM_{10} sobrepasó el ECA en Puente Piedra, Carabayllo, Ate, Santa Anita y Villa María del Triunfo, entre 01 am-12 pm y 5 pm-12 am se presentó la calidad del aire “mala” y “valor umbral de cuidado”. La distribución espacial de la concentración mayor al ECA sobrepasó el 60% del área evaluada. La relación del material particulado con las variables meteorológicas es no lineal y su dispersión es favorecida por la velocidad de viento >3.4 m/s, humedad relativa $>77\%$ y temperatura del aire $<19^{\circ}C$. Se concluye que la concentración de partículas supera el ECA y la calidad del aire es insalubre en la mayoría de distritos de Lima Metropolitana; su acumulación o dispersión depende de los parámetros meteorológicos predominantes en cada distrito.

Palabras clave: Material particulado, contaminación, calidad del aire, parámetros meteorológicos.

Abstract

The general objective of the research was to evaluate pollution, air quality due to particulate matter and its relationship with meteorological parameters in Metropolitan Lima between 2017-2020. The research was applied with a non-experimental and longitudinal-correlational design. Data on the concentration of particulate matter was collected from SENAMHI, and analyzed through time series, spatial distribution and non-parametric regression. The results show that the concentration of PM_{2.5} and PM₁₀ exceeded the ECA in Villa María del Triunfo, Ate, Puente Piedra and Santa Anita, between 01 am-12 pm and 5 pm-12 am the “poor” air quality occurred. ” and “care threshold value”. The spatial distribution of the concentration greater than the ECA exceeded 60% of the evaluated area. The relationship of particulate matter with meteorological variables is non-linear and its dispersion has been favored by wind speed >3.4 m/s, relative humidity >77% and air temperature <19°C. It is concluded that the concentration of atmospheric particles is greater than the ECA and the air quality is not optimal in most of the districts of Metropolitan Lima; Its accumulation or dispersion depends directly on the prevailing meteorological parameters in each district.

Keywords: Particulate matter, pollution, air quality, meteorological parameters.

I. INTRODUCCIÓN

El efecto de la polución atmosférica en la salud humana es significativo (Fotourehchi, 2016). La incidencia de los picos en la concentración de los contaminantes del aire ha aumentado en frecuencia durante la última década (Cohen et al., 2017; Shukla et al., 2017). Alrededor de 800,000 muertes y la pérdida de 4.6 millones de años de vida, son provocadas anualmente por la contaminación del aire (OMS, 2016). Alrededor del 91% de la población mundial reside en zonas donde los contaminantes exceden los estándares de calidad ambiental (ECA) del aire, la contaminación por partículas es un riesgo para la salud principalmente de los grupos sensibles (personas asmáticas o que padecen de cardiopatías) y para la población en general cuando estas concentraciones son excesivas respecto a los límites establecidos en los ECA.

Como realidad problemática tenemos que el estudio y evaluación de la contaminación del aire ha sido el propósito de numerosas investigaciones publicadas en los repositorios de las universidades nacionales; la mayoría de dichas publicaciones se han limitado a la ejecución de monitoreos de los contaminantes, revisiones metodológicas de los índices de la calidad del aire, análisis comparativos de la concentración de los contaminantes y variaciones de la calidad del aire en zonas urbanas, rurales y diversas regiones del Perú. La metodología frecuentemente empleada ha sido el análisis estadístico y la comparación del promedio de la concentración por horas, días de la semana mensual o anual de los contaminantes atmosféricos entre dos o más estaciones de monitoreo y por regiones con el ECA del aire.

Cabe destacar en estas investigaciones, la aplicación deficiente del límite del ECA para 24 horas y el criterio de evaluación de no excederlo siete veces durante un año, establecido por el D.S. N°003-2017-MINAM. Asimismo, para determinar la relación entre los contaminantes del aire y los parámetros ambientales se utilizaron los modelos estadísticos tradicionales, como el coeficiente de Pearson y los Modelos Lineales Generalizados (GML). Según Robles (2020) las pruebas de Shapiro-Wilk y Kolmogorov-Smirnov demostraron que dichas variables no presentan distribución normal, recurriendo a técnicas

estadísticas no paramétricas como el coeficiente de correlación de Spearman, ya que el coeficiente de Pearson solo mide relaciones lineales (Hernández et al., 2018). Sin embargo, el coeficiente de Spearman para datos ordinales solo permite medir relaciones monótonas. Por lo tanto, a pesar que estos coeficientes sean nulos, puede existir una relación significativa entre las variables (Minitab, 2023).

El uso correcto de las diversas herramientas estadísticas es muy importante para la interpretación correcta de los datos, las variables ambientales rara vez cumplen con los supuestos de los modelos lineales, lo que puede provocar resultados no completamente ciertos o que expliquen un menor porcentaje de variación que a través de herramientas más flexibles (Ravindra et al., 2019). Los modelos estadísticos tradicionales no son adecuados para explicar la relación entre las variables bióticas y abióticas del medio ambiente; los parámetros meteorológicos o atmosféricos presentan una alta variabilidad temporal: horaria, diaria, mensual, estacional e interanual. Por lo tanto, no presentan distribución normal y el uso de promedios puede producir un sesgo o error de subestimación o sobreestimación de los valores.

Esta investigación presenta información de la contaminación y calidad del aire por partículas a las personas que visitan, laboran o habitan en los distritos en Lima Metropolitana; que tan excesiva ha sido la concentración de material particulado respecto a los límites establecidos por el ECA y los horarios donde la calidad del aire es óptima o insalubre, lo que puede afectar a la población en general. Desde el punto de vista académico, se difunde a los estudiantes de la carrera de Ingeniería Ambiental, la aplicación del programa R en el procesamiento masivo de datos, la construcción de series temporales, el análisis espacial a través de la interpolación y métodos estadísticos más sofisticados como los modelos aditivos generalizados y árboles de decisión que permiten mejorar la capacidad de interpretación de los datos ambientales.

El problema general fue ¿Cuál es la contaminación y la calidad del aire por partículas y como influyen los parámetros meteorológicos en Lima Metropolitana?.

Los problemas específicos fueron:

1. ¿Cómo es la contaminación del aire por partículas en Lima Metropolitana?
2. ¿Cuál es la distribución geográfica de la concentración de partículas que excede el límite del ECA en Lima Metropolitana?
3. ¿Cómo es la calidad del aire horaria respecto a la concentración de partículas en Lima Metropolitana?
4. ¿Cómo influyen los parámetros meteorológicos en la concentración de partículas en Lima Metropolitana?

El objetivo general fue evaluar la contaminación y la calidad del aire por material particulado, así como su relación con los parámetros meteorológicos en Lima Metropolitana entre el 2017 y 2020.

Los objetivos específicos fueron:

1. Evaluar la concentración de material particulado con diámetro menor de 2.5 micrómetros ($PM_{2.5}$) y menor de 10 micrómetros (PM_{10}), conforme a los valores y criterios de los ECA para periodos de 24 horas y anual, en los distritos de Lima Metropolitana entre el 2017 y 2020.
2. Determinar la distribución espacial de la concentración del $PM_{2.5}$ y PM_{10} , así como la superficie (km^2) donde se excedió el ECA y su proporción respecto del área total comprendida por las estaciones de monitoreo del SENAMHI en Lima Metropolitana, entre el 2017 y 2020.
3. Caracterizar la calidad del aire, su evolución horaria conforme a la calificación del INCA para la concentración de $PM_{2.5}$ y PM_{10} , e identificar los distritos donde es mayor la exposición a la calidad insalubre en Lima Metropolitana entre el 2017 y 2020.
4. Identificar las condiciones que favorecen la acumulación o dispersión del $PM_{2.5}$ y PM_{10} , así como la correlación con los parámetros meteorológicos: temperatura del aire, humedad relativa, velocidad del viento y dirección del viento, con modelos aditivos generalizados y árboles de regresión como alternativa a los coeficientes de Pearson y Spearman, en los distritos de Lima Metropolitana entre el 2017 y 2020.

La hipótesis general fue: “La contaminación y calidad del aire respecto al material particulado ha sido aceptable y óptima según el ECA e INCA, y estas concentraciones son influidas por los parámetros meteorológicos en los distritos de Lima Metropolitana entre el 2017 y 2020”.

Las hipótesis específicas fueron:

1. La concentración de partículas según el ECA es aceptable: para $PM_{2.5}$ en 24 horas es menor o igual a $50 \mu\text{g}/\text{m}^3$ y anualmente menor o igual a $25 \mu\text{g}/\text{m}^3$. Para PM_{10} en 24 horas es menor o igual a $100 \mu\text{g}/\text{m}^3$ y anualmente menor o igual a $50 \mu\text{g}/\text{m}^3$, en los distritos de Lima Metropolitana entre el 2017 y 2020.
2. El área de la distribución geográfica de la concentración de $PM_{2.5}$ y PM_{10} mayores al ECA anual de $25 \mu\text{g}/\text{m}^3$ y $50 \mu\text{g}/\text{m}^3$ respectivamente, es menor al 50% del área comprendida por las estaciones de monitoreo del SENAMHI en Lima Metropolitana entre el 2017 y 2020.
3. La evolución horaria de la calidad del aire según la concentración del $PM_{2.5}$ y PM_{10} es menor o igual al valor 100 del INCA, que es considerado óptimo con una calidad de aire “buena” o “moderada” en los distritos de Lima Metropolitana entre el 2017 y 2020.
4. Los parámetros meteorológicos influyen en la acumulación y dispersión de las partículas en suspensión en los distritos de Lima Metropolitana entre el 2017 y 2020.

II. MARCO TEÓRICO

En los **antecedentes nacionales**, las publicaciones más recientes dirigidas a la evaluación de la contaminación del aire relacionado con las variaciones climáticas son las siguientes:

Rojas et al. (2022) simularon numéricamente la concentración del PM₁₀ y PM_{2.5} vinculado con las condiciones atmosféricas en Lima Metropolitana, utilizando el programa de código abierto CCATT-BRAMS (INPE/CPTEC Brasil). Las concentraciones simuladas se obtuvieron mediante el software CCATT y las variables atmosféricas del modelo numérico BRAMS. Los resultados indican la influencia del incremento de la velocidad del viento y la humedad relativa sobre la disminución del PM₁₀ tanto en verano como en invierno.

Robles (2020) investigó la relación que existe entre el PM₁₀ y las infecciones respiratorias agudas (IRAS), en Cerro de Pasco entre el 2010 y 2016; cuantificando la concentración de partículas en las estaciones de monitoreo y las atenciones hospitalarias durante la estación de clima seco. Determinó que los valores de material particulado no tienen distribución normal recurriendo al coeficiente de correlación de Spearman. Según los resultados existe relación significativa entre el PM₁₀ y las IRAS en la ciudad de Cerro de Pasco entre 2010-2016.

Rojas y Medina (2022) pronosticaron las concentraciones de PM₁₀ con Neuronales Artificiales (RNA) para el distrito de Ate, usando datos simulados del modelo WRF-CHEM e información de los contaminantes del aire y variables meteorológicas obtenidas de la Estación de Monitoreo de Ate. Las distintas arquitecturas de RNA se evaluaron mediante el coeficiente de determinación (R^2), Error Cuadrático Medio (ECM) y la precisión (BIAS). Se concluyó que la arquitectura más óptima para la estimación de la concentración del PM₁₀ en Ate, posee 19 neuronas en la capa oculta, con valores de 0,823 para el R^2 , 0,0230 para el ECM y 0,5308 para la BIAS.

Pacsie y Murriel (2018) investigaron las fluctuación temporal y espacial del PM_{2.5}, en el campus de la Universidad Nacional Agraria La Molina y la relación con las condiciones meteorológicas en el período julio–noviembre del 2016, se

colectaron muestras en cinco puntos de monitoreo y en el análisis de estadístico se aplicó el coeficiente de Spearman. Se concluyó que los valores promedios que superaron el ECA de aire en el 21 % de los días muestreados y el 98,5% el valor guía de la OMS. Estos valores promedios son mayores en el invierno austral con respecto a la primavera y tienen correlación positiva con la humedad relativa y negativa con la temperatura del aire, radiación solar y velocidad del viento.

Copes (2022) estudió la turbulencia atmosférica para determinar el efecto en la dispersión del material particulado entre enero y marzo 2019, caracterizando el comportamiento horario de la estabilidad atmosférica y de la dispersión del material particulado para las 7, 11, 18 horas, utilizando un modelo gaussiano. De los resultados se concluyó que el material particulado incrementa su concentración de acuerdo al grado de estabilidad atmosférica y la distancia.

Fernández et al. (2014), evaluaron la eficiencia estadística del modelado de la climatología de la precipitación mensual entre 1981 a 2010, aplicando con ArcGIS los métodos de interpolación: Kriging Ordinario, Co-Kriging (Co-Ko), Kriging con deriva externa (KED) y ponderación de distancia inversa (IDW); determinando que Co-Ko usando como covariable la elevación, es el mejor método para caracterizar la distribución espacial de la precipitación.

En los **antecedentes internacionales**, el estudio del efecto adverso y la exposición-respuesta de la contaminación por partículas y gases asociado a las variaciones climatológicas, ha sido el interés de innumerables publicaciones científicas en disciplinas como la salud pública, el turismo, la ecología y el ambiente.

Yoon (2019), investigó la relación entre los ingresos minoristas y el nivel de PM₁₀, en el contexto de Seúl, Corea del Sur durante aproximadamente dos años, del 2015 al 2017, utilizando el modelo aditivo generalizado (GAM) de series temporales. Esta investigación fue el primer intento de descifrar el efecto de PM en el turismo y la industria minorista recreativa.

Wang et al. (2019), en su investigación del impacto de las partículas ambientales en las visitas ambulatorias al hospital por enfermedades del sistema circulatorio

y respiratorio en una población urbana china, utilizó una combinación de modelo aditivo generalizado logístico (GAM) para evaluar las asociaciones exposición-respuesta. Los resultados demuestran que la asociación exposición-respuesta fue aproximadamente lineal para la concentración de PM entre 0-150 $\mu\text{g}/\text{m}^3$ y no lineal en todo el rango de exposiciones. Los efectos de PM en la enfermedad del sistema respiratorio y enfermedad del sistema cardio-cerebrovascular fueron sensibles al ajuste adicional por co-contaminantes, lo que indica las consecuencias en la salud resultado de la mezcla del aire contaminado en la ciudad de Nanjing.

Ravindra et al. (2019), efectuó una revisión de la aplicación de modelos aditivos generalizados (GAM) dentro de un campo variado, centrándose en el medio ambiente y los datos meteorológicos, discutiendo además las ventajas y complicaciones al aplicar GAM a los datos ambientales. Los resultados demuestran que el GAM se ha utilizado como una herramienta estadística importante durante décadas para medir las consecuencias de la contaminación del aire. Para analizar los efectos a corto plazo, el GAM con spline penalizado permite un mejor enfoque al estudio de la relación entre medio ambiente, clima y salud, siendo un método sólido para los conjuntos de datos ambientales.

Analitis et al. (2020), en su investigación presentó un modelo para la predicción del $\text{PM}_{2.5}$ en las estaciones de monitoreo de PM_{10} y NO_x en Londres, usando modelos generalizados y métodos de aprendizaje automático. Los resultados propusieron una metodología para mejorar la base de datos de mediciones de $\text{PM}_{2.5}$ para su uso en estudios de salud y aplicarla en el caso de Londres. La combinación con GAM funcionó mejor que las dos predicciones separadas del modelo, lo que demuestra que la combinación de métodos de aprendizaje automático y los modelos de regresión más tradicionales conduce a predicciones mejoradas en todo el rango de concentraciones.

Shukla et al. (2020), efectuó una investigación con el objetivo de mapear el $\text{PM}_{2.5}$ utilizando la interpolación espacial Kriging y ponderación de distancia inversa (IDW), para predecir sus concentraciones en distintas ubicaciones no monitoreadas en la ciudad de Delhi. La primera fase del estudio comprendió la predicción de las concentraciones con datos entre 2015–2017. En la segunda

fase, se obtuvieron proporciones de mezcla de contaminantes en 17 estaciones de monitoreo durante cuatro meses de invierno entre noviembre de 2017 y febrero de 2018. Los resultados indican que, en ambas fases el error promedio de Kriging e IDW fue 22 % y 24 %, respectivamente. La magnitud del cambio en la concentración diaria fue relativamente insignificante y se pudo identificar la tendencia anual.

Garrido et al. (2021), evaluaron los índices de estancamiento del aire (ASI) modelando la variabilidad de PM_{10} en Europa, con modelos aditivos generalizados en el módulo pyGAM de Python (Servén y Brummitt, 2018). Los índices de estancamiento del aire (ASI) se definen en función de diferentes variables meteorológicas y su aplicabilidad puede depender de la ubicación geográfica, permitiendo identificar condiciones favorables para la acumulación de contaminantes en las capas bajas de la atmósfera. Los resultados indican que en verano es particularmente relevante la circulación a gran escala, explicando los diferentes comportamientos del ASI durante esta estación. También identificaron algunas mejoras a ser aplicadas en dos de los ASI; sin embargo, los tres ASI se basan en umbrales fijos y no pueden manejar relaciones no lineales, lo que limita su capacidad para explicar la conversión de PM_{10} .

Wang et al. (2021), investigaron el impacto del monóxido de carbono en los años de vida perdidos y el efecto modificado por las características a nivel individual y de la ciudad, con un estudio de serie temporal a nivel nacional en China utilizando modelos aditivos generalizados con distribución de la familia gaussiana. Los resultados proporcionaron un panorama completo de las asociaciones entre la exposición a corto plazo a la contaminación por CO y los años de vida perdidos debido a todas las causas no accidentales o específicas en China. Estos hallazgos de vulnerabilidad a nivel de ciudad y características a nivel individual pueden ser significativos para mejorar las prácticas clínicas y de salud pública, y la formulación de políticas ambientales y sociales para reducir la carga de enfermedad asociada con la contaminación por CO, incluso en baja concentración.

Lequy et al. (2022), modelizaron la exposición a metales en el aire mediante el bio-monitoreo de musgo en cementerios de dos áreas urbanas alrededor de

París y Lyon en Francia, aplicando el coeficiente de correlación de Spearman y modelos aditivos generalizados considerando cualquier relación no lineal al incluir funciones spline para todas las variables continuas. Los resultados indican que el bio-monitoreo de musgo en cementerios es una alternativa válida para la estimación rápida de la exposición a metales en el aire en áreas urbanas, y la regresión del uso de la tierra proporcionó mapas de concentración confiables para los metales en el aire emitidos por fuentes humanas.

Tudor (2022), investigó la contaminación por ozono, características espacio-temporales, tendencias, transporte e impacto de las medidas de control de la COVID-19 en Londres y Edimburgo; utilizando los modelos aditivos generalizados (Wood, 2017) que, debido a su mayor flexibilidad en el análisis de datos no normales, han sido aplicados cada vez más en la investigación ambiental (Jacob y Winner, 2009; Ravindra et al., 2019). Los resultados confirman la tendencia decreciente entre el 2016-2022 de todos los contaminantes en Londres y Edimburgo, a excepción del ozono (O_3), que presentó tendencia ascendente en Londres y ascendente más leve en Edimburgo. Durante el período pandémico, las concentraciones de O_3 muestran una tendencia negativa significativa, a pesar que en la primera ola pandémica (marzo-junio de 2020), estas concentraciones aumentaron significativamente por hora, día y mes.

Ramírez-Cando et al. (2018), analizaron la distribución espacial del PM_{10} en el aeropuerto de Tababela del Distrito Metropolitano de Quito, con un diseño de muestreo lineal formado por 16 estaciones o puntos ubicados de sur a norte; en ArcGIS aplicaron tres métodos de interpolación: IDW, Kriging y Spline; determinando que el método IDW es más eficiente en el modelado del PM_{10} por su proximidad a la concentración real y la visualización didáctica de los mapas generados.

Haro et al. (2019) aplicaron la técnica de árbol de clasificación para identificar variables meteorológicas predominantes a ocho intervalos de tiempo; mediante un script en el software estadístico R; los datos correspondieron a información horaria del año 2016 de las variables temperatura del aire, humedad relativa, velocidad de viento, presión barométrica, radiación solar difusa, radiación solar

global, temperatura del suelo. Se concluyó que el árbol identificó variables meteorológicas relevantes en determinadas horas, la radiación solar global mayor o igual a 120 w/m^2 entre las 06h00 a 08h00 permite predecir la presión barométrica de 09h00 a 11h00 y cuando es mayor o igual de 709 w/m^2 se puede pronosticar la temperatura del aire.

Las emisiones de **material particulado** consisten en sustancias líquidas o sólidas suspendidas en la atmósfera y son la principal consecuencia del aumento en el parque automotor (Suárez, 2017). Las **partículas con diámetro inferior de 10 micrómetros** representan entre 30% y 60% de total de emisiones de material particulado. El PM_{10} , al acumularse en los pulmones y contribuir al deterioro de la función pulmonar afectando a la salud, acrecentando el asma y dañando el tejido pulmonar (Feng et al., 2019; De Rooij et al. 2019). Cabe destacar el peligro para la salud de las **partículas con diámetro inferior de 2.5 micrómetros**, que pueden penetrar muy profundamente y con mucha facilidad en el pulmón, produciendo inflamación, incrementando la mortalidad cardiovascular y respiratoria (Gong, 2023). Además, la toxicidad de la composición de las partículas puede variar de acuerdo al tipo de industria que las emita, particularmente los metales como: Plomo (Pb), Vanadio (V), Titanio (Ti), Níquel (Ni), Cromo (Cr), etc. La exposición a material particulado puede provocar hiperreactividad bronquial incrementando la virulencia de las infecciones respiratorias y como consecuencia el incremento en los índices de mortalidad (Dong, 2021).

El **Estándar de calidad ambiental (ECA)** mide el nivel de concentración de sustancias, elementos o parámetros físicos, químicos y biológicos presentes en el suelo, agua y aire que no representa peligro para el ambiente ni la salud de las personas (SENAMHI, 2021). Por lo cual, las concentraciones de cada uno de los contaminantes del aire no deben exceder su respectivo ECA (D.S. N° 003-2017-MINAM).

El **Índice de la calidad del aire (INCA)**, muestra el estado de la calidad del aire en valores y colores para facilitar que se conozcan los niveles de exposición a determinados contaminantes. El INCA del aire ha sido establecido a través de la

R.M. N°181-2016-MINAM. La determinación matemática del INCA, se basa en la relación entre el valor del ECA para cada caso y su respectivo valor de la concentración del contaminante.

El **análisis de regresión** es una técnica estadística para comprender la dependencia entre dos o más variables; estos modelos estadísticos han evolucionado desde el modelo lineal como el método de mínimos cuadrados y la correlación lineal, el modelo lineal generalizado, hasta los modelos de regresión no paramétrica donde el predictor no tiene una forma predeterminada, como los árboles de decisión, los bosques aleatorios, el modelo aditivo generalizado y el modelo aditivo generalizado de posición, escala y forma.

Los Modelos Aditivos Generalizados (GAM) fueron introducido por Hastie y Tibshirani en 1990 (Hastie, 2023), que incorpora formas no lineales de los predictores (splines, polinomios, o funciones step, etc.). A comparación de los Modelos Lineales Generalizados, donde la relación entre la variable respuesta y la explicativa está restringida a la forma lineal, en los GAM las funciones pueden asumir cualquiera forma funcional, lo cual aporta mayor información respecto a la relación entre la variable objetivo o dependiente y variable explicativa o independiente (Wood, 2017). El GAM es una modificación del Modelo Aditivo que permite relacionar una variable respuesta Y , que tenga cualquier distribución de la Familia Exponencial (gaussian, gamma, poisson, binomial, etc.), y varias variables explicativas a través del ajuste de funciones suaves y no necesariamente lineales sobre cada variable explicativa de manera individual. La estructura del modelo es:

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}) + \dots, \quad i = 1, 2, \dots, n$$

Este modelo permite el ajuste sobre la variable explicativa X_j de una función f_j no lineal, para modelizar relaciones no lineales entre Y y X_1, X_2, \dots, X_p , lo que no es posible en la regresión lineal. Con estos ajustes no lineales se pueden obtener predicciones más precisas de la variable respuesta (Wood, 2017).

Los GAM tienen muchos tipos de suavizadores, siendo los más utilizados los splines penalizados de regresión ya que se estiman como parte del ajuste del modelo (Wood, 2017). Los splines penalizados de regresión se expresan por:

$$\|Y - X\beta\|^2 = \lambda \int f''(x)^2 dx$$

Dónde $(\lambda \int f''(x)^2 dx)$ es una penalización, por esto se denominan spline penalizados y miden el grado de “rugosidad” de la función (Ravindra, 2019). El parámetro del suavizado (λ) determina el equilibrio entre la precisión y el suavizado, λ es afectado por la función de suavizado f que indica lo suave que es la curva. Un valor alto de f'' representa una curva muy flexible y es altamente no-lineal, mientras que un valor de cero resulta en una línea recta. Para obtener un equilibrio entre suavizado óptimo y baja penalización, Wood (2017) efectuó modificaciones en el GAM con el GCV para la estimación de los parámetros de suavizado y un término denominado grados efectivos de libertad (edf) con un valor entre 0 e infinito, un edf alto indica que f es menos lineal.

Los árboles de decisión son métodos utilizados frecuentemente por la comunidad científica debido a su fácil implementación e interpretación de los resultados, que permiten manipular variables numéricas y categóricas. Sus principales ventajas son su interpretabilidad, la robustez a outliers o valores atípicos y la invarianza a transformaciones monótonas o no lineales de las variables. Los árboles de regresión exploran la relación entre una variable respuesta continua y múltiples variables predictoras que pueden ser discretas o continuas (Cordova, 2019). Su organización es fácil de interpretar y conceptualizar, aspecto relevante para la construcción de modelos de hábitat (Flores et al., 2016). Así pueden permitir hallar interacciones entre las variables que otros métodos obvian (Schiltz et al., 2018; Gao, 2021). Su desempeño puede ser superior a los métodos paramétricos clásicos cuando la verdadera forma funcional entre la variable respuesta y variables predictoras es altamente no lineal y compleja (Cordova, 2019).

Una **serie temporal** también llamada serie de tiempo o serie cronológica, es un conjunto de observaciones de una variable ordenadas de forma secuencial en el

tiempo (Gallardo et al., 2019). En una serie temporal los valores están relacionados a instantes de tiempo, de manera que se analizan de manera conjunta dos variables: la variable en estudio y el tiempo. Las series según los periodos de tiempo donde fueron recopilados los datos que la componen, pueden tener periodicidad diaria, mensual, trimestral, semestral, anual, etc.

La distribución geográfica es un método de análisis espacial, que permite la localización precisa y objetiva de ciertos elementos a nivel espacial, a través del tratamiento estadístico de la posición y los valores de una variable en estudio. Los datos georreferenciados o con coordenadas conocidas permiten generar mapas temáticos o mapas de contornos mediante la interpolación de un parámetro o variable en estudio dentro de una zona o área determinada. En el análisis espacio-temporal de los contaminantes del aire, los dos métodos de interpolación espacial empleados para visualizar los contaminantes y predecir los valores en sitios no monitoreados, han sido el kriging ordinario y la ponderación de distancia inversa (Shukla et al., 2020, Kamboj et al., 2022, Jin et al., 2022).

El **kriging** es un método de interpolación, que a partir de una muestra de puntos distribuidos regular o irregularmente, permite estimar valores en lugares donde no hay información y que cuenta con una variancia mínima conocida (Mamani, 2020). Con el Kriging se elige el variograma o semivariograma que permita explicar mejor la variabilidad espacial obteniendo un error de estimación de la cota en la superficie topográfica. Existen dos métodos de kriging: kriging ordinario donde la variación en los valores de elevación está libre de cualquier componente estructural (Drift) y kriging universal donde dicha tendencia de variación espacial está presente en los datos y puede ser ajustada con un modelo lineal o cuadrático (Mamani, 2020). Un variograma o semivariograma es un gráfico de la varianza (semivarianza) y el rango, utilizado para representar la tasa de cambio de una variable en una orientación específica o región. Las principales características del variograma son:

- El nugget o efecto pepita es la discontinuidad del semivariograma en el origen, debida a errores en la escala de la variable o en la medición de la

misma, donde se sigue presentando variabilidad a distancias insignificantes de separación de muestras.

- El sill o meseta se muestra cuando el variograma alcanza un valor constante.
- El rango es la distancia donde la semivarianza deja de aumentar, indicando la distancia donde las muestras son espacialmente independientes unas de otras.

En el ajuste de un variograma experimental a uno teórico, el modelo esférico es utilizado ampliamente ya que corresponde a los fenómenos observados más frecuentemente, tiene un comportamiento con tendencia casi lineal hasta alcanzar el rango en donde se estabiliza el fenómeno. El modelo exponencial, tiene comportamiento muy similar al esférico, pero no alcanza el valor de la meseta al ser su rango práctico del 95%. El modelo gaussiano tiene la forma típica de la curva gaussiana y se utiliza cuando el fenómeno de estudio es extremadamente continuo (Mamani, 2020).

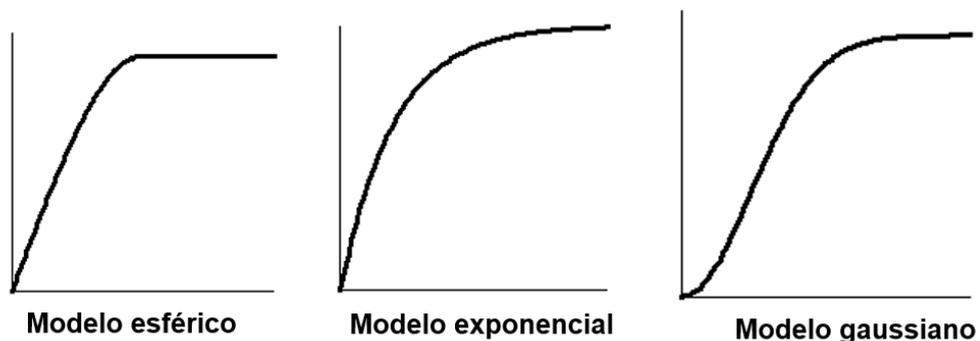


Figura 1. Modelos para el ajuste del variograma teórico

El **co-kriging** es un método de interpolación que utiliza una segunda variable correlacionada densamente muestreada para mejorar la estimación de una variable primaria. Para que el cokriging sea efectivo, las dos variables deben estar correlacionadas y la variable primaria debe estar submuestreada con respecto a la variable secundaria. Existen dos tipos de Cokriging: cokriging simple que no impone restricciones a la suma de los pesos de las variables y el cokriging ordinario que establece la suma de las ponderaciones aplicadas a la variable primaria en uno (1) y la suma de las ponderaciones aplicadas a la variable secundaria en cero (0), tendiendo a limitar severamente la influencia de la variable secundaria (Golden software, 2023).

La **distancia inversa ponderada (IDW)** es un método de interpolación que supone que los valores próximos entre sí son más similares a los más alejados. Para estimar el valor en una posición sin medición, emplea las mediciones alrededor del punto de estimación. IDW asume que la influencia local de cada punto medido disminuye con la distancia; es decir, los valores medidos vecinos a la ubicación de estimación tienen mayor influencia que los valores más alejados (Esri, 2023).

III. METODOLOGÍA

3.1 Tipo y diseño de investigación

Tipo de investigación: Aplicada, donde la finalidad ha sido adquirir conocimiento con la implementación y sistematización de la práctica basada en investigación. El enfoque es cuantitativo, ya que se analizan variables continuas.

Diseño de investigación: No experimental, ya que las variables se observaron en su realidad o medio ambiente y el diseño longitudinal descriptivo y correlacional; la información fue recopilada repetidamente durante un tiempo prolongado y se analizó la dependencia estadística entre las variables.

3.2 Variables y operacionalización

Las variables son de tipo cuantitativas:

- Variable independiente: Concentración de material particulado y parámetros meteorológicos.
- Variable dependiente: Contaminación del aire y Calidad del Aire.

- **Definición conceptual:**

Variables independientes:

- Concentración de material particulado: Conjunto de partículas líquidas y/o sólidas suspendidas en la atmósfera (Suárez, 2017).
- Variables meteorológicas: Es toda característica o propiedad de la atmósfera, que en su conjunto determina el estado del tiempo (a corto plazo) o el clima (a largo plazo). También son conocidas como parámetros meteorológicos (SENAMHI, 2021)

Variables dependientes:

- Contaminación del aire: Es la alteración de las propiedades del aire, que puede afectar nocivamente la vida humana o demás organismos vivos.
- Calidad del Aire: Es la composición del aire y de su idoneidad para determinadas aplicaciones o actividades (INEI, 2016).

- **Definición operacional:**

Variables independientes:

Concentración de material particulado: Se recopiló de la Plataforma Nacional de datos abiertos del estado peruano, donde están publicadas las mediciones de los contaminantes del aire en las estaciones automáticas de monitoreo del SENAMHI localizadas en Lima Metropolitana.

Dimensiones:

- Partículas menores de 2.5 micrómetros y 10 micrómetros: Se clasifican por su tamaño y sus niveles se expresan como número de partículas o concentración de masa por unidad de volumen de aire.

Parámetros meteorológicos: La información se recopiló de la Web del SENAMHI “Datos Hidrometeorológicos a nivel nacional” donde se publican los datos de las estaciones automáticas de monitoreo del SENAMHI.

Dimensiones:

- Variables atmosféricas: El estado del tiempo o de la atmosfera en un sitio determinado es definido por los valores de la temperatura, humedad, presión y vientos:

Temperatura del aire: Cantidad de energía de calor que tiene el aire en un momento específico. Determina la sensación de calor y frío, y se mide a través de termómetros en grados Celsius (°C), Fahrenheit, etc.

Humedad Relativa: Es la relación entre la fracción molar de vapor de agua en un espacio dado y la fracción molar del vapor de agua en su condición de saturación (Tejeda, 2018).

Velocidad del viento: Es la intensidad del desplazamiento del aire en la atmósfera, que son desplazadas desde las zonas de alta presión a las de baja presión. La velocidad del viento se mide con un anemómetro en m/s o km/h.

Dirección del viento: Es la orientación desde la cual fluye el viento, se expresa en grados a partir del norte geográfico.

Variables Dependientes

Contaminación del aire: Los ECA del aire, establecen las concentraciones máximas para los contaminantes del aire.

Dimensiones:

- Estándares de calidad ambiental del aire: Valor máximo para material particulado atmosférico y criterios de evaluación para 24 horas y anual.
- Distribución geográfica: Análisis espacial para generar mapas de contornos de un área determinada, a partir de la concentración de los contaminantes atmosféricos.

Calidad del Aire: El índice que permite informar del estado de la calidad del aire a la población y los probables efectos en la salud.

Dimensiones:

- Índice de calidad del aire: Rangos de concentración de los contaminantes y colores que dan información del estado de la calidad del aire.

- **Indicadores:**

Variables independientes

Concentración de material particulado

- Concentración de masa: $\mu\text{g}/\text{m}^3$

Parámetros meteorológicos

- Variables atmosféricas:
Temperatura del aire: $^{\circ}\text{C}$.
Humedad relativa: %
Velocidad del viento: m/s
Dirección del viento: $^{\circ}$

Variables dependientes

Contaminación del aire

- Estándares de calidad ambiental:
PM_{2.5}: Mayor que $50 \mu\text{g}/\text{m}^3$ en 24 horas

PM_{2.5}: Mayor que 25 µg/m³ anual.

PM₁₀: Mayor que 100 µg/m³ en 24 horas.

PM₁₀: Mayor que 50 µg/m³ anual.

- Distribución geográfica: Área en km²

Calidad del aire

- Estado de calidad del aire:

PM_{2.5}: 0 - 12.5 µg/m³ (Buena).

PM_{2.5}: 12.6 - 25 µg/m³ (Moderada).

PM_{2.5}: 25.1 - 125 µg/m³ (Mala).

PM_{2.5}: >125 µg/m³ (Umbral de cuidado).

PM₁₀: 0 - 75 µg/m³ (Buena).

PM₁₀: 76 - 150 µg/m³ (Moderada).

PM₁₀: 151 - 250 µg/m³ (Mala).

PM₁₀: >250 µg/m³ (Umbral de cuidado).

- **Escala de medición:** Es nominal y de razón, los indicadores de las variables tienen magnitudes físicas o escalares como temperatura, masa, velocidad, tiempo, área, etc.

3.3 Población, muestra y muestreo

Población: Total de datos meteorológicos del SENAMHI.

Muestra: Datos meteorológicos de las estaciones automáticas del SENAMHI en Lima Metropolitana desde enero del 2017 hasta agosto 2020. (Tabla 1, Figura 2).

Tabla 1. Estaciones de monitoreo del SENAMHI en Lima Metropolitana.

ZONA	ESTACIONES DE MONITOREO	DISTRITO	LATITUD	LONGITUD
Lima Norte	Parque ecológico de San Martín de Porres	SMP	12° 0' 33.832" S	77° 5' 8.102" W
	Cerca piscina municipal de Puente Piedra	Puente Piedra	11° 52' 0.596" S	77° 4' 37.815" W
	Piscina Municipal de Carabaylo	Carabaylo	11° 53' 23.871" S	77° 1' 36.972" W
Lima Oeste	Parque Campo de Marte	Jesús María	12° 4' 11.462" S	77° 2' 35.263" W
	Polideportivo Lima Tambo	San Borja	12° 5' 50.185" S	76° 59' 30.891" W
Lima Este	Palacio municipal de Santa Anita	Santa Anita	12° 2' 34.262" S	76° 58' 19.196" W
	Plaza de armas de Ate	Ate	12° 0' 57.862" S	76° 53' 49.724" W
	Oficinas administrativas UCV	San Juan de Lurigancho	11° 58' 53.89" S	76° 59' 57.29" W
Lima Sur	Parque Virgen de Lourdes	Villa María del Triunfo	12° 9' 59.108" S	76° 55' 12.224" W

Nota: Fuente SENAMHI

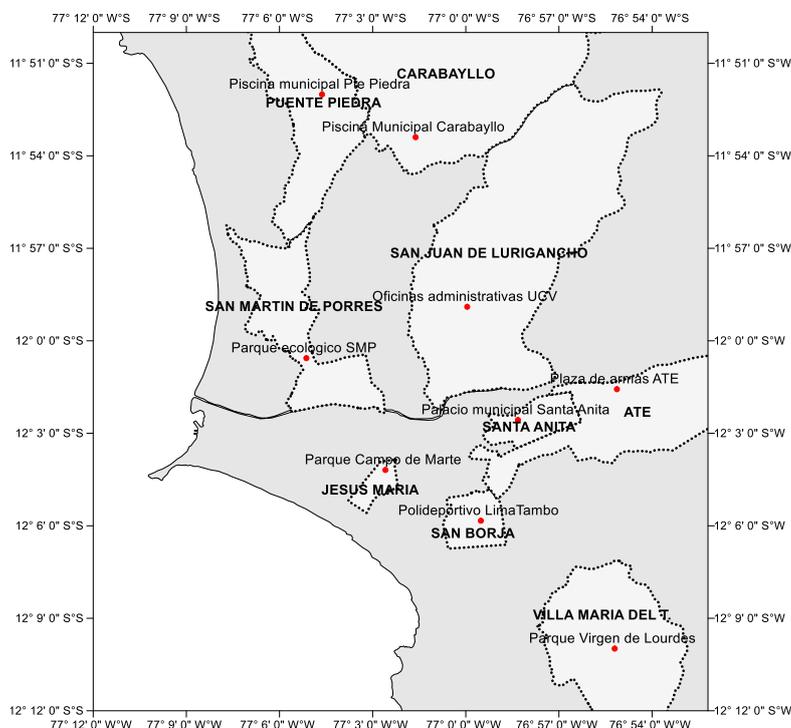


Figura 2. Estaciones meteorológicas del SENAMHI en Lima Metropolitana.

Muestreo: No probabilístico, conformado por puntos de monitoreo llamados “Estaciones Meteorológicas de Monitoreo de la calidad del aire”.

3.4 Técnicas e instrumentos de recolección de datos

Para la colección de datos se recurrió a fuentes abiertas de información pública y gratuita en páginas gubernamentales, se recopiló información de la concentración de partículas y parámetros meteorológicos publicados por el Servicio Nacional de Meteorología e Hidrografía del Perú (SENAMHI).

Se colectaron datos de la Plataforma Nacional de datos abiertos del estado peruano, que permite el acceso a la información pública en formato CSV de los “Datos Horarios de Contaminantes del Aire en Lima Metropolitana”, que tiene información del SENAMHI de los contaminantes del aire entre los años 2010 y 2020. Los parámetros meteorológicos fueron recopilados de la página web del SENAMHI “Datos Hidrometeorológicos a Nivel Nacional”, que permite la descarga de la información pública de las estaciones meteorológicas e hidrológicas a nivel nacional a partir del año 2018.

3.5 Procedimientos

La información del material particulado atmosférico en formato CSV del SENAMHI, fue georreferenciada asignando la posición geográfica a cada estación de monitoreo y ordenada por filas y columnas según el siguiente formato: Fecha, Hora, Estación, Posición geográfica, Contaminante (PM_{2.5} y PM₁₀) y Concentración ($\mu\text{g}/\text{m}^3$).

Los datos de los parámetros meteorológicos de la página Web del SENAMHI “Datos Hidrometeorológicos a Nivel Nacional”, fueron descargados automáticamente y ordenados cronológicamente como tablas en formato XLSX a través de un script en el lenguaje del programa R, de acuerdo al siguiente formato: Estación, Fecha, Hora, Temperatura, Humedad, velocidad del viento, dirección del viento.

El script en lenguaje de R, requiere las librerías “rvest”, “openxlsx” y “openair” para la descarga y almacenamiento en masa de la información de la web del SENAMHI “Datos Hidrometeorológicos a Nivel Nacional”, generando automáticamente a partir de un rango de fechas (2017:2022,12) y el código de la estación de monitoreo (por ejemplo, Carabaylo=111286, Puente piedra=111287), una matriz de datos ordenados en filas y columnas (Script 01). El conjunto de datos horarios de los contaminantes del aire e hidrometeorológicos de las estaciones de monitoreo en Lima Metropolitana supera los 2 millones de registros, por lo cual el procesamiento se efectuó en el programa R, donde las estructuras de datos ordenados como tablas se denominan data frames. Estos data frames fueron fusionados con un script de R que requiere las librerías “openxlsx” y “plyr”, la información de

las columnas "Fecha" y "Hora" en cada data frame se convertirá al formato usado por R: "YY/mm/dd HH:MM:SS", siendo almacenado en una nueva columna denominada "FECHA_HORA". La función "merge" permite unir dos data frames por columnas comunes o por nombres de fila, a partir de las columnas "ESTACION" y "FECHA_HORA" fueron fusionadas ambas bases de datos en un nuevo data frame que integra los datos de material particulado y los parámetros meteorológicos medidos horariamente en cada estación de monitoreo (Script 02).

3.6 Métodos de análisis de datos

Se emplearon funciones y tablas dinámicas de Excel para obtener reportes de los valores para periodos de 24 horas, anuales y las fluctuaciones horarias de la concentración del PM_{2.5} y PM₁₀, que fueron comparados el límite ($\mu\text{g}/\text{m}^3$) del ECA para periodos de 24 horas, anual y los criterios de evaluación descritos en el Decreto Supremo N° 003-2017-MINAM.

Se graficó la concentración del material particulado a través de series temporales con intervalos de 24 horas y anual, el análisis se efectuó por comparación con los valores máximos y los criterios de evaluación descritos en los ECA del aire: Para PM_{2.5} el límite de $50 \mu\text{g}/\text{m}^3$ para un periodo 24 horas no se debe exceder de 7 veces al año y el valor máximo de $25 \mu\text{g}/\text{m}^3$ para un periodo anual determinado por la media aritmética anual. Para PM₁₀ el límite de $100 \mu\text{g}/\text{m}^3$ para un periodo 24 horas no se debe exceder de 7 veces al año y el límite de $50 \mu\text{g}/\text{m}^3$ para un periodo anual determinado por la media aritmética anual.

Evolución horaria de la calidad del aire

Se caracterizó la evolución por horas del día de la calidad del aire de acuerdo a la calificación del INCA para la concentración de PM_{2.5} y PM₁₀. Con funciones de Excel se asignó la calidad del aire: buena, moderada, mala o valor umbral del estado de cuidado, correspondiente a la concentración obtenida cada hora en las estaciones de monitoreo y ordenada de acuerdo a los siguientes intervalos horarios: 00-04, 04-08, 08-12, 12-16, 16-20 y 20-24. Se determinó el tiempo de exposición a cada estado de calidad del aire, a través del conteo del número de horas totales

anuales con la función tablas dinámicas de Excel. Dicha información es representada en series de temporales anuales por cada estación de monitoreo del SENAMHI localizada en Lima metropolitana.

Análisis estadístico de los datos

Se efectuó un análisis exploratorio para contrastar la no linealidad y normalidad de la distribución de los datos de PM_{2.5}, PM₁₀, temperatura del aire, humedad relativa, velocidad del viento y dirección del viento. En el programa estadístico R se efectuó la prueba de normalidad de Kolmogorov-Smirnov que es recomendada para más de 50 observaciones con la función “ks.test” de la librería “tidyverse” y la prueba de heterocedasticidad de Breusch-Pagan con la función “ncvTest” de la librería “car”. Asimismo, se determinó la asimetría, la curtosis y la prueba de Jarque-Bera, para la evaluación de la asimetría y la curtosis, con las funciones “skewness”, “kurtosis” y “jarque.test” de las librerías “broom” y “moments” (Script 03). El test de Kolmogorov-Smirnov indica que los datos no poseen una distribución normal, el valor-p <2.2e-16 es menor al nivel de significancia de 0.05 y se rechaza la hipótesis nula que los residuos de la variable respuesta y predictoras se distribuyen de manera normal (Tabla 2).

Tabla 2. Test de Kolmogorov-Smirnov para material particulado y los parámetros meteorológicos.

Variables	D	Valor p	Nº obser.
PM ₁₀ - temperatura	0.43621	<2.2e-16	59135
PM ₁₀ - humedad	0.92115	<2.2e-16	59135
PM ₁₀ - velocidad	0.99878	<2.2e-16	59135
PM ₁₀ - dirección	0.98182	<2.2e-16	59135
PM _{2.5} - temperatura	0.39972	<2.2e-16	59135
PM _{2.5} - humedad	0.99013	<2.2e-16	59135
PM _{2.5} - velocidad	0.98541	<2.2e-16	59135
PM _{2.5} - dirección	0.99364	<2.2e-16	59135

La prueba de Breusch-Pagan, en el modelo de regresión lineal para PM₁₀ y PM_{2.5} presenta no linealidad en relación a las variables meteorológicas. El gráfico de residuales (residual vs fitted) con forma de embudo, indica que no se cumple el supuesto de homocedasticidad de las varianzas. Los registros 69310, 72894 y 72895 son valores atípicos u “outliers” para PM₁₀; los registros 52304, 107418 y 52303 son “outliers” para PM_{2.5}. El gráfico de localización (Q-Q residuals) y escala de los residuos (scale-location)

demuestra la no linealidad de las variables del modelo, no se observa una distribución homogénea aproximada a una línea horizontal con pendiente cero. El gráfico de probabilidad de los residuos (residual vs leverage) muestra asimetría positiva (Figura 3).

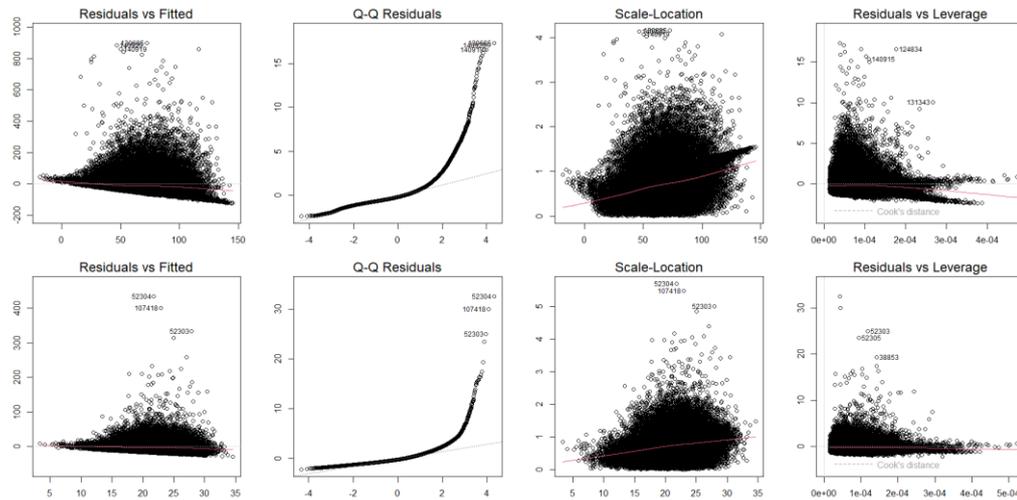


Figura 3. Análisis de normalidad y homocedasticidad de varianzas del modelo lineal: PM_{10} (arriba) y $PM_{2.5}$ (abajo) respecto a las variables meteorológicas.

La prueba de Jarque-Bera, la hipótesis nula se comprueba si son nulos la asimetría y el exceso de curtosis (asimetría = 0 y curtosis = 3), un valor JB grande indica que los errores o residuales no poseen distribución normal y con un valor-p menor de 0.05 se acepta la hipótesis alternativa; en el caso de las partículas y los parámetros meteorológicos tienen asimetría y curtosis, por lo tanto, no se ajustan a la distribución normal (Tabla 3).

Tabla 3. Asimetría, curtosis y test de Jarque-Bera (JB) para material particulado y parámetros meteorológicos

Variables	Asimetría	Curtosis	JB	Valor p	N° obser.
PM_{10}	2.52	15.22	159058	<2.2e-16	59135
$PM_{2.5}$	1.38	5.81	14116	<2.2e-16	59135
Temperatura	0.38	2.07	1322.9	<2.2e-16	59135
Humedad	-0.58	2.76	1282.5	<2.2e-16	59135
Velocidad	0.84	3.97	3479.1	<2.2e-16	59135
Dirección	-0.06	3.06	17.581	0.0001522	59135

Con la función “pairs.panels” de la librería “psych”, se obtuvieron matrices de correlación de los coeficientes de Pearson y Spearman, así como gráficas de dispersión con ajuste “LOESS” (regresión polinómica ponderada) e histogramas de distribución de las variables (Script 04). Se analizaron en total 59,135 observaciones obtenidas por fecha y horariamente en las estaciones de monitoreo del SENAMHI localizadas en

Lima Metropolitana entre el 2017-2020. Los coeficientes de Pearson y Spearman presentan resultados similares, siendo muy baja o casi nula entre el material particulado y los parámetros meteorológicos. En las gráficas de dispersión e histogramas de distribución de los datos, se observa asimetría positiva en las variables PM₁₀, PM_{2.5}, temperatura del aire y velocidad del viento, y asimetría negativa en la humedad relativa y dirección del viento (Figura 4 y 5).

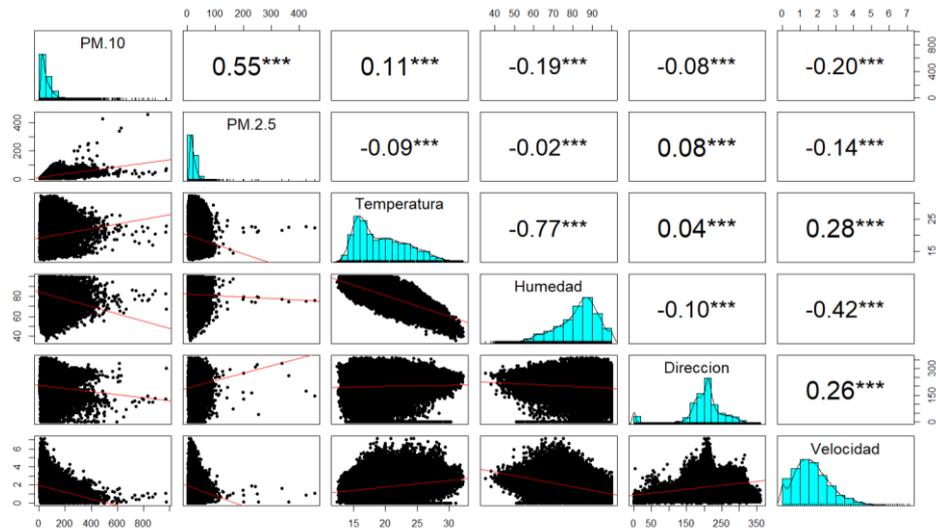


Figura 4. Coeficiente de Pearson entre las partículas y parámetros meteorológicos, de las estaciones del SENAMHI de Lima Metropolitana entre el 2017 y 2020

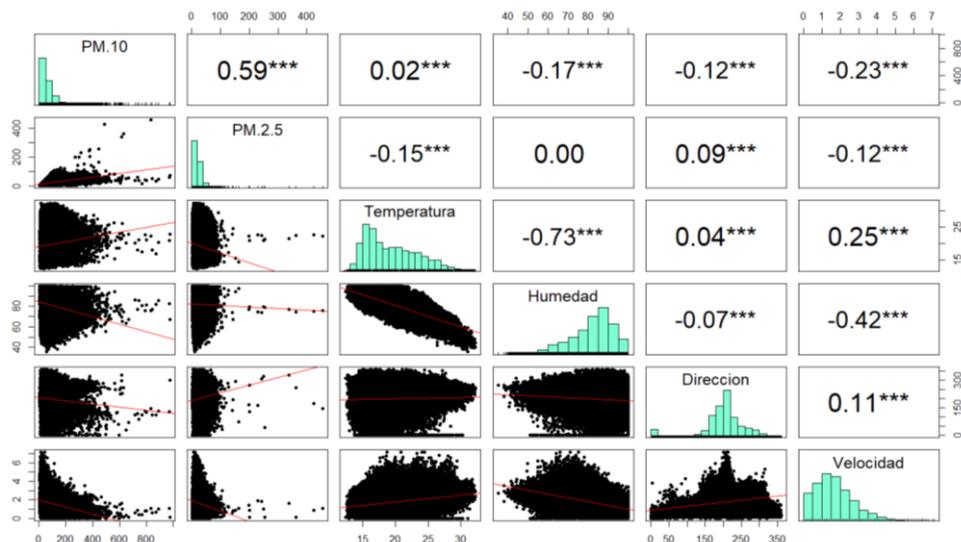


Figura 5. Coeficiente de Spearman entre las partículas y parámetros meteorológicos, de las estaciones del SENAMHI de Lima Metropolitana entre el 2017 y 2020

La Figura 6 presenta las diferencias obtenidas con el coeficiente de Pearson, al analizar promedios mensuales, la distribución de los datos asumió relativa normalidad al eliminar los valores atípicos y se sobreestimó

el coeficiente de Pearson, la relación del PM_{2.5} y PM₁₀ con la temperatura y humedad relativa aumentó a negativa o positiva alta ($0.6 < r < 0.8$) y negativa o positiva baja ($0.2 < r < 0.4$); la relación con la velocidad del viento es negativa muy baja ($0 < r < 0.2$) o casi nula.

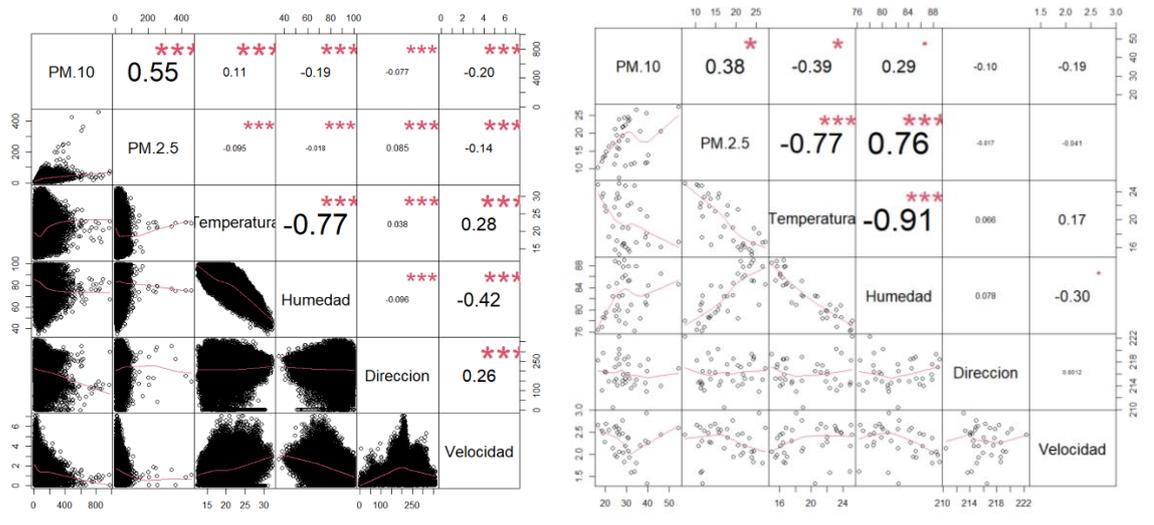


Figura 6. Coeficiente Pearson entre material particulado y los parámetros meteorológicos con observaciones por horas (derecha) y promedios mensuales (izquierda).

Modelos Aditivos Generalizados

Para evaluar la existencia de relaciones no lineales entre las partículas en suspensión y los parámetros atmosféricos, en el programa estadístico R con la función “gam” (Hastie, 2018) de la librería “mgcv”, se ajustaron modelos aditivos generalizados a los datos de concentración de PM_{2.5}, PM₁₀, temperatura del aire, humedad relativa, velocidad del viento y dirección del viento, con distribuciones de la familia gaussiana, gamma y binomial negativa con funciones de enlace identidad (Identity) y logarítmica (log). La función de suavizado se basó en p-splines penalizados con grados de libertad k de 4 a 10 nodos y polinomios de segundo orden ($m=2$). Para la selección del modelo con mejor grado de bondad de ajuste, se comparó la medida de la calidad relativa con el Criterio de Información de Akaike (AIC), que evita el sobreajuste penalizando los modelos más complejos en favor de los sencillos. Para el correcto ajuste de la función de suavizado, se tuvo como referencia los valores del coeficiente de determinación ajustado (R^2), la validación cruzada generalizada (GCV) y % de devianza explicada (Script 05).

Validación de los modelos para material particulado

Para la relación de PM_{2.5} con los parámetros meteorológicos, se ajustaron modelos de las familias gaussiana, gamma y binomial negativa con funciones de enlace identidad: $f(z) = z$, logarítmica: $\log(z)$ e inversa: $f(z) = 1/z$, el número de observaciones fue 56,205 y la devianza (varianza de los residuos) explicada o desviación explicada varió entre 15.9% y 16.9% (Tabla 4).

Tabla 4. Modelos Aditivos Generalizados para PM_{2.5} y parámetros meteorológicos

N°	Fórmula	Familia	Enlace	GCV	UBRE	Devianza	N°
1	log(PM _{2.5})-s(Direccion)+s(Velocidad)+s(Humedad)+s(Temperatura)	Gauss	Identidad: $f(z) = z$	0.2346	-	16.9%	56205
2	log(PM _{2.5})-s(Direccion)+s(Velocidad)+s(Humedad)+s(Temperatura)	Gauss	Log: $f(z) = \log(z)$	0.2356	-	16.6%	56205
3	log(PM _{2.5})-s(Direccion)+s(Velocidad)+s(Humedad)+s(Temperatura)	Gamma	Identidad: $f(z) = z$	0.0285	-	16.5%	56205
4	log(PM _{2.5})-s(Direccion)+s(Velocidad)+s(Humedad)+s(Temperatura)	Gamma	Inversa: $f(z) = 1/z$	0.0289	-	16.1%	56205
5	log(PM _{2.5})-s(Direccion)+s(Velocidad)+s(Humedad)+s(Temperatura)	Binomial negativa	Log: $f(z) = \log(z)$	-	-0.957	16.3%	56205
6	log(PM _{2.5})-s(Direccion)+s(Velocidad)+s(Humedad)+s(Temperatura)	Binomial negativa	Inversa: $f(z) = 1/z$	-	-0.957	15.9%	56205

El modelo N°1 de la familia gaussiana y la función de enlace “identidad”, presenta el mejor grado de bondad de ajuste para la concentración del PM_{2.5} y los cuatro predictores empleados, de acuerdo al menor valor del AIC de 78017 (Tabla 5).

Tabla 5. AIC de los modelos para PM_{2.5} y los parámetros meteorológicos

N°	df	AIC	N°
1	32.3	78017.	56205
2	33.2	78276.	56205
3	33.3	79185.	56205
4	33.0	79429.	56205
5	18.2	170554.	56205
6	19.1	170576.	56205

La Tabla 6, presenta el resumen estadístico del modelo N°1, la significancia aproximada de los términos de suavizado para cada variable predictora (valor-p<2e-16), grados de libertad efectivos (edf>7), R² ajustado, GCV y desviación explicada.

Tabla 6. Resumen del modelo de la familia gaussiana con enlace identidad para PM_{2.5} y parámetros meteorológicos

Family: gaussian
Link function: identity

Formula:
log(PM.2.5) ~ s(Direccion, k = 10, m = c(2, 3), bs = "ps") +
s(Velocidad, k = 10, m = c(2, 3), bs = "ps") + s(Humedad,
k = 10, m = c(2, 3), bs = "ps") + s(Temperatura, k = 10,
m = c(2, 3), bs = "ps")

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Direccion)	8.027	8.538	210.6	<2e-16 ***
s(Velocidad)	7.614	8.006	502.4	<2e-16 ***
s(Humedad)	8.453	8.807	267.1	<2e-16 ***
s(Temperatura)	7.244	7.647	378.4	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-sq.(adj) = 0.169 Deviance explained = 16.9%
GCV = 0.23461 Scale est. = 0.23447 n = 56205

El gráfico cuantil-cuantil (Normal Q-Q) e histograma de los residuos muestran que la distribución es aproximada a la normal teórica. El gráfico de probabilidad normal de los residuales (Residuals vs linear pred.) según los valores ajustados por el modelo N°1, presenta residuos negativos y positivos por la existencia de valores atípicos (Outliers) (Figura 7).

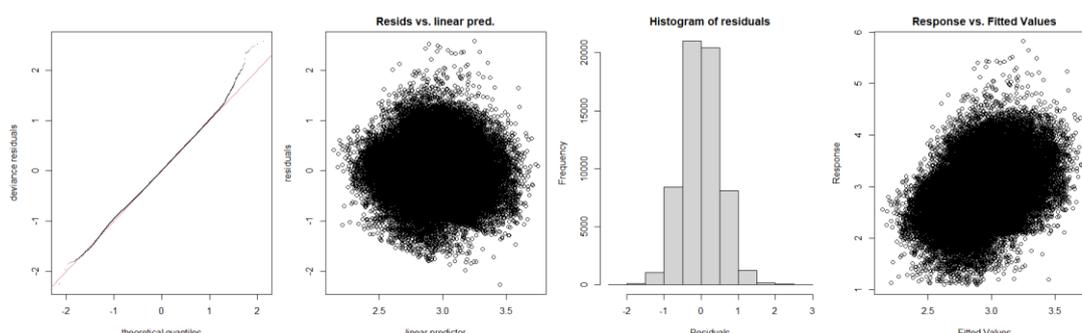


Figura 7. Gráficos de los cuantiles normalizados (Normal Q-Q), histograma de los residuos y probabilidad normal de los residuales (Fitted values vs. Index)

El modelo para la relación del PM_{2.5} con los parámetros meteorológicos, indica que la dirección del viento no influye en el incremento o disminución de la concentración de este contaminante; por lo tanto, se excluye del modelo ya que no es una variable explicativa para este contaminante (Figura 8).

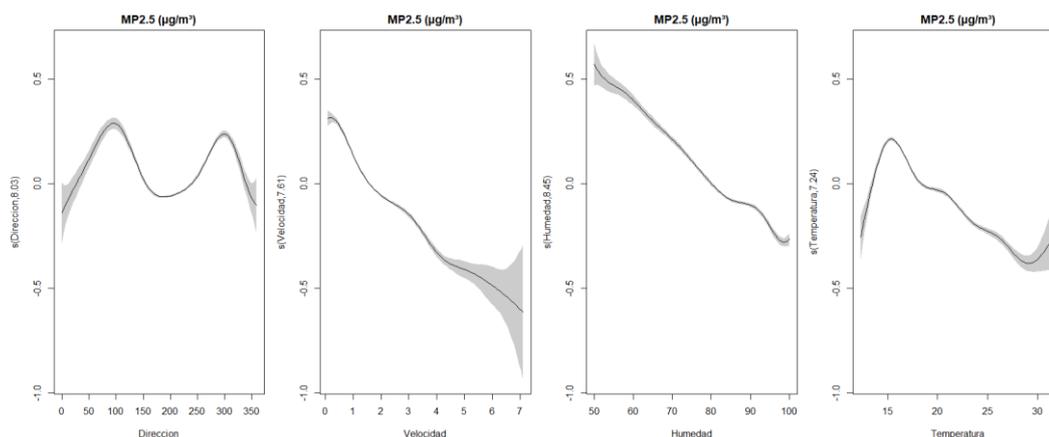


Figura 8. GAM de la concentración de PM_{2.5} en relación a los parámetros meteorológicos en las estaciones de Lima Metropolitana 2017-2020

Para la relación de PM₁₀ con los cuatro parámetros meteorológicos, se ajustaron modelos de las familias gaussiana, gamma y binomial negativa con funciones de enlace de identidad: $f(z) = z$, logarítmica: $\log(z)$ e inversa: $f(z) = 1/z$, el número de observaciones fue 56,205 y la devianza explicada o desviación explicada varió entre 36.7% y 38.2% (Tabla 7).

Tabla 7. Modelos Aditivos Generalizados para PM₁₀ y parámetros meteorológicos

N°	Fórmula	Familia	Enlace	GCV	UBRE	Devianza	N°.
7	$\log(\text{PM}_{10}) \sim s(\text{Dirección}) + s(\text{Velocidad}) + s(\text{Humedad}) + s(\text{Temperatura})$	Gauss	Identidad: $f(z) = z$	0.3370	-	38.2%	56205
8	$\log(\text{PM}_{10}) \sim s(\text{Dirección}) + s(\text{Velocidad}) + s(\text{Humedad}) + s(\text{Temperatura})$	Gauss	Log: $f(z) = \log(z)$	0.3451	-	36.7%	56205
9	$\log(\text{PM}_{10}) \sim s(\text{Dirección}) + s(\text{Velocidad}) + s(\text{Humedad}) + s(\text{Temperatura})$	Gamma	Identidad: $f(z) = z$	0.0243	-	34.8%	56205
10	$\log(\text{PM}_{10}) \sim s(\text{Dirección}) + s(\text{Velocidad}) + s(\text{Humedad}) + s(\text{Temperatura})$	Gamma	Inversa: $f(z) = 1/z$	0.0234	-	37.1%	56205
11	$\log(\text{PM}_{10}) \sim s(\text{Dirección}) + s(\text{Velocidad}) + s(\text{Humedad}) + s(\text{Temperatura})$	Binomial negativa	Log: $f(z) = \log(z)$	-	-0.960	37.6%	56205
12	$\log(\text{PM}_{10}) \sim s(\text{Dirección}) + s(\text{Velocidad}) + s(\text{Humedad}) + s(\text{Temperatura})$	Binomial negativa	Inversa: $f(z) = 1/z$	-	-0.959	36.8%	56205

El modelo N°9 de la familia gaussiana y función de enlace “identidad”, presenta el mejor grado de bondad de ajuste para la concentración del PM₁₀ y los cuatro predictores empleados, de acuerdo al menor valor del AIC de 98384 (Tabla 8).

Tabla 8. AIC de los modelos para PM₁₀ y parámetros meteorológicos

N°	df	AIC	N°
1	33.3	98384.	56205
2	33.4	99719.	56205
3	29.8	100612.	56205
4	33.7	98625.	56205
5	19.2	230803.	56205
6	19.4	230834.	56205

La Tabla 9, presenta el resumen estadístico del modelo N°9, los coeficientes paramétricos, la significancia aproximada de los términos de suavizado para cada variable predictora (valor- $p < 2e-16$), grados de libertad efectivos (edf > 7), R^2 ajustado, GCV y desviación explicada.

Tabla 9. Resumen del modelo de la familia gaussiana con enlace identidad para PM_{10} y parámetros meteorológicos

Family: gaussian
Link function: identity

Formula: $\log(PM_{10}) \sim s(\text{Direccion}, k = 10, m = c(1, 2), bs = "ps") + s(\text{Velocidad}, k = 10, m = c(2, 3), bs = "ps") + s(\text{Humedad}, k = 10, m = c(2, 3), bs = "ps") + s(\text{Temperatura}, k = 10, m = c(2, 3), bs = "ps")$

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Direccion)	8.682	9	1128.3	<2e-16 ***
s(Velocidad)	7.618	9	1133.0	<2e-16 ***
s(Humedad)	8.795	9	969.5	<2e-16 ***
s(Temperatura)	7.249	9	227.9	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-sq.(adj) = 0.382 Deviance explained = 38.2%
GCV = 0.33707 Scale est. = 0.33687 n = 56205

El gráfico de los residuales (Residuals vs linear pred.) según los valores ajustados por el modelo N°9 presenta residuos negativos y positivos por la existencia de valores atípicos (Outliers). En el gráfico cuantil-cuantil (Normal Q-Q) e histograma de los residuos, se observa que la distribución es aproximada a la distribución normal teórica (Figura 9).

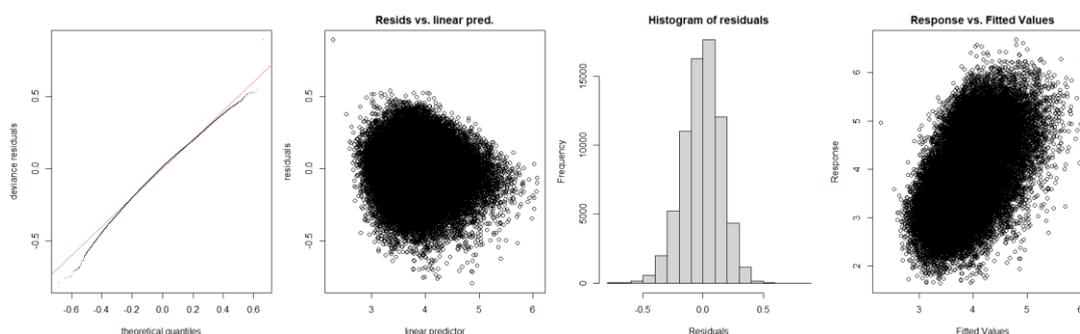


Figura 9. Gráficos de los cuantiles normalizados (Normal Q-Q), histograma de los residuos y probabilidad normal de los residuales (Fitted values vs. Index)

El modelo para la relación del PM_{10} con los parámetros meteorológicos, indica que la concentración de este contaminante se mantuvo casi constante respecto a la dirección del viento; por lo tanto, se excluye del

modelo ya que no es una variable explicativa para este contaminante (Figura 10).

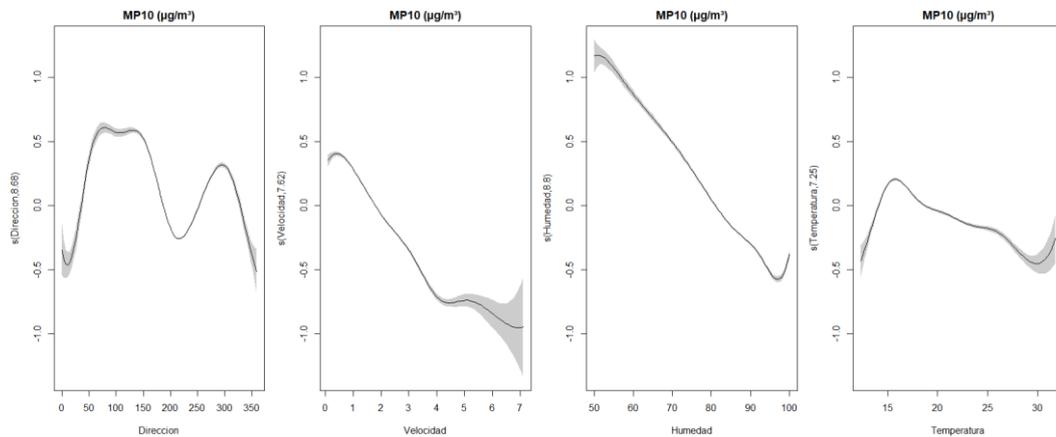


Figura 10. GAM de la concentración de PM₁₀ en relación a los parámetros meteorológicos en las estaciones de Lima Metropolitana 2017-2020

Árboles de regresión

La construcción de los árboles de regresión tiene dos etapas: la fase de construcción y de poda. En la elaboración del árbol, los valores son divididos recursivamente en dos y los grupos formados en cada división son más homogéneos en razón de la variable respuesta. La impureza es la medida de homogeneidad, es decir un grupo menos homogéneo es más impuro y un grupo se considera más homogéneo cuanto más reducida es la suma de cuadrados de los residuos. El resultado final es un árbol dividido la cantidad de veces suficiente para obtener grupos homogéneos en los nodos terminales (Milborrow, 2021).

El árbol cuenta con tres niveles de nodos, tiene un solo nodo en el primer nivel llamado nodo raíz que indica la variable predictora que sido dividido el grupo, nodos internos en el segundo nivel y nodos terminales en la parte inferior con el número de datos y el valor promedio de la variable respuesta. La afirmación es verdadera hacia la rama izquierda y es falsa hacia la rama de la derecha (Figura 11).

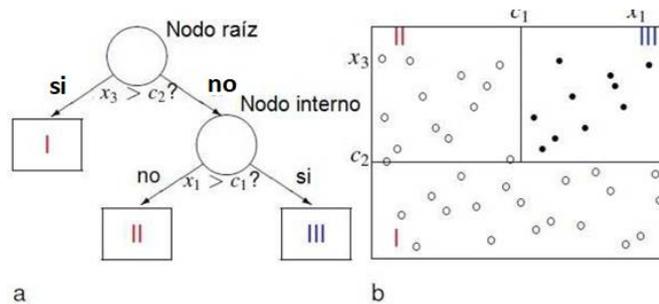


Figura 11. Estructura del árbol de regresión

El algoritmo genera el árbol con una restricción al exceso de divisiones terminando el árbol cuando la reducción de la impureza en el último nodo es inferior al 1% de la impureza inicial o es mayor a 0.01 el parámetro de complejidad (cp). Sin embargo, para conseguir un árbol pequeño de fácil interpretación no es insuficiente esta penalización. En la fase de poda se obtiene un árbol más estable pequeño, a través de validación cruzada con la estimación del número óptimo de divisiones. La validación cruzada está basada en validaciones repetidas en cada división del árbol. Con esta aproximación, se define el árbol óptimo al presentar tasa de error similar en las validaciones y en el árbol inicial, así en el árbol óptimo la tasa de error estimado es menor o igual al error estándar mínimo del total de las validaciones. Para la construcción del árbol se crearon grupos de entrenamiento y prueba (Ferreira et al., 2018), se consideró el 80% de los datos para el conjunto de entrenamiento y 20% para prueba; se utilizaron los paquetes `rpart` versión 4.1.23 (Therneau y Athkinson, 2023) que es el mismo algoritmo descrito por Brieman en 1984 y el paquete `rpart.plot` (Milborrow, 2021) del software R.

Distribución espacial del material particulado

El mapeo de los datos se efectuó con la versión más reciente del programa de interpolación y geoestadística Surfer para Windows (25.2.259 de 64-bit). La predicción se efectuó a través de tres métodos de interpolación: co-kriging ordinario, kriging ordinario y ponderación de distancia inversa (IDW), los atributos fueron: longitud (x), latitud (y) y media aritmética anual (z) de las concentraciones de $PM_{2.5}$ y PM_{10} por punto de muestreo o estación de monitoreo del SENAMHI. Para validar la interpolación de los métodos co-kriging ordinario y kriging ordinario se construyeron variogramas teóricos

para $PM_{2.5}$ y PM_{10} con el modelo esférico, así se explicó la variabilidad espacial dentro del área de estudio. Para el método de IDW se utilizó un valor de ponderación 2, sin alisado o suavizado (Tabla 10 y Figura 12).

Tabla 10. Resumen de los modelos de predicción para la distribución espacial del $PM_{2.5}$ y PM_{10}

Variable	Co-kriging/Kriging		IDW	
	$PM_{2.5}$	PM_{10}	$PM_{2.5}$	PM_{10}
Modelo	Esférico	Esférico	-	-
Rango	0.0857	0.0857	-	-
Efecto pepita (Nugget)	0	0	-	-
Meseta parcial (Sill)	87	1060	-	-
Ponderación	-	-	2	2
Suavizado	-	-	0	0
Coefficiente de determinación R^2	0.759	0.573	0.925	0.486

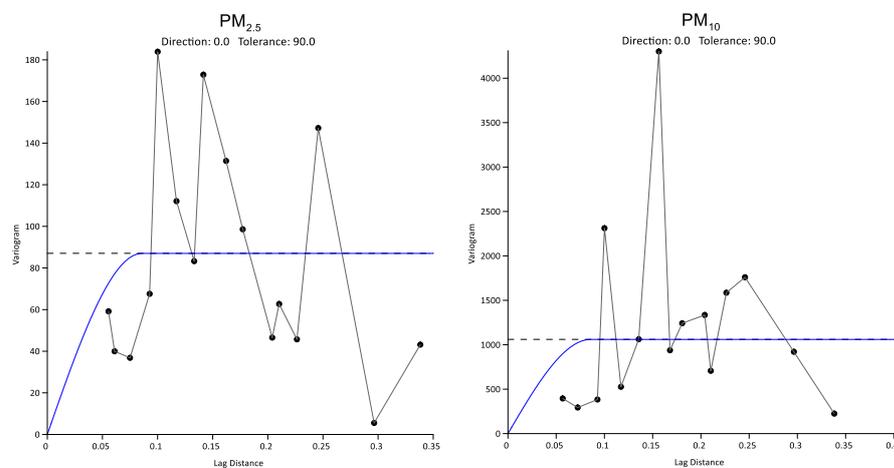


Figura 12. Variograma experimental y teórico de los datos de $PM_{2.5}$ y PM_{10} , para los métodos de interpolación co-kriging ordinario y kriging ordinario.

Para la selección del método de interpolación más eficiente para la predicción de la distribución espacial del PM_{10} y $PM_{2.5}$, se comparó la medida de la desviación estándar, error estándar y coeficiente de determinación (R^2); el co-kriging ordinario se ajustó mejor a la concentración del MP_{10} , donde fue mayor el coeficiente de determinación (R^2) y menor la desviación estándar y error estándar. El kriging ordinario presentó mejor ajuste al comportamiento espacial del $PM_{2.5}$, obtuvo el mayor coeficiente de determinación (R^2) y menor desviación estándar y error estándar. Por su parte, con el IDW no fueron óptimas las medidas del coeficiente de determinación (R^2), desviación estándar y error estándar para la concentración del PM_{10} y $PM_{2.5}$ (Tabla 11).

Tabla 11. Desviación estándar, error estándar y coeficiente de determinación (R^2) de los modelos de predicción para la distribución espacial del $PM_{2.5}$ y PM_{10}

Método	Variable	Medidas estadísticas	2017	2018	2019	2020
Co-kriging ordinario	$PM_{2.5}$	Desviación Estándar	5.07	5.13	5.23	1.38
		Error Estándar	0.05	0.05	0.05	0.01
		Coef. de determinación R^2	0.92	0.70	0.77	0.30
	PM_{10}	Desviación Estándar	14.83	15.64	15.30	13.02
		Error Estándar	0.15	0.16	0.15	0.13
		Coef. de determinación R^2	0.57	0.71	0.88	0.57
Kriging ordinario	$PM_{2.5}$	Desviación Estándar	5.05	4.48	5.23	1.31
		Error Estándar	0.05	0.05	0.05	0.01
		Coef. de determinación R^2	0.94	0.72	0.78	0.48
	PM_{10}	Desviación Estándar	14.92	15.64	15.39	13.64
		Error Estándar	0.15	0.16	0.16	0.14
		Coef. de determinación R^2	0.56	0.72	0.86	0.67
IDW	$PM_{2.5}$	Desviación Estándar	5.89	5.08	7.27	12.02
		Error Estándar	0.06	0.05	0.08	0.12
		Coef. de determinación R^2	0.88	0.61	0.61	0.42
	PM_{10}	Desviación Estándar	16.81	17.58	17.28	12.02
		Error Estándar	0.17	0.18	0.17	0.12
		Coef. de determinación R^2	0.49	0.62	0.82	0.56

Con la función “Volume” del programa Surfer se estimó el área total evaluada (km^2) de las estaciones de monitoreo operativas cada año y la superficie o área donde se excedió el ECA para un periodo anual, con dicha función se pueden calcular áreas y volúmenes, como áreas planas (positivas y negativas) y superficiales a partir de archivos en formato grid (GRD). El área plana positiva representa la proyección del mapa en un plano horizontal, la “Constante Z” que especifica el nivel de la superficie plana a usar como superficie superior o inferior son los valores máximos del ECA anual de $50 \mu g/m^3$ para $PM_{2.5}$ y $25 \mu g/m^3$ para PM_{10} .

3.7 Aspectos éticos

La tesis se basa en información del SENAMHI, el Artículo 6 del código de ética de la UCV obliga a respetar los derechos de propiedad intelectual con las respectivas referencias y citas a investigaciones de otros autores. En cuanto a la elaboración del presente documento, se redactó según lo descrito en la “Guía de Elaboración de Productos de investigación formativa” Resolución de Vicerrectorado de Investigación N°062-2023-VI-UCV de 16 marzo del 2023, que establece el formato y estructura del trabajo de investigación, siendo finalmente revisado con la herramienta Turnitin para revisar la coincidencia del contenido con otros trabajos de investigación, siendo el porcentaje de similitud menor al 20%.

IV. RESULTADOS

4.1. Evaluación de la concentración de partículas.

Partículas con diámetro menor de 2.5 micrómetros

Entre los años 2017 y 2020, el límite del ECA de $50 \mu\text{g}/\text{m}^3$ para la concentración de $\text{PM}_{2.5}$ durante un periodo de 24 horas, no fue excedido de 7 veces al año en las estaciones del SENAMHI de Lima Metropolitana. La estación del Ate excedió el ECA 5 veces en el 2018 y 4 veces en el 2019. (Tabla 12, Figura 13).

Tabla 12. Número de veces excedido el ECA para la concentración de $\text{PM}_{2.5}$ en un periodo de 24 horas, en Lima Metropolitana 2017-2020.

ESTACIÓN	2017	2018	2019	2020	Total
ATE	-	5	4	-	9
CARABAYLLO	-	-	2	-	2
SAN JUAN DE LURIGANCHO	-	-	1	-	1
Total	0	5	7	0	12

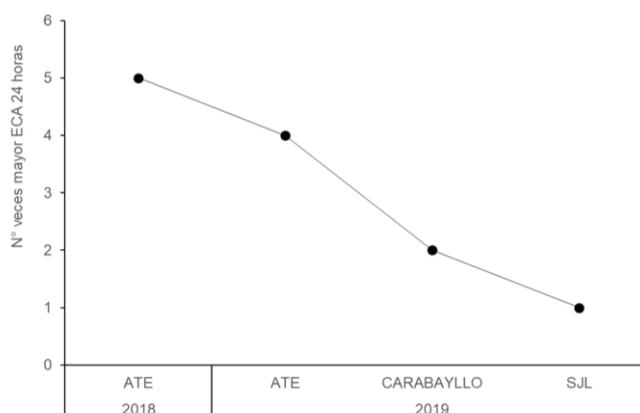


Figura 13. Número de veces excedido el ECA para la concentración de $\text{PM}_{2.5}$ en 24 horas en Lima Metropolitana 2017-2020

Respecto al ECA anual ($25 \mu\text{g}/\text{m}^3$), en el periodo 2017-2019 se excedió principalmente en las estaciones de Ate, San Juan de Lurigancho y Santa Anita donde fue mayor a $30 \mu\text{g}/\text{m}^3$. Durante el año 2020, la concentración de $\text{PM}_{2.5}$ en las estaciones del SENAMHI de Lima Metropolitana no superó el ECA anual (Tabla 13, Figura 14).

Tabla 13. Concentración promedio anual para $\text{PM}_{2.5}$, en Lima Metropolitana entre 2017-2020.

ESTACIÓN	2017	2018	2019	2020
ATE	42.23	46.29	48.76	
CAMPO DE MARTE	13.37	20.02	19.47	17.46
CARABAYLLO		25.11	33.30	17.93
PUENTE PIEDRA	30.44	34.15	30.86	
SAN BORJA	15.35	16.99	13.95	14.74
SAN JUAN DE LURIGANCHO	32.66	33.55	35.70	21.26
SAN MARTIN DE PORRES	17.60	18.91	21.09	
SANTA ANITA	30.87	32.11	33.30	
VILLA MARIA DEL TRIUNFO	25.60	24.86	28.91	23.81

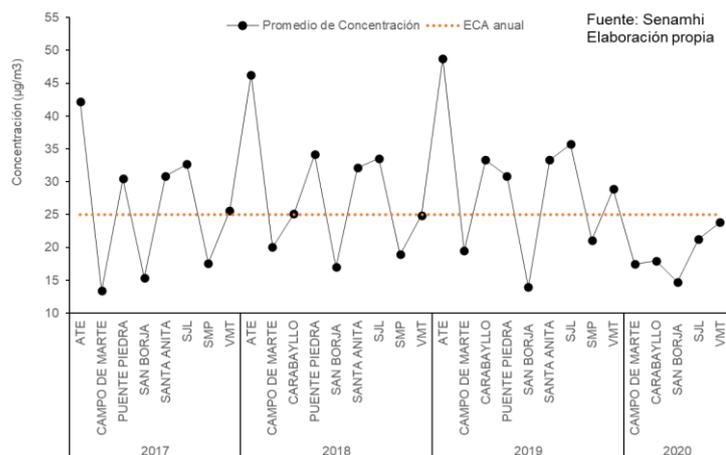


Figura 14. Concentración promedio y ECA anual para PM_{2.5}, en las estaciones de monitoreo del SENAMHI de Lima Metropolitana entre 2017-2020.

Partículas con diámetro menor de 10 micrómetros

Entre los años 2017-2019, se excedió de 7 veces al año el ECA durante 24 horas (100 µg/m³ año) en las estaciones de Ate con 93 veces, Villa María del Triunfo con 88 veces, Puente Piedra con 23 veces y San Juan de Lurigancho con 9 veces. En el 2020, en San Borja se excedió 11 veces el ECA durante 24 horas (Tabla 14, Figura 15).

Tabla 14. Número de veces excedido el ECA para la concentración de PM₁₀ durante un periodo de 24 horas, en Lima Metropolitana 2017-2020.

ESTACIÓN	2017	2018	2019	2020	Total
ATE	27	41	25		93
CARABAYLLO	1				1
PUENTE PIEDRA	23	3			26
SAN BORJA				11	11
SAN JUAN DE LURIGANCHO	9	2			11
SANTA ANITA	3	1			4
VILLA MARIA DEL TRIUNFO	41	27	2	1	71
Total	104	74	27	12	217

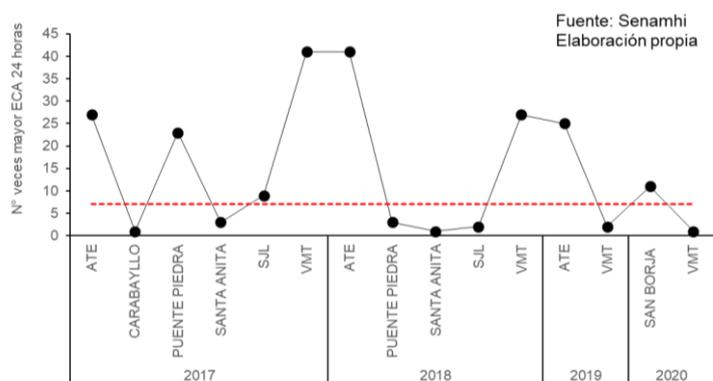


Figura 15. Número de veces excedido el ECA para la concentración de PM₁₀ en 24 horas en Lima Metropolitana 2017-2020

Entre los años 2017-2019, el ECA anual (50 µg/m³) para la concentración de PM₁₀ se excedido principalmente en las estaciones de Villa María del Triunfo, Ate y Carabayllo donde fue superior a 100 µg/m³. Durante el año 2020, las

estaciones que excedieron el ECA anual fueron Villa María del Triunfo y Carabayllo con $94.9 \mu\text{g}/\text{m}^3$ y $62 \mu\text{g}/\text{m}^3$ respectivamente (Tabla 15, Figura 16).

Tabla 15. Concentración promedio anual para PM_{10} , en las estaciones del SENAMHI de Lima Metropolitana entre 2017-2020.

ESTACION	2017	2018	2019	2020
CAMPO DE MARTE	38.56	28.73	25.84	25.24
CARABAYLLO	82.54	87.38	108.73	62.01
PUENTE PIEDRA	110.13	85.19		
SAN BORJA	50.81	54.12	51.71	53.60
SAN JUAN DE LURIGANCHO	90.58	80.36	66.06	51.45
SAN MARTIN DE PORRES	52.20	43.62		
SANTA ANITA	75.08	67.84	78.23	34.56
VILLA MARIA DEL TRIUNFO	131.31	129.01	97.99	94.96

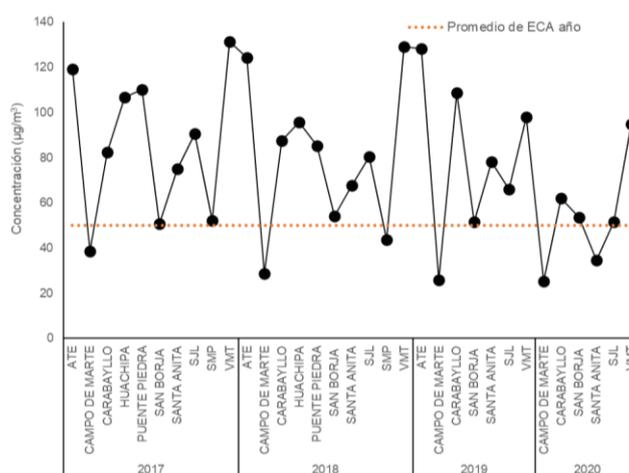


Figura 16. Concentración promedio anual y ECA anual para la concentración de PM_{10} , en Lima Metropolitana entre 2017-2020.

4.2. Distribución espacial de la concentración de partículas

Partículas con diámetro menor de 2.5 micrómetros

El método de interpolación kriging ordinario se ajustó mejor a la variabilidad espacial de la concentración de $\text{PM}_{2.5}$; que durante el periodo 2017-2019 presentó valores superiores al límite de $25 \mu\text{g}/\text{m}^3$ del ECA anual principalmente en las estaciones localizadas al norte y este de Lima Metropolitana. El área geográfica donde la concentración fue mayor a $25 \mu\text{g}/\text{m}^3$, abarcó la estación de Ate donde se incrementó gradualmente de $42.2 \mu\text{g}/\text{m}^3$ a $48.8 \mu\text{g}/\text{m}^3$, así como las estaciones de Puente Piedra, Carabayllo, San Juan de Lurigancho y Santa Anita donde la concentración anual de $\text{PM}_{2.5}$ fue superior a $30 \mu\text{g}/\text{m}^3$. Respecto al área total de 576.2 km^2 comprendida por las estaciones de Lima Metropolitana, la superficie donde fue excedido el ECA se incrementó anualmente con 388.5 km^2 en el año 2017 (67.4%), 429.0 km^2 en el año 2018 (74.4%) y 444.9 km^2 (77.2%) en el año 2019 (Tabla 16).

Tabla 16. Área donde la concentración promedio de PM_{10} fue mayor al ECA anual en Lima Metropolitana.

Año	Nº Estaciones	Superficie $PM_{10} > 50 \mu\text{g}/\text{m}^3$ (Km^2)	Área total (Km^2)	%
2017	9	559.4	576.2	97.1
2018	7	525.4	576.2	91.2
2019	6	398.1	416.4	95.6
2020	4	246.2	412.5	59.7

La menor concentración anual de $PM_{2.5}$ se presentó en las estaciones de Lima Oeste donde no superó $20 \mu\text{g}/\text{m}^3$ y $15 \mu\text{g}/\text{m}^3$. En el año 2020, no fue excedido el límite de $25 \mu\text{g}/\text{m}^3$ del ECA anual para la concentración de $PM_{2.5}$ (Figura 17).

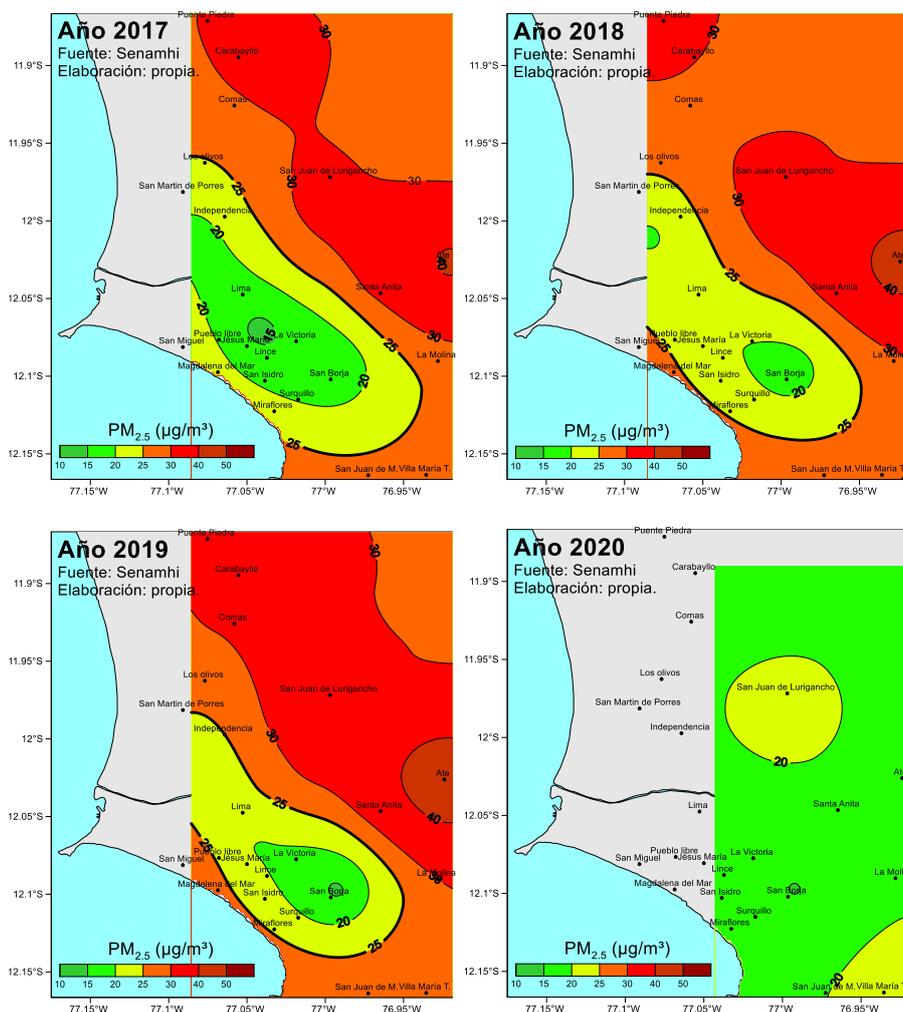


Figura 17. Distribución espacial de la concentración de $PM_{2.5}$ con el método de interpolación kriging ordinario en Lima Metropolitana entre 2017-2020.

Partículas con diámetro menor de 10 micrómetros

El método de interpolación co-kriging ordinario usando como covariable la velocidad del viento permitió predecir mejor la distribución espacial de la concentración de PM_{10} . Entre los años 2017 y 2020 los valores superiores al límite de $50 \mu\text{g}/\text{m}^3$ del ECA anual se presentaron en las estaciones localizadas

al norte, este y sur de Lima Metropolitana, las concentraciones que excedieron el ECA abarcaron una superficie de 559.4 km² en el año 2017 y 525.4 km² en el año 2018, que representó el 97.1% y 91.2% del área total de 576.2 km² comprendida por las estaciones de Lima Metropolitana. Entre los años 2019 y 2020, la superficie con concentración mayor a 50 µg/m³ disminuyó de 398.1 km² a 246.2 km², representando el 95.6% y 59.7 % del área total de 416.4 km² comprendida por las estaciones de Lima Metropolitana (Tabla 17).

Tabla 17. Área donde la concentración promedio de PM_{2.5} fue mayor al ECA anual en las estaciones del SENAMHI de Lima Metropolitana.

Año	N° Estaciones	Superficie PM _{2.5} >25 µg/m ³ (Km ²)	Área total (Km ²)	%
2017	9	388.5	576.2	67.4
2018	9	429.0	576.2	74.4
2019	9	444.9	576.2	77.2
2020	5	0.0	412.5	0.0

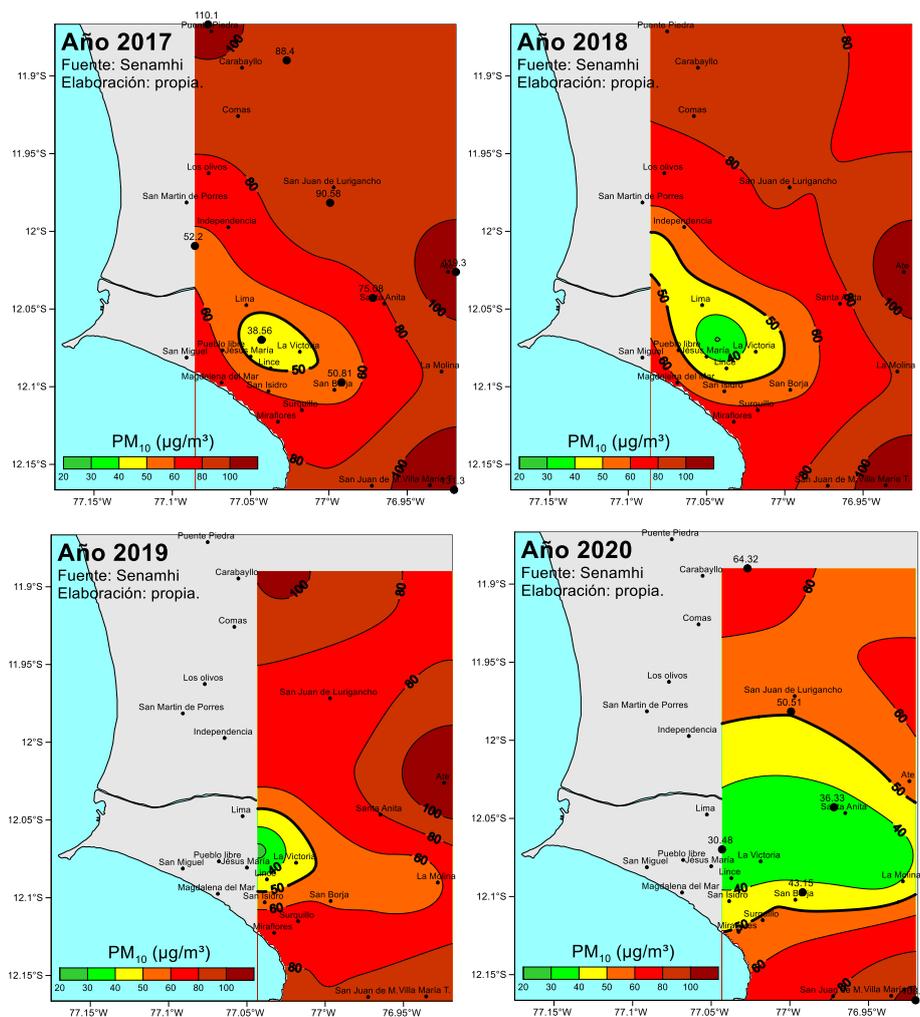


Figura 18. Distribución espacial de la concentración de PM₁₀ con el método de interpolación co-kriging ordinario en Lima Metropolitana entre 2017-2020.

En los años 2017 y 2019, las estaciones localizadas en Puente Piedra, Ate, La Molina y Villa María del Triunfo presentaron concentración promedio anual de PM₁₀ mayor a 100 µg/m³, disminuyendo ligeramente en el 2020. La estación de monitoreo de Jesús María registró la menor concentración de PM₁₀, disminuyendo progresivamente de 38.56 µg/m³ a 25.24 µg/m³ entre el 2017 y 2020 (Figura 18).

4.3. Evolución horaria de la calidad del aire según el INCA Partículas con diámetro menor de 2.5 micrómetros

La concentración del PM_{2.5} fue óptima durante las 24 horas en las estaciones de Jesús María donde la calidad “buena” prevaleció en 9,062 horas (32.7%) y “moderada” en 14,044 horas (50.5%), en San Borja la calidad “buena” se registró en 8,826 horas (42.5%) y “moderada” en 9,636 horas (46.4%), en San Martín de Porres la calidad “buena” predominó en 4,706 horas (34%) y “moderada” en 6,345 horas (45.9%).

Las condiciones insalubres del aire se presentaron entre las 01 am-12 pm y 9 pm-12 am; la calidad del aire “mala” prevaleció en San Juan de Lurigancho con 12,648 horas (57.1 %), Ate con 11,219 horas (77.6 %), Puente Piedra con 8,955 horas (62.8 %) y Santa Anita con 8,679 horas (59.9 %). La exposición a la calidad del aire “Valor del umbral de cuidado” se presentó en Ate con 238 horas (Tabla 18 y Figura 19).

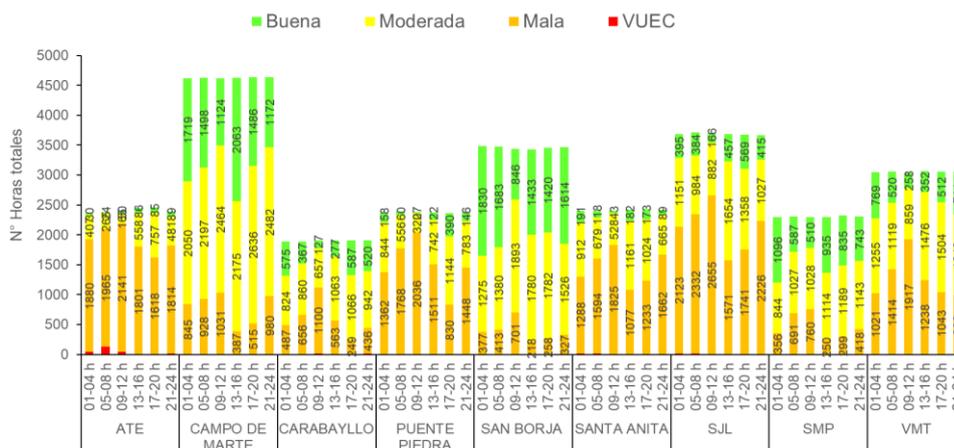


Figura 19. Calidad del aire según el INCA para la concentración de PM_{2.5}, las estaciones de SENAMHI en Lima Metropolitana entre 2017-2020.

Tabla 18. Calidad del aire según el INCA para la concentración de PM_{2.5}, en Lima Metropolitana entre 2017-2020.

ESTACION	INCA	01-04	05-08	09-12	13-16	17-20	21-24	TOTAL	%
ATE	Buena	30	24	50	86	85	89	364	2.5
	Moderada	407	265	165	558	757	481	2633	18.2
	Mala	1880	1965	2141	1801	1618	1814	11219	77.6
	VUEC	45	132	48		1	12	238	1.6
CAMPO DE MARTE	Buena	1719	1498	1124	2063	1486	1172	9062	32.7
	Moderada	2050	2197	2464	2175	2636	2482	14004	50.5
	Mala	845	928	1031	387	515	980	4686	16.9
CARABAYLLO	Buena	575	367	127	277	587	520	2453	21.5
	Moderada	824	860	657	1063	1066	942	5412	47.5
	Mala	487	656	1100	563	249	436	3491	30.6
	VUEC	2	5	19	4	8	11	49	0.4
PUENTE PIEDRA	Buena	158	60	27	122	390	146	903	6.3
	Moderada	844	556	320	742	1144	783	4389	30.8
	Mala	1362	1768	2036	1511	830	1448	8955	62.8
	VUEC	8	6	4			1	19	0.1
SAN BORJA	Buena	1830	1683	846	1433	1420	1614	8826	42.5
	Moderada	1275	1380	1893	1780	1782	1526	9636	46.4
	Mala	377	413	701	218	258	327	2294	11.1
	VUEC	2						2	0.0
SAN JUAN DE LURIGANCHO	Buena	395	384	166	457	569	415	2386	10.8
	Moderada	1151	984	882	1654	1358	1027	7056	31.9
	Mala	2123	2332	2655	1571	1741	2226	12648	57.1
	VUEC	14	11	5	3	10	1	44	0.2
SAN MARTIN DE PORRES	Buena	1096	587	510	935	835	743	4706	34.0
	Moderada	844	1027	1028	1114	1189	1143	6345	45.9
	Mala	356	691	760	250	299	418	2774	20.1
	VUEC	4	1				2	7	0.1
SANTA ANITA	Buena	191	118	43	182	173	89	796	5.5
	Moderada	912	679	528	1161	1024	665	4969	34.3
	Mala	1288	1594	1825	1077	1233	1662	8679	59.9
	VUEC	14	13	8			3	38	0.3
VILLA MARIA DEL TRIUNFO	Buena	769	520	258	352	512	711	3122	17.0
	Moderada	1255	1119	859	1476	1504	1348	7561	41.3
	Mala	1021	1414	1917	1238	1043	996	7629	41.6
	VUEC	4	4	6	1		1	16	0.1

Partículas con diámetro menor de 10 micrómetros

La concentración del PM₁₀ presentó la calidad “buena” durante las 24 horas en Jesús María con 28,602 horas (98.2%), en San Borja con 1,508 horas (84.4%) y en San Martín de Porres con 12,535 horas (88.2%). Las condiciones insalubres del aire se registraron entre las 05 am-12 pm y 5 pm-12 am; la exposición a la calidad “mala” fue mayor en Villa María del Triunfo con 4,061 horas (21.4 %), Ate con 3,782 horas (21%) y Puente Piedra con 1,810 horas (11.8 %). La calidad del aire “Valor del umbral de cuidado” se presentó en la estación de Villa María del Triunfo con 1,527 horas y Ate con 613 horas (Tabla 19 y Figura 20)

Tabla 19. Calidad del aire según el INCA para la concentración de PM₁₀, en las estaciones del SENAMHI de Lima Metropolitana entre 2017-2020.

ESTACION	INCA	01-04	05-08	09-12	13-16	17-20	21-24	TOTAL	%
ATE	Buena	946	546	326	752	489	471	3530	19.6
	Moderada	1471	1198	1682	1980	1967	1767	10065	55.9
	Mala	496	903	861	263	545	714	3782	21.0
	VUEC	79	347	118	14	9	46	613	3.4
CAMPO DE MARTE	Buena	4776	4706	4686	4830	4806	4798	28602	98.2
	Moderada	68	142	160	19	59	73	521	1.8
CARABAYLLO	Buena	2613	2039	930	1200	1668	1623	10073	50.9
	Moderada	659	1120	1826	1904	1574	1621	8704	44.0
	Mala	17	128	537	196	52	40	970	4.9
	VUEC	8	9	7	2	2	3	31	0.2
PUENTE PIEDRA	Buena	1319	1032	735	825	615	682	5208	33.9
	Moderada	1153	1117	1343	1474	1554	1510	8151	53.1
	Mala	74	323	428	246	391	348	1810	11.8
VUEC	Buena	9	86	50	9	8	16	178	1.2
	Buena	2910	2813	2598	2926	2578	2683	16508	84.4
	Moderada	349	438	633	304	667	567	2958	15.1
Mala	Buena	15	17	21	5	6	13	77	0.4
	VUEC	2	2	2	2	1	7	0.0	
	Buena	1920	1759	1200	1754	1378	1402	9413	49.6
Moderada	Buena	1182	1244	1638	1330	1706	1688	8788	46.3
	Mala	66	173	316	63	74	69	761	4.0
	VUEC	10	6	6	2	4	28	0.1	
SAN MARTIN DE PORRES	Buena	2187	2048	1809	2154	2159	2178	12535	88.2
	Moderada	149	302	536	201	180	165	1533	10.8
	Mala	33	19	19	5	26	23	125	0.9
VUEC	Buena	12	1	8	8	6	6	27	0.2
	Buena	3236	2790	2442	3404	2873	2640	17385	68.0
	Moderada	916	1257	1616	832	1377	1531	7529	29.4
Mala	Buena	100	185	191	19	23	92	610	2.4
	VUEC	8	31	7	3	49	0.2		
	Buena	2045	1444	708	533	737	1364	6831	36.0
Moderada	Buena	780	848	1221	1650	1154	915	6568	34.6
	Mala	243	542	926	851	895	604	4061	21.4
	VUEC	74	337	307	133	396	280	1527	8.0

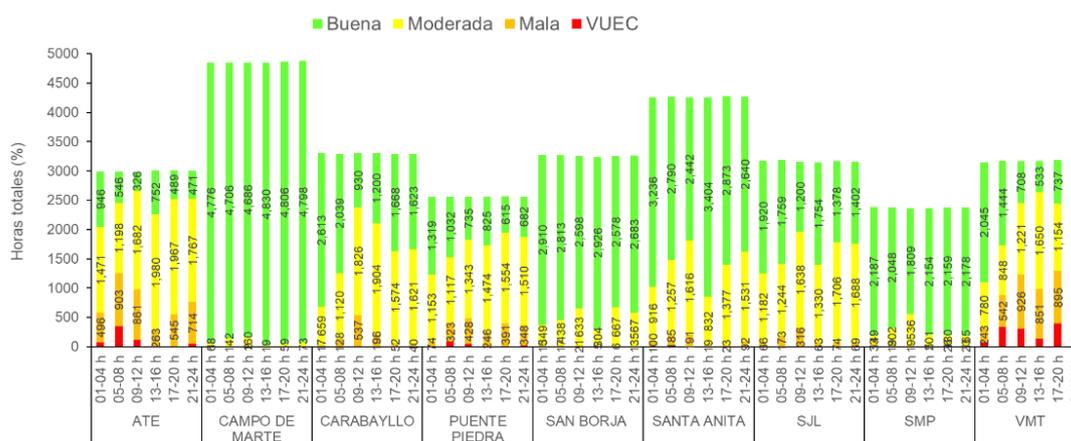


Figura 20. Calidad del aire según el INCA para la concentración de PM₁₀, las estaciones de SENAMHI en Lima Metropolitana entre 2017-2020.

4.4. Material particulado relacionado a los parámetros meteorológicos

Partículas con diámetro menor de 2.5 micrómetros

El GAM ajustado es $\log(\text{PM}_{2.5}) = s(\text{Velocidad}) + s(\text{Humedad}) + s(\text{Temperatura})$, la relación entre el $\text{PM}_{2.5}$ y los parámetros meteorológicos es no lineal ($\text{edf} > 1$). la devianza explicada fue 16.9% y fue significativo ($p\text{-valor} < 2e-16$) el efecto de las variables predictoras en explicar el comportamiento de la concentración de $\text{PM}_{2.5}$. La relación es negativa significativa con la velocidad del viento ($\text{edf} > 3$) y la humedad relativa ($\text{edf} > 5$), con la temperatura del aire la relación es significativa ($\text{edf} > 4$) pero la tendencia varió en cada distrito (Figura 21, Tabla 6).

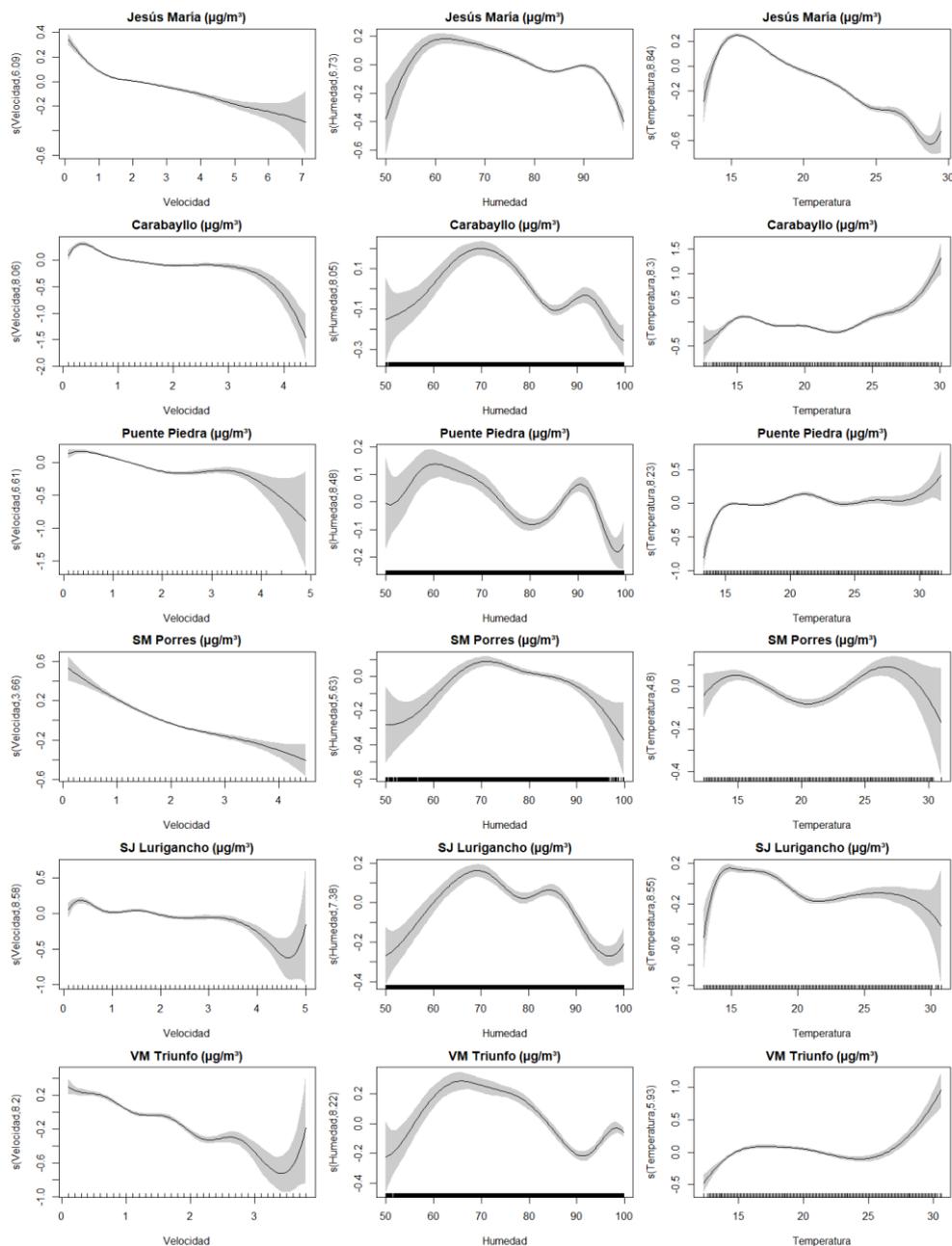


Figura 21. Modelos aditivos generalizados de la relación entre $\text{PM}_{2.5}$ y los parámetros meteorológicos en Lima Metropolitana entre el 2017 y 2020.

Las condiciones ambientales que favorecieron la acumulación del PM_{2.5} fueron la velocidad del viento menor de 1.5 m/s y la humedad relativa menor del 80%. En Carabaylo, Puente Piedra y Villa María del Triunfo, donde la velocidad del viento no supera los 4.3 m/s, la acumulación fue favorecida por la temperatura del aire mayor a 23°C. Sin embargo, en Jesús María, San Juan de Lurigancho y San Martín de Porres donde la velocidad del viento supera los 4.5 m/s, la temperatura del aire mayor a 23°C no influyó sobre la concentración de este contaminante. Los parámetros meteorológicos que contribuyeron a la dispersión del PM_{2.5} son la velocidad del viento mayor de 2 m/s y la humedad relativa mayor del 90% (Figura 21).

Árbol de regresión entre el PM_{2.5} y los parámetros meteorológicos

El modelo del árbol de regresión es $PM_{2.5} \sim \text{Velocidad} + \text{Humedad} + \text{Temperatura}$, excluyendo la dirección del viento que según el análisis GAM no es una variable explicativa de la concentración del PM_{2.5}. El árbol total sin podar ($cp=0.01$) tiene 04 divisiones, el primer corte del árbol lo determina la velocidad del viento, el segundo corte la temperatura del aire y el tercer corte la humedad relativa (Figura 22).

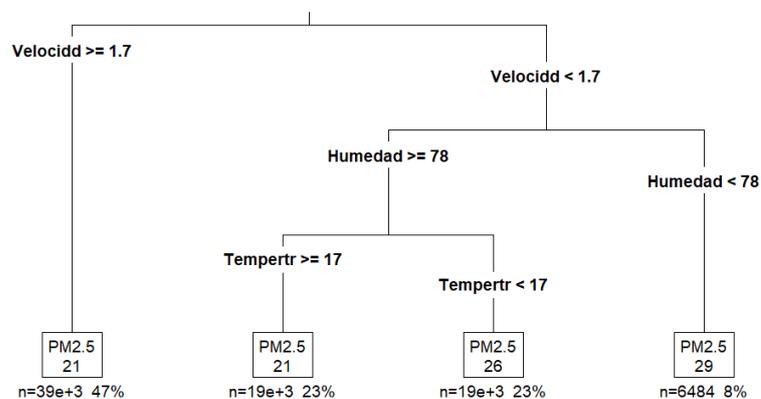


Figura 22. Árbol sin podar de la concentración de PM_{2.5} en relación a los parámetros meteorológicos

Después de la validación cruzada, el árbol presentó tres divisiones y un parámetro de complejidad (cp) de 0.0083; sin embargo, se obtuvo un error relativo ($\text{rel error}=1 - R^2$) en la última división de 0.96145, lo que indica que la variabilidad explicada de los datos fue 3.9% ($R^2=0.039$) (Tabla 20).

Tabla 20. Resumen de la validación cruzada para la poda del árbol de PM_{2.5} y los parámetros meteorológicos

Regression tree:
rpart(formula = PM.2.5 ~ ., data = train, cp = 0)

Variables actually used in tree construction:
[1] Humedad Temperatura Velocidad

Root node error: 18028961/74525 = 241.92
n= 74525

	CP	nsplit	rel error	xerror	xstd
1	0.0106953	0	1.00000	1.00001	0.040467
2	0.0064595	3	0.96791	0.96819	0.039545
3	0.0061712	4	0.96145	0.96207	0.039533

Para ajustar el modelo se adicionaron dos variables explicativas, el año y la calificación para PM_{2.5} según el INCA: “buena”, “moderada”, “mala” y “VUEC” (PM_{2.5} ~ Velocidad + Humedad + Temperatura + AÑO + INCA). Con la validación cruzada se construyeron dos árboles óptimos. En el primer árbol el nodo inicial correspondió a la calidad del aire “buena” y “moderada”, se obtuvo cinco divisiones y un parámetro de complejidad (cp) de 0.043, el error relativo (rel error= 1 - R²) en la última división (0.37759) indica que el árbol explica el 62% de la variabilidad de los datos (R²=0.622), en cada división se reduce el error (xerror) y la desviación estándar (xstd) de la validación cruzada (Tabla 21).

Tabla 21. Validación cruzada para la poda del árbol de regresión entre PM_{2.5} y los parámetros meteorológicos según el INCA “buena” y “moderada”

Regression tree:
rpart(formula = PM.2.5 ~ ., data = train, cp = 0)

Variables actually used in tree construction:
[1] Humedad INCA Temperatura Velocidad

Root node error: 1038943/41128 = 25.261
n= 41128

	CP	nsplit	rel error	xerror	xstd
1	0.6141366	0	1.00000	1.00008	0.0052351
2	0.0029737	1	0.38586	0.38590	0.0021678
3	0.0022592	2	0.38289	0.38293	0.0021514
4	0.0015192	3	0.38063	0.38129	0.0021551
5	0.0011719	5	0.37759	0.37860	0.0021560

El segundo árbol, donde el primer nodo correspondió a la calidad del aire “mala” y “VUEC”, presentó seis divisiones y un parámetro de complejidad (cp) de 0.048, el error relativo (rel error= 1 - R²) en la última división (0.77497) indica que el árbol explica el 22.5% de la variabilidad de los datos (R²=0.225) (Tabla 22).

Tabla 22. Validación cruzada para la poda del árbol entre PM_{2.5} y los parámetros meteorológicos según el INCA “mala” y “VUEC”

Regression tree:
`rpart(formula = PM.2.5 ~ ., data = train, cp = 0)`

Variables actually used in tree construction:
 [1] ANO Humedad INCA Temperatura
 [5] Velocidad

Root node error: 4142058/19857 = 208.59

n= 19857

CP	nsplit	rel error	xerror	xstd
1	0	1.00000	1.00017	0.044331
2	1	0.80077	0.80214	0.017282
3	2	0.78918	0.79060	0.016776
4	4	0.78155	0.79078	0.016788
5	5	0.77826	0.78589	0.016832
6	6	0.77497	0.78058	0.016725

Después de la poda de ambos árboles, el primer nodo estuvo dado por la calificación INCA, el segundo nodo por la velocidad del viento y temperatura del aire, y el tercer nodo por la humedad relativa. En el primer árbol, la concentración de PM_{2.5} presentó calidad “buena”, con la velocidad del viento >1.7 m/s (PM_{2.5}=9.3 µg/m³) y calidad “moderada” con la temperatura del aire >=19°C, la humedad relativa >83% y la velocidad del viento >=3.3 m/s (PM_{2.5}=17 µg/m³). En el segundo árbol, la concentración de PM_{2.5} presento calidad “mala” con la velocidad del viento <1.3 m/s y la humedad relativa <76% (PM_{2.5}=43 µg/m³). Las altas concentraciones de PM_{2.5} (VUEC) correspondieron a la temperatura del aire <23°C (PM_{2.5}=217 µg/m³) (Figura 23).

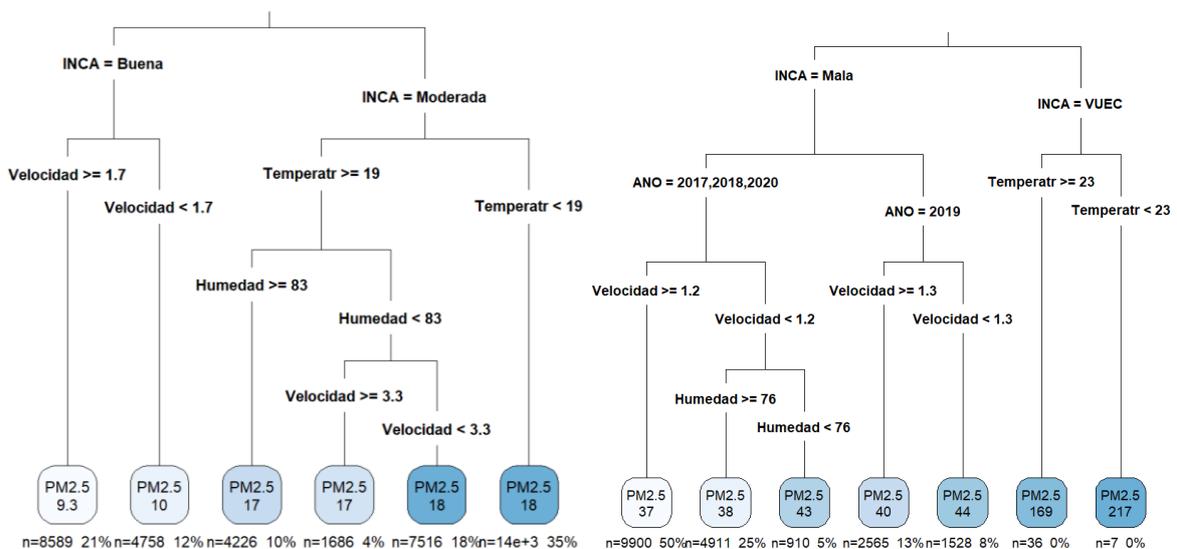


Figura 23. Árboles de regresión entre PM_{2.5} y los parámetros meteorológicos según la calificación del INCA 2017-2020

Partículas con diámetro menor de 10 micrómetros

El GAM ajustado es $\log(\text{PM}_{10}) = s(\text{Velocidad}) + s(\text{Humedad}) + s(\text{Temperatura})$, la relación entre el PM_{10} y los parámetros meteorológicos es no lineal ($\text{edf} > 1$). La devianza explicada fue 38.2% y el efecto de las variables predictoras es significativo ($p\text{-valor} < 2e-16$) en explicar el comportamiento de la concentración de PM_{10} . La relación es negativa significativa con la velocidad del viento ($\text{edf} > 3$) y la humedad relativa ($\text{edf} > 5$), con la temperatura la relación es significativa ($\text{edf} > 4$) pero la tendencia no fue similar en cada distrito (Figura 24, Tabla 9).

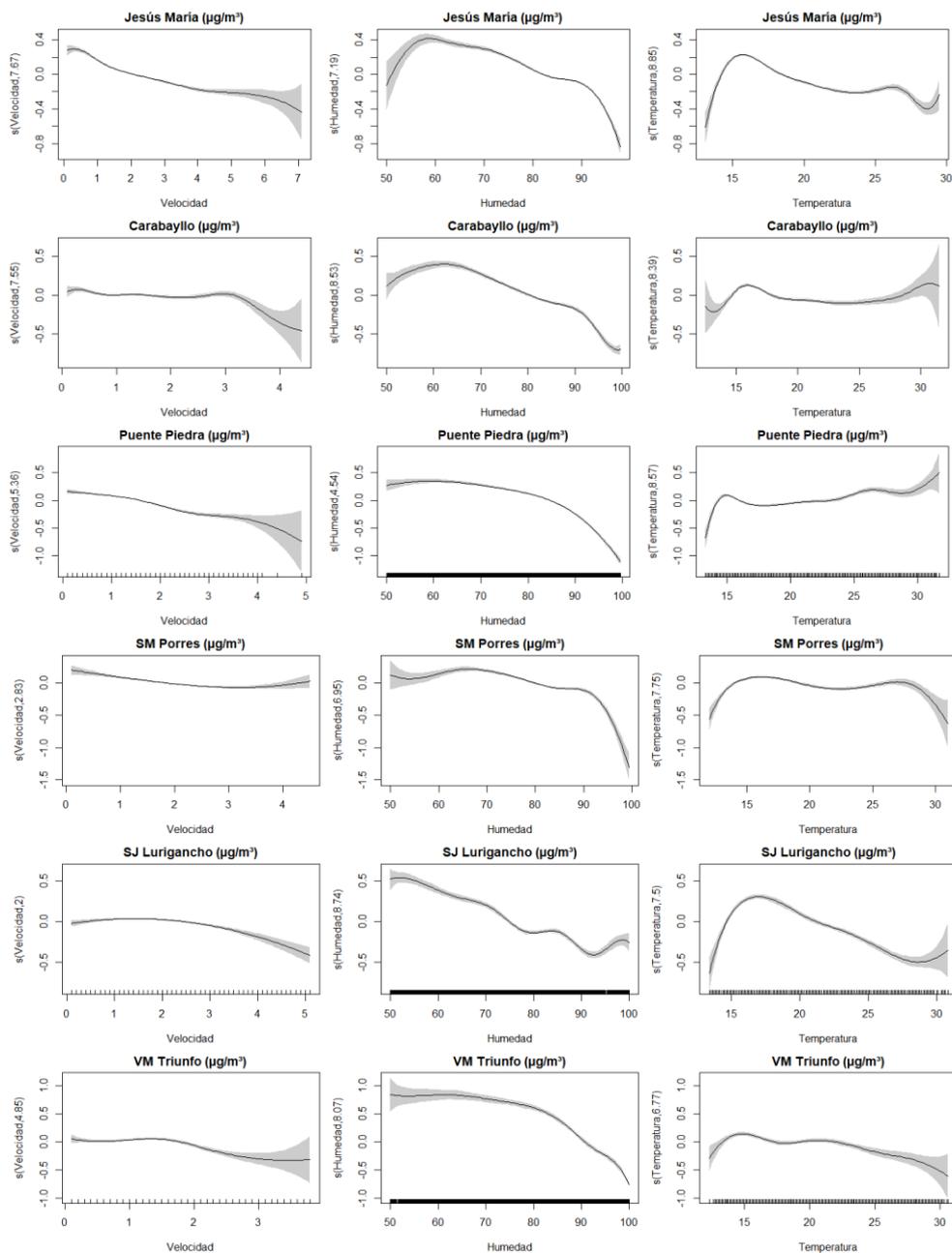


Figura 24. Modelos aditivos generalizados de la relación entre el PM_{10} y los parámetros meteorológicos en Lima Metropolitana entre el 2017 y 2020.

El estancamiento del PM₁₀ en el aire estuvo relacionado a la velocidad del viento menor a 2 m/s, humedad relativa menor del 80% y temperatura del aire mayor a 25°C. En la estación de San Martín de Porres la acumulación del PM₁₀ fue favorecida al disminuir a humedad relativa del 90%. Las condiciones ambientales que contribuyeron en la disminución significativa del PM₁₀ son la velocidad del viento mayor de 2 m/s y la humedad relativa mayor del 90%. Sin embargo, en las estaciones de Carabayllo y San Juan de Lurigancho con el incremento de la velocidad del viento de 3 m/s, descendió significativamente la concentración de PM_{2.5} (Figura 24).

Árbol de regresión entre PM₁₀ y los parámetros meteorológicos

El árbol de regresión se construyó con las variables velocidad del viento, humedad relativa y temperatura del aire (PM₁₀ ~ Velocidad + Humedad + Temperatura), excluyendo la dirección del viento que de acuerdo al análisis GAM no es una variable explicativa de la concentración de material particulado. El árbol total sin podar (cp=0.01) presentó 04 divisiones, el primer corte del árbol estuvo determinado por la humedad relativa y el segundo corte por la velocidad del viento (Figura 25).

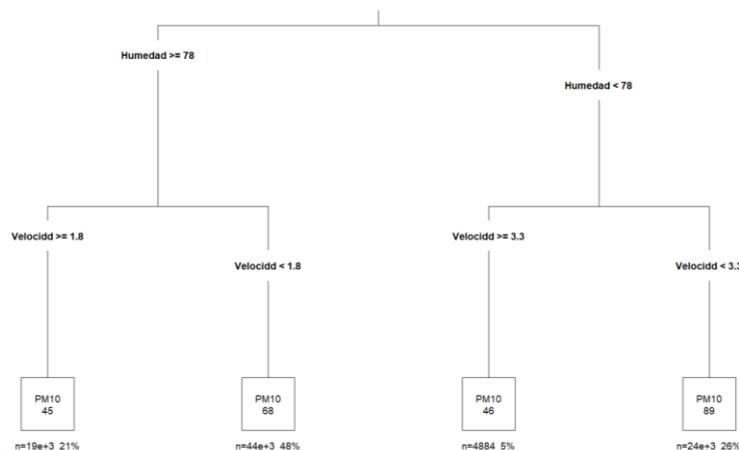


Figura 25. Árbol sin podar de la concentración de PM₁₀ en relación a los parámetros meteorológicos

Después de la validación cruzada, el árbol presentó tres divisiones y un parámetro de complejidad (cp) de 0.028; sin embargo, se obtuvo un error relativo (rel error=1 - R²) en la última división de 0.91056, lo que indica que la variabilidad explicada de los datos fue 8.9% (R²=0.089) (Tabla 39).

Tabla 23. Resumen de la validación cruzada para la poda del árbol de PM₁₀

Regression tree:
 rpart(formula = PM.10 ~ ., data = train, cp = 0)

Variables actually used in tree construction:
 [1] Humedad Velocidad

Root node error: 226596326/73575 = 3079.8

n=73575 (552 observations deleted due to missingness)

CP	nsplit	rel error	xerror	xstd
1 0.0293945	0	1.00000	1.00003	0.017789
2 0.0267535	1	0.97061	0.97066	0.017631
3 0.0255742	2	0.94385	0.94755	0.017452
4 0.0077192	3	0.91828	0.91943	0.017210
5 0.0064922	4	0.91056	0.91252	0.017025

Se adicionó la calificación para PM₁₀ según el INCA: “buena”, “moderada”, “mala” y “VUEC” como variable explicativa (PM₁₀ ~ Velocidad + Humedad + Temperatura + INCA); con la validación cruzada el árbol óptimo presentó 10 divisiones y un parámetro de complejidad (cp) de 0.34, el error relativo (rel error= 1 - R²) en la última división (0.14649) indica que el árbol explica el 85% de la variabilidad de los datos (R²=0.85), en cada división se reduce el error (xerror) y la desviación estándar (xstd) de la validación cruzada (Tabla 40).

Tabla 24. Resumen de la validación cruzada para la poda del árbol de PM₁₀

Regression tree:
 rpart(formula = PM.10 ~ ., data = train, cp = 0)

Variables actually used in tree construction:
 [1] Humedad INCA Temperatura Velocidad

Root node error: 163112935/63031 = 2587.8

n= 63031

CP	nsplit	rel error	xerror	xstd
1 0.57845146	0	1.00000	1.00002	0.0146644
2 0.20266968	1	0.42155	0.42158	0.0100013
3 0.06170030	2	0.21888	0.21897	0.0058171
4 0.00246672	3	0.15718	0.15731	0.0028345
5 0.00174025	4	0.15471	0.15489	0.0028353
6 0.00122299	6	0.15123	0.15214	0.0028337
7 0.00122009	7	0.15001	0.15103	0.0027648
8 0.00080242	8	0.14879	0.14983	0.0026860
9 0.00074592	9	0.14799	0.14934	0.0026965
10 0.00068840	11	0.14649	0.14796	0.0026509

Con la poda del árbol, el primer nodo estuvo dado por la calificación INCA, el segundo por la velocidad del viento y el tercero por la humedad relativa y temperatura del aire. La concentración de PM₁₀ presentó calidad del aire “buena”

(<75 $\mu\text{g}/\text{m}^3$), cuando la velocidad del viento se incrementó de 3.4 m/s ($\text{PM}_{10}=30$ $\mu\text{g}/\text{m}^3$); asimismo, las condiciones fueron optimas cuando la velocidad del viento se incrementó de 1.7 a 3.4 m/s, con la humedad relativa mayor a 77% y la temperatura menor a 19°C ($\text{PM}_{10}=30$ $\mu\text{g}/\text{m}^3$). La concentración de PM_{10} presentó la calidad del aire “VUEC” (>250 $\mu\text{g}/\text{m}^3$) al disminuir la velocidad de viento de 1.3 m/s y la temperatura del aire de 23°C ($\text{PM}_{10}=357$ $\mu\text{g}/\text{m}^3$) (Figura 26).

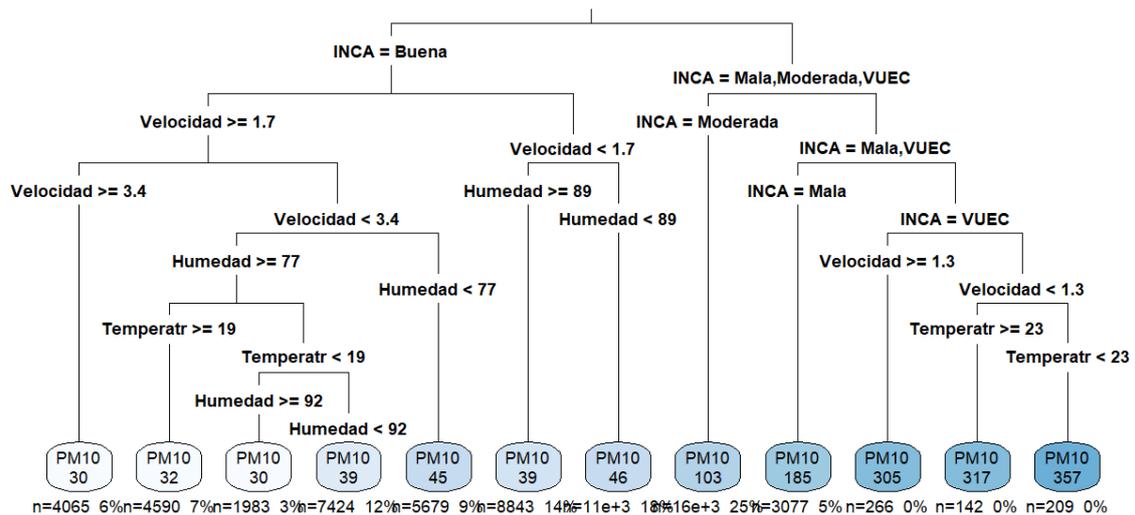


Figura 26. Árbol de regresión entre PM_{10} y los parámetros meteorológicos en las estaciones de Lima Metropolitana

V. DISCUSIÓN

Entre los años 2017 y 2020, las estaciones de Puente Piedra, Carabayllo, Ate, San Juan de Lurigancho, Santa Anita y Villa María del Triunfo sobrepasaron el criterio de evaluación para la concentración de PM_{10} y $PM_{2.5}$, a excepción del ECA para 24 horas del $PM_{2.5}$ que no fue excedido en Lima Metropolitana. Las estaciones cercanas al mar como Jesús María, San Borja y San Martín de Porres no sobrepasaron el criterio de evaluación del ECA para material particulado. La contaminación en Lima Metropolitana proviene de los combustibles quemados por el parque automotor y las zonas industriales; sin embargo, se observó diferencias en los niveles de contaminación dependiendo de la proximidad al mar, que se pueden atribuir a los 6 u 8 microclimas de Lima Metropolitana, que varían por el factor topográfico, la presencia de cerros o laderas, así como su conversión en cuencas donde los vientos confinan la humedad proveniente del mar.

Respecto a la evolución horaria de la calidad del aire, la concentración del $PM_{2.5}$ es insalubre principalmente entre las 01 am-12 pm y 9 pm-12 am en las estaciones de San Juan de Lurigancho, Ate, Puente Piedra y Santa Anita, donde la calidad “mala” superó el 60% del total de horas; la calidad “Valor del umbral de cuidado” estuvo presente en la estación de Ate. La concentración del PM_{10} es insalubre entre las 5 am-12 pm y 5 pm-12 am, en las estaciones de Villa María del Triunfo, Ate y Puente Piedra donde el tiempo de exposición a la calidad del aire “mala” representó entre 11.8 y 21.4% del total de horas. La calidad “Valor del umbral de cuidado” se registró en Villa María del Triunfo y Ate. La mala calidad del aire observada entre las 5 am-12 am y 5 pm-12 pm coincide con las horas punta, donde regularmente se produce congestión o sobredemanda de las autopistas o avenidas principales por la saturación del transporte público debido al ingreso y salida de la población económicamente activa de sus centros laborales.

Los métodos de interpolación co-kriging ordinario con la covariable velocidad del viento y kriging ordinario han demostrado que la distribución espacial de la concentración de material particulado mayor al ECA abarcó entre 59.7% y 97.1% del área donde se localizan las estaciones de monitoreo en Lima Metropolitana,

ambos métodos son más eficientes en la caracterización de la distribución espacial del $PM_{2.5}$ y PM_{10} , el método de interpolación IDW tendió a generar alrededor de los puntos muestreados patrones poco reales como tipo “ojo de buey”. La eficiencia de un método en la predicción de la distribución espacial depende de la ubicación adecuada de las estaciones o puntos de colección de datos, las estaciones de monitoreo del SENAMHI se localizan con relativa homogeneidad y vecinas entre sí. Ibrahimy y Abdelkarim (2017), determinaron que el método de interpolación kriging da mejores resultados que el método IDW en términos de precisión, cuando hay un número de puntos vecinos óptimo y una selección adecuada del modelo para el variograma.

La alta variabilidad de la concentración de partículas y las interacciones atmosféricas, demuestran que entre ellos existe una relación compleja y multivariante que es difícil de predecir con las técnicas estadísticas tradicionales como los modelos de regresión lineal simple, regresión lineal múltiple o GLM. A través de los GAM se ha determinado que la relación entre el material particulado y los parámetros meteorológicos es no lineal según los grados de libertad efectivos ($edf > 1$) obtenidos en cada estación de monitoreo. La concentración del material particulado tiene relación negativa significativa con la velocidad del viento ($edf > 2$, $p_valor < 2e-16$) y la humedad relativa ($edf > 4$, $p_valor < 2e-16$), sin embargo, la relación con la temperatura del aire es significativa ($edf > 4$, $p_valor < 2e-16$) pero no tuvo el mismo patrón o tendencia en cada estación. Los edf indican el número de polinomios de la función de suavizado, es decir la rugosidad o complejidad del modelo, la función representa una línea recta o es lineal con un valor edf igual o menor a 1. Los GAM permitieron precisar las condiciones que influyen en la variación de la concentración de partículas; el estancamiento se produce cuando la velocidad del viento es menor de 1.5 m/s, la humedad relativa es menor del 80% y la temperatura del aire es mayor de 25°C, las partículas se disipan cuando la velocidad del viento es mayor de 2 m/s y la humedad relativa es mayor del 90%. Los GAM modelan la forma funcional de la variable respuesta en relación a una o varias variables explicativas con funciones de suaves, el modelo se ajusta como una regresión por pedazos o splines que son fragmentos de polinomios unidos por nodos, el método de

suavizado por splines es una técnica que representa la variación del promedio de la variable respuesta.

Con la técnica por árboles de regresión se categorizaron los valores de las condiciones ambientales relacionadas con la variación de la concentración de $PM_{2.5}$ y PM_{10} ; los resultados obtenidos son similares a los GAM, las partículas se acumulan en el aire alcanzando condiciones insalubres cuando la velocidad del viento es menor de 1.3 m/s, la humedad relativa es menor del 77% y la temperatura es mayor de 23°C; por el contrario, las partículas se dispersan siendo la calidad del aire optima cuando la velocidad del viento es mayor o igual de 3.4 m/s, la humedad relativa es mayor o igual al 77% y la temperatura menor a 19°C. Cuando aumenta la temperatura atmosférica, el aire se vuelve más seco y al disminuir es más húmedo, la alta temperatura disminuye la humedad relativa del aire y el material particulado se mantiene suspendido tendiendo a acumularse en la atmósfera. Según Li et al. (2017), en China la fuerte velocidad del viento favoreció la dispersión de la concentración de material particulado y la acumulación en la capa límite fue favorecida por la alta temperatura del aire, siendo inversa la relación entre la concentración de material particulado y las variables meteorológicas.

Las investigaciones que utilizaron el coeficiente de Pearson no encontraron una tendencia o patrón en la relación entre el nivel de concentración de material particulado y los parámetros meteorológicos, siendo directa u opuesta fuerte, débil o muy débil entre la temperatura del aire y la humedad relativa con el nivel de concentración de PM_{10} y $PM_{2.5}$, la relación con la velocidad del viento fue negativa baja, muy baja o nula. Según el análisis exploratorio de los datos el uso de promedios altera la relación entre las variables, que se aproximan a una línea recta, sin embargo, esto puede ocasionar sobreestimación o subestimación en el valor del índice de correlación. Ortiz y Ortiz (2021) destacan que, si la verdadera relación entre dos variables no es lineal, al utilizar el coeficiente de Pearson como un indicador del grado en que ambas se relacionan, podemos cometer todo tipo de errores. Además, advierten sobre los sesgos de recomendar el coeficiente de Spearman como “la alternativa no paramétrica” al coeficiente de Pearson al provenir los datos de distribuciones no normales,

siendo un indicador de tendencia monótona que corresponde a un enfoque diferente o complementario del análisis.

Las relaciones lineales entre los organismos y el medio ambiente se han usado a menudo sin evidencia o sentido crítico (Ravindra et al., 2019), los modelos aditivos generalizados se utilizan con frecuencia para evaluar las relaciones no lineales entre la ocurrencia y las variables ambientales (Knudby et al., 2010, Parra et al., 2017, Schismenou et al., 2017). El buen desempeño y capacidad predictiva de los GAM en el modelado de la relación entre variables biológicas, ecológicas y los parámetros del hábitat o ambientales ha sido destacado en numerosas disciplinas de investigación, como la salud pública, el turismo, la ecología, la biología y el ambiente (Knudby et al., 2010, Parra et al., 2017, Schismenou et al., 2017). Debido a sus ventajas el GAM se ha aplicado cada vez más en la investigación ambiental (Jacob y Winner, 2009, Ravindra et al., 2017, Ravindra et al., 2019, Tudor, 2022).

En las entidades científicas y universidades públicas han utilizado el GAM y el método de agrupamiento basado en árboles de clasificación y regresión (CART), para relacionar índices biológicos con las variables ambientales del mar peruano. Bertrand et al. (2004) evaluó las relaciones no lineales y multivariantes entre la abundancia de peces pelágicos y las variables abióticas del mar durante El Niño y La Niña, aplicando un enfoque de GAM para estimar estas funciones no paramétricas. Gutiérrez et al. (2007) utilizó CART para relacionar tres índices biológicos: concentración, ocupación del espacio y agregación de gran abundancia de peces con las variables ambientales, temporales y espaciales. Alegre (2011) aplicó un GAM en el estudio de las relaciones espacio-temporales y ontogénicas del calamar gigante, las variables fueron biológicas, temporales, espaciales y ambientales. Adams y Flores (2016) aplicó GAM para evaluar la relación entre el Niño y la disponibilidad de los recursos pesqueros de la pesquería artesanal de ICA. Los GAM y otras técnicas no paramétricas como los árboles de regresión han mejorado la confiabilidad en los modelos estadísticos y la obtención de aproximaciones para cuantificar las relaciones entre el comportamiento espacio-temporal de los organismos marinos y los parámetros oceanográficos, además de identificar mejor el hábitat y construir modelos de predicción de los principales recursos pesqueros del mar peruano.

VI. CONCLUSIONES

- Las estaciones de monitoreo de Puente Piedra, Carabayllo, Ate, San Juan de Lurigancho, Santa Anita y Villa María del Triunfo excedieron el ECA para 24 horas y anual del PM_{10} , así como el ECA anual para $PM_{2.5}$; las estaciones próximas al mar como Jesús María, San Borja y San Martín de Porres no excedieron el ECA para material particulado, tales diferencias se pueden atribuir a los microclimas de Lima Metropolitana, que varían por el factor topográfico, la presencia de cerros o laderas, así como su conversión en cuencas donde los vientos confinan la humedad proveniente del mar.
- La concentración de $PM_{2.5}$ y PM_{10} mayor al ECA-aire abarcó entre 59.7 y 97.1 % del área total de las estaciones de monitoreo. Los métodos de interpolación co-kriging y kriging son más eficientes en la caracterización de la distribución espacial del material particulado, a diferencia del IDW que generó alrededor de los puntos muestreados patrones poco reales como tipo ojo de buey.
- La calidad del aire para material particulado fue “buena” y “moderada” en las estaciones de Jesús María, San Borja y San Martín de Porres, las condiciones insalubres se presentaron entre 01 am-12 pm y 5 pm-12 am, la calidad “mala” predominó en las estaciones de San Juan de Lurigancho, Ate, Puente Piedra, Santa Anita y Villa María del Triunfo, que representó entre 21% y 77.6% del total de horas; la calidad “Valor del umbral de cuidado” se presentó en las estaciones de Ate y Villa María del Triunfo representando entre 1.6% y 3.4 % del total de horas. La calidad del aire es deteriorada principalmente en las horas punta, donde hay sobredemanda de las autopistas o avenidas principales por el ingreso y salida de la población de sus centros laborales.
- La relación entre el material particulado y parámetros meteorológicos es no lineal, con la velocidad del viento y humedad relativa existe una relación negativa, con la temperatura del aire no tuvo una tendencia o patrón. Las partículas se disipan cuando la velocidad del viento es mayor a 3.4 m/s, la humedad relativa es mayor de 77% y la temperatura menor a 19°C; las partículas se acumulan cuando la velocidad del viento es menor de 1.3 m/s, la humedad relativa es menor de 77 % y la temperatura es mayor de 23°C.

VII. RECOMENDACIONES

- Previo al uso de cualquier técnica estadística, es necesario efectuar un análisis exploración de los datos para evaluar la no linealidad, ya que las variables biológicas, ecológicas o ambientales rara vez se ajustan a los supuestos de los modelos lineales, lo que puede provocar resultados no completamente ciertos.
- Para utilizar el coeficiente de correlación de Pearson es necesario comprobar que la relación entre la variable respuesta y predictora es lineal, y para el coeficiente de Spearman los datos deben estar ordenados por rangos.
- Los datos ambientales rara vez provienen de distribuciones normales, por lo cual se recomienda una transformación logarítmica para reducir los valores atípicos u outliers, evitando el uso de promedios diarios, mensuales o anuales, ya que reducen el número de observaciones, alteran la relación funcional entre las variables y provocan sesgos como la sobreestimación de los coeficientes de correlación de Pearson y Spearman.
- Es importante difundir a los estudiantes y profesionales de la carrera de ingeniería ambiental, las ventajas de los métodos de regresión no paramétrica respecto a los modelos estadísticos tradicionales. Los modelos aditivos generalizados y árboles de regresión hace décadas se utilizan para estudiar la relación entre los organismos y los parámetros oceanográficos del mar peruano y se aplican cada vez más en los estudios de la contaminación ambiental.
- Es necesario dar importancia al aprendizaje de programas informáticos y métodos estadísticos aplicados actualmente en el procesamiento y análisis de datos ambientales, ya que la mayoría de los proyectos de investigación de los estudiantes de la escuela profesional de ingeniería ambiental de la UCV están dirigidos principalmente a coleccionar de datos a través de prospecciones o monitoreos.

REFERENCIAS

ADAMS, Grant y FLORES, Daniel. Influencia de El Niño Oscilación del Sur en la disponibilidad y abundancia de recursos hidrobiológicos de la pesca artesanal en Ica, Perú [en línea]. Scielo: 2016 [Fecha de consulta: 15 de diciembre 2023].

Disponible en: https://www.scielo.cl/scielo.php?pid=S0718-19572016000200005&script=sci_abstract

ISSN: 0718-1957.

ALEGRE, Ana. Relaciones ontogénicas y espacio-temporales en la dieta del calamar gigante (*Dosidicus gigas*) en Perú, utilizando un modelo aditivo generalizado. Tesis (Magister Scientiae). Lima: Universidad Nacional Agraria La Molina, 2011. Disponible en:

https://biblioimarpe.imarpe.gob.pe/bitstream/20.500.12958/2343/1/Alegre%2C%20Ana_Mestria.pdf.

ANALITIS, Antonis [et al.]. Prediction of PM2.5 concentrations at the locations of monitoring sites measuring PM10 and NOx, using generalized additive models and machine learning methods: A case study in London. *Atmospheric Environment* [en línea]. Vol. 240, Noviembre 2020. [Fecha de consulta: 30 de abril]. Disponible en: <https://doi.org/10.1016/j.atmosenv.2020.117757>

ISSN: 1352-2310.

BERTHAND, Arnaud. From small-scale habitat loopholes to decadal cycles: a habitat-based hypothesis explaining fluctuation in pelagic fish populations of Peru. *Fish and Fisheries* [en línea]. Diciembre 2004. [Fecha de consulta: 04 de mayo de 2023]. Disponible en: <https://doi.org/10.1111/j.1467-2679.2004.00165.x>

CARDONA, Néstor. Predicción y selección de variables con bosques aleatorios en presencia de variables correlacionadas. Tesis (Maestría en Ciencias – estadística). Colombia: Universidad Nacional de Colombia, 2019. Disponible en: <https://repositorio.unal.edu.co/bitstream/handle/unal/75561/8063120.2019.pdf?isAllowed=y&sequence=1>

ESTIMATES and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015 por J Cohen Aaron [et al.]. The lancet [en línea]. Abril 2017, Vol. 389, n.o1082. [Fecha de consulta: 14 de diciembre del 2023].

Disponible en: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(17\)30505-6/fulltext#%20](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)30505-6/fulltext#%20)

ISSN: S0140-6736(17)31003-6.

COPEES Artica, Rafael. Efecto de la estabilidad atmosférica en la dispersión del material particulado en la ciudad de Huancayo (Enero – Marzo). Tesis (Maestro en Ingeniería Ambiental). Huancayo: Universidad Nacional del Centro del Perú, 2022.

Disponible en: https://repositorio.uncp.edu.pe/bitstream/handle/20.500.12894/8443/T010_7017_6207_M.pdf?sequence=1&isAllowed=y

Cada año mueren 12,6 millones de personas a causa de la insalubridad del medio ambiente [Comunicado de prensa]. Ginebra (15 de marzo de 2016). [Fecha de consulta: 12 de julio del 2023]. Recuperado de: <https://www.who.int/es/news/item/15-03-2016-an-estimated-12-6-million-deaths-each-year-are-attributable-to-unhealthy-environments>

Cokriging [en línea]. Surfer Help (2009). [Fecha de consulta: 04 de mayo de 2023]. Recuperado de: <https://surferhelp.goldensoftware.com/griddata/cokriging.htm>

Cómo funciona la interpolación de distancia inversa ponderada [en línea]. esri, (2023). [Fecha de consulta: 05 de mayo de 2023]. Recuperado de: <https://pro.arcgis.com/es/pro-app/latest/help/analysis/geostatistical-analyst/how-inverse-distance-weighted-interpolation-works.htm>

DONG, Jiyuan, LIU, Yurong y BAO Hairong. Revalue associations of short-term exposure to air pollution with respiratory hospital admissions in Lanzhou, China after the control and treatment of current pollution [en línea]. Science direct 2021,

Vol. 231. [Fecha de consulta: 15 de diciembre 2023]. Disponible en:
<https://www.sciencedirect.com/science/article/pii/S1438463920306040>
ISSN 1438-4639.

FERNÁNDEZ, Carlos y LAVADO, Waldo. Análisis comparativo de métodos de interpolación espacial para la elaboración de mapas climáticos de precipitación. SENAMHI [en línea]. Septiembre 2014. [Fecha de consulta: 04 de mayo de 2023]. Disponible en:
<https://repositorio.SENAMHI.gob.pe/bitstream/handle/20.500.12542/947/Análisis-comparativo-de-métodos-de-interpolación-espacial-para-la-elaboración-de-mapas-climáticos-de-precipitación.pdf?sequence=1&isAllowed=y>

FERREIRA, Jorge, COELHO, Arnaldo y MOUTINHO, Luis. Dynamic capabilities, creativity and innovation capability and their impact on competitive advantage and firm performance: The moderating role of entrepreneurial orientation [en línea]. Research gate: 2018 [Fecha de consulta: 15 de diciembre 2023]. Disponible en:
https://www.researchgate.net/publication/329602816_Dynamic_capabilities_creativity_and_innovation_capability_and_their_impact_on_competitive_advantage_and_firm_performance_The_moderating_role_of_entrepreneurial_orientation
DOI:[10.1016/j.technovation.2018.11.004](https://doi.org/10.1016/j.technovation.2018.11.004)

FOTOUREHCHI, Zahra. Health effects of air pollution: An empirical analysis for developing countries. Atmospheric Pollution Research [en línea]. Enero 2016, Vol. 7(1). [Fecha de consulta: 04 de mayo de 2023]. Disponible en:
<https://doi.org/10.1016/j.apr.2015.08.011>

SHORT-term PM₁₀ and emergency department admissions for selective cardiovascular and respiratory diseases in Beijing, China por Wei Feng [et al]. Science direct [en línea]. Marzo 2019, Vol. 657. [Fecha de consulta: 15 de diciembre 2023]. Disponible en:
<https://www.sciencedirect.com/science/article/abs/pii/S0048969718349064>
ISSN 0048-9697.

MODELO de Árbol de Decisión: una herramienta para el manejo de la pradera por Flores Paulina [et al]. Agro Sur [en línea]. Noviembre 2016. [Fecha de consulta: 15 de diciembre 2023].

Disponible en: <http://revistas.uach.cl/pdf/agrosur/v44n2/art02.pdf>

DOI:[10.4206/agrosur.2016.v44n2-02](https://doi.org/10.4206/agrosur.2016.v44n2-02)

USING regression tree ensembles to model interaction effects: a graphical approach por Schiltz Fritz [et al]. [Tandf online](https://doi.org/10.1080/00036846.2018.1489520) [en línea]. Julio 2018, Vol. 50, n.o 58. [Fecha de consulta: 15 de diciembre 2023]. Disponible en: <https://doi.org/10.1080/00036846.2018.1489520>

ASSESSING the value of air stagnation indices to reproduce PM10 variability in Europe. Atmospheric Research por Garrido Pérez José [et al.]. Sciedirect [en línea]. Enero 2021, Vol. 248. [Fecha de consulta: 30 de abril 2023].

Disponible en: <https://doi.org/10.1016/j.atmosres.2020.105258>

ISSN 0169-8095.

GALLARDO, Henry [et al.]. Modelación de series temporales en el sector productivo del norte de Santander [en línea]. Bogotá: Universidad Francisco de Paula Santander, 2019. Disponible en: [https://repositorio.ufps.edu.co/bitstream/handle/ufps/897/2.Modelaci%C3%B3n%20de%20series%20temporales%20en%20el%20sector%20productivo%20ebook%20\(1\).pdf?sequence=1](https://repositorio.ufps.edu.co/bitstream/handle/ufps/897/2.Modelaci%C3%B3n%20de%20series%20temporales%20en%20el%20sector%20productivo%20ebook%20(1).pdf?sequence=1)

ASSESSING neighborhood variations in ozone and PM_{2.5} concentrations using decision tree method por Gao Ya [et al.]. Science Direct [en línea]. Enero 2021, Vol. 188, [Fecha de consulta: 15 de diciembre del 2023].

Disponible en: <https://www.sciencedirect.com/science/article/pii/S0360132320308465>

ISSN: 1873-684X.

THE mortality impact of fine particulate matter in China: Evidence from trade shocks por GONG, Yazhen [et al.]. Journal of Enviromental Economics and

Management. [en línea].Vol.177, enero 2023. [Fecha de consulta:14 de diciembre de 2023]. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0095069622001127>

HARO, [Silvia](#). Decision Tree, Application With Meteorological Data [en línea]. Ecuador: ESPOCH, 2020 [Fecha de consulta: 15 de diciembre 2023]. Disponible en: <https://knepublishing.com/index.php/KnE-Engineering/article/view/6217>
DOI: [10.18502/keg.v5i2.6217](https://doi.org/10.18502/keg.v5i2.6217)

HÉRNANDEZ, Lalinde [et al.]. Sobre el uso adecuado del coeficiente de correlación de Pearson: Definición, propiedades y suposiciones. Archivos venezolanos de farmacología y terapeutica.[en línea]. Vol.37, Mayo 2018. [Fecha de consulta 22 de diciembre]. Disponible en: <https://www.redalyc.org/journal/559/55963207025/55963207025.pdf>
[ISSN: 0798- 0264](#)

IBRAHIM, Mohamed y ABDELKARIM Hassan. Comparison between Inverse Distance Weighted (IDW) and Kriging [en línea]. IJSR: 2017, Vol. 6. [Fecha de consulta: 14 de Diciembre 2023]. Disponible en: <https://www.ijsr.net/archive/v6i11/ART20177562.pdf>
ISSN: 2319-7064.

JACOB, Daniel y WINNER, Darrell. Effect of climate change on air quality. Atmospheric Environment [en línea]. Enero 2009, Vol. 43(1). [Fecha de consulta: 04 de mayo de 2023]. Disponible en: <https://doi.org/10.1016/j.atmosenv.2008.09.051>

JIN, Zhexu. Enriched spatial analysis of air pollution: Application to the city of Bogotá, Colombia. Frontiers in Environmental Science [en línea]. Septiembre 2022, Vol. 10. [Fecha de consulta: 04 de mayo de 2023]. Disponible en: <https://doi.org/10.3389/fenvs.2022.966560>

KAMBOJ, Kuldeep [et al.]. Assessment and Spatial Distribution Mapping of Criteria Pollutants. *Water, Air, & Soil Pollution* [en línea]. Febrero 2022, Vol. 233(82). [Fecha de consulta: 04 de mayo de 2023]. Disponible en: <https://link.springer.com/article/10.1007/s11270-022-05522-y>

KNUDBY, Anders, BRENNING, Alexander y LEDREW, Ellsworth. New approaches to modelling fish–habitat relationships. *Ecological Modelling* [en línea]. Febrero 2010, Vol. 221(3). [Fecha de consulta: 04 de mayo de 2023]. Disponible en: <https://doi.org/10.1016/j.ecolmodel.2009.11.008>

LEQUY, Emeline [et al.]. Modeling exposure to airborne metals using moss biomonitoring in cemeteries in two urban areas around Paris and Lyon in France. *Environmental Pollution* [en línea]. Junio 2022, Vol. 303. [Fecha de consulta: 04 de mayo de 2023]. Disponible en: <https://doi.org/10.1016/j.envpol.2022.119097>

LI, Xiaolan [et al.]. Temporal and spatial analyses of particulate matter (PM10 and PM2.5) and its relationship with meteorological parameters over an urban city in northeast China. *Atmospheric Research* [en línea]. Diciembre 2017, Vol. 198. [Fecha de consulta: 04 de mayo de 2023]. Disponible en: <https://doi.org/10.1016/j.atmosres.2017.08.023>

MIBORROW, Stephen. rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'[en línea]. V3.1.1. rdrv, 2021 [Fecha de consulta: 04 de mayo de 2023]. Disponible en: <https://rdrv.io/cran/rpart.plot/>.

MAMANI Turpo, Leonardo. Proceso de estimación de recursos minerales y aplicación de los métodos geoestadísticos aplicado en los cuerpos mineralizados 1, 2 y 5 en la unidad cerro lindo. Tesis (Magíster en Ciencias de la Tierra). Arequipa: Universidad Nacional San Agustín de Arequipa, 2020. [Disponible en: https://repositorio.unsa.edu.pe/items/1c61ae94-40e6-44b3-b6b3-514b27958c71](https://repositorio.unsa.edu.pe/items/1c61ae94-40e6-44b3-b6b3-514b27958c71)

ORTIZ, Jorge y ORTIZ, Andrés. ¿Pearson y Spearman, coeficientes intercambiables? *Comunicaciones en Estadística* [en línea]. 2021, Vol. 14(1). [Fecha de consulta: 12 de julio del 2023]. Disponible en: <https://doi.org/10.15332/23393076.6769>

PACSI, Sergio y MURRIEL, Federico. Evaluación espaciotemporal del material particulado PM_{2.5} y su relación con las variables meteorológicas en la Universidad Nacional Agraria La Molina. [en línea]. Diciembre 2018, Vol.79, n.º 2. [Fecha de consulta: 06 de Diciembre 2023].

Disponible

en:

https://revistas.lamolina.edu.pe/index.php/acu/article/view/992/pdf_109

ISSN: 2519-7398.

PARRA, Hugo [et al]. Predictive modeling of deep-sea fish distribution in the Azores. Deep Sea Research Part II: Topical Studies in Oceanography [en línea]. Noviembre 2017, Vol. 145. [Fecha de consulta: 12 de julio del 2023]. Disponible en: <https://doi.org/10.1016/j.dsr2.2016.01.004> ISSN 0967-0645.

The incidence and severity of air pollution in India is on the rise and urgent action is needed to halt its harmful health impacts, Ravindra Khaiwal writes. Policy forum. 15 de diciembre 2016. Disponible en: <https://www.policyforum.net/air-pollution-india-delhi/>

RAVINDRA, Khaiwal [et al.]. Generalized additive models: Building evidence of air pollution, climate change and human health. *Environment International* [en línea]. Vol. 132, Noviembre 2019. [Fecha de consulta: 30 de abril]. Disponible en: <https://doi.org/10.1016/j.envint.2019.104987> ISSN: 0160-4120.

RAVINDRA, Khaiwal [et al.]. Real-time monitoring of air pollutants in seven cities of North India during crop residue burning and their relationship with meteorology and transboundary movement of air. *Science of The Total Environment* [en línea]. Noviembre 2019, Vol. 690. [Fecha de consulta: 12 de julio del 2023]. Disponible en: <https://doi.org/10.1016/j.scitotenv.2019.06.216>

ROBLES Morales, Eder. La contaminación del aire por material particulado y su relación con las enfermedades de tipo respiratorio en la población de Cerro de

Pasco, 2010 y 2016. Tesis (Magíster en Ciencias Ambientales). Lima: Universidad Nacional Mayor de San Marcos, 2020.

Disponible en: <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/11675>

ROJAS Jhojan y MEDINA Elvis. Pronóstico de las concentraciones de material particulado en el aire (p_{10}) utilizando redes neuronales artificiales: caso estudio en el distrito de Ate, Lima [en línea]. Octubre 2022, Vol.88, n.º 3. [Fecha de consulta: 06 de Diciembre 2023].

Disponible

en:

http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1810-634X2022000300265

ISSN: 1810-634X.

ROJAS, Freddy, PACSI, Sergio y SÁNCHEZ, Odón. Simulación computacional e influencia de las variables meteorológicas en las concentraciones de PM₁₀ y PM_{2.5} en Lima Metropolitana [en línea]. Junio 2022, Vol.33, n.º 3. [Fecha de consulta: 06 de Diciembre 2023].

Disponible en: https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-07642022000300223

ISSN: 0718-0764.

ENDOTOXIN and particulate matter emitted by livestock farms and respiratory health effects in neighboring residents Rooij Myrna [et al]. Science direct [en línea]. Noviembre 2019, Vol. 132. [Fecha de consulta: 15 de diciembre 2023].

Disponible

en:

<https://www.sciencedirect.com/science/article/pii/S0160412019311365>

ISSN 0160-4120.

SCHISMENOU, Eudoxia [et al.]. Modelling small pelagic fish potential spawning habitats: eggs vs spawners and in situ vs satellite data. Hydrobiologia [en línea]. Septiembre 2016, Vol. 788. [Fecha de consulta: 12 de julio del 2023]. Disponible en: <https://link.springer.com/article/10.1007/s10750-016-2983-4>

Servicio Nacional de Meteorología e Hidrología del Perú. Boletín hidroclimático regional dirección zonal 10, Huánuco (junio 2021). SENAMHI-Institucional [en línea]. Junio 2021. No. 06. [Fecha de consulta: 12 de julio del 2023]. Disponible en: <https://hdl.handle.net/20.500.12542/1120>

SERVEN, Daniel y BRUMMITT, Charlie. pyGAM: Generalized Additive Models in Python. [en línea]. v0.4.1. Zenodo, 2018 [Fecha de consulta: 12 de julio de 2023]. Disponible en: <https://doi.org/10.5281/zenodo.1476122>

SUÁREZ, Luis[et al.]. Caracterización química del material particulado atmosférico del centro urbano de Huancayo, PERÚ [en línea]. Vol.83, n.º2. [Fecha de consulta: 14 de Diciembre 2023]. Disponible en: <http://www.scielo.org.pe/pdf/rsqp/v83n2/a05v83n2.pdf>

SHUKLA, Komal [et al.]. Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi. *Sustainable Cities and Society* [en línea]. Vol. 54, Junio 2020. [Fecha de consulta: 30 de abril]. Disponible en: <https://doi.org/10.1016/j.scs.2019.101997> ISSN: 2210-6707.

SHUKLA, K. [et al.]. Trend and variability of atmospheric ozone over middle Indo-Gangetic Plain: impacts of seasonality and precursor gases. *Environ Sci Pollut Res* [en línea]. Enero 2017, Vol 4(1). [Fecha de consulta: 12 de julio del 2023]. Disponible en: <https://doi.org/10.1007/s11356-016-7738-2>

TEJEDA, Adalberto. La humedad en la atmósfera bases físicas, instrumentos y aplicaciones [en línea]. Colima: Universidad de Colima, 2018. [Fecha de consulta: 14 de Diciembre 2023]. Disponible en: http://ww.ucol.mx/content/publicacionesenlinea/adjuntos/La-humedad-en-la-atmosfera_466.pdf

THERNEAU, Terry, ATKINSON, Beth y RIPLEY Brian. Recursive Partitioning and Regression Trees [en línea]. Cran: 2023 [Fecha de consulta: 15 de diciembre 2023].

Disponible en: <https://cran.r-project.org/web/packages/rpart/rpart.pdf>

TUDOR, Cristiana. Ozone pollution in London and Edinburgh: spatiotemporal characteristics, trends, transport and the impact of COVID-19 control measures. *Heliyon* [en línea]. Vol. 8(11), Noviembre 2022. [Fecha de consulta: 30 de abril]. Disponible en: <https://doi.org/10.1016/j.uclim.2022.101313>
ISSN: 2405-8440.

WANG, Yu [et al.]. The impact of carbon monoxide on years of life lost and modified effect by individual- and city-level characteristics: Evidence from a nationwide time-series study in China. *Ecotoxicology and Environmental Safety* [en línea]. Vol. 210, Marzo 2021. [Fecha de consulta: 30 de abril]. Disponible en: <https://doi.org/10.1016/j.ecoenv.2020.111884>
ISSN: 0147-6513.

Una comparación de los métodos de correlación de Pearson y Spearman [en línea]. minitab [fecha de consulta: 22 diciembre de 2023]. Disponible en: <https://support.minitab.com/es-mx/minitab/21/help-and-how-to/statistics/basic-statistics/supporting-topics/correlation-and-covariance/a-comparison-of-the-pearson-and-spearman-correlation-methods/>

WANG, Ce [et al.]. The impact of ambient particulate matter on hospital outpatient visits for respiratory and circulatory system disease in an urban Chinese population. *Science of The Total Environment* [en línea]. Vol. 666, Mayo 2019. [Fecha de consulta: 30 de abril]. Disponible en: <https://doi.org/10.1016/j.scitotenv.2019.02.256>
ISSN: 0048-9697.

WOOD, Simon. Generalized Additive Models An Introduction with R [en línea]. 2.^a ed. Boca Raton: Chapman and Hall/CRC, 2017 [Fecha de consulta: 12 de julio del 2023]. Disponible en: <https://www.taylorfrancis.com/books/mono/10.1201/9781315370279/generalized-additive-models-simon-wood> ISBN: 978-1315370279.

YOON, Heeyeun. Effects of particulate matter (PM10) on tourism sales revenue: A generalized additive modeling approach. *Tourism Management* [en línea]. Vol 74, Octubre 2019. [Fecha de consulta: 30 de abril]. Disponible en: <https://doi.org/10.1016/j.tourman.2019.04.008>

ISSN: 0261-5177.

ANEXOS

ANEXO 1. MATRIZ DE CONSISTENCIA.

MATRIZ DE CONSISTENCIA				
TÍTULO: Evaluación de la contaminación y calidad del aire por material particulado vinculado a las variables meteorológicas en Lima Metropolitana entre 2017-2020				
PROBLEMA	OBJETIVOS	HIPOTESIS	VARIABLE	METODOLOGIA
Problema General	Objetivo General	Hipótesis General	Variable Independiente	Tipo
¿Cuál es la contaminación y la calidad del aire por partículas y como influyen los parámetros meteorológicos en Lima Metropolitana?	Evaluar la contaminación y la calidad del aire por material particulado, así como su relación con los parámetros meteorológicos en Lima Metropolitana.	La contaminación y calidad del aire respecto al material particulado ha sido aceptable y optima según el ECA e INCA, estas concentraciones son influidas por los parámetros meteorológicos en los distritos de Lima Metropolitana.	Material particulado Parámetros meteorológicos	Investigación aplicada
Problemas Específicos	Objetivo Especifico	Hipótesis Especifica	Dimensiones	Enfoque
¿Cómo es la contaminación del aire por partículas en Lima Metropolitana?	Evaluar la concentración de PM _{2.5} y PM ₁₀ conforme a los valores y criterios de los ECA para periodos de 24 horas y anual, en los distritos de Lima Metropolitana.	La concentración de partículas en el aire según el ECA será aceptable: para PM _{2.5} en 24 horas será menor o igual a 50 µg/m ³ y anualmente menor o igual a 25 µg/m ³ . Para PM ₁₀ en 24 horas será menor o igual a 100 µg/m ³ y anualmente menor o igual a 50 µg/m ³ , en los distritos de Lima Metropolitana.	Concentración de partículas menores de 2.5 y 10 micrómetros	Cuantitativo
			Variables atmosféricas	Nivel No experimental
¿Cuál es la distribución geográfica de las concentraciones de material particulado que exceden los ECA en Lima Metropolitana?	Determinar la distribución espacial de la concentración del PM _{2.5} y PM ₁₀ , para cuantificar el área en km ² donde se excedió el ECA y su proporción respecto al área total comprendida por las estaciones de monitoreo del SENAMHI, en Lima Metropolitana.	El área de la distribución espacial de las concentraciones de PM _{2.5} y PM ₁₀ mayores al ECA anual de 25 µg/m ³ y 50 µg/m ³ respectivamente, será menor al 50% del área comprendida por las estaciones de monitoreo del SENAMHI en Lima Metropolitana.	Variable Dependiente	Diseño
			Contaminación del aire	Longitudinal descriptivo y correlacional
			Calidad del Aire	Población Total de datos meteorológicos del SENAMHI.
¿Cómo es la calidad del aire respecto a la concentración de materia particulado en los distritos de Lima Metropolitana?	Caracterizar la calidad del aire, su evolución horaria conforme a la calificación del INCA para la concentración de PM _{2.5} y PM ₁₀ , e identificar los distritos donde es mayor la exposición a la calidad insalubre del aire en Lima Metropolitana entre el 2017 y 2020.	La evolución horaria de la calidad del aire según la concentración del PM _{2.5} y PM ₁₀ será menor o igual al valor 100 del INCA, que es considerado óptimo con una calidad de aire "buena" o "moderada" en los distritos de Lima Metropolitana.	Dimensiones	Muestra
			Estándares de Calidad Ambiental del aire Distribución espacial	Datos meteorológicos del SENAMHI desde enero del 2017 hasta agosto 2020.
¿Cómo influyen los parámetros meteorológicos en la concentración del material particulado en los distritos de Lima Metropolitana?	Identificar las condiciones que favorecen la acumulación o dispersión del PM _{2.5} y PM ₁₀ , así como la correlación con los parámetros meteorológicos: temperatura del aire, humedad relativa, velocidad del viento y dirección del viento, con modelos aditivos generalizados y árboles de regresión como alternativa a los coeficientes de Pearson y Spearman, en los distritos de Lima Metropolitana entre el 2017 y 2020.	Los parámetros meteorológicos influyen en la concentración de las partículas en suspensión en los distritos de Lima Metropolitana.	Índice de calidad Aire	

Anexo 2. Matriz de operacionalización de variables.

VARIABLES INDEPENDIENTES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ESCALA / UNIDADES MEDIDA
Material particulado	Conjunto de partículas sólidas y/o líquidas que se presentan suspendidas en la atmósfera (Mészáros, 1999).	Información de los contaminantes del aire publicados por SENAMHI en la Plataforma Nacional de datos abiertos del estado peruano.	Concentración de partículas menores de 2.5 y 10 micrómetros	Concentración de masa	µg/m ³
Parámetros meteorológicos	Es toda característica de la atmósfera, que en su conjunto determina el estado del tiempo (a corto plazo) o el clima (a largo plazo) (SENAMHI, 2021)	Información de las estaciones en Lima metropolitana de la Web del SENAMHI "Datos Hidrometeorológicos a nivel nacional"	Variables atmosféricas	Temperatura del aire	°C
				Humedad relativa	%
				Velocidad del viento	m/s
				Dirección del viento	°
VARIABLES DEPENDIENTES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ESCALA / UNIDADES MEDIDA
Contaminación del aire	Alteración de las propiedades físicas, químicas o biológicas del aire, que puede afectar nocivamente la vida humana y demás organismos vivos.	Los estándares de calidad ambiental del aire, establecen las concentraciones máximas para los contaminantes del aire	Estándares de Calidad Ambiental del aire	Concentración de masa	µg/m ³
			Distribución espacial	Superficie	km ²
Calidad del Aire	Es la composición del aire y de su idoneidad para determinadas aplicaciones o actividades (INEI, 2016).	La calidad del aire nos indica cuando el aire está exento de contaminantes atmosféricos y por lo tanto apto para ser respirado (INEI, 2016)	Índice de calidad Aire	Estados de calidad del aire	Nominal: buena, moderada, mala, umbral de cuidado

Anexo 3. Instrumento de colección de datos.

Ficha 1: Ubicación y Concentración de material particulado

Ficha 1: Instrumento, ficha de observación de datos de material particulado.									
		Título	Evaluación de la contaminación y calidad del aire por material particulado vinculado a las variables meteorológicas en Lima Metropolitana entre 2017-2020						
		Línea de investigación	Calidad y Gestión de Recursos Naturales						
		Responsables	Rubi Katherine Mendez Reyes Anayca Nayra Urquia Ochavano						
		Asesor	Dc. Rubén Munive Cerrón						
ESTACION	CODIGO ESTACION	LONGITUD	LATITUD	AÑO	MES	DIA	HORA	PM ₁₀	PM _{2.5}

Dr. RUBEN MUNIVE CERRON
 CIP N° 38103

MSc. FIORELLA VANESSA
 GUERE SALAZAR,
 CIP N° 131344

OSWALDO FRANCISCO FLORES SALDAÑA
 INGENIERO PESQUERO
 Reg. CIP N° 816421
 DR. OSWALDO FRANCISCO
 FLORES SALDAÑA
 CIP N° 016421

MSc. MARÍA MELITTA
 HURTADO ZAMORA
 CIP N° 060083

Anexo 4. Validación de instrumento

VALIDACIÓN DE INSTRUMENTO

I. DATOS GENERALES

- 1.1 Apellidos y Nombres: Dr. Munive Cerrón Rubén
- 1.2 Cargo e institución donde labora: Docente UCV
- 1.3 Especialidad o Línea de Investigación: Ingeniería y Ciencias Ambientales
- 1.4 Nombre del instrumento motivo de evaluación: **Ficha 1: Ubicación y concentración de material particulado**
- 1.5 Autor(A) de Instrumento: **Rubi Katherine Mendez Reyes y Anayca Nayra Urquía Ochavano**

II. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE					MÍNIMAMENTE ACEPTABLE			ACEPTABLE				
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.										X			
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.										X			
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.										X			
4. ORGANIZACIÓN	Existe una organización lógica.										X			
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales										X			
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.										X			
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.										X			
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.										X			
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.										X			
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.										X			

III. OPINIÓN DE APLICABILIDAD

- El Instrumento cumple con los Requisitos para su aplicación
- El Instrumento no cumple con los requisitos para su aplicación

X

IV. PROMEDIO DE VALORACIÓN:

85 %



Dr. RUBEN MUNIVE CERRON
CIP N° 38103

FIRMA DEL EXPERTO INFORMANTE

Lima, 19 de junio del 2023

VALIDACIÓN DE INSTRUMENTO

I. DATOS GENERALES

- 1.1 Apellidos y Nombres: Güere Salazar Fiorella Vanessa
- 1.2 Cargo e institución donde labora: Docente UCV
- 1.3 Especialidad o Línea de Investigación: Conservación de suelo y agua
- 1.4 Nombre del instrumento motivo de evaluación: **Ficha 1: Ubicación y Concentración de material particulado.**
- 1.5 Autor(A) de Instrumento: **Rubi Katerine Mendez Reyes y Anayca Nayra Urquia Ochavano**

II. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE						MINIMAMENTE ACEPTABLE			ACEPTABLE			
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.											X		
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.											X		
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.											X		
4. ORGANIZACIÓN	Existe una organización lógica.											X		
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales											X		
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.											X		
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.											X		
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.											X		
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.											X		
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.											X		

III. OPINIÓN DE APLICABILIDAD

- El Instrumento cumple con los Requisitos para su aplicación
- El Instrumento no cumple con Los requisitos para su aplicación

SI

IV. PROMEDIO DE VALORACIÓN:

90%

Lima, 10 de julio del 2023



MSc. FIORELLA VANESSA
GUERE SALAZAR
CIP N°131344

FIRMA DEL EXPERTO INFORMANTE

VALIDACIÓN DE INSTRUMENTO

I. DATOS GENERALES

- 1.1 Apellidos y Nombres: Flores Saldaña Oswaldo Francisco
- 1.2 Cargo e institución donde labora: Docente Principal Universidad Nacional José Faustino Sánchez Carrión
- 1.3 Especialidad o Línea de Investigación: Ingeniería y ciencias ambientales
- 1.4 Nombre del instrumento motivo de evaluación: **Ficha 1. Observación concentración de material particulado.**
- 1.5 Autor(A) de Instrumento: Rubí Katerine Mendez Reyes y Anayca Nayra Urquia Ochavano

II. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE						MINIMAMENTE ACEPTABLE			ACEPTABLE			
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.												X	
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.												X	
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.												X	
4. ORGANIZACIÓN	Existe una organización lógica.												X	
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales												X	
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.												X	
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.												X	
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.												X	
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.												X	
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.												X	

III. OPINIÓN DE APLICABILIDAD

- El Instrumento cumple con los Requisitos para su aplicación
- El Instrumento no cumple con Los requisitos para su aplicación

SI

95 %

IV. PROMEDIO DE VALORACIÓN:


OSWALDO FRANCISCO FLORES SALDAÑA
 INGENIERO PESQUERO
 Reg. CIP N° 016421

 DR. OSWALDO FLORES SALDAÑA
 CIP N° 016421

Lima 26 junio 2023

FIRMA DEL EXPERTO INFORMANTE

VALIDACIÓN DE INSTRUMENTO

I. DATOS GENERALES

- 1.1 Apellidos y Nombres: Hurtado Zamora María Melitta
- 1.2 Cargo e institución donde labora: Docente Universidad Nacional José Faustino Sánchez Carrión
- 1.3 Especialidad o Línea de Investigación: Ingeniería y ciencias ambientales
- 1.4 Nombre del instrumento motivo de evaluación: **Ficha 1. Observación concentración de material particulado.**
- 1.5 Autor(A) de Instrumento: **Rubí Katerine Mendez Reyes y Anayca Nayra Urquía Ochavano**

II. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE						MINIMAMENTE ACEPTABLE			ACEPTABLE			
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.												X	
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.												X	
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.												X	
4. ORGANIZACIÓN	Existe una organización lógica.												X	
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales												X	
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.												X	
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.												X	
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.												X	
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.												X	
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.												X	

III. OPINIÓN DE APLICABILIDAD

- El Instrumento cumple con los Requisitos para su aplicación
- El Instrumento no cumple con los requisitos para su aplicación

SI

90 %

IV. PROMEDIO DE VALORACIÓN:

Lima 01 julio 2023



MSc. MARÍA HURTADO ZAMORA
CIP N° 060083

FIRMA DEL EXPERTO INFORMANTE

Anexo 5. Instrumento de colección de datos

Ficha 2: Observación de los parámetros meteorológicos

Ficha 2: Instrumento, ficha de observación de datos de los parámetros meteorológicos.									
	Título		Evaluación de la contaminación y calidad del aire por material particulado vinculado a las variables meteorológicas en Lima Metropolitana entre 2017-2020						
	Línea de investigación		Calidad y Gestión de Recursos Naturales						
	Responsables		Rubi Katerine Mendez Reyes Anayca Nayra Urquia Ochavano						
	Asesor		Rubén Munive Cerrón						
ESTACION	CODIGO ESTACION	AÑO	MES	DIA	HORA	TEMPERATURA AIRE (°C)	HUMEDAD RELATIVA (%)	VELOCIDAD VIENTO (m/s)	DIRECCION VIENTO (°)


 Dr. RUBEN MUNIVE CERRON
 CIP N° 38103


 MSc. FIORELLA VANESSA
 GUERE SALAZAR,
 CIP N°131344


 OSWALDO FRANCISCO FLORES SALDAÑA
 INGENIERO PESQUERO
 Reg. CIP N° 016421
 DR. OSWALDO FRANCISCO
 FLORES SALDAÑA
 CIP N° 016421


 MSc. MARÍA MELITTA
 HURTADO ZAMORA
 CIP N° 060083

Anexo 6. Validación de instrumento.

VALIDACIÓN DE INSTRUMENTO

I. DATOS GENERALES

- 1.1 Apellidos y Nombres: Dr. Munive Cerrón Rubén
- 1.2 Cargo e institución donde labora: Docente UCV
- 1.3 Especialidad o Línea de Investigación: Ingeniería y Ciencias Ambientales
- 1.4 Nombre del instrumento motivo de evaluación: **Ficha 2: Observación de parámetros meteorológicos**
- 1.5 Autor(A) de Instrumento: **Rubi Katerine Mendez Reyes y Anayca Nayra Urquia Ochavano**

II. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE						MÍNIMAMENTE ACEPTABLE			ACEPTABLE			
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.										X			
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.										X			
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.										X			
4. ORGANIZACIÓN	Existe una organización lógica.										X			
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales										X			
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.										X			
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.										X			
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.										X			
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.										X			
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.										X			

III. OPINIÓN DE APLICABILIDAD

El instrumento cumple con los requisitos para su aplicación

El Instrumento no cumple con los requisitos para su aplicación

X

IV. PROMEDIO DE VALORACIÓN:

85 %



Dr. RUBEN MUNIVE CERRON
CIP N° 38103

FIRMA DEL EXPERTO INFORMANTE

Lima, 19 de junio del 2023

VALIDACIÓN DE INSTRUMENTO

I. DATOS GENERALES

- 1.1 Apellidos y Nombres: Güere Salazar Fiorella Vanessa
 1.2 Cargo e institución donde labora: Docente UCV
 1.3 Especialidad o Línea de Investigación: Conservación de suelo y agua
 1.4 Nombre del instrumento motivo de evaluación: **Ficha 2: Observación parámetros meteorológicos.**
 1.5 Autor(A) de Instrumento: **Rubi Katerine Mendez Reyes y Anayca Nayra Urquia Ochavano**

II. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE						MINIMAMENTE ACEPTABLE			ACEPTABLE			
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.											X		
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.											X		
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.											X		
4. ORGANIZACIÓN	Existe una organización lógica.											X		
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales											X		
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.											X		
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.											X		
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.											X		
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.											X		
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.											X		

III. OPINIÓN DE APLICABILIDAD

El instrumento cumple con los requisitos para su aplicación
 El Instrumento no cumple con los requisitos para su aplicación

X

IV. PROMEDIO DE VALORACIÓN:

90%

Lima, 10 de julio del 2023



MSc. FIORELLA VANESSA
 GUERE SALAZAR
 CIP N°131344

FIRMA DEL EXPERTO INFORMANTE

VALIDACIÓN DE INSTRUMENTO

I. DATOS GENERALES

- 1.1 Apellidos y Nombres: Flores Saldaña Oswaldo Francisco
 1.2 Cargo e institución donde labora: Docente Principal Universidad Nacional José Faustino Sánchez Carrión
 1.3 Especialidad o Línea de Investigación: Ingeniería y ciencias ambientales
 1.4 Nombre del instrumento motivo de evaluación: **Ficha 2: Observación parámetros meteorológicos.**
 1.5 Autor(A) de Instrumento: **Rubí Katherine Mendez Reyes y Anayca Nayra Urquía Ochavano**

II. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE						MÍNIMAMENTE ACEPTABLE			ACEPTABLE			
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.												X	
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.												X	
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.												X	
4. ORGANIZACIÓN	Existe una organización lógica.												X	
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales												X	
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.												X	
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.												X	
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.												X	
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.												X	
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.												X	

III. OPINIÓN DE APLICABILIDAD

El instrumento cumple con los requisitos para su aplicación
 El Instrumento no cumple con los requisitos para su aplicación

X

IV. PROMEDIO DE VALORACIÓN:

95 %

Lima 26 junio 2023


OSWALDO FRANCISCO FLORES SALDAÑA
 INGENIERO PESQUERO
 Reg. CIP N° 016421
DR. OSWALDO FLORES SALDAÑA
 CIP N° 016421

FIRMA DEL EXPERTO INFORMANTE

VALIDACIÓN DE INSTRUMENTO

I. DATOS GENERALES

- 1.1 Apellidos y Nombres: Hurtado Zamora María Melitta
 1.2 Cargo e institución donde labora: Docente Universidad Nacional José Faustino Sánchez Carrión
 1.3 Especialidad o Línea de Investigación: Ingeniería y ciencias ambientales
 1.4 Nombre del instrumento motivo de evaluación: **Ficha 2: Observación parámetros meteorológicos.**
 1.5 Autor(A) de Instrumento: Rubí Katerine Mendez Reyes y Anayca Nayra Urquia Ochavano

II. ASPECTOS DE VALIDACIÓN

CRITERIOS	INDICADORES	INACEPTABLE					MINIMAMENTE ACEPTABLE			ACEPTABLE				
		40	45	50	55	60	65	70	75	80	85	90	95	100
1. CLARIDAD	Esta formulado con lenguaje comprensible.											X		
2. OBJETIVIDAD	Esta adecuado a las leyes y principios científicos.											X		
3. ACTUALIDAD	Esta adecuado a los objetivos y las necesidades reales de la investigación.											X		
4. ORGANIZACIÓN	Existe una organización lógica.											X		
5. SUFICIENCIA	Toma en cuenta los aspectos metodológicos esenciales											X		
6. INTENCIONALIDAD	Esta adecuado para valorar las variables de la Hipótesis.											X		
7. CONSISTENCIA	Se respalda en fundamentos técnicos y/o científicos.											X		
8. COHERENCIA	Existe coherencia entre los problemas objetivos, hipótesis, variables e indicadores.											X		
9. METODOLOGÍA	La estrategia responde una metodología y diseño aplicados para lograr probar las hipótesis.											X		
10. PERTINENCIA	El instrumento muestra la relación entre los componentes de la investigación y su adecuación al Método Científico.											X		

III. OPINIÓN DE APLICABILIDAD

El Instrumento cumple con los requisitos para su aplicación
 El Instrumento no cumple con los requisitos para su aplicación

X

IV. PROMEDIO DE VALORACIÓN:

90 %

Lima 01 julio 2023



MSc. MARÍA HURTADO ZAMORA
 CIP N° 060083

FIRMA DEL EXPERTO INFORMANTE

Anexo 8. Script en lenguaje del programa R.

Script 01. Código para la descarga y almacenamiento masivo de la información de la Web “Datos hidrometeorológicos del SENAMHI a nivel nacional”.

```
# Cargamos las librerías
library(rvest)
library(openxlsx)
library(openair)

# Borramos todo
rm(list = ls(all=TRUE)) #borra workspace
cat("\f") #borra la consola
dev.off() #Borra los plot

# Creamos un rango de fechas
años <- rep(2017:2022,12) # asigna el rango de fechas "2017:2022,12"
meses <- mes <- sort(sprintf("%02d",rep(1:12,5)))
fechas <- sort(paste0(años,meses))

# Generando los links
Estacion_meses <- list()
for (i in fechas) {
  Estacion_meses[i] <- paste0("https://www.SENAMHI.gob.pe/mapas/mapa-estaciones-
2/_dato_esta_tipo02.php?CBOFiltro=",i,"&estaciones=111287&t_e=M&estado=AUTOMATICA&
cod_old=&cate_esta=EAA&soloAlt=292")
}

# iniciando la descarga y almacenamiento en masa
# el siguiente código genera una lista de las tablas extraídas
# datos del año 2017 al 2022
villa <- list()
for (i in 1:length(Estacion_meses)) {
  villa[[i]]<-html_table(
    html_nodes(
      read_html(
        as.character(Estacion_meses[i])
      ), "table"
    )[2],
    header = TRUE,na.strings = "S/D"
  )
}

# Combinando todas las tablas descargadas en una sola:
vm <- Reduce(function(...) merge(..., all=TRUE), villa)
colnames(vm) <- c("date1", "hora", "temp", "pp", "hum", "wd", "ws")

# Guardamos la tabla final en un archivo xlsx:
openxlsx::write.xlsx(vm,"C:/Directorio destino /Nombre_archivo.xlsx")
```

Script 02. Código para fusionar las bases de datos de los contaminantes del aire y parámetros meteorológicos.

```
# Cargamos las librerías
library(openxlsx)
library(plyr)

# Borrar todo
rm(list = ls(all=TRUE)) #borra workspace
cat("\f") #borra la consola
dev.off() #Borra los plot

# Leer CSV Contaminantes del aire
contaminantes_aire <- read.csv("C:/Tesis Rubi Calidad del aire/Bases de
datos/datos_horarios_contaminacion_lima.csv", na.strings = c("-9999" , "NA" ))
attach(contaminantes_aire)

# Leer XLSX parámetros ambientales
Campo_Marte <- read.xlsx("C:/Tesis Rubi Calidad del aire/Bases de datos/Datos
meteorologicos/Campo_Marte.xlsx", na.strings = c("-9999" , "NA" ))
attach(Campo_Marte)

# Crear Columna ESTACION
estacion_meteo<-paste("CAMPO","DE","MARTE",sep=" ")
Campo_Marte$ESTACION<-estacion_meteo

# Formato de hora
hora.parametros2<-(paste(Campo_Marte$hora,"00",sep = ":"))
Fecha.hora2<-paste(Campo_Marte$date1,hora.parametros2,sep=" ")
Campo_Marte$FECHA_HORA <- (as.POSIXct(Fecha.hora2,format="%Y/%m/%d
%H:%M:%S"))
attach(Campo_Marte)

# Crear Columna FECHA y FECHA_HORA
año<- (contaminantes_aire$ANO)
mes<- (contaminantes_aire$MES.ID)
dia<- (contaminantes_aire$DIA)
#hora<-(contaminantes_aire$HORA)
hora<-(paste(contaminantes_aire$HORA,00,00,sep = ":"))
Fecha<-paste(dia,mes,año,sep="/")
Fecha.hora<-paste(Fecha,hora,sep=" ")
contaminantes_aire$FECHA <- (as.Date(Fecha, tryFormats="%d/%m/%Y"))
contaminantes_aire$FECHA_HORA <- (as.POSIXct(Fecha.hora,format="%d/%m/%Y
%H:%M:%S"))
attach(contaminantes_aire)

# Fusionar tablas contaminantes vs parámetros ambientales
nueva_tabla<-merge(x = contaminantes_aire, y = Campo_Marte,
  by = c("ESTACION", "FECHA_HORA")) # fusiona los datos por dos columnas
openxlsx::write.xlsx(nueva_tabla,"C:/Tesis Rubi Calidad del aire/Bases de
datos/base_campomarte.xlsx")
nueva_tabla = rename(nueva_tabla,
  c(ws="Velocidad",
    wd="Direccion",
    temp="Temperatura",
    hum="Humedad"))
names (nueva_tabla)
```

Script 03. Código para pruebas de normalidad.

```
# Cargando las librerías
library(broom)
library(moments)

# Borrar todo
rm(list = ls(all=TRUE)) # borra workspace
cat("\f") # borra la consola
dev.off() # borra los plot

# Leer XLSX material particulado y parametros ambientales
Base_Total <- read.xlsx("C:/Directorio destino/Nombre_archivo.xlsx", na.strings = c("-9999",
"NA" ))

# Borrar columnas
"ESTACION","FECHA_HORA","CODIGO.ESTACION","ANO","MES","DIA","HORA","MES.ID",
"SO2","NO2","O3","CO","FECHA","date1","hora","pp")
borrar <-
c("ESTACION","FECHA_HORA","CODIGO.ESTACION","ANO","MES","DIA","HORA","MES.ID",
"SO2","NO2","O3","CO","FECHA","date1","hora","pp")
Campo_Marte <- Base_Total[ , !(names(Base_Total) %in% borrar)]

# test normalidad de Kolmogorov-smirnov PM10
ks.test(Campo_Marte$PM.10,Base_Total$Temperatura)
ks.test(Campo_Marte$PM.10,Base_Total$Humedad)
ks.test(Campo_Marte$PM.10,Base_Total$Velocidad)
ks.test(Campo_Marte$PM.10,Base_Total$Direccion)

# test normalidad de Kolmogorov-smirnov PM2.5
ks.test(Campo_Marte$PM.2.5,Base_Total$Temperatura)
ks.test(Campo_Marte$PM.2.5,Base_Total$Humedad)
ks.test(Campo_Marte$PM.2.5,Base_Total$Velocidad)
ks.test(Campo_Marte$PM.2.5,Base_Total$Direccion)

# Omitir "na"
test_forma <-na.omit(Base_Total)

# Asimetría
skewness(test_forma$PM.10)
skewness(test_forma$PM.2.5)
skewness(test_forma$Temperatura)
skewness(test_forma$Humedad)
skewness(test_forma$Velocidad)
skewness(test_forma$Direccion)

# Curtosis
kurtosis(test_forma$PM.10)
kurtosis(test_forma$PM.2.5)
kurtosis(test_forma$Temperatura)
kurtosis(test_forma$Humedad)
kurtosis(test_forma$Velocidad)
kurtosis(test_forma$Direccion)

# Test de Jarque-Bera para la asimetría y curtosis de una distribución normal
jarque.test(test_forma$PM.10)
jarque.test(test_forma$PM.2.5)
jarque.test(test_forma$Temperatura)
jarque.test(test_forma$Humedad)
jarque.test(test_forma$Velocidad)
jarque.test(test_forma$Direccion)
```

Script 04. Código para análisis exploratorio de los datos.

```
# Cargando las librerías
library(psych)

# Borrar todo
rm(list = ls(all=TRUE)) # borra workspace
cat("\f") # borra la consola
dev.off() # borra los plot

# Leer XLSX material particulado y parametros ambientales
Base_Total <- read.xlsx("C:/Directorio destino/Nombre_archivo.xlsx", na.strings = c("-9999" ,
"NA" ))

# Borrar columnas
"ESTACION","FECHA_HORA","CODIGO.ESTACION","ANO","MES","DIA","HORA","MES.ID",
"SO2","NO2","O3","CO","FECHA","date1","hora","pp")
borrar <-
c("ESTACION","FECHA_HORA","CODIGO.ESTACION","ANO","MES","DIA","HORA","MES.ID",
"SO2","NO2","O3","CO","FECHA","date1","hora","pp")
Campo_Marte <- Base_Total[ , !(names(Base_Total) %in% borrar)]

# Gráfico de correlación de Pearson con ajuste LOESS
pairs.panels(Campo_Marte,
  method = "pearson", # Método para el cálculo de las correlaciones
  hist.col = "aquamarine",
  density = TRUE, # Mostrando las densidades
  ellipses = FALSE, # Ignorando las elipses de correlación
  lm = TRUE, # TRUE dibuja ajuste lineal, FALSE ajuste LOESS
  main = "Coeficiente de correlación de Pearson"
)

# Gráfico de correlación de Spearman con ajuste LOESS
pairs.panels(Campo_Marte,
  method = "spearman", # Método para el cálculo de las correlaciones
  hist.col = "aquamarine",
  density = TRUE, # Mostrando las densidades
  ellipses = FALSE, # Ignorando las elipses de correlación
  main = "Coeficiente de correlación de Spearman"
)

# Análisis de normalidad y homocedasticidad de varianzas del modelo lineal
par(mfrow = c(2, 4))
par(mar = c(3, 3, 2, 2))
model1 <- lm(PM.10~Temperatura+Humedad+Velocidad+Direccion, data=Campo_Marte)
plot(model1)

model2 <- lm(PM.2.5~Temperatura+Humedad+Velocidad+Direccion, data=Campo_Marte)
plot(model2)

# Contraste de Breusch-Pagan
ncvTest(model1)
ncvTest(model2)
```

Script 05. Código para modelos aditivos generalizados.

```
# Cargando las librerías
library(openxlsx)
library(mgcv)
library(nlme)
library(broom)
library(dplyr)
library(visreg)

# Borrar todo
rm(list = ls(all=TRUE)) #borra workspace
cat("\f") #borra la consola
dev.off() #Borra los plot

# Leer XLSX material particulado y parámetros ambientales
Base_Total <- read.xlsx("C:/Nombre del directorio/Nombre_archivo.xlsx", na.strings = c("-9999"
, "NA" ))

# GAM MP vs parámetros meteorológicos
# gaussian, gamma, binomial negativa

# PM2.5
# modelo familia gausianna con función link identity
gam1 <- gam(log(PM.2.5)~s(Direccion,k=10,m=c(2,3),bs="ps")+
  s(Velocidad,k=10,m=c(2,3),bs="ps")+
  s(Humedad,k=10,m=c(2,3),bs="ps")+
  s(Temperatura,k=10,m=c(2,3),bs="ps"),
  family = gaussian(link = "identity"),
  data=Campo_Marte,
  na.action=na.exclude)
plot(gam1,main="MP2.5 (µg/m³)",se=T, shade = TRUE, scale = -1)

# Resumen del GAM
summary(gam1)

# modelo familia gausianna con función link log
gam2 <- gam(log(PM.2.5)~s(Direccion,k=10,m=c(2,3),bs="ps")+
  s(Velocidad,k=10,m=c(2,3),bs="ps")+
  s(Humedad,k=10,m=c(2,3),bs="ps")+
  s(Temperatura,k=10,m=c(2,3),bs="ps"),
  family = gaussian(link = "log"),
  data=Campo_Marte,
  na.action=na.exclude)
plot(gam2,main="MP2.5 (µg/m³)",se=T)

# Resumen del GAM
summary(gam2)
gam2$coefficients # Mostrar coeficientes del GAM

# modelo familia Gamma con función identity
gam3 <- gam(log(PM.2.5)~s(Direccion,k=10,m=c(2,3),bs="ps")+
  s(Velocidad,k=10,m=c(2,3),bs="ps")+
  s(Humedad,k=10,m=c(2,3),bs="ps")+
  s(Temperatura,k=10,m=c(2,3),bs="ps"),
  family = Gamma(link = "identity"),
  data=Campo_Marte,
  na.action=na.exclude)
plot(gam3,main="MP2.5 (µg/m³)",se=T, shade = TRUE, scale = -1)

# Resumen del GAM
```

```

summary(gam3)
gam3$coefficients # Mostrar coeficientes del GAM

# modelo familia Gamma con función link inversa
gam4 <- gam(log(PM.2.5)~s(Direccion,k=10,m=c(2,3),bs="ps")+
  s(Velocidad,k=10,m=c(2,3),bs="ps")+
  s(Humedad,k=10,m=c(2,3),bs="ps")+
  s(Temperatura,k=10,m=c(2,3),bs="ps"),
  family = Gamma(link = "inverse"),
  data=Campo_Marte,
  na.action=na.exclude)
plot(gam4,main="MP2.5 (µg/m³)",se=T, shade = TRUE, scale = -1)

# Resumen del GAM
summary(gam4)
gam4$coefficients # Mostrar coeficientes del GAM

# modelo familia binomial negativa función link log
gam5 <- gam(log(PM.2.5)~s(Direccion,k=10,m=c(2,3),bs="ps")+
  s(Velocidad,k=10,m=c(2,3),bs="ps")+
  s(Humedad,k=10,m=c(2,3),bs="ps")+
  s(Temperatura,k=10,m=c(2,3),bs="ps"),
  family = negbin(theta = 3, link = "log"),
  data=Campo_Marte,
  na.action=na.exclude)
plot(gam5,main="MP2.5 (µg/m³)",se=T)
# Resumen del GAM
summary(gam5)
gam5$coefficients # Mostrar coeficientes del GAM

# modelo familia binomial negativa inversa
gam6 <- gam(log(PM.2.5)~s(Direccion,k=10,m=c(2,3),bs="ps")+
  s(Velocidad,k=10,m=c(2,3),bs="ps")+
  s(Humedad,k=10,m=c(2,3),bs="ps")+
  s(Temperatura,k=10,m=c(2,3),bs="ps"),
  family = negbin(theta = 3, link = "inverse"),
  data=Campo_Marte,
  na.action=na.exclude)
plot(gam6,main="MP2.5 (µg/m³)",se=T)

# Resumen del GAM
summary(gam6)
gam6$coefficients # Mostrar coeficientes del GAM

# Anova de los modelos
anova(gam1, gam2, gam3, gam4, gam5, gam6, test = "F")

# Comparación de modelos AIC y BIC
g1 <- glance(gam1)
g2 <- glance(gam2)
g3 <- glance(gam3)
g4 <- glance(gam4)
g5 <- glance(gam5)
g6 <- glance(gam6)
as_tibble(rbind(g1, g2, g3, g4, g5, g6))

# PM10
# modelo familia gaussiana con función link identity
gam7 <- gam(log(PM.10)~s(Direccion,k=10,m=c(2,3),bs="ps")+
  s(Velocidad,k=10,m=c(2,3),bs="ps")+

```

```

s(Humedad,k=10,m=c(2,3),bs="ps")+
s(Temperatura,k=10,m=c(2,3),bs="ps"),
data=Campo_Marte,
family = gaussian(link="identity"),
na.action=na.exclude,select = TRUE)
plot(gam7,main="MP10 (µg/m³)",se=T)

# Resumen del GAM
summary(gam7)

# Modelo familia gaussiana con función link log
gam8 <- gam(log(PM.10)~s(Direccion,k=10,m=c(2,3),bs="ps")+
s(Velocidad,k=10,m=c(2,3),bs="ps")+
s(Humedad,k=10,m=c(2,3),bs="ps")+
s(Temperatura,k=10,m=c(2,3),bs="ps"),
family = gaussian(link="log"),
data=Campo_Marte,
na.action=na.exclude)
plot(gam8,main="MP10 (µg/m³)",se=T)

# Resumen del GAM
summary(gam8)
gam8$coefficients # Mostrar coeficientes del GAM

# Modelo familia gamma
gam9<- gam(log(PM.10)~s(Direccion,k=5,m=c(1,2),bs="ps")+
s(Velocidad,k=10,m=c(2,3),bs="ps")+
s(Humedad,k=10,m=c(2,3),bs="ps")+
s(Temperatura,k=10,m=c(2,3),bs="ps"),
family = Gamma(link="identity"),
data=Campo_Marte,
na.action=na.exclude)
plot(gam9,main="MP10 (µg/m³)",se=T, shade = TRUE, scale=-1)
# Resumen del GAM
summary(gam9)
gam9$coefficients # Mostrar coeficientes del GAM

# modelo familia Gamma con función link inversa
gam10 <- gam(log(PM.10)~s(Direccion,k=10,m=c(2,3),bs="ps")+
s(Velocidad,k=10,m=c(2,3),bs="ps")+
s(Humedad,k=10,m=c(2,3),bs="ps")+
s(Temperatura,k=10,m=c(2,3),bs="ps"),
family = Gamma(link = "inverse"),
data=Campo_Marte,
na.action=na.exclude)
plot(gam10,main="MP2.5 (µg/m³)",se=T, shade = TRUE, scale = -1)

# Resumen del GAM
summary(gam10)
gam10$coefficients # Mostrar coeficientes del GAM

# Modelo familia binomial negativa con función link log
gam11 <- gam(log(PM.10)~s(Direccion,k=10,m=c(2,3),bs="ps")+
s(Velocidad,k=10,m=c(2,3),bs="ps")+
s(Humedad,k=10,m=c(2,3),bs="ps")+
s(Temperatura,k=10,m=c(2,3),bs="ps"),
family = negbin(theta = 3, link = "log"),
data=Campo_Marte,
na.action=na.exclude)
plot(gam11,main="MP10 (µg/m³)",se=T, shade = TRUE, scale=-1)

```

```

# Resumen del GAM
summary(gam11)
gam11$coefficients # Mostrar coeficientes del GAM

# modelo familia binomial negativa con función link inversa
gam12 <- gam(log(PM.10)~s(Direccion,k=10,m=c(2,3),bs="ps")+
  s(Velocidad,k=10,m=c(2,3),bs="ps")+
  s(Humedad,k=10,m=c(2,3),bs="ps")+
  s(Temperatura,k=10,m=c(2,3),bs="ps"),
  family = negbin(theta = 3, link = "inverse"),
  data=Campo_Marte,
  na.action=na.exclude)
plot(gam12,main="MP2.5 (µg/m³)",se=T)

# Resumen del GAM
summary(gam12)
gam12$coefficients # Mostrar coeficientes del GAM

# Anova de los modelos
anova(gam7, gam8, gam9, gam10, gam11, gam12, test = "F")

# Comparación de modelos AIC y BIC
g7 <- glance(gam7)
g8 <- glance(gam8)
g9 <- glance(gam9)
g10 <- glance(gam10)
g11 <- glance(gam11)
g12 <- glance(gam12)
as_tibble(rbind(g7, g8, g9, g10, g11, g12))
}

```

Script 06. Código para árbol de regresión para PM10.

```

#Cargar librerías
library(rpart)
library(rpart.plot)
library(openxlsx)

#Borrar todo
rm(list = ls(all=TRUE)) #borra workspace
cat("\f") #borra la consola
dev.off() #Borra los plot

#Leer XLSX material particulado y parámetros ambientales
Base_Total <- read.xlsx("C:/Tesis Rubi Calidad del aire/Bases de
datos/Base_parametros_meteo.xlsx", na.strings = c("-9999" , "NA" ))

#Eliminar columnas
borrar <-
c("ESTACION","FECHA_HORA","CODIGO.ESTACION","ANO","MES","DIA","HORA","MES.I
D","SO2","NO2","O3","CO","FECHA","date1","hora","pp","PM.2.5","Direccion")
Campo_Marte <- Base_Total[ , !(names(Base_Total) %in% borrar)]
Campo_Marte <- na.omit(Campo_Marte)

#Asignar columna y calificación INCA
x<-ifelse(Campo_Marte$PM.10 <= 75.99,
  ifelse(Campo_Marte$PM.10 <= 75.99, "Buena", Campo_Marte$PM.10),
  ifelse (Campo_Marte$PM.10 >= 76 & Campo_Marte$PM.10 <= 150.99, "Moderada",
Campo_Marte$PM.10))
y<-ifelse(Campo_Marte$PM.10 >= 150.9,

```

```

        ifelse (Campo_Marte$PM.10 >= 151 & Campo_Marte$PM.10 <= 250.99, "Mala", x),
        ifelse (Campo_Marte$PM.10 > 251 & Campo_Marte$PM.10 <= 900, "VUEC",x)
z<-ifelse(Campo_Marte$PM.10>251,"VUEC",y)
Campo_Marte$INCA<-z)

skim(Campo_Marte)

##Arboles de regresión rpart
set.seed(500) # Establecimiento de la semilla

##selecciona el 80% de los datos como muestra de entrenamiento y el 20% como muestra
de test
nobs <- nrow(Campo_Marte)
itrain <- sample(nobs, 0.8 * nobs)
train <- Campo_Marte[itrain, ]
test <- Campo_Marte[-itrain, ]

## árbol de decisión con las opciones por defecto
arbol_regresion<-rpart(PM.10 ~ .,
                      method = "anova",
                      data = Campo_Marte)
arbol_regresion

## Grafica con rpart.plot
prp(arbol_regresion,
    type = 3,
    nn = FALSE,
    clip.right.labs = FALSE,
    fallen.leaves = TRUE,
    faclen = 8,
    varlen = 9,
    under = TRUE,
    extra = 101,
    prefix = "PM10\n",
    round = 0,
    under.cex = 1,
    cex = 0.8)
rpart.rules(arbol_regresion)

## Funciones printcp() (o plotcp()) para obtener (representar) los valores de CP para los
árboles (óptimos) de menor tamaño junto con su error de validación cruzada
printcp(arbol_regresion)
par(mar = c(5, 5, 5, 5))
plotcp(arbol_regresion)

## imprimir las reglas
rpart.rules(arbol_regresion, style = "tall")

## Construccion del arbol completo
arbol_regresion <- rpart(PM.10 ~ ., data = train, cp = 0)

## Funciones printcp() (o plotcp()) para obtener (representar) los valores de CP para los
árboles (óptimos) de menor tamaño junto con su error de validación cruzada
printcp(arbol_regresion)
plotcp(arbol_regresion)

## Tabla con los valores de las podas (óptimas, dependiendo del parámetro de complejidad)
head(arbol_regresion$cpstable, 10)

```

```

## valor óptimo de forma automática, siguiendo el criterio de un error estándar de Breiman et
al. (1984)
xerror <- arbol_regresion$cptable["xerror"]
imin.xerror <- which.min(xerror)
# Valor óptimo
arbol_regresion$cptable[imin.xerror, ]

# Límite superior "oneSE rule" y complejidad mínima por debajo de ese valor
upper.xerror <- xerror[imin.xerror] + arbol_regresion$cptable[imin.xerror, "xstd"]
icp <- min(which(xerror <= upper.xerror))
cp <- arbol_regresion$cptable[icp, "CP"]

##Para obtener el modelo final podemos el árbol con el valor de complejidad obtenido
arbol_regresion_final <- prune(arbol_regresion, cp = cp)

## Grafica con rpart.plot
par(mar = c(0.1, 0.1, 0.1, 0.1))
rpart.plot(arbol_regresion_final,
  type = 3,
  nn = FALSE,
  clip.right.labs = FALSE,
  fallen.leaves = TRUE,
  faclen = 8,
  varlen = 9,
  under = TRUE,
  extra = 101,
  prefix = "PM10\n",
  round = 0,
  under.cex = 1,
  cex = 0.75,
  compress=FALSE)
rpart.rules(arbol_regresion_final)
print(arbol_regresion_final)

## Funciones printcp() (o plotcp()) para representar los valores de CP para los árboles
(óptimos) de menor tamaño junto con su error de validación cruzada
printcp(arbol_regresion_final)
par(mar = c(5, 5, 5, 5))
plotcp(arbol_regresion_final)

#Resumen arbol
summary(arbol_regresion_final)

#Gráfico del R-cuadrado
par(mar = c(5, 4, 2, 2))
par(mfrow=c(1,2))
rsq.rpart(arbol_regresion_final)

```

Script 07. Código para árbol de regresión para PM2.5.

```

#Leer XLSX material particulado y parámetros ambientales
Base_Total <- read.xlsx("C:/Tesis Rubi Calidad del aire/Bases de
datos/Base_parametros_meteo.xlsx", na.strings = c("-9999" , "NA" ))

#Eliminar columnas
borrar <-
c("ESTACION","FECHA_HORA","CODIGO.ESTACION","EST","MES","DIA","HORA","MES.I
D","SO2","NO2","O3","CO","FECHA","date1","hora","pp","PM.10","Direccion")
Campo_Marte <- Base_Total[ ,!(names(Base_Total) %in% borrar)]
Campo_Marte$ANO <- as.character(Campo_Marte$ANO)
Campo_Marte <- na.omit(Campo_Marte)

```

```

#Asignar columna y calificacion INCA
x<-ifelse(Campo_Marte$PM.2.5 <= 12.5,
  ifelse(Campo_Marte$PM.2.5 <= 12.5, "Buena", Campo_Marte$PM.2.5),
  ifelse (Campo_Marte$PM.2.5 >= 12.51 & Campo_Marte$PM.2.5 <= 25, "Moderada",
Campo_Marte$PM.2.5))
y<-ifelse(Campo_Marte$PM.2.5 >= 25.01,
  ifelse (Campo_Marte$PM.2.5 >= 25.01 & Campo_Marte$PM.2.5 <= 125, "Mala", x),
  ifelse (Campo_Marte$PM.2.5 > 125 & Campo_Marte$PM.2.5 <= 900, "VUEC",x))
z<-ifelse(Campo_Marte$PM.2.5>125,"VUEC",y)
Campo_Marte$INCA<-(z)

##Establecimiento de la semilla
set.seed(500)

##selecciona el 80% de los datos como muestra de entrenamiento y el 20% como muestra
de test
nobs <- nrow(Campo_Marte)
itrain <- sample(nobs, 0.8 * nobs)
train <- Campo_Marte[itrain, ]
test <- Campo_Marte[-itrain, ]

## árbol de decisión con las opciones por defecto
arbol_regresion<-rpart(PM.2.5 ~ .,
  method = "anova",
  data = Campo_Marte)
arbol_regresion

## Grafica con rpart.plot
par(mar = c(5, 5, 5, 5))
prp(arbol_regresion,
  type = 3,
  nn = FALSE,
  clip.right.labs = FALSE,
  fallen.leaves = TRUE,
  faclen = 8,
  varlen = 9,
  under = TRUE,
  extra = 101,
  prefix = "PM2.5\n",
  round = 0,
  under.cex = 1,
  cex = 0.8)
rpart.rules(arbol_regresion)

## Funciones printcp() (o plotcp()) para representar los valores de CP para los árboles
(óptimos) de menor tamaño junto con su error de validación cruzada
printcp(arbol_regresion)
plotcp(arbol_regresion)

## imprimir las reglas
rpart.rules(arbol_regresion, style = "tall")

## Construccion del arbol completo
arbol_regresion <- rpart(PM.2.5 ~ ., data = train, cp = 0)

## Funciones printcp() (o plotcp()) para representar los valores de CP para los árboles
(óptimos) de menor tamaño junto con su error de validación cruzada
printcp(arbol_regresion)
par(mar = c(5, 5, 5, 5))
plotcp(arbol_regresion)

```

```

## Tabla con los valores de las podas (óptimas, dependiendo del parámetro de complejidad)
head(arbol_regresion$cptable, 10)

## valor óptimo de forma automática, siguiendo el criterio de un error estándar de Breiman et
al. (1984)
xerror <- arbol_regresion$cptable["xerror"]
imin.xerror <- which.min(xerror)

# Valor óptimo
arbol_regresion$cptable[imin.xerror, ]

# Límite superior "oneSE rule" y complejidad mínima por debajo de ese valor
upper.xerror <- xerror[imin.xerror] + arbol_regresion$cptable[imin.xerror, "xstd"]
icp <- min(which(xerror <= upper.xerror))
cp <- arbol_regresion$cptable[icp, "CP"]

##Para obtener el modelo final podemos el árbol con el valor de complejidad obtenido
arbol_regresion_final <- prune(arbol_regresion, cp = cp)

## Grafica con rpart.plot
par(mar = c(1, 1, 1, 1))
rpart.plot(arbol_regresion_final,
  type = 3,
  nn = FALSE,
  clip.right.labs = FALSE,
  fallen.leaves = TRUE,
  faclen = 4,
  varlen = 9,
  under = TRUE,
  extra = 101,
  prefix = "PM2.5\n",
  round = 0,
  under.cex = 1,
  cex = 0.8,
  compress=FALSE)
rpart.rules(arbol_regresion_final)
print(arbol_regresion_final)

## Funciones printcp() (o plotcp()) para representar los valores de CP para los árboles
(óptimos) de menor tamaño junto con su error de validación cruzada
printcp(arbol_regresion_final)

par(mar = c(5, 5, 5, 5))
plotcp(arbol_regresion_final)

#Gráfico del R-cuadrado
par(mar = c(5, 4, 2, 2))
par(mfrow=c(1,2))
rsq.rpart(arbol_regresion_final)

```