

Manuscript_for_Atmosphere - JuanCarlos.pdf

 My Files My Files Universidad Peruana Union

Detalles del documento

Identificador de la entrega

trn:oid:::29566:469176676

Fecha de entrega

23 jun 2025, 9:31 a.m. GMT-5

Fecha de descarga

23 jun 2025, 9:43 a.m. GMT-5

Nombre de archivo

Manuscript_for_Atmosphere - Copy.pdf

Tamaño de archivo

3.0 MB

17 Páginas

5208 Palabras

25.586 Caracteres

14% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Fuentes principales

- 12%  Fuentes de Internet
- 8%  Publicaciones
- 10%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Fuentes principales

- 12% Fuentes de Internet
- 8% Publicaciones
- 10% Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	Internet	cris.upeu.edu.pe	1%
2	Publicación	Hasnain Iftikhar, Murad Khan, Justyna Żywiołek, Mehak Khan, Javier Linkolk Lópe...	<1%
3	Internet	dom.lndb.lv	<1%
4	Internet	hdl.handle.net	<1%
5	Internet	www.unisucre.edu.co	<1%
6	Internet	www.mdpi.com	<1%
7	Trabajos entregados	Canterbury Christ Church University on 2024-01-02	<1%
8	Internet	ouci.dntb.gov.ua	<1%
9	Internet	journals.e-palli.com	<1%
10	Internet	thenutritioninsider.com	<1%
11	Publicación	Bajracharya, Jwala. "Genetic Diversity Study in Landraces of Rice (Oryza Sativa L.) ...	<1%

12	Trabajos entregados	Universidad Tecnológica de Bolívar,UTB on 2025-05-28	<1%
13	Internet	www.researchgate.net	<1%
14	Internet	openjournal.insm.gob.pe	<1%
15	Internet	dspace.udla.edu.ec	<1%
16	Internet	www.andi.com.co	<1%
17	Publicación	Hasnain Iftikhar, Justyna Zywiólek, Javier Linkolk López-Gonzales, Olayan Albalaw...	<1%
18	Trabajos entregados	President University on 2015-02-02	<1%
19	Internet	repository.ubharajaya.ac.id	<1%
20	Trabajos entregados	UNIV DE LAS AMERICAS on 2019-08-29	<1%
21	Internet	dialnet.unirioja.es	<1%
22	Trabajos entregados	University of Limerick on 2022-11-28	<1%
23	Internet	portusonline.org	<1%
24	Trabajos entregados	Stevens Institute of Technology on 2016-04-10	<1%
25	Trabajos entregados	Universidad Santo Tomas on 2018-11-03	<1%

26	Trabajos entregados	ipn on 2025-05-07	<1%
27	Internet	ijbes.utm.my	<1%
28	Internet	patents.google.com	<1%
29	Internet	www.sap.org.ar	<1%
30	Trabajos entregados	Universidad Internacional de la Rioja on 2023-05-10	<1%
31	Trabajos entregados	University of Birmingham on 2025-06-13	<1%
32	Internet	archive.org	<1%
33	Internet	ascelibrary.org	<1%
34	Internet	docplayer.es	<1%
35	Internet	pt.scribd.com	<1%
36	Internet	www.coursehero.com	<1%
37	Internet	www.inf.utfsm.cl	<1%
38	Trabajos entregados	Infile on 2024-02-19	<1%
39	Internet	core.ac.uk	<1%

40	Internet	repositorio.deusto.es	<1%
41	Internet	repository.usergioarboleda.edu.co	<1%
42	Internet	www.agrociencia-colpos.org	<1%
43	Internet	www.slideshare.net	<1%
44	Trabajos entregados	University of Hull on 2008-06-20	<1%
45	Trabajos entregados	Fundacion Universitaria Juan de Castellanos on 2018-08-29	<1%

Abordaje estadístico para evaluar los niveles de ozono en relación con las variables meteorológicas en Lima Metropolitana

Juan Carlos Roque Quezada y Javier Linkolk López-Gonzales

Introducción

El ozono troposférico (O_3) es un contaminante secundario cuya presencia en el aire se debe principalmente a reacciones fotoquímicas entre óxidos de nitrógeno (NO_x) y compuestos orgánicos volátiles (COV), en presencia de luz solar. Aunque su existencia en la estratósfera es fundamental para bloquear la radiación ultravioleta, a nivel del suelo constituye un riesgo significativo para la salud humana, contribuyendo a enfermedades respiratorias, disminución de la función pulmonar y aumento de hospitalizaciones y mortalidad prematura (EPA, 2024; European Environment Agency [EEA], 2023). Las condiciones meteorológicas, como la radiación solar, temperatura, velocidad y dirección del viento, juegan un rol crucial en la formación y acumulación de ozono (Rai & Agrawal, 2020).

En este contexto, la ciudad de Lima Metropolitana, debido a su compleja geografía y dinámica urbana, enfrenta condiciones propicias para la acumulación de contaminantes atmosféricos. Ubicada entre la costa del Pacífico y la cordillera de los Andes, Lima experimenta fenómenos de inversión térmica durante el invierno que reducen la dispersión de contaminantes, favoreciendo la concentración de ozono. Estudios recientes identifican a Lima como una de las ciudades sudamericanas con mayores niveles de polución, especialmente en estaciones críticas como el invierno (Carbo-Bustinza et al., 2023; Quispe et al., 2024).

En el análisis de la contaminación por ozono, los métodos estadísticos y computacionales han demostrado ser herramientas clave para comprender patrones temporales, detectar anomalías y desarrollar modelos predictivos. En primer lugar, los modelos tradicionales como el Autorregresivo Integrado de Media Móvil (ARIMA) han sido aplicados con éxito a series temporales ambientales, gracias a su capacidad para capturar tendencias y estacionalidades (Box et al., 2015). No obstante, suponer linealidad puede ser una limitante al analizar fenómenos atmosféricos inherentemente complejos y no lineales.

Por ello, se han incorporado enfoques más avanzados como los modelos no paramétricos autorregresivos, redes neuronales artificiales (ANN), y modelos de aprendizaje profundo como el Extreme Learning Machine (ELM), que permiten modelar relaciones no lineales entre múltiples variables. Estas técnicas han mostrado gran eficacia en predicciones de alta precisión al incorporar variables meteorológicas como predictores exógenos (Gardner & Dorling, 2000; Huang et al., 2006). Modelos híbridos, que combinan ARIMA con redes

neuronales, también han sido implementados con éxito para mejorar la capacidad de predicción en series de datos ambientales complejos (Zhang, 2003).

En el caso específico de Lima, estudios recientes han utilizado una combinación de estos enfoques para modelar el comportamiento del ozono en función de datos meteorológicos y espaciales. Se han desarrollado sistemas de predicción a corto plazo que integran aprendizaje automático y estadísticas multivariadas para anticipar concentraciones horarias de O_3 , permitiendo una toma de decisiones más oportuna en la gestión de la calidad del aire (Carbo-Bustanza et al., 2023; Quispe et al., 2024; Scientific Reports, 2022).

Este estudio tiene como objetivo evaluar la relación entre los niveles de ozono y las variables meteorológicas en Lima Metropolitana mediante el uso de técnicas estadísticas avanzadas y modelos de predicción basados en series temporales. El análisis se enfoca en los datos obtenidos durante los inviernos de 2017 a 2019 en cuatro estaciones de monitoreo urbano: Ate, Campo de Marte, San Borja y Santa Anita. La investigación busca identificar patrones estacionales, componentes determinísticos y estocásticos de la serie, y evaluar la capacidad predictiva de diversos modelos utilizando métricas como MAE, MAPE, MDAE y MASE. Así, se pretende contribuir con evidencia empírica robusta al diseño de políticas públicas en salud y medio ambiente en el contexto limeño.

Metodología

Area de estudio

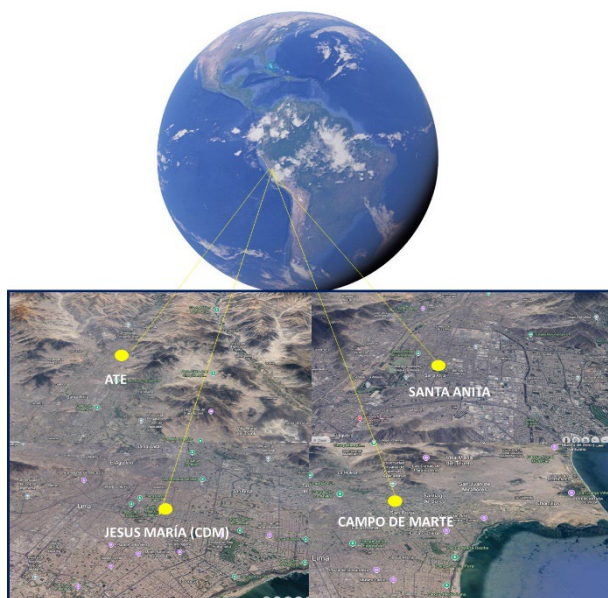


Figura 1. Sitios de monitoreo de parámetros de calidad De aire y meteorológicos en lima Metropolitana a cargo de SENAMHI.

Predicción de ozono troposférico usando MLR y SRM

Se aplicaron los modelos de MLR y SRM para predecir la concentración de ozono troposférico en Lima Metropolitana a partir de los datos observados horarios durante la estación primaveral de los años 2018-2019. Se hicieron tres ejercicios que dividieron los datos en tres grupos: i) se aplicaron todos los datos en general para aplicar el modelo MLR; ii) se usaron los valores de los factores meteorológicos que correspondían a concentraciones de ozono inferiores a los 60 mg/m³ y iii) se aplicaron todos los datos de los factores meteorológicos que correspondieron a concentraciones de ozono superiores a los 60 mg/m³

La ecuación (1) muestra la regresión en términos de coeficientes no estandarizados; B se utilizó para generar la ecuación de regresión de ozono en el periodo evaluado dentro del área de estudio:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad (1)$$

$$Y = \beta_0 + \beta_1 T + \beta_2 RH + \beta_3 WV + \beta_4 WD \quad (2)$$

La ecuación 2 presenta los parámetros meteorológicos aplicados en esta investigación. Donde β_0 es la constante de la ecuación de regresión y $\beta_1, \beta_2, \beta_3$, y β_4 , representa las constantes explicativas de cada variable meteorológica (X_1, X_2, X_3, X_4), respectivamente. El ajuste de este modelo solo mostro un 528% de rendimiento logrado expresado por el coeficiente de determinación R ajustado

Análisis de Superficie de Respuesta - Response Surface Methodology (RSM)

Evaluación de modelos

La evaluación de los modelos se ha efectuado mediante el cálculo de la capacidad predictiva representada por el factor de correlación de Pearson (r), usado para determinar la dependencia lineal entre los valores de ozono observados y calculados mediante los modelos, este coeficiente varía entre -1 y 1, tales valores extremos denotan correlaciones perfectas negativas y positivas respectivamente.

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}} \quad (3)$$

Donde:

X_i, Y_i = Valores individuales de las variables X e Y; $\bar{X}, \bar{Y}$ = Valores medios de X e Y;

$$\sum = \text{Sumatoria de los valores en el conjunto de datos}$$

La precisión predictiva, este parámetro estadístico se determina por el cálculo de Error Cuadrático Medio de la raíz que es la medida del error de predicción (RMSE) como una medida de precisión del modelo frente a los valores observados, donde:

$$RMSE = \sqrt{\frac{\sum(Y_{pred} - Y_{obs})^2}{n}} \quad (4)$$

donde: Yobs = Valores observados, Ypred= Valores predichos, n = Número de observaciones.

El índice de Willmott (IA) que mide cuan preciso un modelo expresa la variabilidad de los datos describiendo un patrón de perturbación comparado con el valor medio. Este parámetro oscila entre 0 y1, este último señala una concordancia perfecta entre los valores observados y los resultados predichos en el modelo. La ecuación explica este parámetro:

$$IA = 1 - \frac{\sum(Y_{pred} - Y_{obs})^2}{\sum(|Y_{pred} - \bar{Y}_{obs}| + |Y_{obs} - \bar{Y}_{obs}|)^2} \quad (5)$$

donde: Yobs, valores observados; Ypred, valores predichos por el modelo; $\bar{Y}_{obs}$, promedio de los valores observados.

También se aplicó el estadístico Fractional Bias (FB) que calcula la concordancia entre los valores observados y los valores calculados por el modelo, este parámetro presenta una tendencia sistemática ya que puede sobrestimar o subestimar los valores observados. Un FB =0 significa que se ha logrado una concordancia perfecta entre ambos grupos de valores observados y calculados para el modelo.

$$FB = \frac{(Y_{pred} - Y_{obs})}{0.5 \times (Y_{pred} + Y_{obs})} \quad (6)$$

donde: Yobs, valor medio de las observaciones; Ypred, valor medio de las predicciones.

Resultados y Discusiones

La tabla 1 muestra los resultados generales del bulk de valores observados en la estación de primavera desde 2017 a 2019 en los 04 sitios evaluados.

Statistical Values	Qzone (mg/m ³)	Temp (°C)	Relative Humidity (%)	Wind Speed (m/s)	Wind Direction (degrees)
Mean	21	15.8	89.2	1.5	216.3
desv stan	25.3	1.5	7.3	0.9	40.6
min	0.1	12.3	57	0	0

max	165.8	24	100	7.1	360
-----	-------	----	-----	-----	-----

La primavera ha presentado una velocidad del viento discreta (media: 1.5 m/s), siendo CDM el que presenta una mayor intensidad y Ate el más estable (CDM>Sta. Anita>San Borja>Ate); con una dirección predominante al suroeste y un ligero sesgo hacia el oeste. Asimismo, se presenta una temperatura media de 15.8 °C, se alcanza los menores valores en agosto y las temperaturas más cálidas ocurren en setiembre, estos valores generaron un orden decreciente de valores medios desde San Borja>CDM>Ate-Santa Anita. Por otro lado, Lima es una ciudad muy húmeda, esto se refleja en la humedad relativa elevada (mean 89.2%) que llega alcanzar el 100%, en el siguiente orden: San Borja>Ate-Sta Anita>CDM.

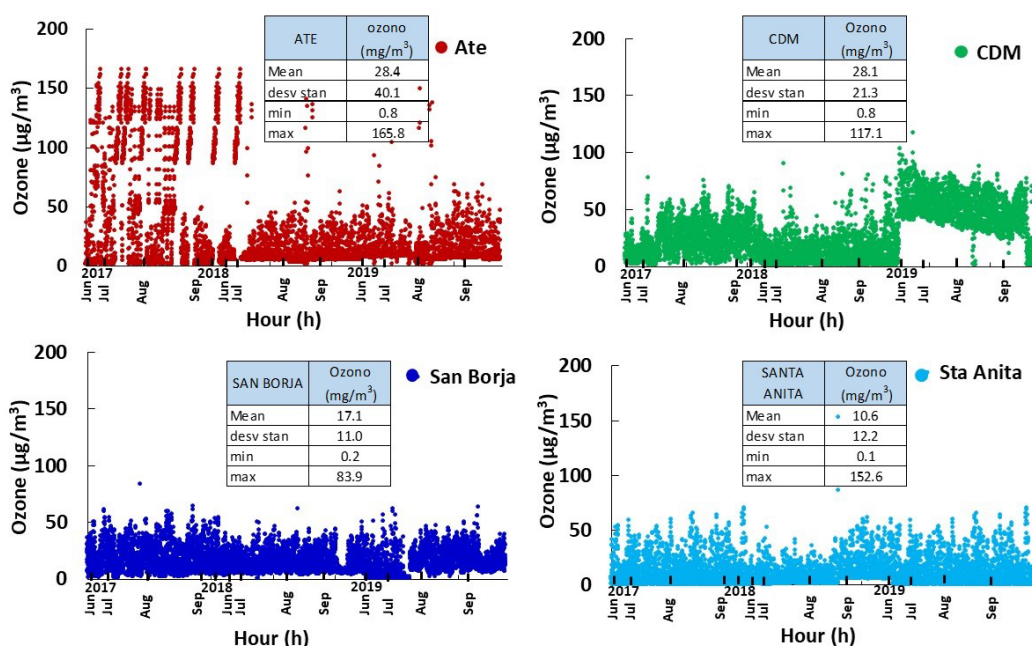


Figura 2. Variación horaria de Ozono troposférico en cuatro sitios de monitoreo de Lima metropolitana: a) Ate; b) Campo de Marte (CDM); c) San Borja y d) Santa Anita

La figura 2 muestra la variabilidad horaria del ozono troposférico en los sitios seleccionados de Lima Metropolitana. De acuerdo a esta figura 2a y 2b, el orden de concentraciones medidas de mayor a menor sigue el siguiente orden Ate > CDM > San Borja > Sta Anita. De acuerdo a la figura 3, los datos fueron divididos en dos grandes grupos: grupo a) para los factores meteorológicos que incluyeron en concentraciones altas de ozono > 60 mg/m³ y b) para los factores meteorológicos que influyeron en concentraciones de ozono < 60 mg/m³. Concentraciones menores de 60 mg/m³ mostraron una mayor humedad relativa (cuartil 2:3 de 85-95%), mientras que concentraciones superiores de ozono se asociaron a menores porcentajes de humedad relativa (cuartil 2:3 de 81-92.5%)

El distrito de Ate se caracteriza por tener medianas y grandes empresas productivas que representan el 10% de las actividades económicas del distrito, seguidas del 81.6% de medianas y pequeñas empresas, le

sigue el distrito de Santa Anita, además el parque automotor agudiza los problemas ambientales al lado de una gestión limitada de residuos sólidos.

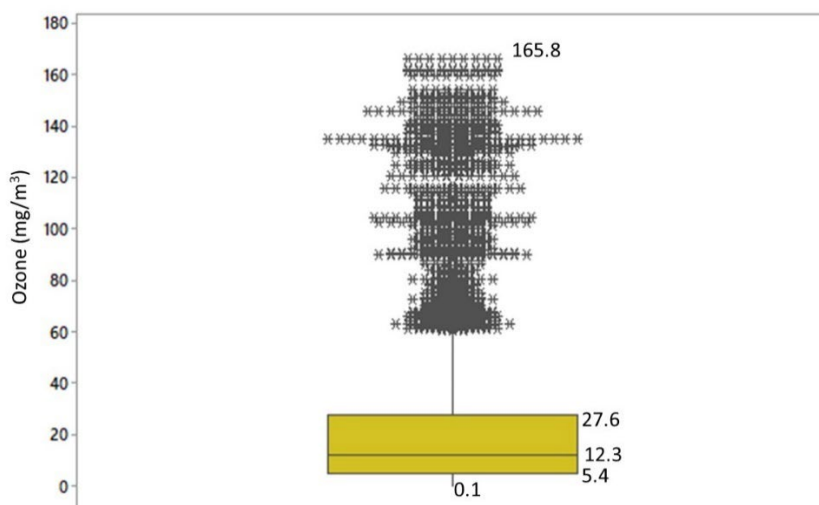


Figura 3. Box plot de las concentraciones ozono en los cuatro sitios evaluados de Lima metropolitana.

De acuerdo a este patrón las pruebas de correlación de Pearson (Tabla 1a) demostraron principalmente a la humedad relativa e intensidad De los vientos, seguida de la temperatura como factores principales que influyen sobre concentraciones de ozono inferiores a 60 mg/m³. En cambio, para concentraciones de ozono superiores (tabla 1b) a los 60mg/m³, es la intensidad del viento seguida de la dirección la que predomina.

De acuerdo a la tabla 1 ciertas condiciones de las variables meteorológicas presenta una correlación significativa con las concentraciones de ozono (> 60 mg/m³), en cambio para concentraciones superiores a los 60 mg/m³ solo el viento presenta correlaciones más significativas (p=0). Para concentraciones de ozono menores a 60 mg/m³ se observa una correlación positiva con la temperatura y positiva para valores mayores de ozono, así las tendencias para RH y WV también mostraron tendencias opuestas entre sí. En ambos casos solo la dirección del viento mostro una correlación positiva en ambos casos.

Tabla 1. Correlación entre variables meteorológicas y ozono troposférico

a) Ozone < 60 mg/m ³	O3 (mg/m ³)	Temp (°C)	Rel (%)	Hum	WV (m/s)
Temp (°C)	0.197				
p-value	0				
Rel Hum (%)	-0.244	-0.818			
p-value	0	0			
WV (m/s)	0.227	0.375	-0.425		
p-value	0	0	0		
WD (degrees)	0.023	-0.017	-0.076	-0.108	

p-value	0	0.006	0	0
b) Ozone > 60 mg/m ³	O ₃ (mg/m ³)	Temp (°C)	Rel Hum (%)	WV (m/s)
Temp (°C)	-0.006			
p-value	0.81			
Rel Hum (%)	0.043	-0.893		
p-value	0.083	0		
WV (m/s)	-0.489	0.395	-0.419	
p-value	0	0	0	
WD (degrees)	0.249	0.102	-0.075	-0.163
p-value	0	0	0.002	0

Prediction of O₃ (mg/m³) using MLR

MLR usando todos los datos observados (MLR):

La tabla 2 muestra los resultados de del ajuste de los valores meteorológicos usados como predictores de ozono. La ecuación MLRt representa el modelo general aplicado a 108292 datos meteorológicos usados en esta investigación. Se hicieron predicciones para concentraciones de ozono inferiores a 60 mg/m³ (MLR1<60, MLR2<60, MLR3<60, MLR4<60); y mayores a 60 mg/m³: MLR5>60, MLR6>60, MLR7>60.

Tabla 2. Modelos de predicción (MLR) para concentraciones de ozono troposférico.

Type Model	Model	RMSEs	IA	FB
MLRt	O ₃ (mg/m ³) = 130.06 - 1.962 T - 0.9914 RH + 0.677 WV + 0.04361 WD	18.048	0.976	-5.02E-04
MLR1<60	O ₃ (mg/m ³) = 45.47 - 0.132 T - 0.3717 RH + 2.618 WV + 0.00872 WD	13.714	0.366	-5.10E-04
MLR2<60	O ₃ (mg/m ³) = 49.77 - 0.186 T - 0.3880 RH + 2.550 WV	13.718	0.364	-5.54E-05
MLR3<60	O ₃ (mg/m ³) = 41.36 - 0.3498 RH + 2.613 WV + 0.00909 WD	13.725	0.363	-4.40E-05
MLR4<60	O ₃ (mg/m ³) = 44.11 - 0.3574 Hum + 2.539 WV	13.719	0.364	3.30E-02
MLR5>60	O ₃ (mg/m ³) = 43.8 + 3.483 T - 0.011 RH - 16.462 WV + 0.1134 WD	25.696	0.667	-1.13E-04
MLR6>60	O ₃ (mg/m ³) = 153.26 - 0.6551 Hum - 16.255 WV + 0.1199 WD	25.831	0.66	-4.58E-05
MLR7>60	O ₃ (mg/m ³) = 61.73 + 4.029 Temp>60 - 17.482 WV	26.036	0.649	4.11E-05

Aunque los parámetros estadísticos de evaluación de los modelos son muy cercanos, no obstante, MLR1 se presenta con mayor precisión para predecir el ozono troposférico con valores inferiores a los 60 mg/m³, mientras que para concentraciones mayores MLR5 muestra una mejor respuesta. Los valores de FB

tienden a ser negativos mostrando una subestimación de la precisión y el Índice IA resulta más preciso para ajustes que generan concentraciones de ozono superiores a los 60 mg/m³.

Concentraciones menores de 60 mg/m³ mostraron una mayor humedad relativa (cuartil 2:3 de 85-95%), mientras que concentraciones superiores de ozono se asociaron a menores porcentajes de humedad relativa (cuartil 2:3 de 81-92.5%). De acuerdo a la tabla 1 ciertas condiciones de las variables meteorológicas presenta una correlación significativa con las concentraciones de ozono (> 60 mg/m³), en cambio para concentraciones superiores a los 60 mg/m³ solo el viento presenta correlaciones más significativas (p=0). Para concentraciones de ozono menores a 60 mg/m³ se observa una correlación positiva con la temperatura y positiva para valores mayores de ozono, así las tendencias para RH y WV también mostraron tendencias opuestas entre sí. En ambos casos solo la dirección del viento mostro una correlación positiva en ambos casos.

Respecto al MLRt aunque el RMSE es relativamente alto (18.048), el IA es excelente, indicando que las predicciones son muy precisas en general, y el FB cercano a cero indica que no hay sesgo importante, por lo que este modelo tiene un ajuste confiable. Los demás ajustes de MLR1 a 4 para concentraciones menores a 60 mg/m³ resultan con RMSE bajo, lo que sugiere que el error general es relativamente pequeño, pero el ajuste (IA = 0.3) es pobre lo cual muestra limitaciones del modelo. Respecto a los modelos lineales para concentraciones de ozono mayores a 60 mg/m³ es el MLR5 (RMSE = 25.696, IA = 0.667, FB = -1.13E-04), que presenta un mejor modelo genera un ajuste moderado con un bajo sesgo y un error relativamente menor. La tabla 3 presenta los resultados de la regresión de variables meteorológicas para predecir las concentraciones de ozono troposférico, como se aprecia el MLRt muestra un Rsq(adj) inferior (5.28%) frente a los demás modelos; las regresiones de MLR7 (Rsq_{adj}=28%) resulta más apropiado ya que el modelo se ajusta mejor a los datos.

Tabla 3. Resultados del ajuste de los modelos MLR

Model	S	R-sq	R-sq(adj)	R-sq(pred)
MLRt	24.6	5.29%	5.28%	5.26%
MLR1<60	13.7151	7.87%	7.85%	7.83%
MLR2<60	13.7192	7.81%	7.80%	7.78%
MLR3<60	13.7153	7.86%	7.85%	7.83%
MLR4<60	13.7199	7.80%	7.79%	7.77%
MLR5>60	25.7	29.95%	29.78%	29.51%
MLR6>60	25.8619	29.21%	29.08%	28.85%
MLR7>60	26.0591	28.08%	28.00%	27.84%

El gráfico Q-Q (Figura 4) o normal probability plot para el modelo MLRt muestra asimetría y colas más pesadas en los extremos, señalando una distribución que no es completamente normal.

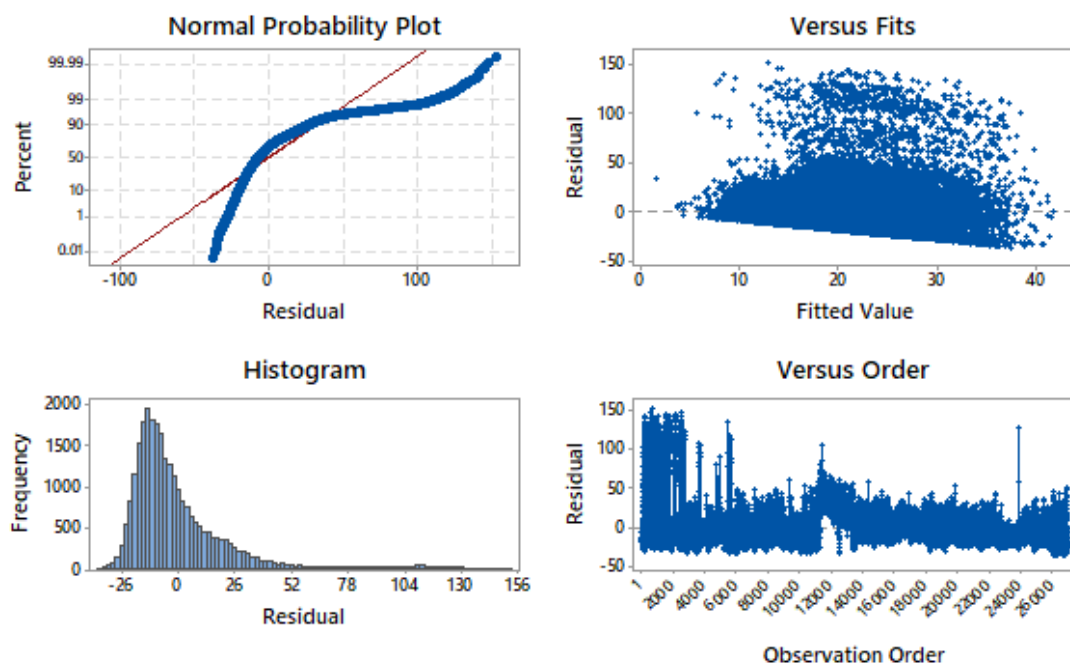


Figura 4. Residual Plots for ozono en cuatro sitios de Lima Metropolitana.

Los resultados del análisis "fitted value vs. residual" evaluar la calidad de un modelo de regresión, donde "fitted values" está conformado por aquellos valores calculados por el modelo "predichos" con el MLR y los residuos representa las diferencias encontradas entre los valores observados y los ajustados, dado por: $\text{Residuos} = \text{Valores observados} - \text{Valores ajustados}$.

Se observa una tendencia general que sugiere una relación lineal debido a que se muestra una mayor concentración de valores sobre la línea, pero el modelo subestima los valores en esa región, aunque no es estrictamente lineal y que un modelo cuadrático podría ajustarse mejor, la dispersión mostrada puede indicar la existencia de valores atípicos o extremos lo cual influye en la línea de ajuste. El análisis de "frecuencia vs. residuos" verifica la calidad del modelo de regresión si los residuos cumplen con los supuestos de normalidad y homocedasticidad, como se aprecia desviaciones significativas de la normalidad, lo cual señala considerar transformación de los datos, la asimetría y sesgo mostrado por la concentración de residuos a la izquierda de cero, junto con un punto máximo negativo, señala un sesgo en el modelo, y se podría incluir un modelo no lineal.

El análisis de variancias (ANOVA) (Tabla 4), muestra significancia estadística para la aplicación de la ecuación general MLRt, sin embargo, cuando se trata de predicciones para concentraciones s de ozono superiores a los 60 mg/m^3 la temperatura no resulta significativa ($p = 0.198$); mientras que la variación de la humedad relativa no ejerce influencia significativa para concentraciones de ozono mayores a 60 mg/m^3 . Para aquellas concentraciones de ozono inferiores a los 60 mg/m^3 los valores de temperaturas oscilaron entre 14.8 (cuartil 1)-16.6 (cuartil Q3) con un promedio de 15.81°C , con humedades relativas entre 85

(cuartil 1) y 95% (cuartil 3) y una mediana de 90%. El 61% de los datos de ozono correspondieron a valores superiores a 60 mg/m³, en este caso la RH no resulto significativa y más bien la temperatura ($p > 0.05$) y el viento jugaron un rol preponderante en el modelo.

Tabla 4. ANOVA de los principales modelos calculados para la predicción de ozono troposférico

a) Source MLRt	DF	AdjSS	Adj MS	F-Value	P-Value
Regression	4	916962	229241	378.1	0
Temp (°C)	1	73760	73760	121.66	0
RH (%)	1	423321	423321	698.2	0
WV (m/s)	1	7257	7257	11.97	0.001
WD (degrees)	1	80847	80847	133.34	0
Error	27067	16410799	606		
Lack-of-Fit	26900	16362778	608	2.12	0
Pure Error	167	48021	288		
Total	27071	17327761			
b) Source MLR1<60	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	408221	102055	542.54	0
Temp (°C)	1	312	312	1.66	0.198
RH (%)	1	55487	55487	294.98	0
WV (m/s)	1	99259	99259	527.68	0
WD (degrees)	1	3044	3044	16.18	0
Error	25409	4779553	188		
Lack-of-Fit	25253	4748880	188	0.96	0.667
Pure Error	156	30674	197		
Total	25413	5187774			
c) Source MLR5>60	DF	AdjSS	Adj MS	F-Value	P-Value
Regression	4	467995	116999	176.66	0
Temp (°C)	1	11529	11529	17.41	0
RH (%)	1	2	2	0	0.952
WV (m/s)	1	361014	361014	545.12	0
WD (degrees)	1	29148	29148	44.01	0
Error	1653	1094726	662		
Total	1657	1562721			

Principal Component Analysis: Temp; Hum; WV; WD

La tabla 5 provee los resultados del PCA, se ha considerado como relevante el eigenvlor > 1.0 porque mide la cantidad de varianza que cada componente principal captura de los datos originales, de tal modo que PC1 y PC2 resultan significativos al lado de los demás componentes.

Tabla 5. a) Eigenanalysis of the Correlation Matrix

Eigenvalue	2.111	1.0474	0.6714	0.1702
Proportion	0.528	0.262	0.168	0.043
Cumulative	0.528	0.79	0.957	1

b) Componentes principales

Variable	PC1	PC2	PC3
Temp	-0.624	-0.067	0.372

Hum	0.635	0.149	-0.233
WV	-0.455	0.299	-0.836
WD	0.001	-0.94	-0.329

La humedad tiene una relación positiva con PC1 mostrando una asociación favorable, mientras que la dirección del viento con una carga alta y negativa en PC2 indica que, a medida que la dirección del viento cambia PC2 disminuye, esto implicaría que ciertas direcciones del viento podrían influenciar en el comportamiento del ozono. De acuerdo con los resultados PC1 y PC2 explican el 79% de la variabilidad, siendo la temperatura y la humedad las más relevantes en PC1 dado que captura la mayor parte de la variabilidad en el conjunto de datos evaluados. En PC2, se muestra la captura ortogonal de datos no explicados por PC1, de tal forma que la dirección del viento parece ser un factor significativo en este componente. A diferencia del MRA el PCA gestiona variables independientes; porque permite la reducción de la dimensionalidad, elimina la multicolinealidad, porque transforma las variables iniciales en componentes no correlacionados, por los nuevos componentes ortogonales entre sí, esto significa que el PCA no es sólido cuando se usan predictores en una ecuación de regresión, donde los factores están fuertemente correlacionados.

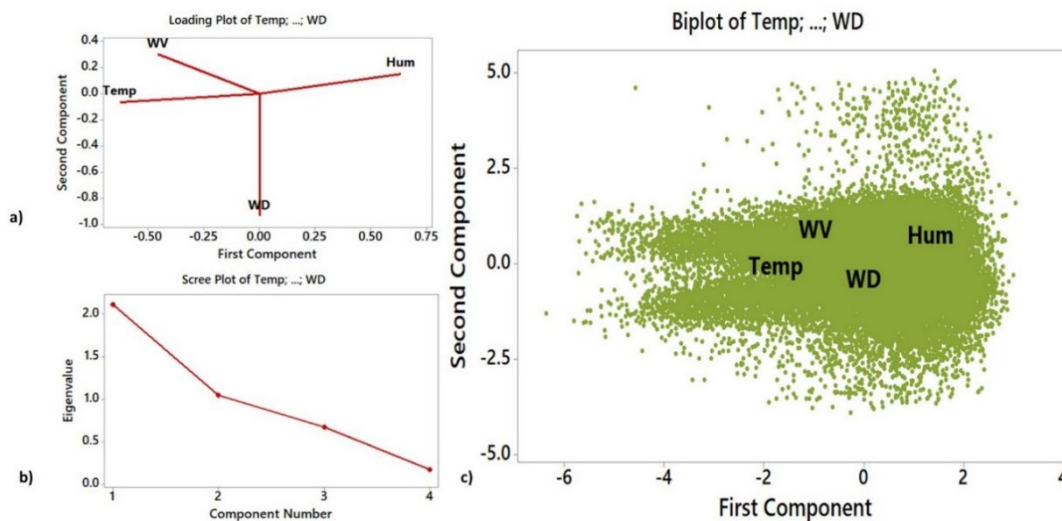


Figura 5. Análisis de componentes principales a) loading plot; b) Scree plot; c) biplot

Response Surface Regression: O₃ versus Temp; Hum; WV; WD

El gráfico tridimensional del volumen los dos principales factores: temperatura y humedad y de la variable respuesta (O₃) se observa en la figura 1. La figura 1 muestra la distribución tridimensional de las variables temperatura-humedad, b) dirección-intensidad de vientos, c) intensidad de vientos-temperatura y d) dirección de vientos –humedad relativa.

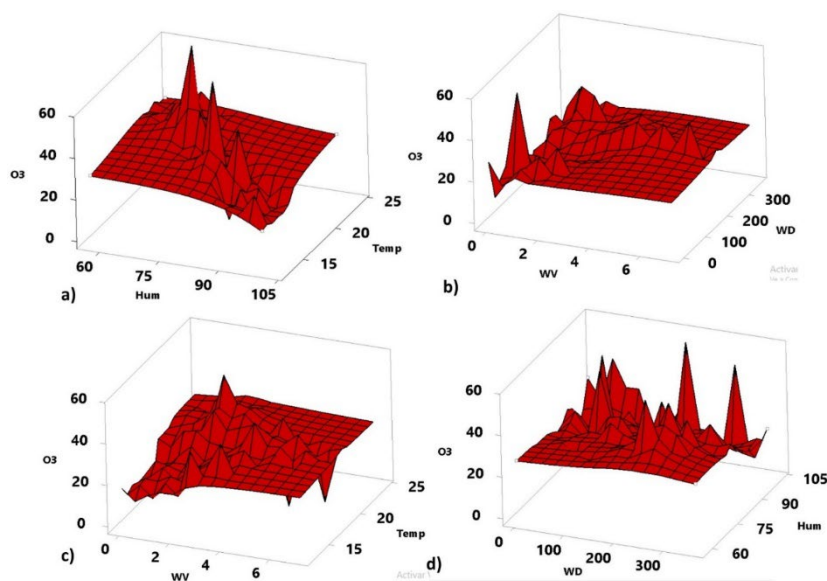


Figura 6. Representación tridimensional de la distribución de los factores meteorológicos que influyen en la concentración de ozono troposférico.

El ajuste de los datos muestra que la ecuación que RSM1 presenta un mayor Rsq adj (29.35%) para concentraciones de ozono mayores a 60 mg/m³ mientras que la ecuación RSM8 mostro un mejor ajuste Rsq adj (7.79%).

Tabla 6. Modelos de predicción (MSM) para concentraciones de ozono troposférico.

Model	Ecuación	Rsq adj
RSM1<60	$O_3 \text{ (mg/m}^3\text{)} = 120.6 - 2.126 RH + 2.07 WV + 0.01017 RH \cdot RH + 0.0774 WV \cdot WV + 0.0021 RH \cdot WV$	8.03%
RSM2<60	$O_3 \text{ (mg/m}^3\text{)} = 368.6 - 11.60 T - 5.32 RH + 0.1382 T \cdot T + 0.02019 RH \cdot RH + 0.0777 T \cdot RH$	6.22%
RSM3<60	$O_3 \text{ (mg/m}^3\text{)} = 60.26 - 7.741 T + 9.24 WV + 0.2955 T \cdot T + 0.4246 WV \cdot WV - 0.4777 T \cdot WV$	7.00%
RSM4<60	$O_3 \text{ (mg/m}^3\text{)} = 117.0 - 2.049 RH + 0.0883 WD + 0.00961 RH \cdot RH - 0.000111 WD \cdot WD - 0.000453 RH \cdot WD$	6.26%
RSM5<60	$O_3 \text{ (mg/m}^3\text{)} = 32.21 - 4.018 T + 0.0147 WD + 0.1620 T \cdot T - 0.000114 WD \cdot WD + 0.00257 T \cdot WD$	4.19%
RSM6<60	$O_3 \text{ (mg/m}^3\text{)} = 14.84 - 3.466 WV - 0.0258 WD + 0.1346 WV \cdot WV + 0.000020 WD \cdot WD + 0.03278 WV \cdot WD$	5.76%
RSM7>60	$O_3 \text{ (mg/m}^3\text{)} = -5352 + 299.8 T + 69.4 RH - 3.934 T \cdot T - 0.2131 RH \cdot RH - 1.982 T \cdot RH$	2.68%
RSM8>60	$O_3 \text{ (mg/m}^3\text{)} = 60.3 + 8.46 T - 64.35 WV - 0.263 T \cdot T + 0.417 WV \cdot WV + 2.747 T \cdot WV$	29.77%
RSM9>60	$O_3 \text{ (mg/m}^3\text{)} = 388.8 - 24.69 T - 1.086 WD + 0.467 T \cdot T + 0.001787 WD \cdot WD + 0.0372 T \cdot WD$	12.21%
RSM10>60	$O_3 \text{ (mg/m}^3\text{)} = 98.2 + 0.89 RH + 13.4 WV - 0.00617 RH \cdot RH + 0.740 WV \cdot WV - 0.404 RH \cdot WV$	28.23%
RSM11>60	$O_3 \text{ (mg/m}^3\text{)} = 51 - 0.86 RH + 0.258 WD + 0.01805 RH \cdot RH + 0.001739 WD \cdot WD - 0.00845 RH \cdot WD$	12.36%

$$\text{RSM12}>60 \quad \text{O}_3 \text{ (mg/m}^3\text{)} = 133.31 - 65.73 \text{ WV} - 0.0744 \text{ WD} + 2.862 \text{ WV*WV} + 0.000166 \text{ WD*WD} + 0.1893 \text{ WV*WD} \quad 29.90\%$$

La ecuación de RSM1<60 ha presentado un mayor ajuste ($R^2 = 8.03\%$) en concentraciones inferiores a 60 mg/m^3 eso implica que cuando la RH es baja ($< 70\%$) con mayores intensidades de viento ($> 3\text{m/s}$), ayuda a mantener condiciones aceptables de ozono troposférico en el sistema atmosférico (Figura 7^a). Las mayores concentraciones de ozono ($> 60 \text{ mg/m}^3$) se presentan para intensidades $< 6 \text{ m/s}$ con vientos que vienen del oeste y sur oeste y soplan al este y nor este, inclusive vientos del norte oeste dirigidos hacia el sur este (figura 7b).

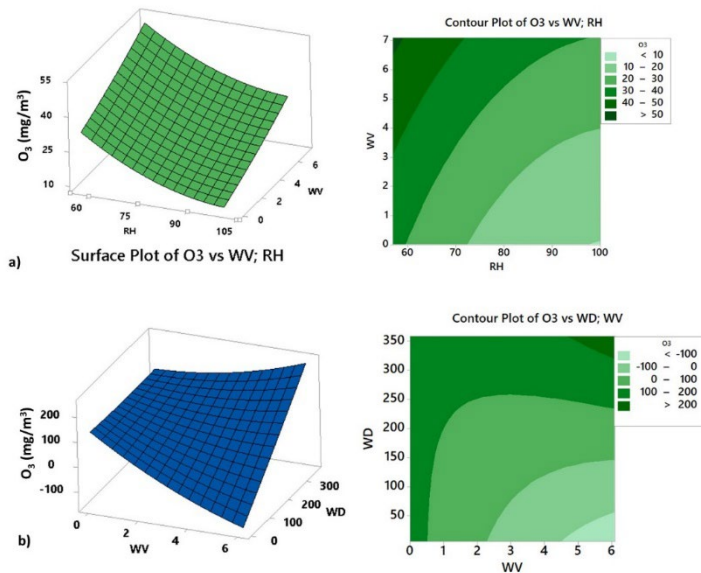


Figura 7. Gráfico de superficie de respuesta (RSM) para a) WD-WD, $< 60 \text{ mg/m}^3$ de ozono b) WV-DW, $> 60 \text{ mg/m}^3$ de ozono

La bondad del ajuste entre los valores observados y los valores calculados o predichos, se muestra en las figuras 9 y 9, la RH y WV muestran el mayor coeficiente (valor absoluto) demostrando una mayor influencia en la presencia de concentraciones de ozono inferiores a los 60 mg/m^3 (Figura 8), mientras que la temperatura y la velocidad del viento dominan la variación de ozono superiores a los 60 mg/m^3 (figura 9).

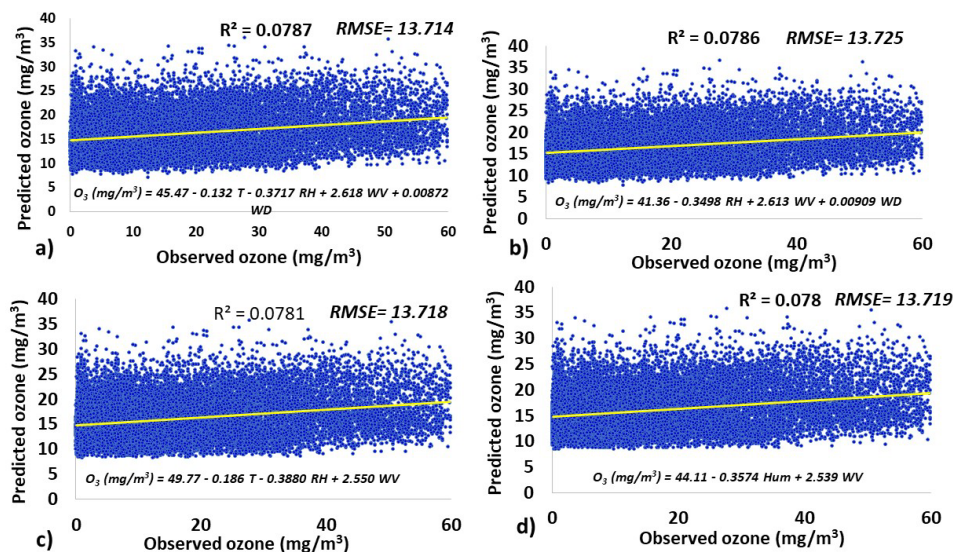


Figura 8. Predicción de ozono ($< 60 \text{ mg/m}^3$) basada en los valores observados de ozono en cuatro sitios de lima metropolitana durante la primavera 2017-2019 usando MLR: (a) T-RH-WV-WD, (b) RH-WV-WD (c) T-RH-WV (d) RH-WV

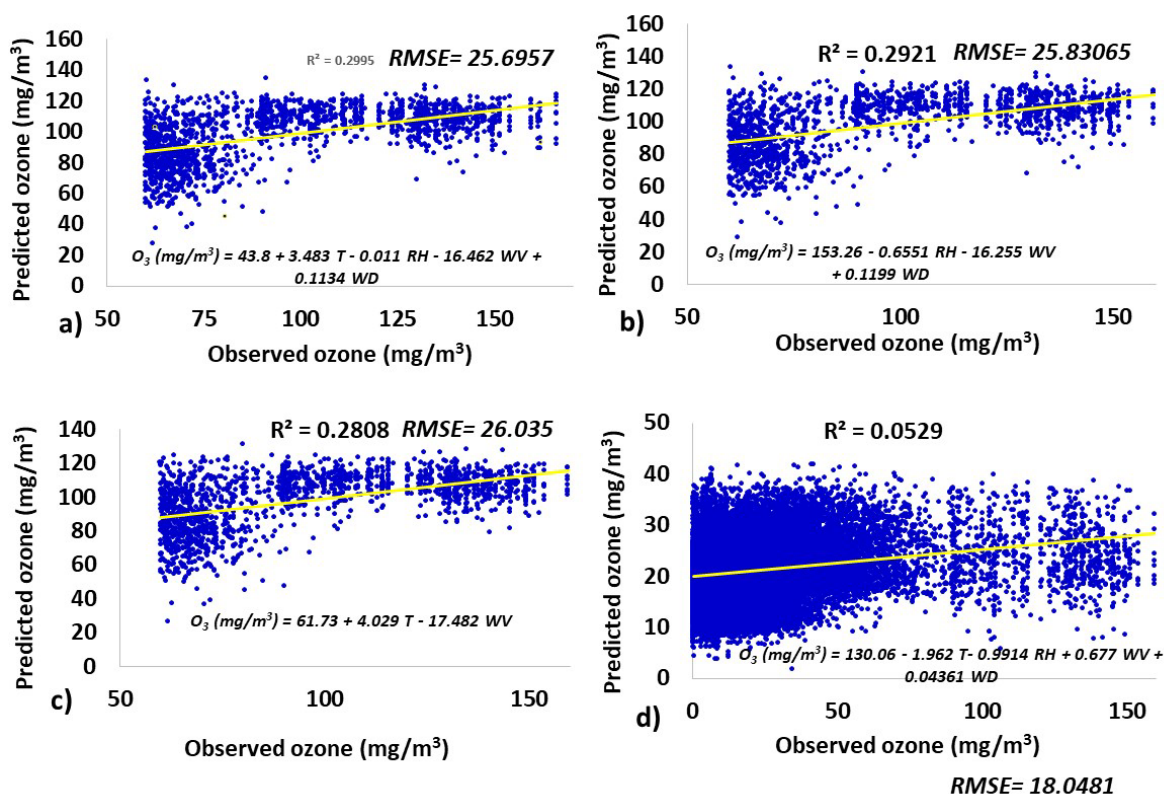


Figura 9. Predicción de ozono $> 60 \text{ mg/m}^3$, basada en los valores observados de ozono en cuatro sitios de Lima metropolitana durante la primavera 2017-2019: (a) MLR (T-RH-WV-WD), (b) MLR (RH-WV-WD) (c) MLR (T-WV), (d) MLRt; (T-RH-WV-WD)

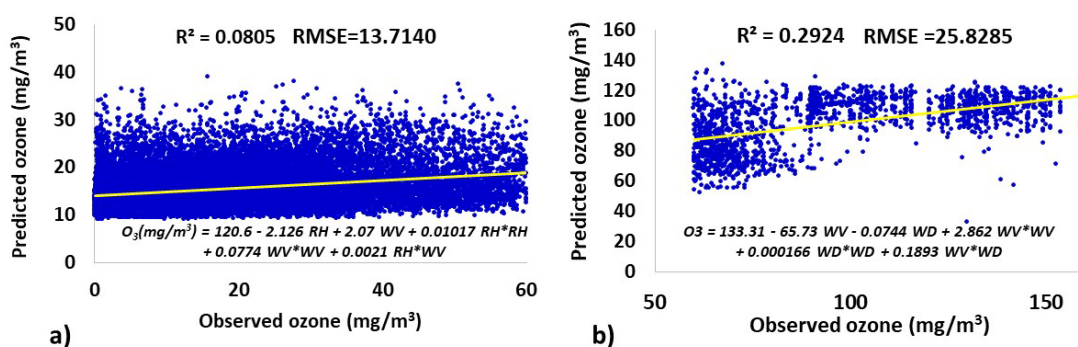


Figura 10. Predicción de ozono por RSM: a) concentraciones $< 60 \text{ mg/m}$ de ozono y b) concentraciones $> 60 \text{ mg/m}^3$ de ozono

De acuerdo con la figura 10, la ecuación para factores que generan concentraciones de ozono $< 60 \text{ mg/m}^3$ presentó un menor error con el mayor porcentaje de determinación ($R^2=0.0805$) a diferencia del modelo para mayores concentraciones de ozono.

Las diferencias entre los distritos son relevantes, el distrito de Ate se caracteriza por ser un área industrial de grandes empresas (textiles, cervezas, etc) que constituyen el 10% de la economía del lugar seguido de las medianas y pequeñas empresas, mucho comercio formal e informal y un extenso parque automotor, el área es plana sin embargo se encuentra cerrada por cerros, le sigue el distrito de Santa Anita que además presenta importantes centros comerciales y es uno de los distritos con mayor tráfico vehicular. Ambos distritos se encuentran más alejados de la línea costera y circunscritos con áreas montañosas que influyen en la circulación atmosférica y que no ha sido considerado en esta investigación. Los distritos de Jesús maría (donde se encuentra “Campo de Marte”) se ubican mas apegados al mar en zonas planas presentado áreas más abiertas para la circulación atmosférica., las altitudes son bajas en el siguiente orden: Ate (355msnm)>Santa Anita (234 msnm)>San Borja (170 msnm)>Jesús María (115 msnm).

Conclusiones

En esta investigación se ha evidenciado dos grandes grupos de concentración del ozono troposférico, el primer grupo presenta valores inferiores a los 60 mg/m³ y el segundo muestra valores superiores que alcanzaron una máxima concentración de 167 mg/m³, siendo Ate el área con un mayor contenido de ozono troposférico, seguido de Santa Anita. En cambio, los distritos de San Borja y Jesús María (Campo de Marte) mostraron menores valores de ozono. Existe una diferencia entre las actividades humanas separando a este grupo en aquellos con mayor actividad industrial, comercial (Ate y Santa Anita) con aquellos donde el primordial agente contaminante solo es el parque automotor. Sin embargo, existen otras características de geolocalización y topográficas que no han sido estudiadas en este documento que generan particularidades en cada sitio de muestreo.

Respecto a los modelos evaluados MLR y RSM, el análisis se ha basado en los parámetros estadísticos RMSE, AI, FB, MLR1, correlaciones de Pearson y coeficientes de determinación. El modelo general MLR resultó más representativo y confiable basado en un RMSE relativamente alto (18.048), pero con un valor IA (0.976) muy alto y sesgo insignificante. Las divisiones para de los demás modelos que predicen concentraciones menores y mayores de 60 mg/m³ mostraron ajuste IA (< 0.67) moderados a bajos lo cual disminuyó la precisión o confiabilidad de los modelos. En cambio, el análisis de superficie de respuesta, ha generado modelos con menores errores para concentraciones de ozono menores a 60 mg/m³, pero con menor valor del coeficiente de determinación entre los valores observados y modelados.

De otro lado, el análisis de componentes principales, ha mostrado que los dos primeros componentes lograron explicar el 79% de la variabilidad, esto implica que es necesario tomar en cuenta otros factores

13 como el topográfico o fisicoquímico atmosférico, entre otros, no considerado en esta investigación. A pesar de ello, se evidenció una relación positiva y favorable entre la humedad relativa y PC1 y una respuesta negativa entre la dirección del viento con PC2, señalando que, a medida que la dirección del viento varía este podría influenciar en el comportamiento del ozono.

Referencias

1. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). John Wiley & Sons.
2. Carbo-Bustinza, N., Iftikhar, H., Belmonte, M., Cabello-Torres, R. J., De La Cruz, A. R. H., & López-Gonzales, J. L. (2023). Short-term forecasting of ozone concentration in metropolitan Lima using hybrid combinations of time series models. *Applied Sciences*, 13(18), 10514. <https://doi.org/10.3390/app131810514>
3. European Environment Agency. (2023). *Ground-level ozone effects on human health under the changing climate*. <https://climate-adapt.eea.europa.eu/en/observatory/evidence/health-effects/ground-level-ozone/ground-level-ozone>
4. Gardner, M. W., & Dorling, S. R. (2000). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14–15), 2627–2636. [https://doi.org/10.1016/S1352-2310\(97\)00444-0](https://doi.org/10.1016/S1352-2310(97)00444-0)
5. Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3), 489–501. <https://doi.org/10.1016/j.neucom.2005.12.126>
6. Quispe, F., Salcedo, E., Iftikhar, H., Zafar, A., Khan, M., Turpo-Chaparro, J. E., Rodrigues, P. C., & López-Gonzales, J. L. (2024). Multi-step ahead ozone level forecasting using a component-based technique: A case study in Lima, Peru. *AIMS Environmental Science*, 11(3), 401–425. <https://doi.org/10.3934/environsci.2024020>
7. Rai, R., & Agrawal, M. (2020). Urban air pollution and meteorological interactions: A global review. *Atmospheric Environment*, 237, 117679. <https://doi.org/10.1016/j.atmosenv.2020.117679>
8. Scientific Reports. (2022). A machine learning approach to analyse ozone concentration in metropolitan area of Lima, Peru. *Scientific Reports*. <https://www.nature.com/articles/s41598-022-26575-3>
9. U.S. Environmental Protection Agency. (2024). *Ground-level ozone basics*. <https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics>
10. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)