



ESCUELA UNIVERSITARIA DE POSGRADO

EVALUACIÓN DEL IMPACTO DE LA INVERSIÓN TÉRMICA EN LA
CONCENTRACIÓN DE DIÓXIDO DE AZUFRE (SO₂) PARA OPTIMIZAR
LA FISCALIZACIÓN AMBIENTAL MINERA EN EL SECTOR DE
CAJAMARQUILLA DEL DISTRITO LURIGANCHO-CHOSICA, LIMA

Línea de Investigación:

Biodiversidad, ecología y conservación

Tesis para optar el Grado Académico de Maestro en Gestión Ambiental

Autor

Fajardo Vargas, Lázaro Walther

Asesora

Ángeles Lazo, Ana María

ORCID: 0000-0003-1815-6700

Jurado

Zambrano Cabanillas, Abel Walter

Mendoza García, José Tomás

Altez Rodríguez, José Félix

Lima - Perú

2026

EVALUACIÓN DEL IMPACTO DE LA INVERSIÓN TÉRMICA EN LA CONCENTRACIÓN DE DIÓXIDO DE AZUFRE (SO₂) PARA OPTIMIZAR LA FISCALIZACIÓN AMBIENTAL MINERA EN EL SECTOR DE CAJAMARQUILLA DEL DISTRITO LURIGANCHO-CHO

INFORME DE ORIGINALIDAD

17 %

INDICE DE SIMILITUD

15 %

FUENTES DE INTERNET

11 %

PUBLICACIONES

6 %

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	hdl.handle.net Fuente de Internet	2 %
2	visorsig.oefa.gob.pe Fuente de Internet	1 %
3	repositorio.oefa.gob.pe Fuente de Internet	1 %
4	repositorio.unfv.edu.pe Fuente de Internet	<1 %
5	cdn.www.gob.pe Fuente de Internet	<1 %
6	Submitted to Universidad Nacional Federico Villarreal Trabajo del estudiante	<1 %
7	www.coursehero.com Fuente de Internet	<1 %
8	sinia.minam.gob.pe Fuente de Internet	<1 %
9	Submitted to Organismo de Evaluación y Fiscalización Trabajo del estudiante	<1 %
10	HATCH ASOCIADOS S.A.. "ITS del Proyecto Lavadores de Gases de SO ₂ en Plantas de	<1 %



Universidad Nacional
Federico Villarreal

VRIN | VICERRECTORADO
DE INVESTIGACIÓN

ESCUELA UNIVERSITARIA DE POSGRADO

EVALUACIÓN DEL IMPACTO DE LA INVERSIÓN TÉRMICA EN LA
CONCENTRACIÓN DE DIÓXIDO DE AZUFRE (SO₂) PARA OPTIMIZAR LA
FISCALIZACIÓN AMBIENTAL MINERA EN EL SECTOR DE CAJAMARQUILLA DEL
DISTRITO LURIGANCHO-CHOSICA, LIMA.

Línea de Investigación:

Biodiversidad, ecología y conservación

Tesis para optar el Grado Académico de Maestro en Gestión Ambiental

Autor

Fajardo Vargas, Lázaro Walther

Asesora

Ángeles Lazo, Ana María

ORCID: 0000-0003-1815-6700

Jurado

Zambrano Cabanillas, Abel Walter

Mendoza García, José Tomás

Altez Rodríguez, José Félix

Lima – Perú

2026

DEDICATORIA

A mis padres, **Juan Fajardo e Hilda Vargas**, por ser la base de todo lo que soy. Su amor incondicional, sacrificio silencioso y sabias enseñanzas han sido el pilar que ha sostenido cada uno de mis pasos a lo largo de la vida y, especialmente, en este camino académico. Este logro también es suyo.

A mi esposa, **Edith Venero**, por su paciencia inagotable, su comprensión en los momentos difíciles y su compañía constante en las largas jornadas de estudio y desvelo. Su apoyo firme y motivación diaria han sido clave para mantenerme enfocado y seguir adelante.

A mis hijos, **Walther, Jonathan y Priscila**, cuyo amor, palabras de aliento y alegría han sido un refugio emocional en los momentos de mayor exigencia. Gracias por enseñarme, sin proponérselo, el verdadero significado de la perseverancia.

A mis tíos, **Graciela, Esther, Jesús, Leoncio y Virgilio**, por su respaldo incondicional y constante presencia en mi vida. Su cariño y apoyo han sido siempre una fuerza silenciosa pero poderosa.

A mis hermanos, **Aydee, Juan y Hilda**, por su apoyo incondicional, por creer en mí y por estar siempre presentes, incluso en la distancia.

A todos ellos, mi gratitud profunda y eterna.

AGRADECIMIENTOS

A Dios

Mi más profundo agradecimiento a Dios, fuente de sabiduría, fortaleza y esperanza, por haber sido mi guía espiritual a lo largo de este proceso. En los momentos de dificultad y en cada paso de este camino, su presencia fue un sostén indispensable.

Al OEFA

Expreso mi sincero reconocimiento al **Organismo de Evaluación y Fiscalización Ambiental (OEFA)**, por haberme brindado no solo las herramientas técnicas necesarias, sino también el acceso a información clave que hizo posible el desarrollo riguroso y fundamentado de esta investigación. Su labor institucional fue esencial para el cumplimiento de los objetivos planteados.

A mi asesora

A la Dra. Ana María Ángeles Lazo, mi asesora, mentora y guía académica, le expreso mi más profundo agradecimiento por su acompañamiento constante, su rigurosidad metodológica y su generosa disposición. Sus valiosos aportes enriquecieron sustancialmente este trabajo, y su compromiso con la investigación ha sido para mí una permanente fuente de inspiración.

A mis maestros

A mis docentes y formadores a lo largo de esta maestría, quienes, con su conocimiento, dedicación y vocación por la enseñanza, no solo transmitieron contenido académico, sino que también sembraron en mí el sentido de responsabilidad, la pasión por la gestión ambiental y el compromiso con la excelencia profesional.

ÍNDICE DEL CONTENIDO

RESUMEN	16
ABSTRACT.....	17
I. INTRODUCCIÓN.....	18
1.1. Planteamiento del Problema	18
1.2. Descripción del Problema	20
1.3. Delimitación del problema.....	22
1.3.1. Delimitación espacial.....	22
1.3.2. Delimitación temporal	23
1.3.3. Delimitación temática y unidad de análisis	23
1.4. Formulación del Problema.....	24
1.4.1. Problema General.....	25
1.4.2. Problema Específico	25
1.5. Antecedentes	26
1.5.1. Antecedentes Internacionales.....	26
1.5.2. Antecedentes Nacionales	28
1.6. Justificación de la investigación	29
1.6.1. Justificación científica	29
1.6.2. Justificación técnica	31
1.6.3. Justificación socioeconómica.....	31
1.6.4. Justificación ambiental.....	32
1.6.5. Impacto en la salud y la agricultura	32
1.6.6. Relevancia estratégica.....	32
1.7. Limitaciones de la investigación.....	33
1.7.1. Delimitación espacial.....	33
1.7.2. Delimitación temporal	33
1.7.3. Identificación de inversiones térmicas.....	33
1.7.4. Limitaciones del modelo predictivo.....	34
1.7.5. Alcance del estudio	34
1.8. Objetivos	34

1.8.1. Objetivo General.....	34
1.8.2. Objetivos Específicos.....	34
1.9. Hipótesis	35
1.9.1. Hipótesis General.....	35
1.9.2. Hipótesis Específicas	35
1.10. Coherencia de la investigación	36
II. MARCO TEÓRICO	40
2.1. Marco Conceptual.....	40
2.1.1. Inversión térmica	40
2.1.2. Altura de la Capa Límite Atmosférica	41
2.1.3. Dióxido de azufre (SO ₂).....	41
2.1.4. Modelamiento predictivo con Random Forest (RF)	42
2.1.5. Área de estudio	43
2.1.6. Refinería de Zinc de Cajamarquilla	44
2.1.7. Estándar de Calidad Ambiental para Aire (ECA – Aire).....	48
2.1.8. Protocolo Nacional de Monitoreo de la Calidad Ambiental del Aire	49
2.1.9. Fiscalización ambiental.....	49
2.1.10. Zona de barlovento y sotavento	50
2.1.11. Monitoreo automático en tiempo real	50
2.2. Bases Teóricas	52
2.3. Bases Legales.....	56
2.4. Aspectos de Responsabilidad Social y Medio Ambiente	57
III. MÉTODO.....	61
3.1. Tipo de Investigación.....	61
3.1.1. Por su Naturaleza	61
3.1.2. Por su Profundidad.....	62
3.1.3. Por su Metodología	63
3.1.4. Por su Diseño	64
3.2. Población y Muestra	65
3.2.1. La población.....	65

3.2.2. La muestra.....	68
3.3. Operacionalización de Variables	69
3.3.1. Matriz de consistencia.....	69
3.3.2. Matriz de Operacionalización de Variables.....	72
3.3.2.1. Variable dependiente (VD1): Concentración horaria de dióxido de azufre (SO ₂).....	73
3.3.2.2. Variable independiente (VI1): Condiciones de inversión térmica.....	74
3.3.2.3. Variable de interacción: Coincidencia entre inversión térmica y contaminación crítica.....	74
3.3.2.3. Variable dependiente (VD3): Estimación de concentración horaria de SO ₂ mediante modelamiento predictivo.....	74
3.4. Instrumentos.....	77
3.4.1. Instrumentos para la variable dependiente.....	78
3.4.2. Instrumentos para la variable independiente.....	79
3.4.3. Instrumentos Complementarios	81
3.4.4. Herramientas computacionales de modelamiento predictivo	83
3.5. Procedimientos.....	84
3.5.1. Etapa 1: Recopilación y Preparación de Datos	84
3.5.2. Etapa 2: Identificación y caracterización de fuentes de error	84
3.5.3. Etapa 3: Diseño e Implementación del Modelo Estadístico y Predictivo basado en Random Forest.....	85
3.5.3.1. Modelo estadístico estocástico de caja negra.....	85
3.5.3.2. Modelo de aprendizaje automático basado en Random Forest.....	86
A. Preparación del conjunto de datos	86
B. Entrenamiento del modelo RF.....	86
C. Evaluación del desempeño	86
D. Análisis de importancia de variables	86
E. Simulación de escenarios de alerta temprana	86
3.5.4. Etapa 4: Validación y evaluación del modelo.....	87
3.5.5. Etapa 5: Procedimiento para la detección, corrección y relleno de datos.....	88
3.5.5.1. Detección de valores faltantes y atípicos.	88

3.5.5.2. Criterios para la corrección de valores anómalos	88
3.5.5.3. Relleno de datos faltantes	89
3.5.5.4. Verificación posterior al tratamiento	89
3.6. Análisis de Datos	90
3.6.1. Análisis estadístico descriptivo	90
3.6.2. Identificación y Clasificación de Fuentes de Error	91
3.6.3. Aplicación del Modelo Matemático y Predictivo	91
3.6.4. Evaluación del Modelo Predictivo	91
3.6.5. Presentación de Resultados	92
3.7. Consideraciones Éticas	93
3.7.1. Uso responsable de datos	93
3.7.2. Veracidad y honestidad en la recolección y análisis de información	93
3.7.3. Reconocimiento de fuentes y derechos de autor	94
3.7.4. Confidencialidad y protección de la información	94
3.7.5. Enfoque ético en la interpretación y uso de resultados	94
IV. RESULTADOS	96
4.1. Análisis Descriptivo de los Datos	96
4.1.1. Caracterización general de los datos obtenidos	96
4.1.2. Análisis meteorológicos asociado	103
4.1.2.1. Temperatura del aire	104
4.1.2.2. Velocidad y dirección del viento	105
4.1.2.3. Altura de la Capa Límite Atmosférica (BLH)	107
4.1.2.4. Frecuencia e intensidad de eventos de inversión térmica	108
4.1.3. Análisis comparativo de concentraciones a barlovento y sotavento	112
4.1.3.1. Tendencias generales por estación	112
4.1.3.2. Comparación horaria y diaria	113
4.1.3.3. Relevancia del análisis comparativo	115
4.1.4. Análisis de correlación entre SO ₂ y variables meteorológicas (Estación Sotavento, 2022–2024)	117
4.1.4.1. Correlaciones estratificadas por régimen	121

4.1.4.2.	Correlaciones con rezagos (1–6 h).....	122
4.1.4.3.	Dispersogramas con LOWESS (no linealidades)	125
4.1.4.4.	Caso ilustrativo: día con inversión térmica.....	127
4.1.4.5.	Implicancias para la modelación y la fiscalización.....	128
4.2.	Análisis de cumplimiento normativo	129
4.2.1.	Evaluación respecto a los Estándares de Calidad Ambiental (ECA).....	130
4.2.1.1.	Frecuencia anual de superación del ECA (promedio 24 horas).....	130
4.2.1.2.	Representación gráfica del cumplimiento del ECA.....	131
4.2.1.3.	Limitaciones del criterio diario vs. evidencia horaria.....	132
4.2.2	Análisis respecto a niveles de alerta nacionales	133
4.2.2.1.	Umbrales normativos para SO ₂	134
4.2.2.2.	Resultados del análisis por nivel de alerta	134
4.2.3.	Comparación con los valores guía de la OMS para SO ₂ (promedio 24 horas).136	
4.2.3.1.	Valores guía establecidos por la OMS para SO ₂ (24h).....	136
4.2.3.2.	Resultados comparativos	137
4.3.	Resultados del Modelamiento Predictivo (Random Forest)	139
4.3.1	Diseño e implementación del modelo.....	140
4.3.1.1.	Enfoque de modelamiento tipo “caja negra”	140
4.3.1.2.	Justificación del uso de Random Forest.....	141
A.	Alta precisión predictiva	141
B.	Robustez ante ruido y valores atípicos	141
C.	Capacidad para capturar relaciones no lineales.....	141
D.	Manejo eficiente de datos faltantes y variables irrelevantes.....	141
E.	Aplicabilidad comprobada en estudios ambientales.....	141
4.3.1.3.	Variables utilizadas en el modelo	141
4.3.1.4.	Control de fuga de información y construcción de variables	143
4.3.2	Validación y desempeño del modelo	143
4.3.2.1.	Métricas de rendimiento	143
4.3.2.2.	Gráfico de correlación entre valores observados y predichos	145
4.3.2.3.	Importancia relativa de las variables predictoras.....	146

4.3.2.4. Validación cruzada temporal del modelo Random Forest.....	150
4.3.3 Predicciones generadas	153
4.3.3.1. Propósito y enfoque	153
4.3.3.2. Ventana representativa (23-30/08/2023, tramo de prueba).....	153
4.3.3.3. Dispersión global y calibración Figura	156
4.3.3.4. Lectura “tipo alerta”: eventos vs. no eventos (horas $\geq 250 \mu\text{g}/\text{m}^3$ destacadas)”	157
4.3.3.5. Diagnóstico de residuales: tiempo y por niveles de predicho.....	159
4.3.3.6. Implicancias para la fiscalización y operación	160
4.3.4. Análisis de colinealidad entre predictores (VIF)	161
4.3.4.1. Discusión sobre predictores y robustez del Modelo	162
4.3.4.2. Factor de Inflación de la Varianza	163
V. DISCUSIÓN DE RESULTADOS	167
5.1. Discusión del análisis descriptivo de los Datos	167
5.2. Discusión sobre cumplimiento normativo ambiental	168
5.2.1. Implicancias regulatorias y limitaciones institucionales.....	169
5.2.2. Reflexión sobre la suficiencia del marco normativo actual	170
5.3. Discusión del modelo predictivo implementado (Random Forest)	170
5.3.1. Valoración del rendimiento del modelo.....	171
5.3.2. Interpretación de la capacidad predictiva	171
5.3.3. Respecto a la colinealidad.....	171
5.3.4. La jerarquía de importancias del RF (permutación e impureza).....	172
5.3.5. En cuanto a la evaluación del modelo RF.....	172
5.4. Aporte del modelo a la gestión ambiental preventiva.....	173
5.5. Contraste con hipótesis y objetivos planteados	173
5.5.1. Cumplimiento de los objetivos específicos.....	173
5.5.2. Validación de hipótesis específicas	175
5.5.3. Respuesta a la pregunta de investigación general.....	176
5.6. Implicancias prácticas y contribuciones del estudio.....	176
5.6.1. Aportes específicos a la gestión ambiental en Cajamarquilla.....	177

5.6.2.	Propuestas derivadas para mejorar la fiscalización ambiental minera.....	177
5.6.3.	Impacto potencial en otras zonas industriales similares	178
5.7.	Limitaciones y recomendaciones para futuras investigaciones	178
5.7.1.	Limitaciones del estudio	178
5.7.2.	Recomendaciones para futuras investigaciones.....	179
VI.	CONCLUSIONES	181
6.1.	Concentración horaria de SO ₂ y superaciones del umbral	181
6.2.	Excedencias del ECA y evaluación normativa	181
6.3.	Influencia de condiciones meteorológicas e inversión térmica	181
6.4.	Valor del modelo Random Forest en la gestión predictiva.....	181
6.5.	Validación de hipótesis y cumplimiento de objetivos	182
VII.	RECOMENDACIONES	183
7.1.	Implementar alerta temprana predictiva basada en RF y meteorología.....	183
7.2.	Actualizar criterios normativos para contemplar exposiciones horarias	183
7.3.	Ampliar y modernizar el monitoreo (incluida la BLH)	183
7.4.	Gobernanza del modelo y mejora continua	184
7.5.	Capacitación técnica aplicada para fiscalizadores	184
7.6.	Replicabilidad y escalamiento del enfoque metodológico en otras zonas industriales	184
7.7.	Comunicación de riesgo y coordinación territorial.....	185
7.8.	Extensiones técnicas recomendadas (mediano plazo)	185
VIII.	REFERENCIAS.....	187
IX.	ANEXOS	193
	Anexo 1 Matriz de Consistencia	193
	Anexo 2 Matriz de operacionalización de variables	196
	Anexo 3 Mapa de ubicación de las estaciones de monitoreo del OEFA	197
	Anexo 4 Mapa de ubicación del Área de Estudio	198
	Anexo 5 Población por zonas y rango de edades en Lurigancho- Chosica	199
	Anexo 6 Matriz de correlación durante un episodio de inversión térmica entre SO ₂ y variables meteorológicas (5 de julio de 2024) – Estación Sotavento	200

Anexo 7 Diagramas de rosa de los vientos por meses SO ₂ para el año 2024, de la estación de monitoreo ubicada a sotavento.....	201
Anexo 8 Rosa de vientos por horas para el SO ₂ de las 00:00 a las 23 horas, de enero a noviembre de 2024, en la estación ubicada a sotavento	202
Anexo 9 Estándares de calidad ambiental de la calidad para aire, aprobado mediante D.S, N° 03-2017-MINAM.....	203
Anexo 10 Reglamento de los Niveles de Estados de Alerta Nacionales para contaminantes del aire, aprobado mediante D.S. N.º 009-2003-S.A.	204
Anexo 11 Valores Guía establecidos por la OMS para la calidad del aire	205
Anexo 12 Comportamiento de las concentraciones ($\mu\text{g}/\text{m}^3$) de SO ₂ promediadas, en la estación del OEFA ubicada a sotavento de enero a noviembre de 2024.	206
Anexo 13 Ficha de Informe de juicio de experto del instrumento de investigación.....	207
Anexo 14 Ficha de Validación del Instrumento - Experto 1	208
Anexo 15 Ficha de Validación del Instrumento - Experto 2	209
Anexo 16 Ficha de Validación del Instrumento - Experto 3	210
Anexo 17 Ficha de Validación del Instrumento - Experto 4	211
Anexo 18 Método de evaluación de resultados Validación de Instrumento.....	212
Anexo 19 Base de Datos a Barlovento y Sotavento	216
Anexo 20 ¿Cómo funciona Random Forest?	217
Anexo 21 Validación Cruzada temporal: cómo evaluamos el modelo Random Forest ...	218
Anexo 22 Script extendido que corre el Random Forest para sotavento y barlovento, calcula métricas y guarda predicciones e importancias	219
Anexo 23 Formato de recolección de Datos	222
Anexo 24 Certificado de calibración del Analizador automático de Dióxido de Azufre de estación ubicada a Sotavento	223
Anexo 25 Certificado de calibración del Analizador automático de Dióxido de Azufre de estación ubicada a Barlovento	224

ÍNDICE DE TABLAS

Tabla 1 Comparación de superaciones horarias y diarias del SO ₂ en las estaciones ubicadas a sotavento y barlovento (2022–2024)	30
Tabla 2 Matriz de Coherencia: Problemas-Objetivos - Hipótesis.....	38
Tabla 3 Estándares de calidad ambiental (ECA) de SO ₂ para aire.....	49
Tabla 4 Protocolo de monitoreo de la calidad del aire	49
Tabla 5 Ubicación de las estaciones de monitoreo de calidad del aire	51
Tabla 6 Cuadro comparativo internacional sobre inversión térmica y SO ₂	53
Tabla 7 Cuadro comparativo nacional sobre inversión térmica y SO ₂	54
Tabla 8 Población por zonas en Lurigancho-Chosica.....	67
Tabla 9 Matriz de Consistencia.....	70
Tabla 10 Matriz de operacionalización de variables.....	76
Tabla 11 Equipo de SO ₂ de utilizado en la estación a barlovento CA-RZC-01 (Los Andes) ..	78
Tabla 12 Equipo de SO ₂ utilizado en la estación a sotavento CA-RZC-02 (Radio Observatorio de Jicamarca - ROJ)	78
Tabla 13 Equipos meteorológicos utilizados en la estación a barlovento CA-RZC-01 (Los Andes)	80
Tabla 14 Equipos meteorológicos utilizados en la estación a sotavento CA- RZC-02 (ROJ) 80	
Tabla 15 Métodos de análisis de los parámetros evaluados	81
Tabla 16 Registros horarios válidos de SO ₂ para el periodo 2022-2024	97
Tabla 17 Datos estadísticos de las concentraciones horarias de SO ₂ (µg/m ³) y de (BLH, en metros)	97
Tabla 18 Clasificación de los eventos por intensidad según la BLH.....	109
Tabla 19 Frecuencia de eventos de inversión térmica (Sotavento, 2022–2024).....	112
Tabla 20 Tendencias generales por estación.....	113
Tabla 21 Comparación horaria con SO ₂ > 250 µg/m ³ en las estaciones de sotavento y barlovento	114
Tabla 22 Coeficiente de correlación entre SO ₂ y variables meteorológicas (serie completa 2022–2024, estación a sotavento)	118
Tabla 23 Correlación con componentes de viento (u, v)	119
Tabla 24 Máximo r por variable y lente (hora / cuartil BLH / rezago)	121

Tabla 25 Resumen — Mejor $ r $ por hora del día (extracto)	121
Tabla 26 Resumen — Mejor $ r $ por cuartil de BLH (extracto)	122
Tabla 27 Resumen — Mejor $ r $ por rezago (extracto)	123
Tabla 28 Correlaciones con rezados 1-6h (extracto)	123
Tabla 29 Frecuencia anual de superación del ECA periodo 2022-2024.....	130
Tabla 30 Niveles de alerta para SO ₂ de acuerdo al Decreto Supremo N.º 009-2003-SA	134
Tabla 31 Resultados por nivel de alerta periodo 2022-2024	134
Tabla 32 Valores guía establecidos por la OMS para SO ₂ (24h)	136
Tabla 33 Resultados comparativos con datos diarios de concentraciones de SO ₂ de la estación de sotavento en el periodo 2022-2024.	137
Tabla 34 Variables independientes	142
Tabla 35 Métricas de rendimiento (hold-out temporal 20 %, RF=100)	144
Tabla 36 Importancia por permutación (RF=100, hold-out 20 %)	147
Tabla 37 Importancia por impureza (Gini) (RF=100, entrenamiento 80 %)	148
Tabla 38 Validación cruzada temporal (Time Series Split, k = 5, RF=100)	152
Tabla 39 Calibración por deciles del valor predicho: promedio observado vs. promedio predicho – ventana 23-30/08/2023.....	157
Tabla 40 Detección de eventos ($\geq 250 \mu\text{g}/\text{m}^3$) en la semana 23-30 de agosto de 2023: conteos TP, FP, FN y TN.....	159
Tabla 41 Estimación del Factor de Inflación de la Varianza (VIF) para las variables del bloque meteorológico (Todo el periodo, con constante)	164
Tabla 42 Estimación del Factor de Inflación de la Varianza (VIF) para predictores ‘Top’ (Train 80% con constante).....	164
Tabla 43 Cumplimiento de los objetivos específicos.....	174
Tabla 44 Validación de hipótesis específicas	175

ÍNDICE DE FIGURAS

Figura 1 Comparaciones de superaciones horarias y diarias de SO ₂ en la estación ubicada a sotavento y barlovento en el periodo 2022-2024.....	30
Figura 2 Gradiente térmico vertical normal y perturbado	40
Figura 3 Ubicación del Área de Estudio	44
Figura 4 Ubicación de la Refinería de Zinc de Cajamarquilla	45
Figura 5 Modelo conceptual de la fuente de SO ₂ , sus rutas de transporte y receptores	48
Figura 6 Ubicación de las estaciones de monitoreo	52
Figura 7 Distribución porcentual de la población por zonas del distrito de Lurigancho-Chosica	67
Figura 8 Distribución de SO ₂ a barlovento y sotavento correspondiente al periodo 2022-2024	99
Figura 9 Distribución de la altura de la capa límite atmosférica (BLH) a barlovento y sotavento correspondiente al periodo 2022-2024.....	100
Figura 10 Análisis temporal de concentración mensual promedio de SO ₂ . A barlovento y sotavento durante el periodo 2022-2024.....	101
Figura 11 Análisis temporal de la altura mensual promedio de la BLH a barlovento y sotavento durante el periodo 2022-2024	101
Figura 12 Promedio de SO ₂ por hora del día (2022–2024)	103
Figura 13 Mapa de calor de SO ₂ en sotavento (promedio 2022–2024) por mes y hora.	104
Figura 14 Diagramas de rosa de los vientos por meses SO ₂ para el año 2024, de la estación de monitoreo ubicada a sotavento.....	106
Figura 15 Relación temporal entre SO ₂ y BLH - Estación Sotavento (2022-2024)	108
Figura 16 Gradiente térmico vertical: normal e inversión térmica	110
Figura 17 Gráfico de dispersión: Relación entre Altura de Capa Límite (BLH) y SO ₂ horario - Estación Sotavento.....	111
Figura 18 Comparación de SO ₂ y BLH en Barlovento y Sotavento (2022-2024).....	114
Figura 19 Comparación del número de horas con concentraciones horarias de SO ₂ superiores a 250 µg/m ³ por estación y año (2022–2024).	116
Figura 20 Matriz de correlación de Pearson (SO ₂ vs. Meteorología) periodo 2022-2024, estación a sotavento.	120
Figura 21 SO ₂ vs. BLH con LOWESS, período 2022-2024 a sotavento.....	125

Figura 22 SO ₂ vs. velocidad del viento con LOWESS, periodo 2022-2024 a sotavento	126
Figura 23 SO ₂ vs. temperatura con LOWESS, periodo 2022-2024 a sotavento.....	126
Figura 24 SO ₂ vs. humedad relativa con LOWESS periodo 2022-2024 a sotavento	127
Figura 25 Matriz de correlación (Pearson) — 5 de julio de 2024 (día con inversión térmica)	128
Figura 26 Número de superaciones del ECA de 250 µg/m ³ como promedio de 24 horas- Estación Sotavento (2022-2024).....	131
Figura 27 Comparación de superaciones horarias y diarias de SO ₂ (2022-2024)	132
Figura 28 Eventos de superación de niveles de alerta por SO ₂ (promedio móvil 3h)	135
Figura 29 Comparaciones de días con superación de límites diarios de SO ₂ según OMS y ECA Perú- Estación de Sotavento (2022-2024)	138
Figura 30 Comparación entre valores reales y predichos del modelo Random Forest – Estación Sotavento.....	146
Figura 31 Importancias por permutación (RF =100, hold-out)	149
Figura 32 Importancias por impurezas -Gini (RF=100, train 80%)	150
Figura 33 SO ₂ : Zoom matutino 24/08/2023 (00:00 -14:00 h)	154
Figura 34 SO ₂ : Zoom matutino 25/08/2023 (00:00 -14:00 h)	154
Figura 35 SO ₂ : Zoom matutino 29/08/2023 (00:00 -14:00 h)	155
Figura 36 SO ₂ : Zoom matutino 30/08/2023 (00:00 -14:00 h)	155
Figura 37 Dispersión del SO ₂ predicho vs. observado (línea y = x) en la ventana 23–30/08/2023	156
Figura 38 Lectura tipo alerta: serie horaria con horas ≥ 250 µg/m ³ destacadas en la ventana 23–30/08/2023 (Sotavento).....	158
Figura 39 Residuales (observado – predicho) a lo largo del tiempo - ventana 23-30/08/2023	160

RESUMEN

Esta tesis evalúa el impacto de la inversión térmica en las concentraciones horarias de SO₂ para optimizar la fiscalización ambiental en el entorno de la Refinería de Zinc de Cajamarquilla (Lurigancho-Chosica, Lima). Se analizaron datos horarios 2022–2024 de estaciones automáticas, diferenciando barlovento/sotavento. El enfoque fue cuantitativo correlacional-explicativo, integrando variables meteorológicas (temperatura, humedad relativa, presión, velocidad y dirección del viento) y la altura de la capa límite atmosférica (BLH). Se implementó un Random Forest (100 árboles) con validación cruzada temporal, orientado a explicar patrones y soportar alertas tempranas. Las importancias por permutación e impureza muestran que la memoria de corto plazo del contaminante (rezagos 1–24 h) domina la predicción, seguida por la BLH (y su promedio móvil retrasado) y, en menor medida, por viento, temperatura y la estacionalidad intradiaria; esto concuerda con la física de mezcla y ventilación asociada a condiciones de BLH baja. Se identificaron superaciones horarias del umbral 250 µg/m³ y, a escala diaria, 5 excedencias del ECA-Aire (24 h, 250 µg/m³) en 2022 y 6 en 2023 (dentro del límite de ≤ 7 /año); 2024 no registró excedencias. El indicador diario subestima la recurrencia y magnitud de picos horarios, lo que respalda criterios complementarios y un esquema de prealerta basado en pronósticos del modelo. Los resultados apoyan vigilancia focalizada en mañanas con BLH baja y vientos débiles, ofreciendo una ruta operativa para fortalecer la fiscalización ambiental en Cajamarquilla.

Palabras clave: inversión térmica, dióxido de azufre (SO₂), fiscalización ambiental, calidad del aire, minería, alerta temprana, Random Forest.

ABSTRACT

This thesis evaluates the impact of thermal inversion on hourly SO₂ concentrations with the aim of optimizing environmental enforcement around the Cajamarquilla Zinc Refinery (Lurigancho–Chosica, Lima). Hourly data from 2022–2024 were analyzed from automatic monitoring stations, distinguishing windward and leeward conditions. The approach was quantitative and correlational–explanatory, integrating meteorological variables (air temperature, relative humidity, pressure, wind speed and direction) and boundary layer height (BLH). A Random Forest model (100 trees) with temporal cross-validation was implemented to explain patterns and support early warning. Permutation- and impurity-based variable importance shows that the pollutant’s short-term memory (SO₂ lags of 1–24 h) dominates prediction, followed by BLH (and its lagged moving average) and, to a lesser extent, wind, temperature, and intra-daily seasonality; this is consistent with mixing and ventilation physics under low-BLH conditions. Hourly exceedances of the 250 µg/m³ threshold were identified and, at the daily scale, five exceedances of the Air Quality Standard (24 h, 250 µg/m³) in 2022 and six in 2023 (within the limit of ≤ 7 per year); none were recorded in 2024. The daily indicator underestimates the recurrence and magnitude of hourly peaks, which supports complementary criteria and a pre-alert scheme based on model forecasts. The results support targeted surveillance in morning hours with low BLH and weak winds, providing an operational pathway to strengthen environmental enforcement in Cajamarquilla.

Keywords: thermal inversion, sulfur dioxide (SO₂), environmental enforcement, air quality, mining, early warning, Random Forest.

I. INTRODUCCIÓN

1.1. Planteamiento del Problema

La contaminación atmosférica generada por actividades minero-metalúrgicas representa una amenaza constante para la salud pública y el equilibrio ambiental, especialmente en zonas urbanas e industriales de América Latina. En el contexto peruano, uno de los casos más representativos es el de la Refinería de Zinc de Cajamarquilla, ubicada en el distrito de Lurigancho-Chosica, Lima, la cual constituye una fuente significativa de emisiones de dióxido de azufre (SO₂), un gas corrosivo, irritante y precursor de lluvia ácida (Manahan, 2007).

De acuerdo con el Decreto Supremo N.º 003-2017-MINAM, el Estándar de Calidad Ambiental (ECA) para aire establece un valor máximo de 250 µg/m³ de SO₂ como promedio en 24 horas (Ministerio del Ambiente del Perú [MINAM], 2017). No obstante, diversos informes técnicos del Organismo de Evaluación y Fiscalización Ambiental (OEFA) han evidenciado múltiples superaciones de este límite en registros horarios dentro del área de influencia directa de la Refinería de Cajamarquilla durante los años 2022 a 2024 (Organismo de Evaluación y Fiscalización Ambiental [OEFA], 2022, 2023, 2024). Estas superaciones, sin embargo, no son siempre reflejadas en los promedios diarios normativos, lo que limita la capacidad de respuesta ante episodios críticos de contaminación.

Este comportamiento se encuentra estrechamente relacionado con el fenómeno atmosférico de la inversión térmica, definido como un proceso en el que la temperatura del aire aumenta con la altitud dentro de una capa específica, impidiendo el ascenso de masas de aire caliente cargadas de contaminantes (Arya, 1999). Como resultado, se genera una *tapa térmica* que atrapa gases como el SO₂ en niveles bajos de la atmósfera, afectando directamente la calidad del aire y exponiendo a la población a concentraciones agudas de contaminantes. Según Seinfeld y Pandis (2016, p. 495), “las inversiones térmicas actúan como barreras físicas a la dispersión vertical de contaminantes, produciendo acumulaciones temporales significativas de

gases y partículas en superficie”.

A pesar de esta evidencia científica, el marco normativo nacional carece de umbrales horarios específicos o protocolos operativos que permitan activar mecanismos de alerta temprana frente a episodios agudos vinculados a fenómenos meteorológicos como la inversión térmica (Ministerio de Salud [MINSA], 2003). Esta omisión normativa dificulta la aplicación eficaz del principio de prevención ambiental y deja un vacío técnico en los procesos de fiscalización ambiental minera.

Frente a este escenario, se plantea la necesidad de evaluar rigurosamente la influencia de las inversiones térmicas sobre las concentraciones horarias de SO_2 , utilizando datos provenientes de estaciones automáticas del OEFA. Este análisis permitirá identificar patrones de acumulación no visibles a través de los promedios móviles diarios, y fortalecer así los criterios técnicos para la toma de decisiones regulatorias. Tal como señalan Muñoz y Salinas (2018, p. 44), “la inclusión de variables meteorológicas en la gestión de calidad del aire permite una respuesta más anticipada y eficaz frente a episodios críticos de contaminación”.

Adicionalmente, la presente investigación incorpora como componente metodológico el desarrollo de un modelo de predicción de concentraciones horarias de SO_2 utilizando técnicas de aprendizaje automático, específicamente el algoritmo Random Forest (RF). Este enfoque ha sido validado en investigaciones recientes por su alto rendimiento en la estimación de contaminantes atmosféricos y su capacidad para modelar relaciones no lineales entre variables meteorológicas y concentraciones de gases (Liu et al., 2022; Grange et al., 2018). La implementación de este modelo predictivo constituye una herramienta complementaria de vigilancia ambiental, orientada a anticipar episodios agudos de contaminación en escenarios de inversión térmica y a sustentar técnicamente decisiones regulatorias más oportunas y eficaces.

Esta investigación, por tanto, se orienta a generar evidencia empírica para optimizar los mecanismos de fiscalización ambiental minera, en concordancia con el enfoque preventivo del Estado y el cumplimiento del derecho constitucional a un ambiente sano (Congreso de la República, 1993).

Asimismo, se debe considerar que la zona de estudio se encuentra dentro del ámbito del distrito de Lurigancho-Chosica, el cual contaba con una población de 240 814 habitantes según el Censo Nacional 2017 realizado por el Instituto Nacional de Estadística e Informática (INEI). En este contexto, la población de Cajamarquilla representaba aproximadamente el 19 % de la población total del distrito, lo cual equivale a un estimado de 45 755 personas. Según proyecciones realizadas, se estima que la población de Cajamarquilla alcanzará los 70 954 habitantes en el año 2031. Cabe señalar que, al año 2017, la densidad poblacional del distrito era de 1 018 personas por kilómetro cuadrado, mientras que la densidad de hogares alcanzaba los 45 hogares por kilómetro cuadrado. Estos datos demuestran la alta exposición poblacional a los riesgos asociados a la contaminación del aire en esta zona industrial.

1.2. Descripción del Problema

La Refinería de Zinc de Cajamarquilla se ubica en el valle de la quebrada Jicamarca, en el distrito de Lurigancho-Chosica, Lima, una zona caracterizada por condiciones topográficas cerradas y meteorología local que favorecen la ocurrencia de inversiones térmicas, particularmente durante las primeras horas del día y en temporadas con baja radiación solar (Servicio Nacional de Meteorología e Hidrología del Perú [SENAMHI], 2020). Estas condiciones generan estabilidad térmica en la capa inferior de la atmósfera, reduciendo la dispersión vertical y propiciando la acumulación de contaminantes como el dióxido de azufre (SO₂).

De acuerdo con Arya (1999, p. 208), la inversión térmica se manifiesta cuando se invierte el gradiente térmico típico de la atmósfera, generando una capa estable que actúa como

barrera a la dispersión vertical: “los contaminantes quedan atrapados cerca del suelo, generando concentraciones inusualmente elevadas en periodos cortos”. Este fenómeno, al interactuar con fuentes puntuales de emisión como la refinería, puede ocasionar episodios agudos de contaminación atmosférica.

El Organismo de Evaluación y Fiscalización Ambiental (OEFA) ha instalado estaciones automáticas de monitoreo a barlovento y sotavento de la refinería, cuyos registros permiten evaluar la calidad del aire. Durante el periodo 2022–2024, la estación ubicada a sotavento registró cinco superaciones del ECA Aire de 24 horas en 2022 y seis en 2023, mientras que en 2024 no se reportaron excedencias, manteniéndose dentro del límite normativo establecido por el Decreto Supremo N.º 003-2017-MINAM (máximo de siete superaciones anuales del valor de $250 \mu\text{g}/\text{m}^3$). Sin embargo, un análisis más detallado de los registros horarios revela una realidad crítica: en la misma estación sotavento se identificaron 443 superaciones horarias del umbral de $250 \mu\text{g}/\text{m}^3$ en 2022, 346 en 2023 y 413 en 2024. Estas cifras demuestran la magnitud de los episodios de contaminación aguda que permanecen invisibles bajo la evaluación normativa tradicional.

Tal como advierten Seinfeld y Pandis (2016, p. 691), “las métricas basadas exclusivamente en promedios diarios pueden enmascarar eventos breves pero intensos de contaminación que requieren respuestas inmediatas”. Este vacío normativo limita la activación de mecanismos de alerta temprana y la eficacia de la fiscalización ambiental minera, en un contexto donde la salud de la población y la sostenibilidad de los ecosistemas locales pueden verse comprometidas.

El Protocolo Nacional de Monitoreo de la Calidad del Aire reconoce la importancia de integrar variables meteorológicas como temperatura, humedad relativa, presión atmosférica, radiación solar, velocidad y dirección del viento, así como la altura de la capa límite atmosférica (BLH), en el análisis de la dinámica de los contaminantes (MINAM, 2018). No

obstante, en la práctica, estos factores aún no se incorporan plenamente en los procesos regulatorios de decisión ni en la fiscalización ambiental (MINSA, 2003).

Frente a esta realidad, es fundamental caracterizar con mayor precisión el comportamiento horario del SO_2 bajo condiciones de inversión térmica y su relación con variables meteorológicas críticas. Esta aproximación permitirá identificar patrones de acumulación que suelen pasar desapercibidos en los análisis basados únicamente en promedios diarios.

En esa línea, la presente investigación incorpora el uso de modelos de aprendizaje automático, particularmente Random Forest (RF). Esta técnica ha demostrado ser altamente efectiva en escenarios internacionales de predicción de contaminantes atmosféricos, gracias a su capacidad para procesar grandes volúmenes de datos y manejar variables correlacionadas (Zhan et al., 2021; Zhang et al., 2020). Como advierten Muñoz y Salinas (2018, p. 40), “la omisión de variables meteorológicas críticas en los sistemas de alerta compromete la eficacia de las políticas de control de calidad del aire en contextos urbanos-industriales complejos”.

1.3. Delimitación del problema

1.3.1. Delimitación espacial

El estudio se circunscribe al área de influencia directa de la Refinería de Zinc de Cajamarquilla, ubicada en el distrito de Lurigancho-Chosica, provincia y departamento de Lima. Esta zona geográfica se caracteriza por estar emplazada en un valle estrecho de la quebrada Jicamarca, lo cual favorece la ocurrencia de fenómenos de inversión térmica debido a las condiciones topográficas cerradas y a la meteorología local (SENAMHI, 2020).

El ámbito específico de análisis comprende las estaciones automáticas de monitoreo de calidad del aire instaladas por el Organismo de Evaluación y Fiscalización Ambiental (OEFA), estratégicamente localizadas a barlovento y sotavento de la refinería. Estas estaciones permiten evaluar diferencias en las concentraciones de dióxido de azufre (SO_2) en función de la dirección

predominante del viento y de la capacidad de dispersión atmosférica (OEFA, 2022; OEFA, 2023; OEFA, 2024).

1.3.2. Delimitación temporal

El periodo de análisis abarca los años 2022, 2023 y 2024, en los cuales se dispone de registros horarios de SO₂ y de variables meteorológicas asociadas. En este intervalo, la estación sotavento registró cinco superaciones del Estándar de Calidad Ambiental para Aire (ECA Aire) de 24 horas en 2022 y seis en 2023, mientras que en 2024 no se reportaron excedencias. Dichos resultados se encuentran dentro del límite normativo de hasta siete superaciones anuales del valor de 250 µg/m³ establecido por el Decreto Supremo N.º 003-2017-MINAM (MINAM, 2017).

No obstante, el análisis de las series horarias evidencia la ocurrencia de superaciones críticas no reflejadas en la evaluación diaria, lo que pone de relieve la necesidad de contar con criterios técnicos complementarios para la gestión de episodios de contaminación aguda (Seinfeld & Pandis, 2016).

1.3.3. Delimitación temática y unidad de análisis

La investigación se centra en el impacto de la inversión térmica sobre las concentraciones horarias de SO₂ y en su relación con variables meteorológicas locales como temperatura, humedad relativa, presión atmosférica, velocidad y dirección del viento, y altura de la capa límite atmosférica (BLH). El análisis incorpora como unidad de observación los registros horarios provenientes de las estaciones automáticas de monitoreo del OEFA en Cajamarquilla.

Asimismo, se considera la limitación normativa derivada de la aplicación exclusiva del ECA Aire de 24 horas, que puede enmascarar episodios horarios críticos. Frente a esta brecha, se plantea la utilización de un modelo predictivo Random Forest (RF), el cual ha demostrado efectividad en contextos internacionales para la estimación de contaminantes atmosféricos

debido a su robustez ante datos complejos y variables correlacionadas (Zhan et al., 2021; Zhang et al., 2020; Muñoz & Salinas, 2018). Esta aproximación metodológica busca fortalecer la fiscalización ambiental minera mediante la anticipación de episodios de contaminación asociados a condiciones de inversión térmica.

1.4. Formulación del Problema

La presente investigación parte de la necesidad de comprender la influencia de factores meteorológicos locales —en particular, la inversión térmica— sobre los patrones de dispersión del dióxido de azufre (SO_2) en zonas industriales de alto impacto ambiental. Diversos estudios han demostrado que este fenómeno atmosférico puede modificar sustancialmente la calidad del aire a escala horaria, generando acumulaciones agudas de contaminantes que no son captadas por los promedios móviles diarios empleados comúnmente en los marcos regulatorios (Seinfeld & Pandis, 2016). En ese sentido, el vínculo entre episodios de inversión térmica y superaciones horarias del SO_2 representa una problemática aún poco abordada desde el enfoque de fiscalización ambiental en el contexto peruano.

En el área de influencia directa de la Refinería de Zinc de Cajamarquilla, ubicada en el distrito de Lurigancho-Chosica, los registros horarios del OEFA correspondientes al periodo 2022–2024 evidencian la presencia de concentraciones puntuales de SO_2 por encima del límite normativo de $250 \mu\text{g}/\text{m}^3$ establecido en el Decreto Supremo N.º 003-2017-MINAM (MINAM, 2017). No obstante, el uso exclusivo de promedios diarios puede enmascarar la ocurrencia de estos picos críticos, limitando el accionar preventivo del Estado y dificultando la activación de medidas de mitigación inmediatas.

Además, el Reglamento de los Niveles de Estados de Alerta Nacionales para Contaminantes del Aire, aprobado por Decreto Supremo N.º 009-2003-SA, solo contempla protocolos específicos para ciudades como Ilo y La Oroya, sin establecer umbrales horarios aplicables a otras zonas como Lima Metropolitana (MINSA, 2003). Esta brecha normativa

debilita la respuesta institucional frente a eventos de contaminación aguda en zonas industriales relevantes.

Como lo sostienen Muñoz y Salinas (2018, p. 43), “la inclusión de variables meteorológicas en los sistemas de fiscalización y alerta es indispensable para una gestión ambiental eficaz en contextos urbanos-industriales complejos”. Bajo esta perspectiva, la presente investigación busca generar evidencia técnica que permita optimizar los procesos de supervisión ambiental en contextos con alta recurrencia de inversiones térmicas, y fortalecer así el enfoque preventivo consagrado en la Ley N.º 29325 del Sistema Nacional de Evaluación y Fiscalización Ambiental (OEFA, 2009).

En esa línea, se ha considerado incorporar un componente innovador: la implementación de un modelo de aprendizaje automático Random Forest (RF), que permita estimar las concentraciones horarias de SO₂ a partir de variables meteorológicas. Esta técnica ha demostrado alta precisión en estudios de predicción ambiental y es capaz de identificar relaciones no lineales complejas entre parámetros meteorológicos y niveles de contaminación atmosférica (Liu et al., 2022; Grange & Carslaw, 2018; Zhang et al., 2020).

1.4.1. Problema General

¿Cómo influye la inversión térmica en las concentraciones horarias de dióxido de azufre (SO₂) registradas por las estaciones automáticas del Organismo de Evaluación y Fiscalización Ambiental (OEFA) en el área de influencia directa de la Refinería de Zinc de Cajamarquilla, durante el periodo 2022–2024?

1.4.2. Problema Específico

Para abordar el problema general, se identifican preguntas específicas que orientarán el desarrollo de la investigación:

PE1: ¿Con qué frecuencia las concentraciones horarias de SO₂ superan el valor límite de 250 µg/m³ y cómo se comparan estas superaciones con el promedio diario del ECA Aire de 24 horas durante el periodo 2022–2024?

PE2: ¿Cuál es la relación estadística entre las concentraciones horarias de SO₂ y las variables meteorológicas (temperatura, humedad relativa, presión atmosférica, velocidad y dirección del viento, y altura de la capa límite atmosférica – BLH), considerando la diferencia espacial entre estaciones de barlovento y sotavento en el área de estudio?

PE3: ¿De qué manera las condiciones de inversión térmica influyen en la acumulación y variación horaria de SO₂ en el área de estudio?

PE4: ¿En qué medida las concentraciones horarias de SO₂ superan el valor del ECA Aire (promedio 24h) y cuáles son las limitaciones de este estándar para activar mecanismos de alerta temprana en la fiscalización ambiental minera?

PE5: ¿En qué medida el modelo predictivo Random Forest permite estimar con precisión las concentraciones horarias de SO₂, integrando variables meteorológicas locales (temperatura, humedad relativa, presión atmosférica, velocidad y dirección del viento, y BLH)?

1.5. Antecedentes

1.5.1. Antecedentes Internacionales

A nivel internacional, se ha documentado ampliamente la relación entre los fenómenos de inversión térmica y la acumulación de contaminantes atmosféricos en zonas urbanas e industriales. Este vínculo es particularmente crítico en el caso del dióxido de azufre (SO₂), cuyas concentraciones tienden a incrementarse significativamente durante episodios de estabilidad térmica que impiden la dispersión vertical de los contaminantes (Arya, 1999; Seinfeld & Pandis, 2016).

Investigaciones realizadas en Santiago de Chile, Katmandú (Nepal) y regiones industriales de Europa han demostrado que, durante episodios de inversión térmica, los niveles

horarios de SO₂ pueden incrementarse en más del 70 % debido a la retención de contaminantes en la capa límite atmosférica (Gautam et al., 2021; Lenschow et al., 2001; Muñoz & Salinas, 2018). En Katmandú, por ejemplo, Gautam et al. (2021, como se citó en Seinfeld & Pandis, 2016) evidenciaron que la presencia de inversión térmica duplicó los niveles de material particulado y gases contaminantes, con impactos directos en la salud pública. De manera similar, Muñoz y Salinas (2018, p. 43) identificaron que en Santiago de Chile los días con inversión térmica presentaban un incremento sustancial de SO₂ y material particulado fino, señalando que “la persistencia de capas de inversión térmica, especialmente durante los meses de invierno, ha sido el principal factor meteorológico asociado a episodios críticos de contaminación atmosférica, lo que justifica la inclusión de variables meteorológicas en los sistemas de alerta temprana”.

En años recientes, la literatura ha incorporado modelos de aprendizaje automático, como Random Forest (RF), para mejorar la predicción de contaminantes en escenarios de inversión térmica. Estudios en Londres, Beijing y California concluyen que estos modelos superan a los métodos tradicionales en la estimación horaria de contaminantes como SO₂, NO₂ y O₃, especialmente cuando se integran variables meteorológicas como temperatura, presión atmosférica, velocidad del viento y altura de la capa límite (Grange & Carslaw, 2018; Liu et al., 2022; Zhan et al., 2021). Según Liu et al. (2022, p. 7), “la técnica Random Forest resulta altamente efectiva para capturar patrones no lineales y mejorar la predicción horaria de contaminantes en escenarios de complejidad meteorológica, incluyendo inversiones térmicas”. En la misma línea, Zhang et al. (2020) plantean que los modelos basados en inteligencia artificial son herramientas clave para sistemas de alerta temprana, dado que permiten anticipar episodios críticos de contaminación en tiempo real.

Estos antecedentes internacionales coinciden en resaltar que la gestión de la calidad del aire debe considerar no solo los promedios diarios, sino también las variaciones horarias, las

condiciones meteorológicas locales y la incorporación de modelos predictivos avanzados como apoyo a políticas eficaces de control ambiental.

1.5.2. Antecedentes Nacionales

En el contexto peruano, si bien los estudios académicos sobre la relación entre inversión térmica y contaminación atmosférica son aún limitados, existen antecedentes técnicos relevantes de instituciones como el Organismo de Evaluación y Fiscalización Ambiental (OEFA), el Servicio Nacional de Certificación Ambiental para las Inversiones Sostenibles (SENACE) y el Instituto Geofísico del Perú (IGP). Estos informes han documentado episodios de alta concentración de SO₂ en zonas industriales, particularmente en el entorno de la Refinería de Zinc de Cajamarquilla, donde se han registrado superaciones puntuales de los valores establecidos en el Estándar de Calidad Ambiental para Aire (ECA Aire) (MINAM, 2017).

Durante el periodo 2022–2024, los registros horarios de la estación sotavento del OEFA reportaron cinco superaciones del ECA Aire de 24h en 2022 y seis en 2023, mientras que en 2024 no se registraron excedencias (OEFA, 2022; OEFA, 2023; OEFA, 2024). Aunque estos resultados se encuentran dentro del límite normativo de hasta siete superaciones anuales, se identificaron superaciones horarias críticas, lo que confirma la limitación del marco regulatorio vigente para detectar episodios agudos.

El Protocolo Nacional de Monitoreo de Calidad del Aire resalta la importancia de registrar variables meteorológicas junto con las concentraciones de contaminantes (MINAM, 2018); sin embargo, en la práctica, la integración de estos factores en la fiscalización ambiental y la activación de alertas aún no se ha implementado plenamente. Como advierte OEFA (2009), la fiscalización ambiental debe sustentarse en evidencia técnica confiable que permita adoptar medidas preventivas eficaces en protección de la salud y el ambiente.

En el ámbito académico, Ramos et al. (2021) han demostrado que los modelos de aprendizaje automático —especialmente RF— son aplicables en el Perú para la predicción de

contaminantes atmosféricos, resaltando su robustez frente a datos variables y su capacidad para identificar interacciones complejas entre factores meteorológicos. Asimismo, algunos reportes técnicos internos del OEFA han explorado preliminarmente el uso de modelamiento predictivo a partir de datos de estaciones automáticas, lo que abre la posibilidad de fortalecer la fiscalización ambiental con herramientas de inteligencia artificial.

En consecuencia, la presente investigación se enmarca en un campo emergente a nivel nacional, proponiendo la aplicación de un modelo Random Forest para la estimación de concentraciones horarias de SO₂ en Cajamarquilla, con el fin de superar las limitaciones normativas actuales y contribuir al enfoque preventivo establecido en la Ley N.º 29325, que regula el Sistema Nacional de Evaluación y Fiscalización Ambiental.

1.6. Justificación de la investigación

La presente investigación se justifica por la necesidad de abordar una problemática ambiental crítica desde un enfoque multidimensional, que integre perspectivas científicas, técnicas, socioeconómicas, ambientales y sanitarias. Comprender las condiciones que intensifican la contaminación atmosférica en entornos minero-metalúrgicos resulta esencial para fortalecer la fiscalización ambiental en el Perú, proteger la salud pública, preservar los ecosistemas y promover territorios más resilientes.

1.6.1. Justificación científica

El estudio contribuye a llenar un vacío significativo en la literatura nacional sobre la relación entre inversión térmica y concentraciones horarias de dióxido de azufre (SO₂), fenómeno ampliamente documentado en otros países, pero poco explorado en el Perú (Muñoz & Salinas, 2018; Seinfeld & Pandis, 2016). La investigación aporta evidencia empírica inédita: entre 2022 y 2024, en la estación ubicada a sotavento de la Refinería de Zinc de Cajamarquilla se registraron 1 202 superaciones horarias del umbral crítico de 250 µg/m³ de SO₂ (443 en 2022, 346 en 2023 y 413 en 2024) y 11 superaciones del promedio diario del ECA Aire (24 h)

(5 en 2022 y 6 en 2023). En contraste, la estación ubicada a barlovento superó el umbral horario en 38 ocasiones (15 en 2022, 19 en 2023 y 4 en 2024) y no registró superaciones del promedio diario. Esta brecha normativa respalda la necesidad de incorporar enfoques horarios en la vigilancia ambiental.

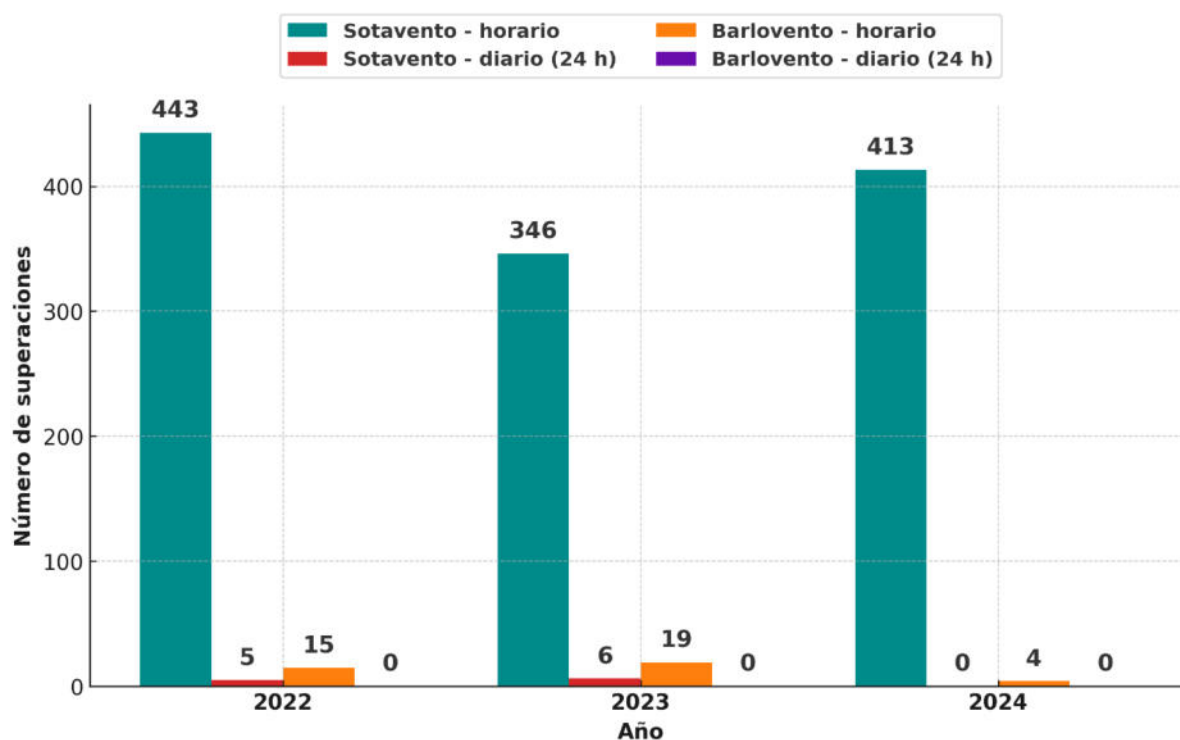
Como innovación metodológica, se aplicará el modelamiento predictivo mediante algoritmos de aprendizaje automático, específicamente Random Forest (RF). Este enfoque ha mostrado elevada precisión en la estimación de contaminantes como el SO₂ en distintos contextos internacionales (Liu et al., 2022; Grange & Carslaw, 2018). Su incorporación permitirá generar propuestas replicables en futuras evaluaciones de calidad del aire en escenarios de alta complejidad atmosférica.

Tabla 1
Comparación de superaciones horarias y diarias del SO₂ en las estaciones ubicadas a sotavento y barlovento (2022–2024)

Año	Estación	Superaciones horarias (> 250 µg/m³)	Superaciones promedio diario ECA Aire (24h)	Cumplimiento normativo (≤ 7 veces/año)
2022	Sotavento	443	5	Sí
	Barlovento	15	0	Sí
2023	Sotavento	346	6	Sí
	Barlovento	19	0	Sí
2024	Sotavento	413	0	Sí
	Barlovento	4	0	Sí

Fuente: Elaboración propia a partir de la información del OEFA

Figura 1
Comparaciones de superaciones horarias y diarias de SO₂ en la estación ubicada a sotavento y barlovento en el periodo 2022-2024



Fuente: Elaboración propia a partir de la base de datos del OEFA

1.6.2. Justificación técnica

Desde la perspectiva técnica, el trabajo optimiza el uso de los registros horarios de SO_2 y su correlación con variables meteorológicas críticas —temperatura, humedad relativa, presión atmosférica, velocidad y dirección del viento, y altura de la capa límite atmosférica (BLH)—, en concordancia con el *Protocolo Nacional de Monitoreo de la Calidad del Aire* (MINAM, 2018).

La implementación de Random Forest permitirá anticipar episodios críticos, fortaleciendo los sistemas de alerta temprana y la capacidad operativa del OEFA, en línea con la Ley N.º 29325 y con las recomendaciones de OEFA (2009) sobre fiscalización preventiva.

1.6.3. Justificación socioeconómica

La exposición frecuente a concentraciones elevadas de SO_2 genera impactos directos en salud pública que se traducen en costos ocultos para el Estado: incremento del gasto en atención médica, reducción de productividad laboral y deterioro de los servicios ecosistémicos.

La investigación aporta a la reducción de pasivos ambientales y promueve comunidades más resilientes frente a riesgos de contaminación (Congreso de la República, 1993; Manahan, 2007).

1.6.4. Justificación ambiental

La investigación responde al mandato constitucional y a la *Ley General del Ambiente* (Ley N.º 28611), que establecen la obligación de proteger la salud y el ambiente. El SO₂ impacta negativamente en la calidad del aire, la cobertura vegetal y la biodiversidad urbana.

El uso de modelos predictivos avanzados permite identificar factores críticos y proponer medidas preventivas y sostenibles, en concordancia con la Ley N.º 29325 (Sistema Nacional de Evaluación y Fiscalización Ambiental).

1.6.5. Impacto en la salud y la agricultura

El SO₂ tiene efectos adversos sobre la salud respiratoria, agravando enfermedades como bronquitis, asma y afecciones pulmonares crónicas (Seinfeld & Pandis, 2016; Manahan, 2007). Asimismo, altera el pH del suelo y reduce la productividad agrícola, generando riesgos para la seguridad alimentaria de las comunidades cercanas.

La investigación busca mitigar estos riesgos mediante un enfoque predictivo que anticipe episodios críticos, fortaleciendo la fiscalización y protegiendo tanto la salud como los ecosistemas locales.

1.6.6. Relevancia estratégica

El estudio se alinea con los compromisos internacionales del Perú y con los Objetivos de Desarrollo Sostenible: ODS 3 (*Salud y Bienestar*), ODS 11 (*Ciudades y Comunidades Sostenibles*), ODS 12 (*Producción y Consumo Responsables*) y ODS 13 (*Acción por el Clima*).

De este modo, la investigación no solo responde a un vacío científico y técnico, sino que contribuye a fortalecer el rol preventivo del Estado y a garantizar el derecho constitucional a un ambiente sano.

1.7. Limitaciones de la investigación

Como toda investigación aplicada en entornos complejos, el presente estudio presenta ciertas limitaciones que deben considerarse al momento de interpretar sus resultados y alcance:

1.7.1. Delimitación espacial

El análisis se circunscribe exclusivamente a los registros de las estaciones automáticas de monitoreo instaladas por el Organismo de Evaluación y Fiscalización Ambiental (OEFA) en el área de influencia directa de la Refinería de Zinc de Cajamarquilla. No se incluyen datos de otras fuentes de emisión ni de estaciones regionales adicionales, lo que restringe la perspectiva del estudio a un entorno industrial específico con condiciones meteorológicas y topográficas particulares.

1.7.2. Delimitación temporal

La base de datos corresponde únicamente al periodo 2022–2024. En este trienio, la estación sotavento registró 443 superaciones horarias del umbral de $250 \mu\text{g}/\text{m}^3$ en 2022, 346 en 2023 y 413 en 2024, lo que refleja una variabilidad interanual significativa. Sin embargo, los resultados expresan únicamente las condiciones de dicho periodo, por lo que no son extrapolables a años con dinámicas meteorológicas o industriales distintas.

1.7.3. Identificación de inversiones térmicas

La detección de eventos de inversión térmica se realizará a partir del análisis de variables meteorológicas indirectas, como la temperatura superficial, la presión atmosférica, la velocidad del viento y la altura estimada de la capa límite atmosférica.

No obstante, no se dispone de perfiles verticales de temperatura obtenidos mediante radiosondeos o torres meteorológicas, instrumentos que permitirían verificar de manera directa la magnitud y duración de las inversiones térmicas (Arya, 1999; MINAM, 2018). A pesar de esta limitación, el enfoque adoptado se respalda en metodologías ampliamente validadas en la literatura internacional.

1.7.4. Limitaciones del modelo predictivo

Si bien el uso del algoritmo Random Forest (RF) constituye un aporte metodológico innovador por su robustez frente a datos no lineales y de alta dimensionalidad, también presenta restricciones. Entre ellas destacan la necesidad de ajustes finos en la selección de hiperparámetros, la sensibilidad frente a variables correlacionadas y su carácter de “caja negra” en la interpretación de resultados (Liu et al., 2022; Zhang et al., 2020). Además, la calidad de las predicciones depende directamente de la integridad, consistencia y resolución temporal de los datos meteorológicos utilizados como variables de entrada.

1.7.5. Alcance del estudio

La investigación no contempla una evaluación directa de los efectos en la salud humana ni en los ecosistemas. Su alcance se limita al análisis ambiental de las concentraciones horarias de dióxido de azufre (SO₂) y su relación con condiciones meteorológicas adversas. Estos resultados buscan aportar insumos técnicos que contribuyan a mejorar la fiscalización ambiental minera y el diseño de sistemas de alerta temprana basados en evidencia científica.

1.8. Objetivos

1.8.1. Objetivo General

Evaluar la influencia de la inversión térmica en las concentraciones horarias de dióxido de azufre (SO₂) registradas por las estaciones automáticas del OEFA en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024, con el fin de optimizar los criterios técnicos de fiscalización ambiental minera en contextos de alta recurrencia de eventos críticos de contaminación.

1.8.2. Objetivos Específicos

OE1: Cuantificar las superaciones horarias y diarias del umbral de 250 µg/m³ de SO₂ registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.

OE2: Analizar la relación estadística entre las concentraciones horarias de SO₂ y las variables meteorológicas (BLH, temperatura, humedad relativa, presión atmosférica, velocidad y dirección del viento), incorporando la diferencia espacial entre estaciones de barlovento y sotavento.

OE3: Evaluar la influencia de las condiciones de inversión térmica en la acumulación y variabilidad horaria de las concentraciones de SO₂ en el área de estudio.

OE4: Comparar las concentraciones horarias de SO₂ registradas en el periodo 2022–2024 con los estándares normativos vigentes (ECA Aire de 24h y protocolos nacionales), e identificar las limitaciones de dichos instrumentos para la activación de mecanismos de alerta temprana en la fiscalización ambiental minera.

OE5: Desarrollar y validar un modelo predictivo Random Forest para estimar concentraciones horarias de SO₂, integrando variables meteorológicas locales (BLH, temperatura, humedad relativa, presión atmosférica y velocidad y dirección del viento), a fin de fortalecer la fiscalización ambiental minera.

1.9. Hipótesis

1.9.1. Hipótesis General

La ocurrencia de eventos de inversión térmica influye de manera significativa en el incremento de las concentraciones horarias de dióxido de azufre (SO₂) registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.

1.9.2. Hipótesis Específicas

HE1. La frecuencia de superaciones horarias del umbral de 250 µg/m³ de SO₂ es significativamente mayor en días caracterizados por condiciones meteorológicas compatibles con inversión térmica.

HE2. Existe una relación estadística significativa entre las concentraciones horarias de SO₂ y las variables meteorológicas locales (altura de la capa límite atmosférica, temperatura,

humedad relativa, presión atmosférica y velocidad y dirección del viento), manifestándose en niveles más elevados en la estación sotavento respecto a la estación barlovento durante episodios de inversión térmica.

HE3. Durante los eventos de inversión térmica se identifican patrones horarios recurrentes de alta concentración de SO₂ —especialmente en las primeras horas del día— que confirman la influencia directa de este fenómeno atmosférico en la acumulación de contaminantes.

HE4a. Las concentraciones horarias de SO₂ registradas en el área de estudio durante 2022–2024 presentan superaciones críticas (443 en 2022, 346 en 2023 y 413 en 2024) que no siempre se reflejan en los valores normativos establecidos en los promedios de 24 horas del ECA Aire.

HE4b. El análisis conjunto de las concentraciones horarias de SO₂ y de variables meteorológicas permite establecer criterios técnicos complementarios que refuercen los mecanismos de alerta temprana y la fiscalización ambiental del OEFA.

HE5. El modelo de aprendizaje automático Random Forest predice con alta precisión las concentraciones horarias de SO₂ al integrar variables meteorológicas locales (BLH, temperatura, humedad relativa, presión atmosférica y velocidad y dirección del viento), mostrando un desempeño superior a los métodos tradicionales de estimación.

1.10. Coherencia de la investigación

La presente investigación mantiene una coherencia integral entre los problemas planteados, los objetivos formulados y las hipótesis propuestas.

En primer lugar, el problema general identifica la necesidad de evaluar la influencia de la inversión térmica en las concentraciones horarias de dióxido de azufre (SO₂), cuestión que se traduce en el objetivo general, orientado a analizar dicha influencia para optimizar la fiscalización ambiental minera. A partir de este planteamiento central, la hipótesis general

establece que la ocurrencia de eventos de inversión térmica incrementa significativamente las concentraciones horarias de SO₂ en la zona de estudio.

En segundo lugar, los problemas específicos permiten desagregar la pregunta central en dimensiones operativas: la frecuencia de superaciones horarias (PE1), la relación entre variables meteorológicas y concentraciones de SO₂ (PE2), los patrones horarios recurrentes (PE3), las limitaciones del estándar normativo (PE4) y la capacidad predictiva de un modelo de inteligencia artificial (PE5). Estos se corresponden directamente con los objetivos específicos (OE1–OE5), los cuales plantean acciones metodológicas concretas de cuantificación, análisis, evaluación, comparación y desarrollo de un modelo predictivo.

De manera paralela, cada objetivo se vincula con una hipótesis específica (HE1–HE5) que anticipa el comportamiento esperado en función de la teoría y los datos preliminares: desde la mayor frecuencia de superaciones horarias bajo inversión térmica (HE1), hasta la predicción robusta de concentraciones mediante Random Forest (HE5). Además, las hipótesis HE4a y HE4b refuerzan la importancia de evidenciar las brechas normativas y la necesidad de criterios técnicos complementarios en la fiscalización ambiental.

Esta estructura garantiza que la investigación no solo mantenga una lógica metodológica clara, sino que también articule el planteamiento del problema con los objetivos de investigación y con los supuestos verificables a través del análisis empírico.

En la siguiente Matriz de Coherencia se sintetizan estas relaciones de manera esquemática, mostrando la correspondencia directa entre cada problema específico, su objetivo derivado y la hipótesis que lo operacionaliza. Este esquema no solo permite visualizar con claridad la alineación interna de la investigación, sino que también constituye una herramienta metodológica que facilita la evaluación de la consistencia lógica del estudio. Asimismo, evidencia la trazabilidad entre la identificación del problema, la formulación de objetivos y la

contrastación de hipótesis, lo que refuerza la validez del diseño y brinda solidez a la propuesta en su fase de ejecución.

Tabla 2

Matriz de Coherencia: Problemas-Objetivos - Hipótesis

Problemas	Objetivos	Hipótesis
Problema General: ¿Cómo influye la inversión térmica en las concentraciones horarias de dióxido de azufre (SO ₂) registradas por las estaciones automáticas del OEFA en el área de influencia directa de la Refinería de Zinc de Cajamarquilla, durante el periodo 2022–2024?	Objetivo General: Evaluar la influencia de la inversión térmica en las concentraciones horarias de SO ₂ registradas por las estaciones automáticas del OEFA en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024, con el fin de optimizar los criterios técnicos de fiscalización ambiental minera en contextos de alta recurrencia de eventos críticos de contaminación.	Hipótesis General: La ocurrencia de eventos de inversión térmica influye de manera significativa en el incremento de las concentraciones horarias de SO ₂ registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.
PE1: ¿Con qué frecuencia las concentraciones horarias de SO ₂ superan el umbral de 250 µg/m ³ y cómo se comparan estas superaciones con el promedio diario del ECA Aire de 24 horas durante el periodo 2022–2024?	OE1: Cuantificar las superaciones horarias y diarias del umbral de 250 µg/m ³ de SO ₂ registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.	HE1: La frecuencia de superaciones horarias del umbral de 250 µg/m ³ de SO ₂ es significativamente mayor en días caracterizados por condiciones meteorológicas compatibles con inversión térmica.
PE2: ¿Cuál es la relación estadística entre las concentraciones horarias de SO ₂ y las variables meteorológicas, considerando la diferencia espacial entre estaciones de barlovento y sotavento?	OE2: Analizar la relación estadística entre las concentraciones horarias de SO ₂ y las variables meteorológicas (BLH, temperatura, humedad relativa, presión atmosférica, velocidad y dirección del viento), incorporando la diferencia espacial entre estaciones de barlovento y sotavento.	HE2: Existe una relación estadística significativa entre las concentraciones horarias de SO ₂ y las variables meteorológicas locales, manifestándose en niveles más elevados en la estación sotavento respecto a la estación barlovento durante episodios de inversión térmica.
PE3: ¿De qué manera las condiciones de inversión térmica influyen en la acumulación y variación	OE3: Evaluar la influencia de las condiciones de inversión térmica en la acumulación y variabilidad horaria de las concentraciones	HE3: Durante los eventos de inversión térmica se identifican patrones horarios recurrentes de alta

Problemas	Objetivos	Hipótesis
horaria de SO ₂ en la zona de estudio?	de SO ₂ en el área de estudio.	concentración de SO ₂ — especialmente en las primeras horas del día— que confirman la influencia directa de este fenómeno en la acumulación de contaminantes.
PE4: ¿En qué medida las concentraciones horarias de SO ₂ superan el valor límite del ECA Aire de 24 horas y cuáles son las limitaciones de este estándar para activar mecanismos de alerta temprana?	OE4: Comparar las concentraciones horarias de SO ₂ registradas en el periodo 2022–2024 con los estándares normativos vigentes (ECA Aire de 24h y protocolos nacionales), e identificar las limitaciones de dichos instrumentos para la activación de mecanismos de alerta temprana en la fiscalización ambiental minera.	HE4a: Las concentraciones horarias de SO ₂ registradas en el área de estudio durante 2022–2024 presentan superaciones críticas (443 en 2022, 346 en 2023 y 413 en 2024) que no siempre se reflejan en los valores normativos establecidos en los promedios de 24 horas del ECA Aire. HE4b: El análisis conjunto de las concentraciones horarias de SO ₂ y de variables meteorológicas permite establecer criterios técnicos complementarios que refuercen los mecanismos de alerta temprana y la fiscalización ambiental del OEFA.
PE5: ¿En qué medida el modelo Random Forest permite estimar con precisión las concentraciones horarias de SO ₂ integrando variables meteorológicas locales?	OE5: Desarrollar y validar un modelo predictivo Random Forest para estimar concentraciones horarias de SO ₂ , integrando variables meteorológicas locales (BLH, temperatura, humedad relativa, presión atmosférica, velocidad y dirección del viento), a fin de fortalecer la fiscalización ambiental minera.	HE5: El modelo de aprendizaje automático Random Forest predice con alta precisión las concentraciones horarias de SO ₂ al integrar variables meteorológicas locales, mostrando un desempeño superior a los métodos tradicionales de estimación.

Fuente: Elaboración propia

II. MARCO TEÓRICO

2.1. Marco Conceptual

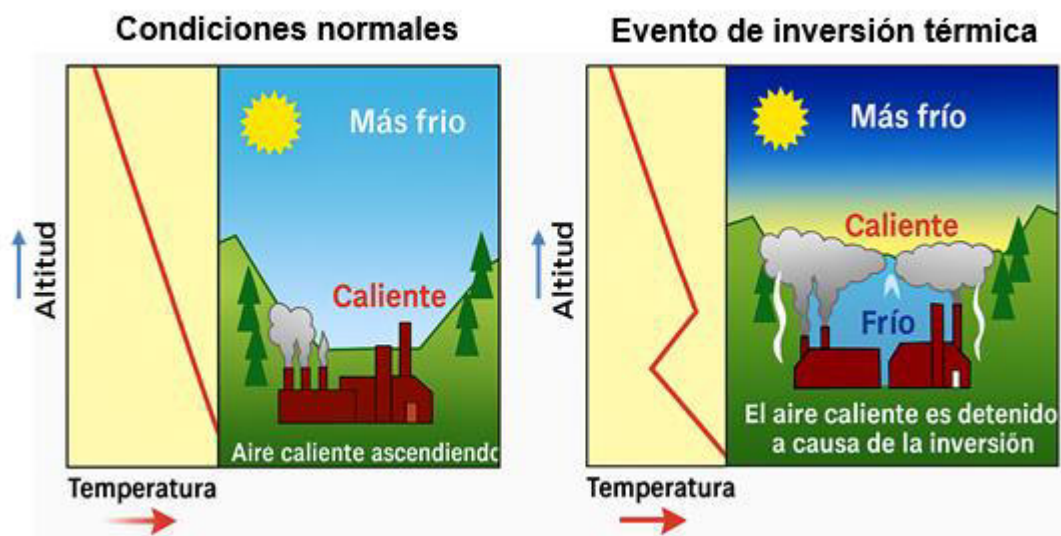
A continuación, se presentan los conceptos fundamentales que sustentan la presente investigación:

2.1.1. *Inversión térmica*

La inversión térmica es un fenómeno meteorológico que se presenta cuando una capa de aire cálido se sitúa por encima de una capa de aire más fría en los niveles bajos de la atmósfera, interrumpiendo el flujo vertical natural del aire. Esta condición atmosférica impide la dispersión de los contaminantes hacia capas superiores, lo que provoca su acumulación cerca de la superficie terrestre, particularmente durante las primeras horas de la mañana y en situaciones de baja radiación solar y escasa circulación de viento.

Figura 2

Gradiente térmico vertical normal y perturbado



Fuente: Elaboración propia

La figura 2, ilustra el comportamiento típico de la temperatura atmosférica en condiciones normales (descenso de temperatura con la altura) y durante una inversión térmica (aumento anómalo de temperatura en un determinado estrato). Este fenómeno genera una

barrera física que impide la dispersión vertical de los contaminantes, como el dióxido de azufre (SO_2), favoreciendo su acumulación en las capas bajas de la atmósfera.

Tal como señalan Seinfeld y Pandis (2016, p. 134), “la inversión térmica interrumpe la circulación vertical del aire, atrapando contaminantes cerca del suelo”. Este fenómeno resulta determinante para explicar la acumulación de contaminantes como el dióxido de azufre (SO_2) en zonas urbanas e industriales. El SO_2 es un gas contaminante generado principalmente por procesos industriales, y cuya exposición prolongada puede ocasionar afecciones en el sistema respiratorio (Manahan, 2007, p. 215).

2.1.2. *Altura de la Capa Límite Atmosférica*

La Altura de la Capa Límite Atmosférica (BLH, por sus siglas en inglés: Boundary Layer Height) se refiere a la altura máxima desde la superficie terrestre hasta donde se extiende la mezcla turbulenta del aire bajo influencia directa de la fricción con la superficie y del calentamiento diurno. Esta capa representa el volumen atmosférico donde ocurren los procesos de dispersión y dilución de contaminantes. Su magnitud varía a lo largo del día, siendo más baja en las primeras horas de la mañana y más elevada durante las horas de máxima radiación solar.

En condiciones de inversión térmica, la BLH tiende a reducirse significativamente, limitando la capacidad de ventilación vertical y provocando acumulación de contaminantes como el dióxido de azufre (SO_2) en las capas más bajas de la atmósfera, lo que incrementa el riesgo de exposición para la población. Según Arya (1999, p. 212), “la altura de la capa límite es una de las variables más críticas para evaluar la capacidad de dispersión atmosférica y la intensidad de los episodios de contaminación”.

2.1.3. *Dióxido de azufre (SO_2)*

El dióxido de azufre (SO_2) es un gas incoloro, irritante y tóxico que se genera principalmente como subproducto de la combustión de combustibles fósiles que contienen

azufre —como el carbón y el petróleo— y de procesos industriales de fundición y refinación de minerales metálicos, en particular durante la tostación de sulfuros. En condiciones atmosféricas normales, el SO_2 es soluble en agua y puede oxidarse en presencia de otros compuestos, contribuyendo a la formación de ácido sulfúrico (H_2SO_4) y, por ende, a la precipitación ácida (lluvia ácida) (Manahan, 2007, p. 215).

Desde el punto de vista de la salud ambiental, el SO_2 es un contaminante primario de alta relevancia por su capacidad para afectar directamente al sistema respiratorio humano. Su inhalación a concentraciones elevadas puede provocar broncoconstricción, irritación pulmonar, agravamiento de cuadros asmáticos y reducción de la función pulmonar, especialmente en poblaciones vulnerables como niños y adultos mayores. Por estas razones, el SO_2 está regulado en numerosos estándares internacionales y nacionales de calidad del aire.

En el Perú, el Decreto Supremo N.º 003-2017-MINAM establece un valor límite de 250 $\mu\text{g}/\text{m}^3$ como promedio en 24 horas para proteger la salud pública. No obstante, los registros horarios pueden revelar episodios críticos de concentración que no son reflejados por los promedios diarios, especialmente bajo condiciones de inversión térmica que restringen la dispersión vertical del contaminante (MINAM, 2017).

2.1.4. Modelamiento predictivo con Random Forest (RF)

El modelamiento predictivo basado en Random Forest (RF) es una técnica de aprendizaje automático supervisado utilizada para predecir valores continuos (regresión) o categorías (clasificación) a partir de múltiples variables explicativas. En el ámbito ambiental, RF ha sido ampliamente aplicado para estimar concentraciones de contaminantes atmosféricos, como el dióxido de azufre (SO_2), a partir de registros históricos y variables meteorológicas relevantes como temperatura, presión atmosférica, viento y altura de la capa límite atmosférica (BLH).

El algoritmo Random Forest se basa en la construcción de múltiples árboles de decisión

(árboles aleatorios) que, en conjunto, mejoran la precisión del modelo y reducen el sobreajuste (overfitting). Su capacidad para capturar relaciones no lineales y manejar datos con alta dimensionalidad lo convierte en una herramienta robusta para estudios en condiciones meteorológicas complejas, como las asociadas a la inversión térmica (Liu et al., 2022; Grange & Carslaw, 2018).

En el presente estudio, el uso del modelo RF permitirá generar una estimación horaria de las concentraciones de SO₂, lo que constituye una innovación metodológica orientada a fortalecer los sistemas de alerta temprana y mejorar la fiscalización ambiental minera con base en evidencia técnica predictiva.

2.1.5. Área de estudio

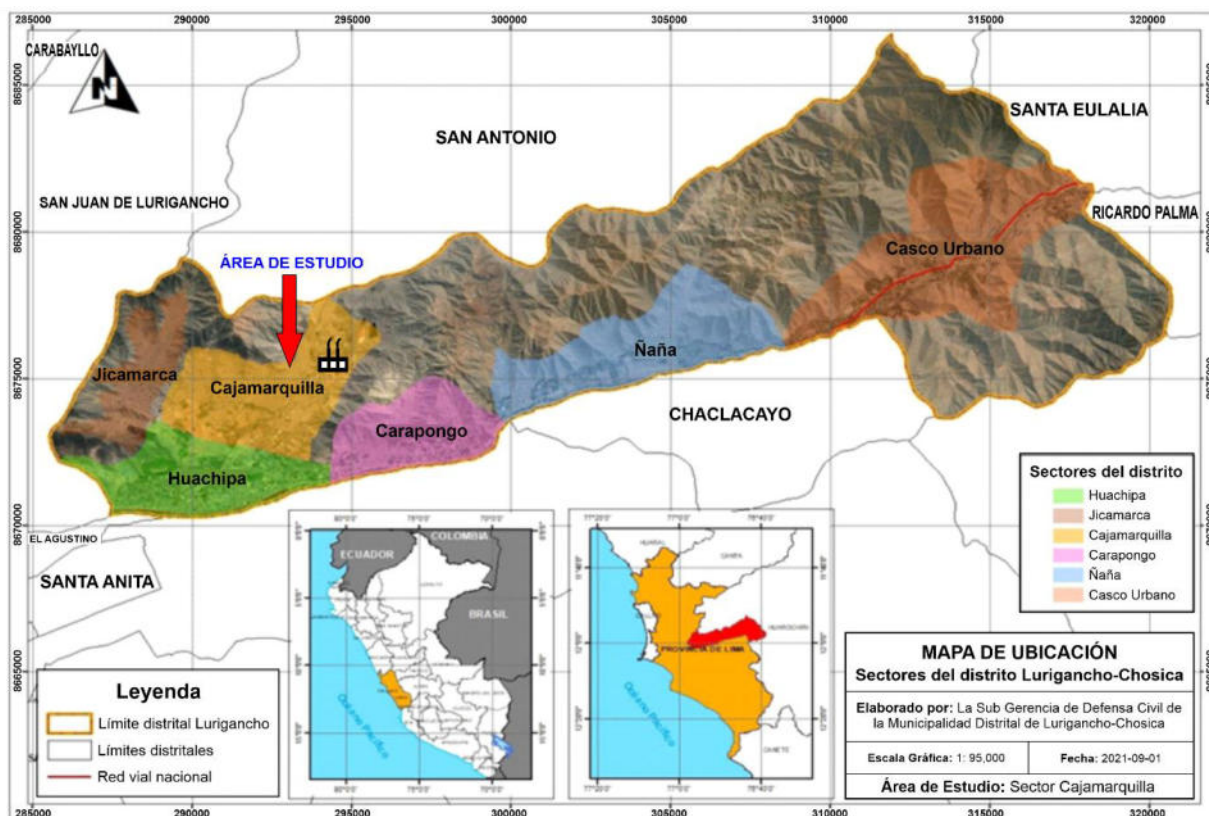
El área de estudio constituye un espacio geográfico delimitado técnicamente con el fin de realizar acciones de fiscalización ambiental. En el presente estudio, dicha área corresponde al ámbito de influencia ambiental directa de la Refinería de Zinc Cajamarquilla, instalación industrial perteneciente al subsector minero, ubicada en el sector Cajamarquilla del distrito de Lurigancho-Chosica, provincia y departamento de Lima.

Cabe señalar que, según el Censo Nacional 2017 realizado por el Instituto Nacional de Estadística e Informática (INEI), el distrito de Lurigancho-Chosica registró una población total de 240 814 habitantes, con una densidad poblacional de 1 018 personas por kilómetro cuadrado y una densidad de 45 hogares por kilómetro cuadrado. En ese marco, se estima que el sector Cajamarquilla concentraba alrededor del 19 % de la población distrital, lo que equivale a aproximadamente 45 755 habitantes al año 2017. Asimismo, para el año 2031, la población estimada de Cajamarquilla asciende a 70 954 habitantes, conforme a proyecciones locales no oficiales. Este crecimiento poblacional refuerza la importancia de evaluar los riesgos ambientales en dicha zona y priorizar acciones de fiscalización ambiental que consideren su densidad demográfica y vulnerabilidad.

A continuación, se presenta la Figura 3, en el que se muestra la ubicación del sector Cajamarquilla dentro de los sectores que conforman el distrito de Lurigancho, señalándose el área correspondiente al Área de estudio:

Figura 3

Ubicación del Área de Estudio



Fuente: Municipalidad distrital Lurigancho-Chosica (2021)

2.1.6. Refinería de Zinc de Cajamarquilla

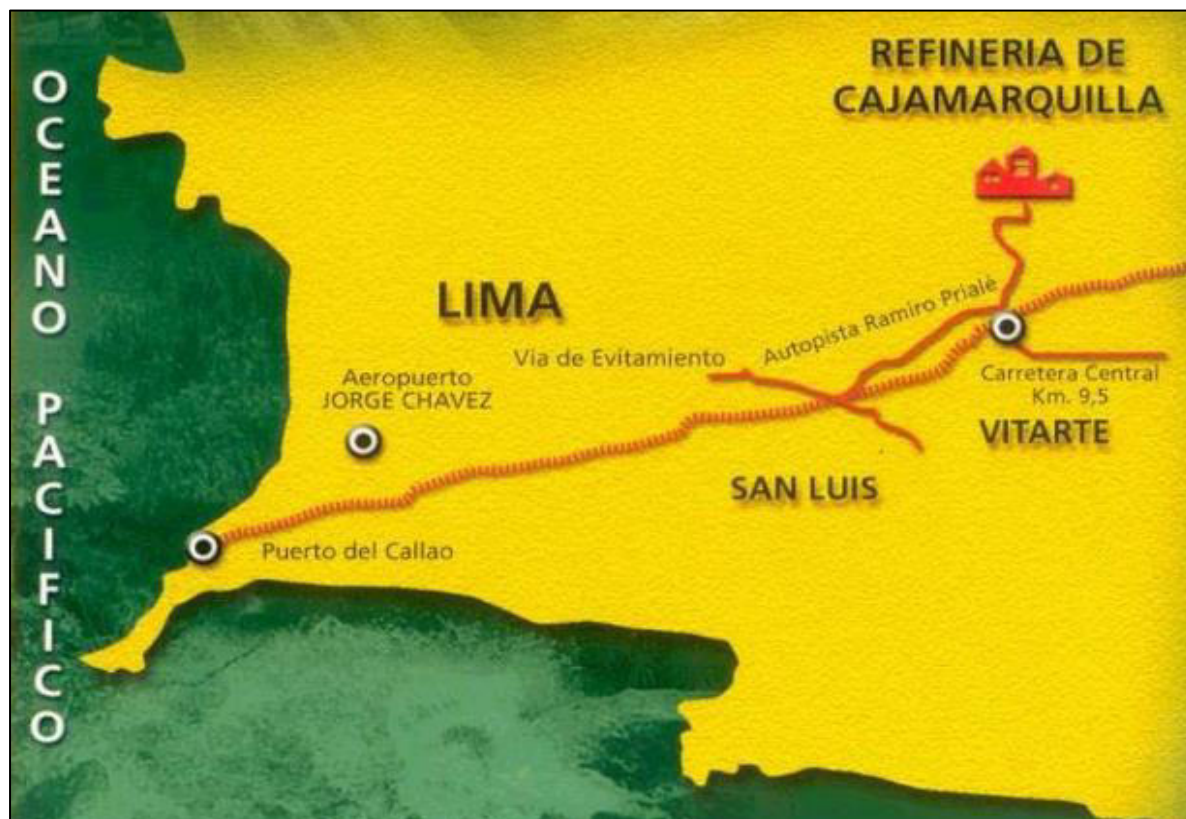
Durante el periodo 2022–2024, la estación de monitoreo ubicada a sotavento del OEFA registró un número considerable de superaciones horarias del umbral de $250 \mu\text{g}/\text{m}^3$ de SO_2 : 443 en 2022, 346 en 2023 y 413 en 2024. Sin embargo, las superaciones del promedio diario del ECA Aire (24 horas) fueron solo 5 en 2022 y 6 en 2023, mientras que en 2024 no se reportaron excedencias. Esta discrepancia pone en evidencia un vacío normativo relevante, ya que el estándar diario enmascara episodios críticos de contaminación que ocurren en intervalos

horarios específicos y que representan riesgos agudos tanto para la salud humana como para el ambiente.

La refinería de zinc Cajamarquilla se ubica en el distrito Lurigancho-Chosica, provincia y departamento Lima, en el valle de la quebrada Jicamarca (afluente del río Rímac), a una altitud de 450 m s. n. m. En este valle, al noreste de la refinería, se encuentra el Radio Observatorio de Jicamarca (ROJ) del Instituto Geofísico del Perú (IGP). Hidrográficamente, está ubicada en el ámbito de la cuenca del río Rímac, perteneciente a la región hidrográfica de la vertiente del Pacífico. Se localiza en la cuenca seca de la quebrada Jicamarca, denominada así a partir de la confluencia de las quebradas Huaycoloro y Río Seco, a la altura del kilómetro 9,5 de la Carretera Central, a 22 km del centro de Lima en dirección noreste.

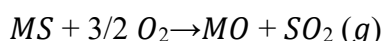
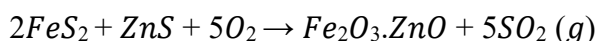
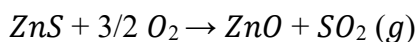
Figura 4

Ubicación de la Refinería de Zinc de Cajamarquilla



Fuente: Refinería de Zinc de Cajamarquilla (Cubero, 2014)

La Refinería de Zinc de Cajamarquilla produce zinc refinado con 99,995 % de pureza a partir de concentrados de zinc, además obtiene subproductos como: ácido sulfúrico (98,5 % de H_2SO_4), cadmio refinado (99,7 % de Cd); cemento de cobre (60% de Cu), concentrado comercializable de plomo-plata (15 % de Pb y 500 onzas de Ag/TC), residuo de cobalto-níquel (2% de Co); polvo de zinc; e indio refinado. Asimismo, el proceso de tostación se realiza para oxidar los sulfuros, remover el 99% del azufre contenido en los concentrados de zinc y obtener una calcina que es un producto intermedio en el proceso de refinación de zinc que contiene óxidos, y gases residuales conteniendo dióxido de azufre (SO_2), siendo las principales reacciones químicas que se producen en esta etapa:



Donde:

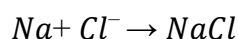
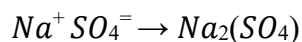
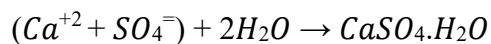
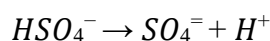
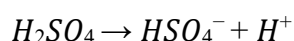
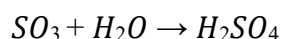
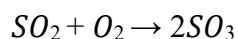
$M = Pb, Cu, Cd, Ag, etc.$

Los gases de tostación son conducidos a una etapa de depuración de gases que consiste en torres en serie, de enfriamiento y lavado, y una batería de 4 filtros electrostáticos. Los gases limpios y saturados con agua, procedentes de la planta de depuración de gases, pasan luego por una torre de secado en el cual, en contracorriente, se ponen en contacto los gases con una solución de ácido sulfúrico al 95%.

El gas residual con SO_2 procedente de las plantas de ácido sulfúrico, con una concentración aproximado de 500 ppm, se envía a la primera etapa del sistema de lavado de gases (Depurador Venturi) con una temperatura de aproximadamente 80 °C. Después de la etapa del Depurador Venturi, el gas de salida conteniendo todavía cantidades residuales de SO_2 , llega a la segunda etapa compuesta de una Torre de Lavado en contracorriente a la irrigación ácido. En esta torre de lavado o riego, el contenido de SO_2 del gas se reduce hasta el valor

deseado por la reacción con H_2O_2 que aún está contenido en el ácido. Finalmente, los gases tratados son descargados al ambiente, mediante dos chimeneas de 60 metros de altura sobre el nivel superficial de la refinería Cajamarquilla.

Jiménez (2017), y Figueruelo y Marino (2014) indicaron que el SO_2 en la atmosfera se oxida a trióxido de azufre y al contacto con la niebla de agua se forma ácido sulfúrico, que posteriormente cae sobre el suelo; este ácido fuerte se disocia en sulfato y 2 hidrógenos. En el suelo el sulfato precipita como yeso, o forma sales más solubles que el yeso (sulfato de sodio y cloruro de sodio). Las reacciones químicas de estos procesos en la atmósfera y en el suelo se detallan a continuación:



Manahan (2007) menciona que durante la deposición de SO_2 también podría estar presente otros gases como Cl_2 , H_2S , NO_2 , CO y O_3 los cuales llegan a precipitarse en el suelo luego de sus reacciones en la atmosfera.

Norma técnica que establece los estándares de calidad ambiental para la concentración de contaminantes en el aire con el fin de proteger la salud y el ambiente. En Perú, el D.S. N.º 003-2017-MINAM fija el valor límite diario de concentración de SO₂ en 250 µg/m³ en promedio de 24 horas.

Tabla 3*Estándares de calidad ambiental (ECA) de SO₂ para aire*

Parámetro	Periodo	Formato del estándar		Método de análisis	Norma
		Valor (µg/m ³)	Criterios de evaluación		
Dióxido de azufre (SO ₂)	24 horas	250	NE más de 7 veces al año	Fluorescencia ultravioleta (método automático)	D.S. N.º 003-2017-MINAM «Estándares de Calidad Ambiental (ECA) para Aire y Disposiciones Complementarias»

Fuente: Elaboración propia

2.1.8. Protocolo Nacional de Monitoreo de la Calidad Ambiental del Aire

El “Protocolo Nacional de Monitoreo de la Calidad Ambiental del Aire”, es un instrumento que permite estandarizar los criterios técnicos para el monitoreo ambiental del aire en el país, a fin de generar información de calidad, comparable, compatible, confiable y representativa.

Tabla 4*Protocolo de monitoreo de la calidad del aire*

Matriz	Protocolo	Sección	País	Institución	Dispositivo legal	Año
Aire	Protocolo nacional de monitoreo de la calidad ambiental del aire	Todo el documento	Perú	Minam	Decreto Supremo N.º 010-2019-MINAM	2019

Fuente: Elaboración Propia

2.1.9. Fiscalización ambiental

Conjunto de acciones ejecutadas por las autoridades competentes para verificar el cumplimiento de la normativa ambiental, condiciones de los instrumentos de gestión ambiental

y obligaciones derivadas de las actividades económicas. Incluye supervisión, monitoreo, seguimiento de obligaciones y aplicación de medidas correctivas o sancionadoras. En el Perú, el Organismo de Evaluación y Fiscalización Ambiental (OEFA) es el organismo responsable de la fiscalización ambiental en el subsector minero.

La fiscalización ambiental puede ser entendida en dos sentidos:

- **Sentido amplio:** Comprende las acciones de evaluación de la calidad ambiental, supervisión, fiscalización y sanción, así como la aplicación de incentivos, realizadas por el OEFA y las Entidades de Fiscalización Ambiental (EFA) con la finalidad de asegurar el cumplimiento de obligaciones ambientales fiscalizables.

- **Sentido estricto:** Comprende la facultad de investigar la comisión de posibles infracciones administrativas en el marco de un procedimiento administrativo sancionador, y la de imponer sanciones y medidas cautelares y correctivas frente al incumplimiento de obligaciones ambientales por parte de los administrados.

2.1.10. Zona de barlovento y sotavento

Términos meteorológicos utilizados para describir la posición relativa de un punto respecto al viento. La zona de barlovento es aquella desde donde proviene el viento, mientras que la de sotavento es aquella hacia donde se dirige. En estudios de dispersión de contaminantes, el análisis entre estaciones ubicadas en estas zonas permite entender la distribución espacial del contaminante.

Las zonas de barlovento y sotavento son esenciales para entender la dispersión espacial de contaminantes, ya que describen respectivamente los puntos desde donde proviene el viento y hacia donde se dirige (Arya, 1999, p. 88).

2.1.11. Monitoreo automático en tiempo real

Sistema de vigilancia ambiental que emplea sensores y equipos especializados para el registro continuo de parámetros de calidad del aire y variables meteorológicas. Las estaciones

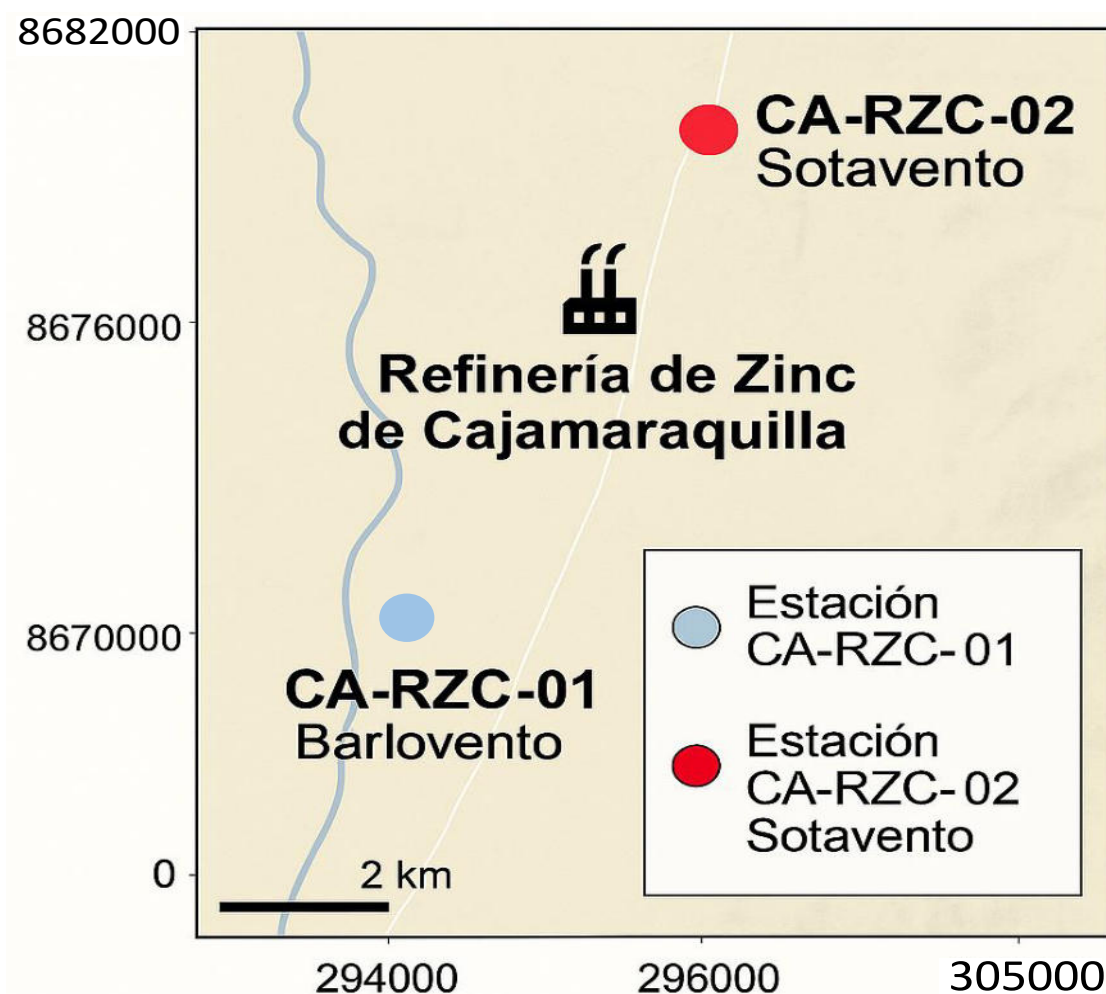
automáticas de monitoreo permiten obtener datos en tiempo real sobre contaminantes atmosféricos y condiciones meteorológicas, proporcionando información con alta resolución temporal (MINAM, 2018, p. 22). Esta información es fundamental para la identificación de eventos críticos, como las superaciones de los Estándares de Calidad Ambiental (ECA) o la acumulación de contaminantes durante episodios de inversión térmica.

Tabla 5

Ubicación de las estaciones de monitoreo de calidad del aire

N.º	Código	Coordenadas UTM WGS-84 Zona 18 L		Altitud (m.s.n.m)	Descripción
		Este (m)	Norte (m)		
1	CA-RZC-01 (Barlovento)	294343	8675073	428	Techo de la casa ubicada en el Jr. Los Andes Lote 109, aproximadamente a 750 m y al sur suroeste (SSO) del ingreso de la refinería de zinc Cajamarquilla.
2	CA-RZC-02 (Sotavento)	295758	8677964	504	Techo de la oficina de medidores de SO ₂ en las instalaciones del Radio Observatorio de Jicamarca (ROJ), ubicada al norte noreste (NNE) de la refinería de zinc Cajamarquilla de Nexa Resources Cajamarquilla S.A.C.

Fuente: Elaboración propia

Figura 6*Ubicación de las estaciones de monitoreo*

Fuente: Elaboración propia

2.2. Bases Teóricas

La literatura científica internacional y nacional ha documentado de manera consistente la relación entre las condiciones de inversión térmica y el incremento de contaminantes atmosféricos como el dióxido de azufre (SO_2). Estos antecedentes son relevantes porque permiten identificar tanto los patrones comunes de acumulación de contaminantes bajo estabilidad atmosférica, como las brechas metodológicas y normativas que persisten en el contexto peruano. A continuación, se presentan cuadros comparativos que sintetizan los hallazgos más significativos a nivel internacional y nacional, con énfasis en la relación entre

meteorología, calidad del aire y herramientas predictivas como el modelo Random Forest.

Tabla 6

Cuadro comparativo internacional sobre inversión térmica y SO₂

País / Ciudad	Fenómeno observado	Impacto en SO₂ u otros contaminantes	Fuente
Santiago de Chile	Inversiones térmicas frecuentes en invierno, asociadas a baja ventilación atmosférica.	Incremento significativo de SO ₂ y material particulado fino durante episodios críticos; se recomendaron indicadores meteorológicos para activar alertas.	Muñoz & Salinas (2018)
Katmandú (Nepal)	Presencia de inversiones térmicas persistentes en valle cerrado.	Duplicación de niveles de SO ₂ y material particulado, con impactos severos en salud pública.	Gautam et al. (2021)
Regiones industriales de Europa	Episodios agudos de contaminación coinciden con condiciones atmosféricas estables (poca dispersión vertical).	Incremento de SO ₂ y otros contaminantes durante periodos de estabilidad térmica.	Lenschow et al. (2001)
Londres (Reino Unido)	Uso de modelos predictivos (RF) para estimar contaminantes bajo condiciones meteorológicas complejas.	Random Forest alcanzó $R^2 > 0.85$ en predicciones horarias de SO ₂ y NO ₂ .	Grange & Carslaw (2018)
Beijing (China)	Alta recurrencia de episodios de inversión térmica con fuerte impacto en calidad del aire urbano.	Modelos RF y ML superan métodos tradicionales en predicción horaria de contaminantes.	Liu et al. (2022)
California (EE. UU.)	Inversiones térmicas en valles costeros y zonas industriales.	Modelos RF predicen episodios críticos de SO ₂ y O ₃ , mejorando la gestión de alertas tempranas.	Zhan et al. (2021)

Fuente: Elaboración propia a partir de Muñoz y Salinas (2018); Gautam et al. (2021); Lenschow et al. (2001); Grange y Carslaw (2018); Liu et al. (2022); Zhan et al. (2021).

Tabla 7

Cuadro comparativo nacional sobre inversión térmica y SO₂

Institución / Autor	Fenómeno observado	Impacto en SO₂ u otros contaminantes	Fuente
OEFA (2020, 2022, 2023, 2024)	Registros horarios en estaciones automáticas en Cajamarquilla.	Superaciones puntuales del límite de 250 µg/m ³ de SO ₂ , especialmente en horas de la mañana, coincidiendo con condiciones de inversión térmica.	OEFA (Reportes técnicos N.º 00007-2022, 00022-2023, 00038-2024)
SENAMHI (2020)	Condiciones meteorológicas locales en Lima Este.	Identificación de recurrencia de inversiones térmicas matinales en valles cerrados como Jicamarca, con baja ventilación atmosférica.	SENAMHI (2020)
IGP – Radio Observatorio de Jicamarca (2019)	Observación de estabilidad atmosférica en la cuenca del Rímac.	Datos meteorológicos y topográficos muestran limitaciones a la dispersión vertical de contaminantes en la quebrada Jicamarca.	IGP (2019)
Ramos et al. (2021)	Uso de modelos de <i>machine learning</i> en calidad del aire en Perú.	Validación de Random Forest para predecir contaminantes atmosféricos, con robustez frente a datos variables y correlacionados.	Ramos et al. (2021)
MINAM (2018, 2019)	Protocolos nacionales de monitoreo de calidad del aire.	Reconoce la importancia de integrar variables meteorológicas al análisis de contaminantes, aunque aún no se aplican plenamente en fiscalización.	MINAM (2018, 2019)

Fuente: Elaboración propia a partir de OEFA (2022, 2023, 2024); SENAMHI (2020); IGP (2019); Ramos et al. (2021); MINAM (2018, 2019).

Del análisis comparativo se observa que, en países como Chile, Nepal, China y Reino Unido, las inversiones térmicas han sido ampliamente estudiadas y reconocidas como factores determinantes en los episodios críticos de contaminación atmosférica, lo que ha impulsado el uso de modelos predictivos avanzados como el Random Forest para la gestión de alertas tempranas. En contraste, en el Perú, aunque existen reportes técnicos del OEFA, SENAMHI, IGP y estudios académicos recientes, la integración operativa de variables meteorológicas en la fiscalización ambiental aún es incipiente. Esta brecha justifica la necesidad de investigaciones como la presente, orientadas a fortalecer la fiscalización ambiental minera mediante el uso de herramientas de inteligencia artificial que permitan anticipar episodios críticos de contaminación por SO₂ en escenarios de inversión térmica.

El modelo Gaussian Plume establece que “la dispersión de contaminantes en la atmósfera depende críticamente de la estabilidad atmosférica, la velocidad del viento y la temperatura” (Arya, 1999, p. 145). Gautam et al. (2021, p. 5) demostraron que “en Katmandú, durante eventos de inversión térmica, los niveles de SO₂ y material particulado se incrementaron en más del 70% debido a la escasa dispersión vertical”. Muñoz y Salinas (2018, p. 37) señalaron que “en Santiago de Chile, las inversiones térmicas son un factor determinante en la ocurrencia de episodios críticos de contaminación atmosférica”, recomendando el uso de indicadores meteorológicos en la activación de alertas sanitarias. Lenschow et al. (2001, p. 53) indicaron que “en regiones industriales de Europa, los episodios agudos de contaminación suelen coincidir con patrones meteorológicos estables”, reforzando la necesidad de integrar estos factores en las políticas de control de calidad del aire.

En los últimos años, las limitaciones de los modelos clásicos deterministas y estadísticos para capturar la complejidad no lineal de las condiciones atmosféricas han motivado el uso de técnicas de aprendizaje automático (machine learning) en el estudio de la contaminación del aire. Entre estas, el algoritmo Random Forest (RF) ha demostrado alta

eficiencia y precisión en la predicción de contaminantes como el dióxido de azufre (SO₂), al modelar relaciones complejas entre variables meteorológicas y niveles de concentración en tiempo real (Grange & Carslaw, 2018; Liu et al., 2022).

El modelo RF consiste en la construcción de múltiples árboles de decisión que se entrenan con subconjuntos aleatorios del conjunto de datos. A partir de la agregación (ensamblaje) de sus resultados, se obtiene una predicción más robusta y menos sensible al sobreajuste. Esta técnica ha sido utilizada con éxito en diversas investigaciones internacionales para estimar la concentración de contaminantes atmosféricos a escalas horarias y subdiarias, especialmente en contextos meteorológicos adversos como la inversión térmica, donde los métodos tradicionales no logran capturar la variabilidad dinámica (Zhang et al., 2020; Zhan et al., 2021).

En estudios de caso desarrollados en ciudades como Londres, Los Ángeles y Beijing, el modelo Random Forest ha sido entrenado con variables como temperatura, presión atmosférica, velocidad del viento y altura de la capa límite atmosférica (BLH), obteniendo coeficientes de determinación (R^2) superiores al 0.85 en escenarios de predicción de SO₂ y NO₂ (Liu et al., 2022). En consecuencia, su aplicación en la presente investigación representa una contribución metodológica sustantiva, al introducir un enfoque de inteligencia artificial orientado a anticipar eventos críticos de contaminación atmosférica y reforzar el principio de prevención en la fiscalización ambiental minera.

2.3. Bases Legales

La Constitución Política del Perú establece en su artículo 2, inciso 22, el derecho de toda persona a gozar de un ambiente equilibrado y adecuado para el desarrollo de la vida (Congreso de la República, 1993, p. 12).

La Ley General del Ambiente, Ley N.º 28611 (Ministerio del Ambiente, 2005, p. 34) señala que “el Estado garantiza el derecho de la población a vivir en un ambiente saludable,

equilibrado y adecuado, estableciendo los principios de prevención y precaución”.

La Ley del Sistema Nacional de Evaluación y Fiscalización Ambiental (Ley N.º 29325) otorga al OEFA la función de “supervisar y fiscalizar el cumplimiento de las obligaciones ambientales” (OEFA, 2009, p. 15).

A nivel normativo, el Decreto Supremo N.º 003-2017-MINAM aprueba los Estándares de Calidad Ambiental (ECA) para aire, fijando un límite diario de 250 $\mu\text{g}/\text{m}^3$ para SO_2 (MINAM, 2017, p. 7),

Mientras que el Decreto Supremo N.º 009-2003-SA establece niveles de alerta para contaminantes críticos (MINSA, 2003, p. 10).

El Protocolo Nacional de Monitoreo de Calidad del Aire define “los estándares técnicos para la operación de las estaciones automáticas, incluyendo criterios de calibración y mantenimiento” (MINAM, 2018, p. 28).

2.4. Aspectos de Responsabilidad Social y Medio Ambiente

La presente investigación incorpora de manera transversal el enfoque de responsabilidad social y la perspectiva ambiental, reconociendo que la gestión adecuada de la calidad del aire no solo responde a obligaciones técnicas y legales, sino también a compromisos éticos e institucionales con las comunidades y el entorno.

Desde el enfoque de responsabilidad social, el estudio contribuye a visibilizar los impactos diferenciados que la contaminación por dióxido de azufre (SO_2) puede generar sobre poblaciones vulnerables, especialmente en zonas urbanas aledañas a complejos industriales como la Refinería de Zinc de Cajamarquilla. Estas poblaciones, a menudo expuestas de manera crónica a condiciones atmosféricas adversas, enfrentan un mayor riesgo de afectación en su salud y calidad de vida debido a la limitada capacidad de respuesta institucional en ausencia de herramientas de vigilancia anticipada.

Tal como lo señala la Ley General del Ambiente, Ley N.º 28611 (2005, art. 6), “el

Estado y las personas naturales y jurídicas, públicas o privadas, están obligadas a prevenir y controlar los riesgos ambientales que puedan afectar a las generaciones presentes y futuras”. En este marco, el fortalecimiento de los mecanismos de fiscalización ambiental, basados en evidencia técnica y criterios meteorológicos, se convierte en una herramienta clave para garantizar el derecho a vivir en un ambiente saludable, en concordancia con el artículo 2, inciso 22, de la Constitución Política del Perú (Congreso de la República, 1993).

En ese sentido, la presente investigación propone la integración de herramientas de inteligencia artificial como el modelo predictivo Random Forest (RF) para estimar concentraciones horarias de SO₂ con base en variables meteorológicas locales. Esta innovación no solo representa un avance metodológico, sino también un acto de responsabilidad institucional, ya que permite anticipar eventos críticos de contaminación antes de que ocurran, reforzando el principio de prevención y reduciendo la exposición innecesaria de la población a riesgos ambientales severos.

Desde la perspectiva ambiental, esta investigación promueve el principio de prevención al proponer alertas anticipadas que permitan reducir la exposición poblacional a concentraciones horarias críticas de SO₂. La investigación se alinea con el objetivo del Sistema Nacional de Evaluación y Fiscalización Ambiental (SINEFA), cuyo fin es asegurar el cumplimiento de las obligaciones ambientales por parte de los administrados (OEFA, 2009). En ese sentido, el fortalecimiento de los sistemas de monitoreo y análisis atmosférico no solo contribuye a mejorar la fiscalización del OEFA, sino que refuerza la gobernanza ambiental y la participación informada de los actores sociales.

El enfoque de sostenibilidad también está presente en el diseño metodológico del estudio, al considerar no solo la concentración de contaminantes, sino también su comportamiento en función de las condiciones meteorológicas locales. El uso del modelo Random Forest (RF) permite además construir escenarios de predicción basados en datos

reales, lo que facilita una gestión adaptativa del riesgo ambiental, compatible con estrategias de desarrollo sostenible y resiliencia territorial.

Esta perspectiva integrada permite adoptar decisiones basadas en ciencia, con sensibilidad social y en armonía con los Objetivos de Desarrollo Sostenible (ODS), en particular los ODS 3 (Salud y bienestar), 11 (Ciudades y comunidades sostenibles), 12 (Producción y consumo responsables) y 13 (Acción por el clima).

En el caso del distrito de Lurigancho-Chosica, la zona de Cajamarquilla constituye un espacio particularmente sensible debido a su carácter urbano-industrial y a su densidad poblacional, que representa cerca del 19 % del total distrital (INEI, 2017; Ipsos Perú, 2016). Esta concentración demográfica, sumada a la presencia de la Refinería de Zinc de Cajamarquilla, refuerza la necesidad de integrar la dimensión social en la gestión ambiental. La exposición continua de comunidades a episodios de contaminación atmosférica, como los derivados de la inversión térmica, plantea no solo un reto técnico, sino también un imperativo ético y social en la fiscalización ambiental, en concordancia con los principios de prevención, precaución y sostenibilidad.

En conclusión, la presente investigación no solo posee un valor técnico y académico, sino que también proyecta un aporte práctico a la construcción de territorios más resilientes, instituciones más eficaces en la fiscalización ambiental y comunidades mejor protegidas frente a los efectos adversos de la contaminación atmosférica. Al integrar el análisis horario de contaminantes con variables meteorológicas críticas y la aplicación de modelos predictivos como Random Forest, el estudio responde a los vacíos normativos actuales y aporta criterios técnicos complementarios que fortalecen los mecanismos de alerta temprana. De esta manera, el Marco Teórico se articula directamente con los objetivos e hipótesis planteados en el Capítulo I, ofreciendo un sustento conceptual y metodológico sólido para el desarrollo de la investigación.

En suma, el Marco Teórico no solo ha permitido identificar los fundamentos conceptuales, normativos y sociales vinculados a la problemática de la contaminación por dióxido de azufre (SO_2) en contextos de inversión térmica, sino que también ha evidenciado la pertinencia del uso de modelos predictivos como Random Forest para fortalecer la fiscalización ambiental. Estos elementos constituyen la base sobre la cual se construye el diseño metodológico del presente estudio. En el siguiente capítulo se desarrolla la estrategia de investigación aplicada, detallando la población y muestra, el enfoque metodológico y las técnicas de análisis empleadas para dar cumplimiento a los objetivos planteados y validar las hipótesis formuladas.

III. MÉTODO

3.1. Tipo de Investigación

3.1.1. *Por su Naturaleza*

La presente investigación es de naturaleza aplicada, puesto que no se limita a la generación de conocimiento teórico, sino que busca ofrecer soluciones prácticas al problema de la fiscalización ambiental minera en la zona de influencia directa de la Refinería de Zinc de Cajamarquilla. Según Hernández et al. (2014), este tipo de investigación permite traducir los hallazgos científicos en mejoras concretas, como lineamientos técnicos y recomendaciones operativas para las entidades competentes, en este caso el Organismo de Evaluación y Fiscalización Ambiental (OEFA).

Tal como señalan los mismos autores:

La investigación aplicada está orientada a resolver problemas específicos y concretos que se presentan en la sociedad, organizaciones o comunidades, utilizando los conocimientos científicos previamente generados. Su finalidad no es únicamente ampliar el conocimiento teórico, sino transformar la realidad práctica a partir de hallazgos comprobables y aplicables, permitiendo que instituciones u organismos desarrollen acciones correctivas, preventivas o de mejora.

(Hernández et al., 2014, p. 126)

En coherencia con este enfoque, el estudio incorpora como innovación metodológica la implementación de un modelo predictivo basado en Random Forest (RF), orientado a estimar con precisión las concentraciones horarias de dióxido de azufre (SO₂) a partir de variables meteorológicas. Esta herramienta, fundamentada en técnicas de aprendizaje automático, permite anticipar escenarios de riesgo ambiental y formular medidas de control más eficaces, reforzando las acciones preventivas y correctivas del OEFA frente a eventos críticos de contaminación.

De este modo, la investigación trasciende el plano descriptivo y se orienta a la generación de instrumentos técnicos aplicables a la fiscalización ambiental minera, contribuyendo a la toma de decisiones basada en evidencia, al fortalecimiento institucional y a la garantía del derecho constitucional a un ambiente sano.

3.1.2. *Por su Profundidad*

El presente estudio se ubica en el nivel explicativo, dado que no se limita a describir las concentraciones horarias de SO₂, sino que busca comprender las relaciones de causa-efecto entre las condiciones de inversión térmica y los episodios críticos de contaminación. De acuerdo con Sampieri et al. (2014), la investigación explicativa constituye el nivel más profundo, pues permite identificar los factores subyacentes que influyen en un fenómeno y determinar las condiciones en las que este se manifiesta.

Tal como señalan los autores:

El nivel explicativo es el más profundo de los niveles de investigación, pues no solo describe fenómenos y establece relaciones entre variables, sino que pretende determinar las causas de los eventos o fenómenos que se estudian. Se enfoca en responder a las preguntas de por qué ocurren los hechos y bajo qué condiciones se manifiestan, permitiendo diseñar modelos que puedan predecir situaciones similares en el futuro.

(Sampieri et al., 2014, p. 143)

En este sentido, el estudio busca establecer cómo variables meteorológicas como la temperatura, la velocidad del viento y la presión atmosférica inciden en el comportamiento horario del dióxido de azufre (SO₂), particularmente bajo condiciones de inversión térmica. Asimismo, incorpora como componente de profundización analítica un modelo predictivo basado en Random Forest (RF), que permite no solo validar empíricamente las relaciones causales identificadas, sino también simular y anticipar concentraciones futuras de SO₂ bajo condiciones atmosféricas similares.

De esta manera, el nivel explicativo, complementado con técnicas de aprendizaje automático, refuerza la capacidad de anticipación frente a episodios críticos de contaminación, proporcionando a las entidades fiscalizadoras herramientas que permiten pasar del diagnóstico al diseño de medidas preventivas basadas en evidencia.

3.1.3. *Por su Metodología*

La presente investigación adopta un enfoque cuantitativo, caracterizado por el uso de mediciones numéricas, análisis estadísticos y modelos matemáticos. Este enfoque resulta idóneo para analizar patrones en datos atmosféricos y establecer relaciones cuantificables entre variables ambientales y meteorológicas. Según Bisquerra (2004):

El enfoque cuantitativo se fundamenta en la cuantificación de los datos recolectados, la aplicación de herramientas estadísticas para su análisis y la obtención de conclusiones que puedan generalizarse a partir de los resultados. Este enfoque permite establecer relaciones numéricas entre las variables y utilizar modelos predictivos que fortalezcan la validez de los hallazgos obtenidos (p. 99).

En coherencia con esta perspectiva, el estudio emplea análisis estadístico de series temporales, correlaciones entre variables meteorológicas y concentraciones horarias de dióxido de azufre (SO₂), así como el cálculo de frecuencias, medias y tendencias. Como innovación metodológica, se incorpora un modelo de aprendizaje automático basado en Random Forest (RF), orientado a predecir concentraciones horarias de SO₂ en función de variables meteorológicas tales como temperatura, presión atmosférica, velocidad y dirección del viento y altura de la capa límite atmosférica (BLH).

El algoritmo Random Forest opera bajo una estructura matemática robusta de múltiples árboles de decisión, cuya combinación se traduce en resultados más estables y precisos, mediante procesos de votación o promedios, según se trate de clasificación o regresión. La integración de este modelo dentro del enfoque cuantitativo fortalece la capacidad explicativa y

predictiva de la investigación, aportando una herramienta rigurosa para la toma de decisiones basadas en evidencia en el ámbito de la fiscalización ambiental.

3.1.4. *Por su Diseño*

El diseño de la presente investigación es no experimental y correlacional, en tanto no se manipulan las variables independientes ni se aplican tratamientos controlados, sino que se observan y analizan las relaciones entre las variables en su contexto natural. De acuerdo con Hernández et al. (2014), este diseño resulta pertinente cuando se trabaja con datos históricos y se busca establecer asociaciones estadísticas sin intervención directa sobre los fenómenos estudiados.

En este caso, se emplean registros horarios reales de concentración de dióxido de azufre (SO_2) y de variables meteorológicas —como temperatura, presión atmosférica, velocidad y dirección del viento, y altura de la capa límite atmosférica (BLH)— obtenidos de las estaciones automáticas del OEFA para el periodo 2022–2024. El análisis se orienta a identificar patrones y asociaciones entre estas variables, con especial énfasis en episodios de inversión térmica.

Dentro de este diseño, se incorpora la implementación de un modelo predictivo basado en Random Forest (RF), el cual, si bien no implica manipulación experimental, permite estimar la variable dependiente (concentración horaria de SO_2) a partir de múltiples predictores meteorológicos. El modelo se ajusta a la lógica del diseño correlacional, en tanto identifica y cuantifica relaciones funcionales entre variables observadas, apoyándose en patrones empíricos extraídos de los datos históricos.

En consecuencia, el diseño no experimental se ve enriquecido mediante el uso de herramientas de inteligencia artificial que amplían la capacidad de interpretación y predicción, sin alterar las condiciones naturales del fenómeno. Este enfoque resulta coherente con la naturaleza observacional y predictiva del estudio, y fortalece el análisis ambiental aplicado en entornos complejos como el de la Refinería de Zinc de Cajamarquilla.

En síntesis, la investigación se caracteriza por ser de naturaleza aplicada, de nivel explicativo, con enfoque cuantitativo y diseño no experimental de tipo correlacional. Este marco metodológico permite abordar con rigor la relación entre la inversión térmica y las concentraciones horarias de SO_2 , integrando tanto el análisis estadístico de datos históricos como la aplicación de modelos predictivos avanzados. Con esta base, en el siguiente apartado se definen la población y muestra que constituyen el objeto empírico del estudio, lo cual asegura la coherencia entre el planteamiento metodológico y los objetivos de investigación.

3.2. Población y Muestra

3.2.1. La población

La población de estudio está constituida por la totalidad de los registros horarios de concentración de dióxido de azufre (SO_2) y de las variables meteorológicas (temperatura, presión atmosférica, velocidad y dirección del viento), obtenidos en las dos estaciones automáticas de monitoreo ambiental instaladas por el Organismo de Evaluación y Fiscalización Ambiental (OEFA) en el área de influencia directa de la Refinería de Zinc de Cajamarquilla. Estas estaciones se encuentran ubicadas estratégicamente a barlovento y sotavento del complejo industrial, lo que permite caracterizar con mayor precisión la dispersión espacial y temporal de contaminantes en función de las condiciones atmosféricas locales.

Adicionalmente, se incorporó como variable independiente la altura de la capa límite atmosférica (BLH), información descargada del repositorio internacional del programa *ERA5 – Reanalysis Single Levels*, disponible en el Copernicus Climate Data Store. Esta variable, que no es registrada directamente por las estaciones del OEFA, resulta clave para estimar la estabilidad atmosférica y la ocurrencia de inversión térmica, factores determinantes en el comportamiento de la contaminación por SO_2 .

Este conjunto integrado de datos constituye la base empírica para el análisis descriptivo, correlacional y predictivo de la investigación. Además, ofrece las condiciones necesarias para

la implementación de modelos de aprendizaje automático como Random Forest (RF), los cuales requieren bases multivariadas y de alta resolución temporal.

Por otro lado, la zona de estudio —correspondiente al sector Cajamarquilla del distrito de Lurigancho-Chosica— concentra aproximadamente 45,755 habitantes, lo que equivale al 19 % de la población total distrital según el Censo Nacional 2017 del Instituto Nacional de Estadística e Informática (INEI, 2017). Esta densidad poblacional otorga especial relevancia al estudio, ya que los impactos de la contaminación atmosférica tienen implicancias directas sobre la salud pública y el bienestar de comunidades urbano-industriales vulnerables.

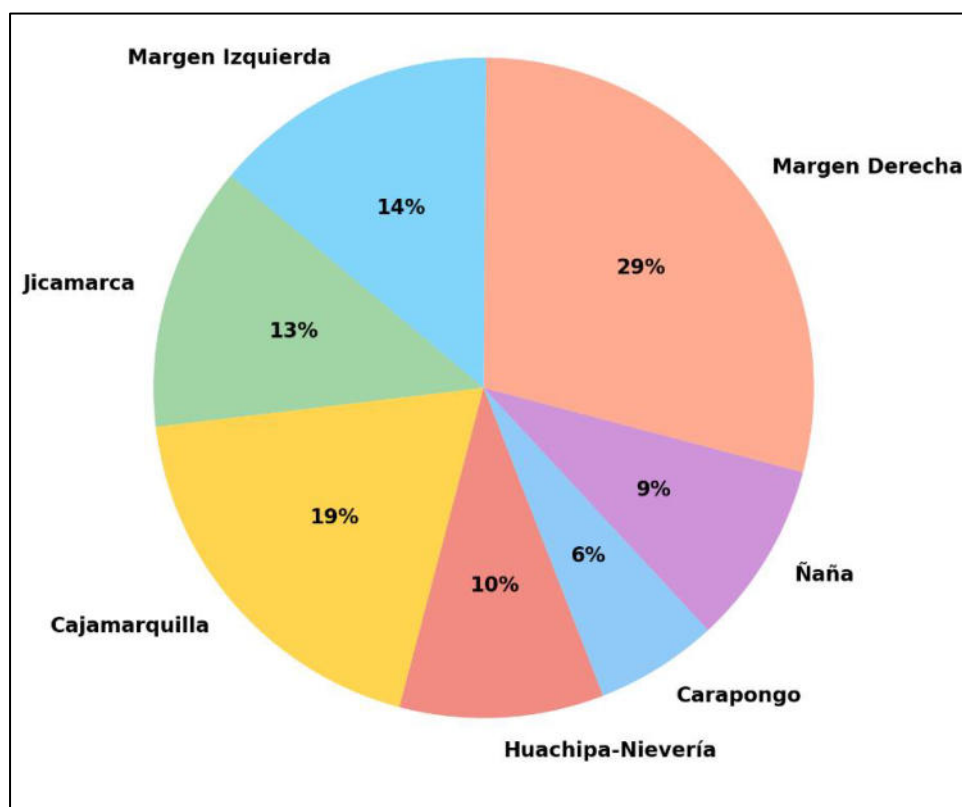
De acuerdo con el estudio *Perfiles Zonales 2016* (Ipsos Perú, 2016), la población distrital de 225,026 habitantes se distribuye de manera heterogénea en seis zonas: Huachipa-Nievería, Jicamarca, Cajamarquilla, Carapongo, Ñaña y las márgenes derecha e izquierda del río Rímac. El análisis porcentual muestra que la zona con mayor densidad es la Margen Derecha (29 %), seguida por Cajamarquilla (19 %), mientras que Jicamarca (13 %), Margen Izquierda (14 %), Huachipa-Nievería (10 %), Ñaña (9 %) y Carapongo (6 %) representan proporciones menores, aunque relevantes para la planificación territorial y ambiental.

Esta distribución confirma la importancia de Cajamarquilla dentro del contexto distrital, no solo por su peso demográfico, sino también por su condición de zona estratégica con alta concentración de actividades industriales. En consecuencia, su selección como área de estudio se justifica tanto por criterios técnicos y geográficos, como por la necesidad de integrar la dimensión poblacional en el análisis ambiental focalizado (ver Tabla 8 y Figura 7).

Tabla 8*Población por zonas en Lurigancho-Chosica*

Zona	Población	%
Jicamarca	29 273	13 %
Cajamarquilla	43 324	19 %
Huachipa-Nievería	23 079	10 %
Carapongo	14 063	6 %
Naña	19 156	9 %
Margen Derecha	65 571	29 %
Margen Izquierda	30 561	14 %
Total	225 026	100 %

Fuente: Elaboración propia, adaptado de “Perfiles zonales 2016”, por Ipsos Perú, 2016

Figura 7*Distribución porcentual de la población por zonas del distrito de Lurigancho-Chosica*

Fuente: Elaboración propia con información de Ipsos Perú, 2016

3.2.2. *La muestra*

La muestra está constituida por los registros horarios comprendidos en el periodo 2022–2024, seleccionado por su relevancia analítica al abarcar tres años consecutivos con comportamiento meteorológico contrastado e incluir episodios críticos de contaminación atmosférica asociados a condiciones de inversión térmica. Este intervalo temporal representa además una serie continua, reciente y representativa, idónea para el entrenamiento, validación y evaluación de modelos predictivos.

La selección es de tipo no probabilística, en tanto se trabaja con una base de datos determinada por la disponibilidad técnica de los registros generados por el sistema de monitoreo del OEFA y por el repositorio oficial de *Copernicus Climate Data Store* para el caso de la variable BLH. De acuerdo con Sampieri, Collado y Lucio (2014, p. 146), “en investigaciones donde se analizan bases de datos estructuradas o fenómenos específicos, se justifica el uso de muestras no probabilísticas por criterios de accesibilidad y delimitación operativa”.

La calidad, continuidad y granularidad horaria de esta muestra garantizan la conformación de conjuntos de datos robustos para la implementación del modelo Random Forest (RF), lo que asegura tanto la validez metodológica del análisis como su pertinencia práctica en el marco de la fiscalización ambiental preventiva.

En síntesis, la definición de la población y la delimitación de la muestra aseguran una base empírica sólida para el análisis estadístico y predictivo del estudio. La disponibilidad de registros horarios confiables y continuos, complementados con variables meteorológicas de alta resolución, permite estructurar conjuntos de datos consistentes y adecuados para la validación de hipótesis y el entrenamiento del modelo Random Forest. Con esta base, en el siguiente apartado se presenta la operacionalización de variables, en la cual se integran los

problemas, objetivos e hipótesis mediante la matriz de consistencia y la matriz de operacionalización, garantizando la coherencia interna del diseño metodológico.

3.3. Operacionalización de Variables

3.3.1. *Matriz de consistencia*

Para este estudio se elaborará una matriz de consistencia que articula de manera lógica y estructurada los componentes centrales de la investigación: el problema general, los objetivos específicos, las hipótesis, las variables (independientes y dependientes), así como sus respectivos indicadores y fuentes de verificación. Esta herramienta metodológica resulta fundamental para garantizar la coherencia interna del diseño y asegurar que todas las partes del estudio se orienten de manera integrada a la resolución del problema identificado.

De acuerdo con Bisquerra (2004, p. 105), “la matriz de consistencia representa un esquema lógico que permite establecer la coherencia interna de una investigación, relacionando el problema, los objetivos, las hipótesis, las variables y los instrumentos de medición, con el fin de cumplir con los estándares de calidad científica”.

En este marco, la matriz incorpora además un componente innovador vinculado al modelamiento predictivo, específicamente al objetivo específico N.º 5 y a la hipótesis H5, orientados a la implementación de un modelo de aprendizaje automático Random Forest (RF). Para ello, se incluye como variable dependiente adicional la precisión del modelo predictivo, medida mediante indicadores estadísticos como el coeficiente de determinación (R^2), el error cuadrático medio (RMSE) y el error absoluto medio (MAE).

De esta manera, la matriz de consistencia no solo cumple una función lógica y organizativa, sino que también refleja la integración metodológica de técnicas predictivas con aplicaciones prácticas en la fiscalización ambiental minera, aportando evidencia científica para la gestión de episodios críticos de contaminación atmosférica.

En este contexto, la matriz de consistencia constituye el eje articulador entre los problemas planteados, los objetivos propuestos y las hipótesis formuladas, asegurando la alineación entre la fase conceptual y la operativa de la investigación. A continuación, se presenta la **Tabla 9 – Matriz de Consistencia**, en la cual se sintetizan de manera esquemática estas relaciones, incorporando tanto las variables de análisis como los indicadores y fuentes de verificación correspondientes.

Tabla 9

Matriz de Consistencia

Problemas	Objetivos	Hipótesis	Variables	Metodología
Problema General: ¿Cómo influye la inversión térmica en las concentraciones horarias de dióxido de azufre (SO ₂) registradas por las estaciones automáticas del OEFA en el área de influencia directa de la Refinería de Zinc de Cajamarquilla, durante el periodo 2022–2024?	Objetivo General: Evaluar la influencia de la inversión térmica en las concentraciones horarias de SO ₂ registradas por las estaciones automáticas del OEFA en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024, con el fin de optimizar los criterios técnicos de fiscalización ambiental minera en contextos de alta recurrencia de eventos críticos de contaminación.	Hipótesis General: La ocurrencia de eventos de inversión térmica influye de manera significativa en el incremento de las concentraciones horarias de SO ₂ registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.	Variable dependiente: $Y =$ Concentración horaria de SO ₂ Variable independiente: $X =$ Inversión térmica (identificada mediante condiciones meteorológicas: temperatura, presión, viento, BLH)	• Tipo de estudio: cuantitativo no experimental, diseño correlacional explicativo. • Población: Estaciones de monitoreo de calidad del aire del OEFA en tiempo real. • Muestra: Datos horarios de SO₂ y variables meteorológicas tomadas del 2022 al 2024. • Técnicas de análisis: Estadística descriptiva, análisis de frecuencias, cruces temporales y análisis de correlación entre variables
PE1: ¿Con qué frecuencia las concentraciones horarias de SO ₂ superan el umbral de 250 µg/m ³ y cómo se comparan estas superaciones con el promedio diario del ECA Aire de 24 horas durante el periodo	OE1: Cuantificar las superaciones horarias y diarias del umbral de 250 µg/m ³ de SO ₂ registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.	HE1: La frecuencia de superaciones horarias del umbral de 250 µg/m ³ de SO ₂ es significativamente mayor en días caracterizados por condiciones meteorológicas compatibles con inversión térmica.		

Problemas	Objetivos	Hipótesis	Variables	Metodología
2022–2024?				meteorológicas y concentraciones horarias de SO₂. • Modelamiento con Random Forest validado con métricas RMSE, MAE y R².
PE2: ¿Cuál es la relación estadística entre las concentraciones horarias de SO ₂ y las variables meteorológicas, considerando la diferencia espacial entre estaciones de barlovento y sotavento?	OE2: Analizar la relación estadística entre las concentraciones horarias de SO ₂ y las variables meteorológicas (BLH, temperatura, humedad relativa, presión atmosférica, velocidad y dirección del viento), incorporando la diferencia espacial entre estaciones de barlovento y sotavento.	HE2: Existe una relación estadística significativa entre las concentraciones horarias de SO ₂ y las variables meteorológicas locales, manifestándose en niveles más elevados en la estación sotavento respecto a la estación barlovento durante episodios de inversión térmica.		
PE3: ¿De qué manera las condiciones de inversión térmica influyen en la acumulación y variación horaria de SO ₂ en la zona de estudio?	OE3: Evaluar la influencia de las condiciones de inversión térmica en la acumulación y variabilidad horaria de las concentraciones de SO ₂ en el área de estudio.	HE3: Durante los eventos de inversión térmica se identifican patrones horarios recurrentes de alta concentración de SO ₂ —especialmente en las primeras horas del día— que confirman la influencia directa de este fenómeno en la acumulación de contaminantes.		
PE4: ¿En qué medida las concentraciones horarias de SO ₂ superan el valor límite del ECA Aire de 24 horas y cuáles son las limitaciones de este estándar para activar	OE4: Comparar las concentraciones horarias de SO ₂ registradas en el periodo 2022–2024 con los estándares normativos vigentes (ECA Aire de 24h y protocolos	HE4a: Las concentraciones horarias de SO ₂ registradas en el área de estudio durante 2022–2024 presentan superaciones críticas (443 en 2022, 346 en 2023 y 413 en 2024) que no siempre se reflejan en los valores normativos		

Problemas	Objetivos	Hipótesis	Variables	Metodología
mecanismos de alerta temprana?	nacionales), e identificar las limitaciones de dichos instrumentos para la activación de mecanismos de alerta temprana en la fiscalización ambiental minera.	establecidos en los promedios de 24 horas del ECA Aire. HE4b: El análisis conjunto de las concentraciones horarias de SO ₂ y de variables meteorológicas permite establecer criterios técnicos complementarios que refuercen los mecanismos de alerta temprana y la fiscalización ambiental del OEFA.		
PE5: ¿En qué medida el modelo Random Forest permite estimar con precisión las concentraciones horarias de SO ₂ integrando variables meteorológicas locales?	OE5: Desarrollar y validar un modelo predictivo Random Forest para estimar concentraciones horarias de SO ₂ , integrando variables meteorológicas locales (BLH, temperatura, humedad relativa, presión atmosférica y velocidad y dirección del viento), a fin de fortalecer la fiscalización ambiental minera.	HE5: El modelo de aprendizaje automático Random Forest predice con alta precisión las concentraciones horarias de SO ₂ al integrar variables meteorológicas locales, mostrando un desempeño superior a los métodos tradicionales de estimación.		

Fuente: Elaboración propia

3.3.2. *Matriz de Operacionalización de Variables*

La operacionalización de variables constituye una etapa clave del diseño metodológico, pues permite traducir los conceptos teóricos en dimensiones observables y medibles, alineadas con los objetivos e hipótesis de investigación (Bisquerra, 2004; Hernández et al., 2014). En el presente estudio se han definido una variable dependiente con tres subdimensiones, una variable independiente y una dimensión de interacción entre ambas, cada una con indicadores cuantificables que serán evaluados mediante herramientas estadísticas apropiadas y modelos

predictivos de aprendizaje automático.

3.3.2.1. Variable dependiente (VD1): Concentración horaria de dióxido de azufre (SO₂). El dióxido de azufre (SO₂) es un contaminante atmosférico generado principalmente por procesos industriales metalúrgicos, cuya acumulación en la atmósfera puede ocasionar efectos negativos en la salud respiratoria y en la calidad ambiental (Manahan, 2007; Seinfeld & Pandis, 2016). En esta investigación, representa el nivel de contaminación atmosférica registrado en intervalos horarios en la zona de influencia directa de la Refinería de Zinc de Cajamarquilla. Esta variable se desagrega en dos dimensiones:

Y1 = Comportamiento temporal de la concentración horaria de SO₂

Indicadores:

I1: Concentración promedio horaria y diaria de SO₂ (µg/m³).

Se calcularán los promedios horarios y diarios de las concentraciones registradas por las estaciones automáticas de monitoreo del OEFA, lo que permitirá identificar variaciones temporales y patrones característicos en la zona.

I2: Número de superaciones de las concentraciones horarias de SO₂ (>250 µg/m³).

Se cuantificarán los registros horarios que superan el umbral de referencia de 250 µg/m³, utilizado como valor técnico para identificar episodios críticos de contaminación (OEFA, 2022; 2023; 2024).

Y2 = Distribución espacial de la concentración horaria de SO₂

Indicadores:

I3: Diferencia de concentración promedio entre estación a barlovento y estación a sotavento.

El análisis permitirá establecer si las concentraciones son significativamente mayores a sotavento, lo cual evidenciaría un patrón de dispersión influido por condiciones meteorológicas locales (Arya, 1999).

3.3.2.2. Variable independiente (VI1): Condiciones de inversión térmica. La inversión térmica es un fenómeno atmosférico que ocurre cuando una capa de aire cálido se sitúa sobre aire frío en la superficie, limitando el ascenso vertical del aire y favoreciendo la acumulación de contaminantes (Stull, 1988; Seinfeld & Pandis, 2016).

X = Condiciones de inversión térmica

Indicadores (I):

I4: Altura de la capa límite planetaria (BLH, en metros), estimada a partir de datos de reanálisis ERA5 del Copernicus Climate Data Store y de variables meteorológicas locales.

I5: Número de horas con BLH < 300 m.

Este umbral permite identificar la duración de episodios de inversión térmica moderada a severa, asociados con mayor riesgo de acumulación de SO₂ (Arya, 1999).

3.3.2.3. Variable de interacción: Coincidencia entre inversión térmica y contaminación crítica. La combinación de condiciones de estabilidad atmosférica (BLH reducida) y emisiones industriales eleva el riesgo de acumulación crítica de contaminantes (Seinfeld & Pandis, 2016).

Z = Coincidencia temporal entre eventos de inversión y episodios críticos de contaminación

Indicador:

I6: Número de coincidencias entre BLH < 300 m y superaciones horarias de SO₂ (>250 µg/m³). Este cruce de datos permitirá evaluar correlaciones empíricas entre inversión térmica y picos de contaminación en el área de estudio.

3.3.2.3. Variable dependiente (VD3): Estimación de concentración horaria de SO₂ mediante modelamiento predictivo. El aprendizaje automático ofrece técnicas robustas para modelar fenómenos ambientales con alta variabilidad y no linealidad. En este caso, se empleará el algoritmo Random Forest (RF), ampliamente utilizado en estudios de predicción

de calidad del aire (Breiman, 2001; Zhang et al., 2018).

Y3 = Concentración horaria de SO₂ estimada por modelo Random Forest (RF)

Indicador:

I7: Precisión del modelo Random Forest en la predicción de concentraciones horarias de SO₂, evaluada mediante métricas estadísticas de desempeño:

- Error Cuadrático Medio (RMSE)
- Error Absoluto Medio (MAE)
- Coeficiente de Determinación (R²)

El modelo se entrenará con registros históricos de concentración horaria de SO₂ y variables meteorológicas (temperatura, humedad relativa, velocidad del viento y BLH). Se aplicarán técnicas de validación cruzada y análisis comparativo temporal entre valores observados y predichos, con el fin de verificar su aplicabilidad como herramienta de alerta temprana.

Es necesario precisar que la hipótesis específica 4 se presenta en dos dimensiones complementarias: HE4a, orientada a evidenciar que las superaciones horarias de SO₂ (443 en 2022, 346 en 2023 y 413 en 2024) no siempre se reflejan en los valores normativos establecidos en el promedio diario del ECA Aire, y HE4b, que plantea la necesidad de criterios técnicos adicionales basados en la integración de variables meteorológicas para reforzar los mecanismos de alerta temprana y la fiscalización ambiental. Esta diferenciación fortalece la coherencia metodológica de la investigación y se incorpora en la matriz de operacionalización de variables, manteniendo los indicadores previamente definidos.

En concordancia con la matriz de consistencia, la matriz de operacionalización de variables constituye una herramienta metodológica clave que permite detallar cómo se medirán y evaluarán las variables planteadas en la investigación. Este instrumento asegura la correspondencia entre las hipótesis y los indicadores seleccionados, especificando

dimensiones, técnicas e instrumentos de análisis, así como las fuentes de verificación. De esta manera, se garantiza que el proceso investigativo se desarrolle con un nivel adecuado de rigurosidad y trazabilidad, facilitando tanto la validación de los resultados como su aplicabilidad en la fiscalización ambiental.

A continuación, se presenta la **Tabla 10 – Matriz de Operacionalización de Variables**, en la cual se sintetiza la relación entre las hipótesis específicas, las variables de estudio, sus indicadores e instrumentos de medición.

Tabla 10

Matriz de operacionalización de variables

Variable	Dimensión	Indicador	Técnica / Instrumento	Fuente de Datos	Escala de Medición
Variable Dependiente (VD1): Concentración horaria de SO ₂	Y1: Comportamiento temporal	I1: Concentración promedio horaria y diaria de SO ₂ (µg/m ³)	Estadística descriptiva, análisis de frecuencias y cruces temporales	OEFA 2022-2024	µg/m ³
		I2: Número de superaciones de las concentraciones horarias de SO ₂ (>250 µg/m ³)			
Variable Dependiente - VD2 (Subdimensión): Distribución espacial de la concentración horaria de SO ₂	Y2: Distribución espacial	I3: Diferencia de concentración promedio entre estación a barlovento y estación a sotavento	Comparación de medias, análisis de dispersión espacial	OEFA 2022-2024	µg/m ³
Variable Independiente (VI1): Condiciones de Inversión térmica	X: Condiciones de inversión térmica	I4: Altura de la capa límite planetaria (BLH, en metros)	Análisis de series temporales, estadística descriptiva	Copernicus ERA5 / OEFA	Metros/Hora
		I5: Número de horas con BLH < 300 m			
Variable de Interacción de VI-VD	Z: Coincidencia inversión-contaminación	I6: N.º de coincidencias entre BLH < 300 m y superaciones de las concentraciones horarias de SO ₂ (>250 µg/m ³)	Cruces temporales, análisis de correlación	OEFA + Copernicus ERA5	Frecuencia de eventos

Variable	Dimensión	Indicador	Técnica / Instrumento	Fuente de Datos	Escala de Medición
Variable Dependiente (VD3): Modelamiento predictivo Random Forest RF	Y3: Concentración estimada por RF	I7: Precisión del modelo Random Forest en la predicción de concentraciones horarias de SO ₂ (RMSE, MAE, R ²)	Modelamiento Random Forest, validación cruzada temporal.	OEFA (2022-2024), variables meteorológicas	Métricas estadísticas
Nota: Los indicadores I2 (superaciones horarias y diarias) y I6 (coincidencia entre episodios de baja BLH y concentraciones críticas de SO ₂) se vinculan específicamente con las hipótesis HE4a y HE4b , respectivamente, reflejando la doble dimensión normativa y técnico-operativa de la fiscalización ambiental.					

Fuente: Elaboración propia

En conclusión, la matriz de operacionalización de variables garantiza la coherencia entre los objetivos, las hipótesis y los indicadores definidos en la investigación. Este esquema metodológico no solo permite estructurar de manera rigurosa el análisis estadístico y predictivo, sino que también asegura que los resultados se vinculen directamente con los criterios técnicos requeridos para fortalecer la fiscalización ambiental. Con esta base, en el siguiente apartado se describen los instrumentos empleados para la recolección, procesamiento y análisis de los datos, a fin de sustentar la validez y confiabilidad del estudio.

3.4. Instrumentos

Para garantizar la validez y confiabilidad de la investigación, resulta indispensable definir los instrumentos utilizados en la recolección, procesamiento y análisis de los datos. Estos instrumentos permiten medir de manera precisa las variables dependientes e independientes planteadas en la matriz de consistencia y de operacionalización, asegurando su correspondencia con los indicadores definidos. Asimismo, su selección responde a criterios de pertinencia científica, disponibilidad técnica y coherencia con los objetivos e hipótesis de estudio.

En ese sentido, se han organizado los instrumentos en cuatro categorías: aquellos destinados a la medición de la variable dependiente, los aplicados a la variable independiente, los instrumentos complementarios que refuerzan la calidad de la información, y las

herramientas computacionales que sustentan el modelamiento predictivo basado en Random Forest.

3.4.1. Instrumentos para la variable dependiente

La medición de la variable dependiente —concentración horaria de dióxido de azufre (SO_2)— se realiza mediante equipos automáticos de monitoreo continuo instalados en las estaciones a sotavento y barlovento en el área de influencia directa de la Refinería de Zinc de Cajamarquilla. Estos analizadores de gases están debidamente calibrados y certificados, conforme a los estándares técnicos establecidos por el Organismo de Evaluación y Fiscalización Ambiental (OEFA) y las normas internacionales de calidad del aire, que incluyen registros validados bajo los lineamientos del Protocolo Nacional de Monitoreo de Calidad del Aire (MINAM, 2018, p. 30).

A continuación, se presenta el detalle de los equipos utilizados para el monitoreo de SO_2 en cada estación:

Tabla 11

Equipo de SO_2 de utilizado en la estación a barlovento CA-RZC-01 (Los Andes)

Parámetro	Equipo	Marca	Modelo	Serie	Certificado de calibración
Dióxido de azufre (SO_2)	Analizador continuo de gases	Thermo Scientific	43iQ-BAN	12123713085	LF-449-2023

Fuente: Elaboración propia con información del OEFA, 2024.

Tabla 12

Equipo de SO_2 utilizado en la estación a sotavento CA-RZC-02 (Radio Observatorio de Jicamarca - ROJ)

Parámetro	Equipo	Marca	Modelo	Serie	Certificado de calibración
Dióxido de azufre (SO_2)	Analizador continuo de gases	Thermo Scientific	43iQ-BAN	12123713083	LF-451-2023

Fuente: Elaboración propia con información del OEFA, 2024.

La utilización de equipos automáticos certificados y calibrados garantiza la confiabilidad de los registros horarios de SO₂, asegurando su validez como insumo principal para el análisis estadístico, normativo y predictivo del estudio. Estos datos constituyen la base empírica fundamental para contrastar las hipótesis planteadas y evaluar la influencia de la inversión térmica en la acumulación de contaminantes atmosféricos.

3.4.2. *Instrumentos para la variable independiente*

Para la caracterización de la inversión térmica como variable independiente, se emplearán los datos meteorológicos obtenidos a través de estaciones automáticas del Organismo de Evaluación y Fiscalización Ambiental (OEFA), las cuales registran parámetros clave para el análisis de las condiciones atmosféricas locales. Estos datos permitirán identificar episodios de estabilidad atmosférica asociados a inversión térmica, a través del comportamiento de variables como la temperatura, humedad, presión barométrica, velocidad del viento y precipitación.

Con el fin de asegurar la trazabilidad técnica de los registros meteorológicos empleados en esta investigación, a continuación, se detallan los equipos instalados en las estaciones automáticas de monitoreo ambiental a barlovento (CA-RZC-01, Los Andes) y sotavento (CA-RZC-02, ROJ), así como los métodos de análisis aplicados a cada uno de los parámetros evaluados. Esta información garantiza la confiabilidad de las mediciones y su alineación con estándares nacionales e internacionales de monitoreo de calidad del aire.

Tabla 13*Equipos meteorológicos utilizados en la estación a barlovento CA-RZC-01 (Los Andes)*

Parámetro	Equipo	Marca	Modelo	Serie	Certificado de calibración
Temperatura (termohigrómetro)	Meteorología	<i>Campbell Scientific</i>	HydroVUE10	E2968	LM-10182023
Humedad (termohigrómetro)	Meteorología	<i>Campbell Scientific</i>	HydroVUE10	E2968	LM-10192023
Anemómetro	Meteorología	<i>Gill Instrument</i>	GMX200	21340004	LM-10202023
Presión barométrica	Meteorología	<i>YOUNG</i>	61402L	BPA11711	LM-10212023
Precipitación	Meteorología	<i>YOUNG</i>	52203	TB16963	LM-10222023
Radiación solar	Meteorología	<i>KIPP & ZONEN</i>	CMP6	214263	LM-10242023

Fuente: OEFA (2024)

Tabla 14*Equipos meteorológicos utilizados en la estación a sotavento CA- RZC-02 (ROJ)*

Parámetro	Equipo	Marca	Modelo	Serie	Certificado de calibración
Temperatura (termohigrómetro)	Meteorología	<i>Campbell Scientific</i>	HydroVUE10	E2966	LM-10242023
Humedad (termohigrómetro)	Meteorología	<i>Campbell Scientific</i>	HydroVUE10	E2966	LM-10252023
Anemómetro	Meteorología	<i>Gill Instrument</i>	GMX200	21340003	LM-10262023
Presión barométrica	Meteorología	<i>YOUNG</i>	61402L	BPA11713	LM-10272023
Precipitación	Meteorología	<i>YOUNG</i>	52203	TB16964	LM-10282023
Radiación solar	Meteorología	<i>KIPP & ZONEN</i>	CMP6	214221	LM-10292023

Fuente: OEFA (2024)

Tabla 15*Métodos de análisis de los parámetros evaluados*

Parámetro	Método	Técnica Empleada
Dióxido de azufre (SO ₂)	Norma Técnica Peruana: NTP-ISO 10498:2017 y NTP-ISO 10498:2017/COR1:2017 (Método automático)	Fluorescencia ultravioleta
Velocidad de viento Dirección de viento Humedad relativa Temperatura ambiente Precipitación Presión barométrica	Método automático	No aplica

Fuente: OEFA (2024)

La disponibilidad de registros meteorológicos provenientes de equipos calibrados y certificados asegura la confiabilidad de los datos empleados para la caracterización de las condiciones de inversión térmica. Estos instrumentos permiten vincular de manera precisa la dinámica atmosférica con los niveles de dióxido de azufre (SO₂), fortaleciendo el análisis causal y predictivo de la investigación. Asimismo, la integración de estos parámetros en el modelo metodológico contribuye a garantizar un enfoque multidimensional y riguroso en la evaluación de la calidad del aire en entornos industriales.

3.4.3. Instrumentos Complementarios

Para complementar el análisis estadístico y fortalecer la interpretación de los resultados, se emplearán modelos estadísticos y matemáticos orientados a representar la dinámica de dispersión de contaminantes atmosféricos bajo condiciones de estabilidad atmosférica como la inversión térmica.

Uno de los modelos teóricos que respaldará el análisis es el modelo de penacho gaussiano (Gaussian Plume), ampliamente utilizado en estudios de calidad del aire para describir el comportamiento de contaminantes emitidos desde fuentes puntuales en función de

variables meteorológicas y parámetros físicos de emisión (Arya, 1999, p. 152). Este modelo permite estimar concentraciones esperadas de contaminantes como el SO_2 a diferentes distancias de la fuente emisora, bajo condiciones de viento, estabilidad atmosférica y altura efectiva de emisión.

Además de su utilidad descriptiva, este tipo de modelamiento servirá como base conceptual para interpretar el comportamiento de las concentraciones horarias de SO_2 en las estaciones ubicadas a barlovento y sotavento, en función de los patrones de inversión térmica observados.

En complemento, se utilizarán herramientas estadísticas como:

- **Análisis de correlación** para establecer relaciones entre las variables meteorológicas (altura de capa límite atmosférica, temperatura, viento) y las concentraciones horarias de SO_2 .
- **Pruebas de hipótesis no paramétricas** (U de Mann-Whitney) para comparar diferencias de concentración entre estaciones de monitoreo.
- **Series temporales y análisis de frecuencias** para caracterizar la distribución horaria y diaria de las concentraciones de SO_2 .

Estas herramientas permitirán una triangulación metodológica robusta entre los datos observados, las condiciones meteorológicas locales y los marcos teóricos de dispersión atmosférica, generando así una interpretación integral orientada a la toma de decisiones en fiscalización ambiental.

En conjunto, estos instrumentos complementarios permiten integrar enfoques teóricos y estadísticos que refuerzan la interpretación de los resultados, asegurando un análisis multidimensional del fenómeno estudiado. Esta triangulación metodológica sienta las bases para la incorporación de herramientas computacionales avanzadas, como el modelamiento predictivo mediante algoritmos de aprendizaje automático, que se desarrollará en el siguiente

apartado.

3.4.4. *Herramientas computacionales de modelamiento predictivo*

Además del análisis estadístico clásico y del marco conceptual del modelo de dispersión Gaussian Plume, la presente investigación incorporará una técnica de aprendizaje automático supervisado basada en el algoritmo Random Forest (RF). Este modelo permite construir predicciones robustas de concentraciones horarias de SO₂ en función de variables meteorológicas como:

- Temperatura
- Humedad
- Velocidad y dirección del viento
- Presión barométrica
- Altura de la capa límite atmosférica (BLH)

El modelo Random Forest consiste en un conjunto de árboles de decisión generados a partir de subconjuntos aleatorios del conjunto de datos, donde el resultado final se obtiene a partir del promedio (para regresión) o votación mayoritaria (para clasificación). Su capacidad para manejar relaciones no lineales y su resistencia al sobreajuste lo convierten en una herramienta poderosa para aplicaciones ambientales con múltiples variables explicativas y ruido en los datos (Breiman, 2001).

La implementación del algoritmo Random Forest se realizará mediante herramientas computacionales especializadas como **Python (scikit-learn, pandas, matplotlib)** y **R (randomForest, caret)**, que permiten entrenar, validar y evaluar el desempeño del modelo con métricas estadísticas como el coeficiente de determinación (R^2), el error cuadrático medio (RMSE) y el error absoluto medio (MAE). De este modo, el enfoque predictivo complementa el análisis estadístico y teórico, aportando una herramienta innovadora y replicable para la fiscalización ambiental minera.

3.5. Procedimientos

El desarrollo metodológico de la presente investigación se estructurará en cinco etapas secuenciales, diseñadas para garantizar la calidad del análisis, la trazabilidad en el procesamiento de datos y la validez del modelo aplicado. Estas etapas integran el tratamiento estadístico, los fundamentos físico-matemáticos de la dispersión de contaminantes atmosféricos y la implementación de técnicas de aprendizaje automático, de manera que los resultados obtenidos tengan rigor científico y aplicabilidad práctica en la fiscalización ambiental.

3.5.1. *Etapas 1: Recopilación y Preparación de Datos*

En esta primera fase se procederá a la recopilación de registros horarios de dióxido de azufre (SO₂) y de variables meteorológicas (temperatura, humedad relativa, velocidad y dirección del viento, presión barométrica y altura de la capa límite atmosférica – BLH) provenientes de las estaciones automáticas ubicadas a barlovento y sotavento de la Refinería de Zinc de Cajamarquilla.

Los datos serán sometidos a procesos de depuración y control de calidad mediante procedimientos de validación interna, eliminando duplicaciones, registros nulos o inconsistentes, así como valores atípicos que no correspondan a condiciones reales de operación. Se aplicarán criterios de control de calidad de datos conforme a lo propuesto por Hernández et al. (2014, p. 150), complementados con protocolos internacionales de aseguramiento de calidad en monitoreo atmosférico (EPA, 2016).

3.5.2. *Etapas 2: Identificación y caracterización de fuentes de error*

Esta fase estará orientada a la detección, análisis y caracterización de posibles fuentes de error asociadas al proceso de medición, registro y calibración de los equipos utilizados. Se distinguirán principalmente:

- **Errores sistemáticos**, derivados de desajustes instrumentales, deriva de sensores,

problemas de calibración o condiciones de instalación que puedan generar sesgos constantes en los registros.

- **Errores aleatorios**, vinculados a la variabilidad natural de la atmósfera, interferencias ambientales no controladas o limitaciones en la resolución temporal de los equipos.

Para su control, se aplicarán criterios de validación y aseguramiento de calidad de datos conforme a lo establecido en el Protocolo Nacional de Monitoreo de Calidad del Aire (MINAM, 2018, p. 33), así como las recomendaciones metodológicas de organismos internacionales como la Environmental Protection Agency (EPA, 2016) y la European Environment Agency (EEA, 2018).

El propósito de esta etapa es garantizar la trazabilidad, fiabilidad y precisión de los registros empleados, de modo que los datos constituyan un insumo sólido y verificable para el modelamiento estadístico y predictivo que se desarrollará en las siguientes fases.

3.5.3. Etapa 3: Diseño e Implementación del Modelo Estadístico y Predictivo basado en Random Forest

En esta etapa se desarrollará un modelo de estimación y predicción de las concentraciones horarias de dióxido de azufre (SO₂), integrando enfoques estadísticos y técnicas avanzadas de aprendizaje automático. El propósito es construir una herramienta robusta de análisis y predicción, capaz de capturar relaciones complejas entre las variables meteorológicas y las concentraciones de SO₂, bajo condiciones de inversión térmica.

3.5.3.1. Modelo estadístico estocástico de caja negra. Como punto de partida, se empleará un modelo estadístico de tipo estocástico de caja negra, el cual representa un sistema donde la estructura interna no se modela explícitamente, sino que se analiza la relación entre variables de entrada (factores meteorológicos) y salida (concentraciones horarias de SO₂). Este modelo utilizará técnicas de análisis de series temporales multivariadas y regresiones,

combinadas con fundamentos del modelo de dispersión Gaussian Plume, ajustado a las condiciones locales (Arya, 1999, p. 158).

Esta estrategia permitirá explorar patrones temporales y realizar simulaciones con base en comportamientos históricos, a fin de identificar momentos críticos de acumulación de contaminantes en relación con variables como la altura de la capa límite atmosférica, temperatura, presión y viento.

3.5.3.2. Modelo de aprendizaje automático basado en Random Forest. De manera complementaria, se implementará un modelo predictivo utilizando la técnica de Random Forest (RF), un algoritmo de aprendizaje supervisado que construye un conjunto de árboles de decisión sobre subconjuntos aleatorios del conjunto de datos y luego promedia los resultados para mejorar la precisión y reducir el riesgo de sobreajuste (Breiman, 2001).

El procedimiento incluirá las siguientes fases:

A. Preparación del conjunto de datos. Selección de registros depurados, normalización de variables y división en subconjuntos de entrenamiento (80%) y prueba (20%).

B. Entrenamiento del modelo RF. Ajuste de hiperparámetros (número de árboles, profundidad máxima, número mínimo de muestras por hoja) mediante validación cruzada.

C. Evaluación del desempeño. Se aplicarán métricas de validación como el Error Cuadrático Medio (RMSE), el Error Absoluto Medio (MAE) y el Coeficiente de Determinación (R^2) para cuantificar el grado de ajuste del modelo.

D. Análisis de importancia de variables. Se determinará el peso relativo de cada variable meteorológica en la predicción de las concentraciones de SO_2 , fortaleciendo la interpretación científica del modelo.

E. Simulación de escenarios de alerta temprana. Se utilizarán datos pronosticados o en tiempo real para generar estimaciones predictivas de las concentraciones de SO_2 , lo que permitirá diseñar un esquema de alerta temprana útil para la fiscalización ambiental.

Este modelo predictivo, una vez validado, representará una herramienta innovadora para la vigilancia ambiental proactiva, facilitando intervenciones oportunas en zonas de riesgo.

La implementación conjunta del modelo estadístico estocástico y del modelo predictivo Random Forest permitirá no solo identificar patrones históricos de acumulación de contaminantes, sino también anticipar escenarios críticos de contaminación atmosférica bajo condiciones de inversión térmica. De este modo, la investigación trasciende el análisis descriptivo, aportando un marco metodológico robusto para la generación de alertas tempranas y la toma de decisiones estratégicas en la fiscalización ambiental minera.

3.5.4. Etapa 4: Validación y evaluación del modelo

La validación del modelo se realizará mediante el contraste entre las concentraciones estimadas y los datos reales observados en las estaciones de monitoreo. Para tal fin, se aplicarán métricas estadísticas de desempeño como:

- **Error Cuadrático Medio (ECM o RMSE):** mide la magnitud promedio del error entre valores observados y predichos.
- **Error Absoluto Medio (MAE):** complementa al RMSE al evaluar la magnitud de los errores sin ponderar al cuadrado las desviaciones.
- **Coefficiente de Determinación (R^2):** estima el grado de ajuste del modelo a los datos observacionales, indicando la proporción de la variabilidad explicada por las variables predictoras.

De acuerdo con Seinfeld y Pandis (2016, pp. 140–141):

La validación de un modelo matemático consiste en comparar las predicciones generadas por el modelo con datos observacionales reales. Este proceso permite identificar las desviaciones, ajustar los parámetros del modelo y evaluar su capacidad predictiva. Un modelo validado proporciona una herramienta robusta para la toma de decisiones y la formulación de políticas basadas en evidencia científica.

Los resultados de esta validación permitirán determinar la aplicabilidad práctica del modelo en escenarios de fiscalización ambiental, facilitando la identificación temprana de episodios críticos de contaminación por SO₂ y fortaleciendo los criterios técnicos del monitoreo regulador en contextos de alta vulnerabilidad atmosférica.

3.5.5. *Etapas 5: Procedimiento para la detección, corrección y relleno de datos*

La calidad de los datos utilizados en el modelamiento predictivo constituye un aspecto crítico para garantizar la confiabilidad y validez de los resultados. Por ello, se implementó una etapa específica destinada a la detección, corrección y tratamiento de valores faltantes, atípicos o anómalos en los registros horarios de concentración de dióxido de azufre (SO₂) y variables meteorológicas obtenidos por las estaciones automáticas del OEFA.

3.5.5.1. Detección de valores faltantes y atípicos. Se identificaron datos faltantes mediante el conteo de celdas vacías o marcadas con indicadores de error (ej. “N/A”, “-999”, “NaN”). Adicionalmente, se aplicaron procedimientos automáticos para la detección de valores atípicos utilizando el método de Tukey (basado en el rango intercuartílico), y análisis gráfico mediante boxplots para identificar registros fuera del rango esperable, especialmente en variables meteorológicas como temperatura, presión atmosférica y altura de capa límite (BLH).

En el caso del SO₂, se utilizó un umbral de 1,000 µg/m³ únicamente como criterio exploratorio para señalar registros que requerían revisión detallada. En la estación sotavento se detectaron 19 registros superiores a este valor, alcanzando un máximo de 1,535.79 µg/m³. Tras su contraste con las condiciones meteorológicas concurrentes (altura de la capa límite, velocidad del viento y estabilidad atmosférica), estos registros fueron validados como episodios reales de contaminación, por lo que fueron preservados como parte integral de la base de datos analizada.

3.5.5.2. Criterios para la corrección de valores anómalos. Una vez detectados los valores anómalos o inconsistentes, se aplicaron los siguientes criterios:

- **Verificación cruzada** con variables correlacionadas (ej. BLH y concentración de SO_2) para descartar errores por falla de sensores.
- **Validación temporal:** Se contrastaron registros con datos horarios anteriores y posteriores (lag ± 3 horas) para identificar errores de transmisión.
- Se descartaron registros que presentaban inconsistencias múltiples y simultáneas en más de una variable.

3.5.5.3. Relleno de datos faltantes. Para preservar la continuidad de las series temporales, especialmente en el entrenamiento del modelo Random Forest, se aplicaron técnicas de imputación de datos faltantes, considerando:

- **Interpolación lineal simple:** Para brechas ≤ 3 horas.
- **Imputación basada en medias horarias móviles:** Cuando los datos faltantes correspondían a períodos homogéneos dentro de una ventana de ± 12 horas.
- **Modelos de regresión múltiples o k-NN imputer:** En casos donde el relleno requería estimaciones multivariadas, se utilizó la imputación por proximidad entre variables correlacionadas (como BLH, temperatura y presión atmosférica) para inferir el valor más probable.

Todas las imputaciones fueron documentadas, conservando una copia de los datos originales.

3.5.5.4. Verificación posterior al tratamiento. Una vez completado el proceso de corrección y relleno, se realizó una nueva revisión estadística y gráfica de los datos para garantizar:

- La **coherencia interna** de los registros horarios.
- La **consistencia entre variables dependientes e independientes.**
- La ausencia de artefactos numéricos generados por interpolaciones excesivas.

Este procedimiento se desarrolló en concordancia con los lineamientos de aseguramiento de calidad de datos establecidos en el Protocolo Nacional de Monitoreo de Calidad del Aire (MINAM, 2018) y las guías internacionales de EPA (2016) y EEA (2018).

En síntesis, la aplicación de procedimientos rigurosos de detección, corrección y validación de datos aseguró que la base de información final sea consistente, continua y estadísticamente confiable. La preservación de los registros extremos validados permitió incorporar en el análisis episodios críticos de contaminación, esenciales para caracterizar con precisión la dinámica del SO_2 en condiciones de inversión térmica. De este modo, el conjunto de datos resultante constituye un insumo sólido para la implementación del modelo de aprendizaje automático Random Forest, garantizando que sus predicciones reflejen de manera fiel la complejidad y variabilidad del fenómeno estudiado.

3.6. Análisis de Datos

El análisis de datos se realizará de forma secuencial y rigurosa, integrando técnicas estadísticas convencionales y herramientas de modelamiento predictivo. Esta etapa se apoya en la base de datos previamente depurada y validada (ver 3.5), y busca establecer relaciones entre las variables meteorológicas y la concentración horaria de dióxido de azufre (SO_2), así como validar el desempeño del modelo predictivo basado en Random Forest.

3.6.1. Análisis estadístico descriptivo

Se calcularán medidas de tendencia central (media, mediana, moda), de dispersión (desviación estándar, rango intercuartílico) y de frecuencia (histogramas, distribuciones por intervalos) para caracterizar el comportamiento de las variables principales: concentración horaria de SO_2 , altura de la capa límite planetaria (BLH), temperatura, presión atmosférica y viento. Este análisis permitirá identificar patrones típicos, valores atípicos y la estructura temporal de los datos.

“El análisis descriptivo es el primer paso en la exploración de datos cuantitativos, facilitando la comprensión de su comportamiento antes de aplicar modelos más complejos” (Bisquerra, 2004, p. 120).

3.6.2. Identificación y Clasificación de Fuentes de Error

Se procederá a la depuración de la base de datos, identificando errores de medición, registros atípicos, datos faltantes y posibles inconsistencias entre estaciones. Estos errores se clasificarán por su naturaleza (sistemáticos, aleatorios, instrumentales) y su magnitud, aplicando criterios establecidos en el Protocolo Nacional de Monitoreo de Calidad del Aire (MINAM, 2018, p. 34). En caso de vacíos críticos, se aplicarán técnicas de interpolación o exclusión justificada.

3.6.3. Aplicación del Modelo Matemático y Predictivo

Se desarrollarán dos enfoques complementarios:

- **Modelo estadístico de relación:** se evaluará la correlación entre variables meteorológicas y la concentración horaria de SO₂ mediante análisis bivariado (coeficientes de Pearson/Spearman) y regresión múltiple.
- **Modelo de aprendizaje automático Random Forest (RF):** se implementará un modelo supervisado para predecir la concentración horaria de SO₂, utilizando como variables explicativas la BLH, temperatura, viento y presión atmosférica. El algoritmo será entrenado con un subconjunto de datos históricos y validado con datos de prueba, siguiendo un esquema de validación cruzada k-fold.

“Los modelos de dispersión como Gaussian Plume son útiles bajo supuestos ideales; sin embargo, en escenarios complejos, los modelos no paramétricos como Random Forest permiten capturar relaciones no lineales y mejorar la precisión predictiva” (Arya, 1999, p. 160).

3.6.4. Evaluación del Modelo Predictivo

Se verificará la capacidad predictiva del modelo Random Forest mediante métricas

cuantitativas de desempeño:

- Coeficiente de Determinación (R^2)
- Error Cuadrático Medio (RMSE)
- Error Absoluto Medio (MAE)
- Curvas de dispersión de valores predichos vs observados

El modelo será considerado robusto si supera umbrales mínimos de $R^2 > 0.70$ y errores dentro de márgenes aceptables. Adicionalmente, se analizará la importancia relativa de cada variable meteorológica en la predicción de SO_2 , permitiendo mejorar la interpretación del fenómeno.

“Los modelos predictivos deben ser validados con datos reales, de forma que su capacidad de generalización no dependa únicamente de la muestra de entrenamiento” (Seinfeld & Pandis, 2016, p. 145).

3.6.5. *Presentación de Resultados*

Los hallazgos serán organizados en gráficos de series temporales, diagramas de dispersión, mapas de calor y tablas estadísticas. Se incluirán análisis interpretativos de los picos de contaminación, su asociación con condiciones meteorológicas y la capacidad del modelo para anticipar estos eventos. Esta presentación estará directamente orientada a contrastar los objetivos específicos y las hipótesis de investigación, facilitando la comunicación técnica hacia entidades como el OEFA y fundamentando propuestas de mejora en los sistemas de alerta ambiental.

“La visualización de datos es clave para la interpretación de fenómenos complejos y para facilitar la toma de decisiones basada en evidencia” (Bisquerra, 2004, p. 130).

En síntesis, el análisis de datos combina enfoques descriptivos, correlacionales y predictivos, articulando tanto técnicas estadísticas convencionales como herramientas avanzadas de aprendizaje automático. Esta integración metodológica no solo permite

caracterizar de manera rigurosa el comportamiento del dióxido de azufre (SO₂) bajo condiciones de inversión térmica, sino también anticipar escenarios de riesgo ambiental con base en la validación de modelos robustos como Random Forest. Con ello, se asegura la generación de resultados confiables y útiles para la fiscalización ambiental preventiva, en coherencia con los objetivos de la investigación. A partir de esta base empírica y técnica, en el siguiente apartado se desarrollan las consideraciones éticas que orientan el uso responsable de los datos y la interpretación de los hallazgos.

3.7. Consideraciones Éticas

La presente investigación se desarrollará respetando estrictamente los principios éticos fundamentales que rigen la producción científica, en cumplimiento de los lineamientos del Comité de Ética de la Investigación y de estándares nacionales e internacionales. Se garantizará la integridad académica, la veracidad en el tratamiento de la información y la responsabilidad en el uso de datos ambientales, asegurando así la transparencia y confiabilidad de los resultados.

3.7.1. *Uso responsable de datos*

Los datos provendrán exclusivamente de fuentes oficiales y verificadas, como el Organismo de Evaluación y Fiscalización Ambiental (OEFA) y la plataforma de reanálisis meteorológico ERA5 de Copernicus. Serán empleados únicamente con fines académicos, bajo criterios de trazabilidad y buenas prácticas científicas, evitando su distorsión o manipulación. Según Hernández et al. (2014, p. 160), “el uso ético de la información implica verificar la validez de las fuentes y garantizar que los datos no sean utilizados para propósitos distintos a los declarados en la investigación”.

3.7.2. *Veracidad y honestidad en la recolección y análisis de información*

Se asegurará la transparencia metodológica en todas las fases del estudio, desde la recolección y depuración de datos hasta su análisis e interpretación. Se evitarán sesgos,

omisiones deliberadas o alteraciones que comprometan la validez de los hallazgos, garantizando la reproducibilidad del estudio.

“Toda investigación debe desarrollarse con apego a principios de veracidad, fidelidad de datos y transparencia metodológica, garantizando así su reproducibilidad y validez” (Sampieri et al., 2014, p. 155).

3.7.3. *Reconocimiento de fuentes y derechos de autor*

Todas las fuentes bibliográficas, técnicas y documentales serán citadas conforme a las normas del estilo **APA, 7.^a edición**. Ello asegura el reconocimiento del trabajo intelectual de terceros y evita cualquier forma de plagio. Se incluirán publicaciones científicas, bases de datos, informes técnicos y marcos legales nacionales e internacionales.

“La integridad académica exige el reconocimiento explícito de todas las fuentes utilizadas mediante un sistema de citación adecuado” (APA, 2020, p. 17).

3.7.4. *Confidencialidad y protección de la información*

Aunque los datos son de carácter público y ambiental, se mantendrán criterios de confidencialidad respecto a cualquier información que pueda comprometer la seguridad operativa de instalaciones industriales o la identidad de personal técnico vinculado a los registros. Se actuará bajo los principios de neutralidad, no discriminación y responsabilidad institucional.

“Aun cuando se utilicen datos públicos, debe evitarse divulgar cualquier información que, directa o indirectamente, vulnere derechos o principios de neutralidad institucional” (OEFA, 2020, p. 48).

3.7.5. *Enfoque ético en la interpretación y uso de resultados*

Los resultados se interpretarán con responsabilidad, evitando conclusiones fuera del alcance de los datos o interpretaciones alarmistas. El uso de los hallazgos se orientará exclusivamente a fines académicos y técnicos, promoviendo la aplicación de la investigación

como insumo para fortalecer las capacidades estatales en la protección ambiental y la salud pública.

En conclusión, la investigación se sustenta en un marco ético sólido que garantiza la validez, transparencia y responsabilidad en el uso de datos y en la interpretación de resultados. Estos principios no solo fortalecen la confiabilidad académica del estudio, sino que aseguran que sus hallazgos puedan convertirse en herramientas útiles para la gestión ambiental y la protección de la salud pública. Con esta base ética y metodológica, en el siguiente capítulo se presentan los resultados del análisis estadístico y del modelamiento predictivo, orientados a responder a los objetivos e hipótesis planteados.

IV. RESULTADOS

El objetivo de este capítulo es presentar de manera clara, objetiva y estructurada los resultados obtenidos tras la aplicación de la metodología descrita en el Capítulo III. Se exponen los hallazgos vinculados a la evaluación del impacto de la inversión térmica en las concentraciones horarias de dióxido de azufre (SO_2) en el entorno de la Refinería de Zinc de Cajamarquilla, considerando tanto el análisis estadístico descriptivo como la validación del modelo predictivo basado en Random Forest (RF). Los resultados se organizan en función de los objetivos e hipótesis específicas, incorporando tablas, figuras y representaciones gráficas que permiten evidenciar la magnitud, recurrencia y patrones de acumulación del contaminante bajo condiciones de estabilidad atmosférica.

4.1. Análisis Descriptivo de los Datos

4.1.1. *Caracterización general de los datos obtenidos*

En el presente apartado se realiza una caracterización cuantitativa y descriptiva de los datos obtenidos de las estaciones automáticas de monitoreo de calidad del aire del OEFA, ubicadas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla. Se analizan los registros horarios de dióxido de azufre (SO_2) comprendidos entre los años 2022 y 2024, así como la altura de la capa límite atmosférica (BLH) obtenida a partir del portal Copernicus, que representa un factor meteorológico clave en la dispersión vertical de contaminantes.

Se consideraron dos estaciones de medición:

- Barlovento: ubicada en dirección contraria a la propagación predominante del viento.
- Sotavento: ubicada en dirección hacia la cual el viento transporta los contaminantes generados por la fuente emisora.

El número de registros válidos para el periodo 2022–2024 se muestra en la siguiente tabla:

Tabla 16*Registros horarios válidos de SO₂ para el periodo 2022-2024*

Estación ubicada	Registros válidos de SO₂ (horarios)
Barlovento	18,593 registros
Sotavento	18,867 registros

Fuente: Elaboración propia

A continuación, se presentan los principales datos estadísticos descriptivos de las concentraciones horarias de SO₂ (µg/m³) y de la altura de la capa límite atmosférica (BLH, en metros):

Tabla 17*Datos estadísticos de las concentraciones horarias de SO₂ (µg/m³) y de (BLH, en metros)*

Parámetro	SO₂ Barlovento	BLH Barlovento	SO₂ Sotavento	BLH Sotavento
N.º de datos (count)	18,593	18,973	18,867	19,159
Promedio (mean)	36.25	257.52	111.04	260.01
Desviación estándar (std)	29.53	141.50	94.58	141.87
Mínimo (min)	3.30	38.31	5.58	38.31
Percentil 25 (Q1)	20.10	143.78	56.57	145.45
Mediana (Q2 / 50%)	27.56	203.39	85.57	205.56
Percentil 75 (Q3)	41.24	364.61	132.12	368.06
Máximo (max)	470.94	859.67	1,535.79	859.67

Fuente: Elaboración propia

Interpretación inicial

- En ambas estaciones, la concentración horaria de SO₂ muestra alta variabilidad.
Sin embargo, la estación sotavento registra valores promedio considerablemente

más altos, con un valor máximo de 1,535.79 $\mu\text{g}/\text{m}^3$, lo que confirma su mayor exposición a la pluma de emisión.

- **La altura de la capa límite atmosférica (BLH)** presenta distribuciones similares en ambas estaciones, con un promedio cercano a 260 metros, aunque con episodios de reducción severa (mínimo de ~ 38 m), que pueden favorecer la acumulación de contaminantes.

Estos resultados permiten establecer una base sólida para los análisis posteriores, especialmente en lo que respecta a la correlación entre eventos críticos de SO_2 y condiciones meteorológicas, como la inversión térmica.

A fin de complementar el análisis estadístico descriptivo de las concentraciones horarias de dióxido de azufre (SO_2) y de la altura de la capa límite atmosférica (BLH), se han elaborado gráficos de distribución para ambas variables, considerando los registros recopilados por las estaciones automáticas del OEFA ubicadas a barlovento y sotavento de la Refinería de Zinc de Cajamarquilla, correspondientes al periodo 2022–2024.

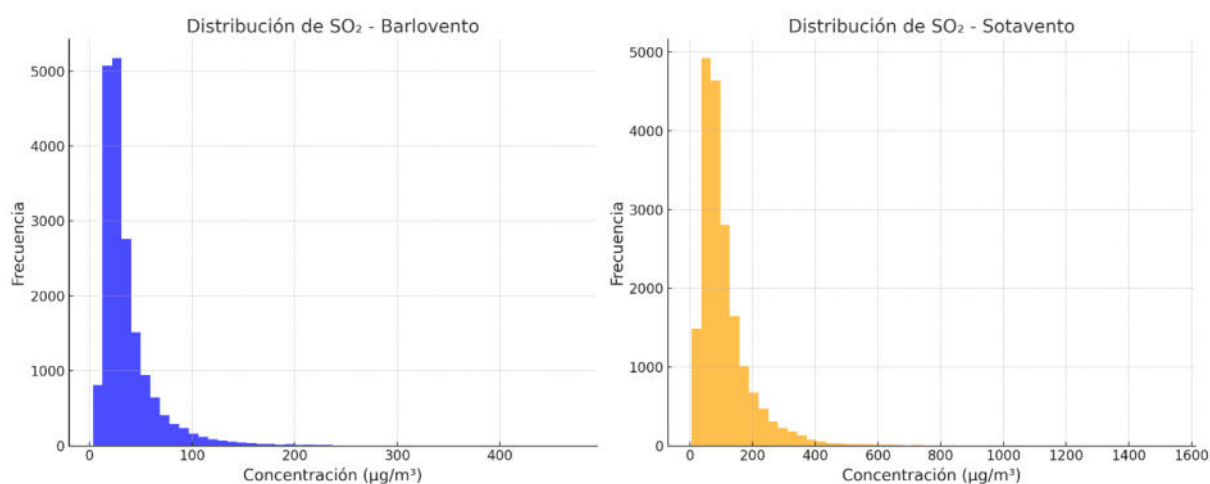
Los histogramas permiten visualizar la frecuencia de ocurrencia de los distintos niveles de concentración de SO_2 y de las alturas de capa límite, lo que facilita identificar patrones de dispersión, acumulación y comportamiento atmosférico. Esta visualización es clave para reconocer eventos críticos, así como para interpretar la dinámica de transporte y confinamiento de contaminantes en la zona de estudio.

A continuación, se presenta los siguientes gráficos:

- **Distribución de SO_2 ($\mu\text{g}/\text{m}^3$):**
 - ✓ Comparación visual entre Barlovento (azul) y Sotavento (naranja).
 - ✓ Evidencia clara de mayores frecuencias de concentraciones elevadas en la estación Sotavento.

Figura 8

Distribución de SO₂ a barlovento y sotavento correspondiente al periodo 2022-2024



Fuente: Elaboración propia

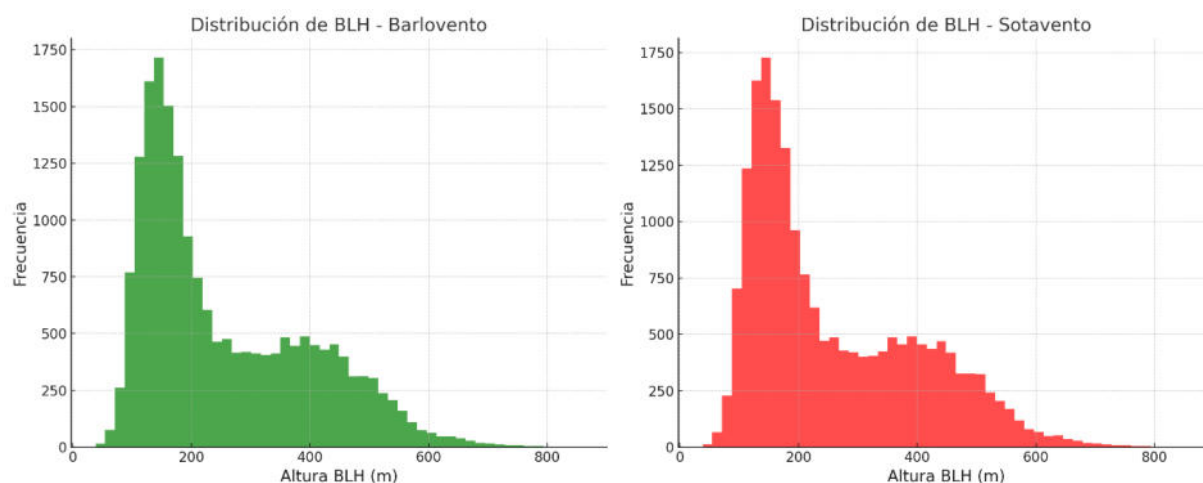
- **Distribución de la altura de la capa límite atmosférica (BLH):**

- ✓ En ambas estaciones, la mayoría de los valores se concentra entre 100 y 400 metros.
- ✓ Esta variable será clave para el análisis de episodios de acumulación de contaminantes.

A fin de identificar el comportamiento temporal de los contaminantes y su relación con las condiciones atmosféricas, se realizó un análisis gráfico de la evolución mensual promedio de la concentración horaria de dióxido de azufre (SO₂) y de la altura de la capa límite atmosférica (BLH) en las estaciones de monitoreo ubicadas a barlovento y sotavento de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.

Figura 9

Distribución de la altura de la capa límite atmosférica (BLH) a barlovento y sotavento correspondiente al periodo 2022-2024

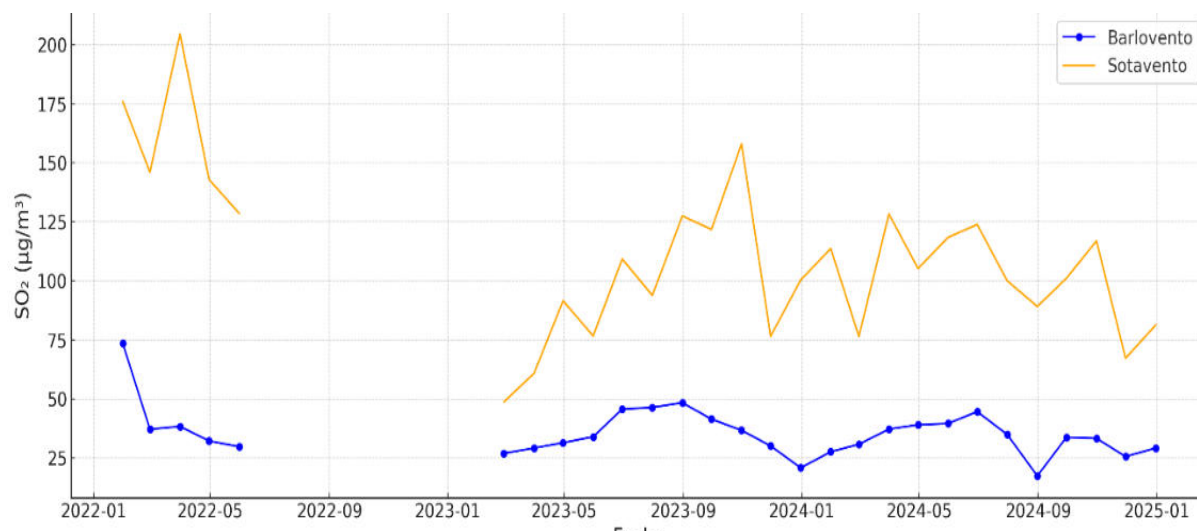


Fuente: Elaboración propia

Este análisis permite visualizar tendencias estacionales, episodios críticos de acumulación y posibles patrones de estancamiento atmosférico vinculados a la variabilidad de la capa límite. La comparación entre ambas estaciones facilita entender la influencia del transporte de contaminantes por efecto del viento y la importancia del comportamiento meteorológico en la dispersión vertical del SO_2 .

Figura 10

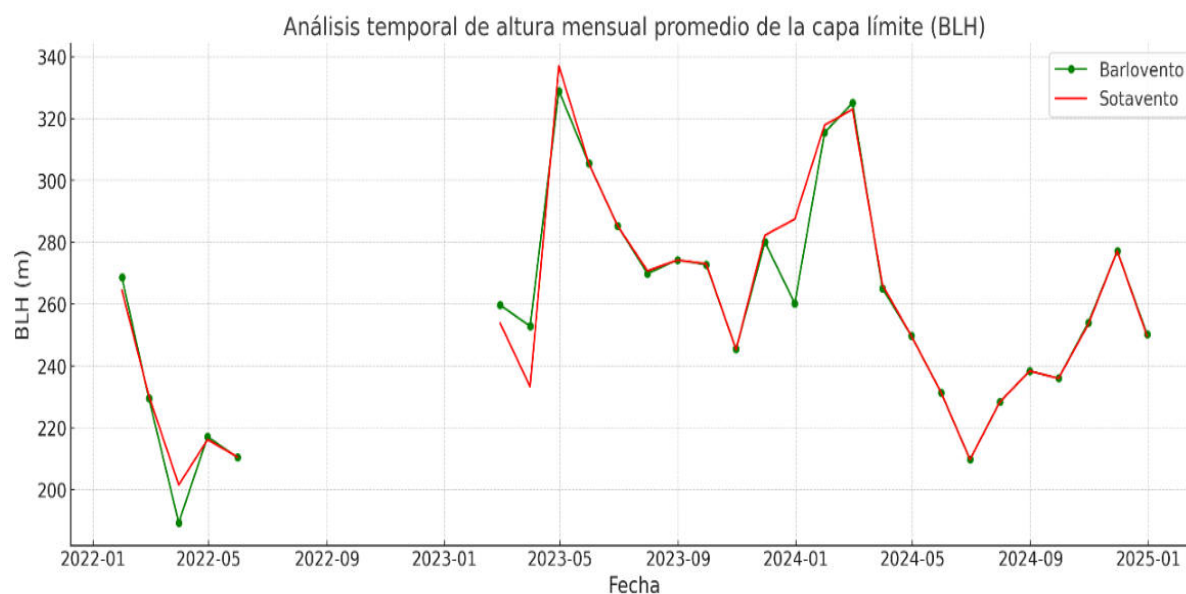
Análisis temporal de concentración mensual promedio de SO₂. A barlovento y sotavento durante el periodo 2022-2024



Fuente: Elaboración propia

Figura 11

Análisis temporal de la altura mensual promedio de la BLH a barlovento y sotavento durante el periodo 2022-2024



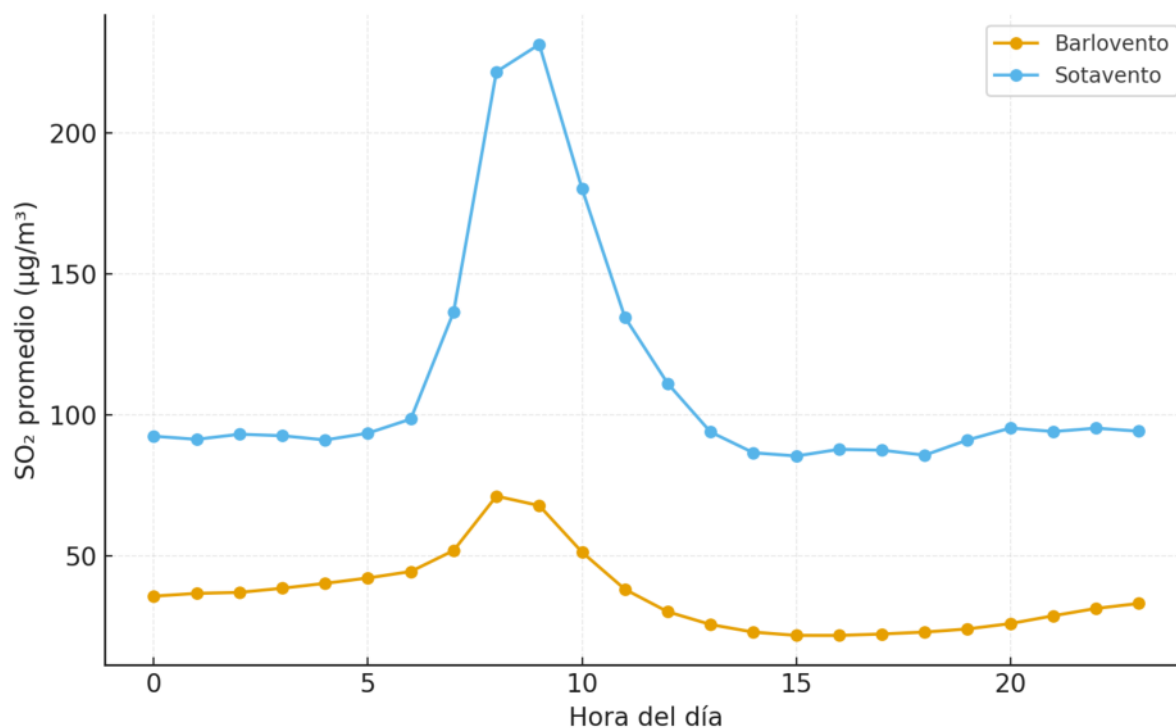
Fuente: Elaboración propia

La distribución espacial y temporal de las concentraciones de SO_2 observadas en las estaciones de barlovento y sotavento resulta consistente con los fundamentos teóricos del modelo de penacho gaussiano (Gaussian Plume), el cual describe cómo los contaminantes emitidos desde una fuente puntual se dispersan en la atmósfera en función de la velocidad del viento, la estabilidad atmosférica y la altura efectiva de emisión (Arya, 1999). Tal como predice este modelo, los registros más elevados se concentran en la estación sotavento, especialmente en las primeras horas de la mañana, cuando la inversión térmica limita la dispersión vertical y propicia acumulaciones críticas de contaminantes. De esta manera, el Gaussian Plume sirve como marco conceptual que complementa los análisis estadísticos, ofreciendo una explicación física de los patrones identificados en los datos empíricos.

En síntesis, los resultados descriptivos muestran diferencias marcadas entre las estaciones de barlovento y sotavento, evidenciando una mayor acumulación de SO_2 en esta última, particularmente durante las primeras horas de la mañana. Este comportamiento es consistente con lo planteado por el modelo de penacho gaussiano, el cual explica la acumulación de contaminantes bajo condiciones de estabilidad atmosférica e inversión térmica. La Figura 12 muestra el promedio horario de SO_2 durante el periodo 2022–2024, destacándose un pico matutino pronunciado en sotavento ($\approx 08:00\text{--}09:00$), lo que respalda empíricamente la influencia de la inversión térmica en la dispersión limitada y el transporte de la pluma hacia sotavento.

Figura 12

Promedio de SO₂ por hora del día (2022–2024)



Fuente: Elaboración propia

4.1.2. *Análisis meteorológicos asociado*

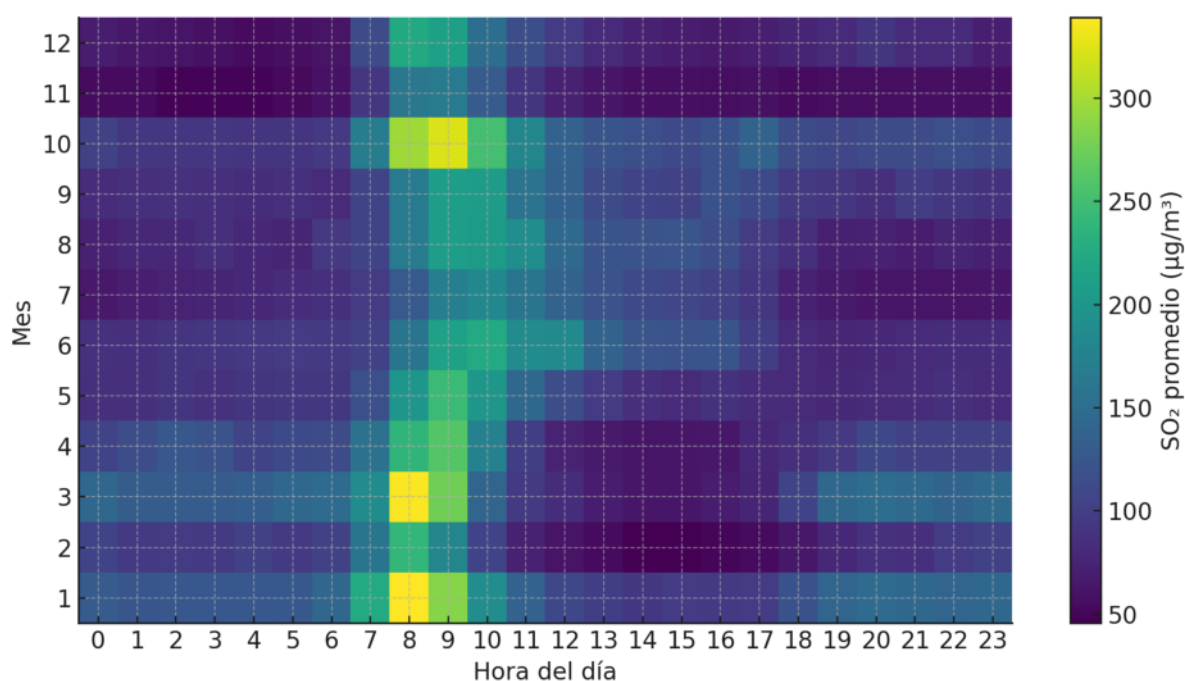
El comportamiento de la concentración de contaminantes atmosféricos, como el dióxido de azufre (SO₂), depende estrechamente de las condiciones meteorológicas locales. Para el periodo 2022–2024, se analizaron las principales variables atmosféricas registradas por las estaciones automáticas del OEFA a barlovento y sotavento, a fin de interpretar su relación con la dispersión o acumulación del contaminante.

Para complementar el análisis descriptivo, resulta necesario examinar la variabilidad horaria y estacional de las concentraciones de SO₂ en la estación sotavento, en función de las condiciones atmosféricas que favorecen la acumulación de contaminantes. La **Figura 13** presenta un mapa de calor que sintetiza el comportamiento promedio de SO₂ por mes y hora

durante el periodo 2022–2024. Se observa un máximo robusto alrededor de las 08:00–09:00 horas, particularmente en meses en los que la altura de la capa límite atmosférica (BLH) tiende a valores reducidos. Este patrón confirma la influencia de la inversión térmica en la dinámica de dispersión vertical y aporta evidencia empírica coherente con lo planteado por el modelo de penacho gaussiano.

Figura 13

Mapa de calor de SO₂ en sotavento (promedio 2022–2024) por mes y hora.



Fuente: Elaboración propia

4.1.2.1. Temperatura del aire. Las temperaturas promedio horarias oscilaron entre los 14 °C y 28 °C, con valores ligeramente superiores durante los meses de verano. Las temperaturas mínimas y máximas no presentan diferencias marcadas entre las estaciones, lo cual es consistente con el clima templado de la zona de Cajamarquilla. Sin embargo, es importante señalar que los gradientes verticales de temperatura, especialmente en horas de la madrugada, contribuyen a la formación de inversiones térmicas, condición clave en el análisis de acumulación de SO₂.

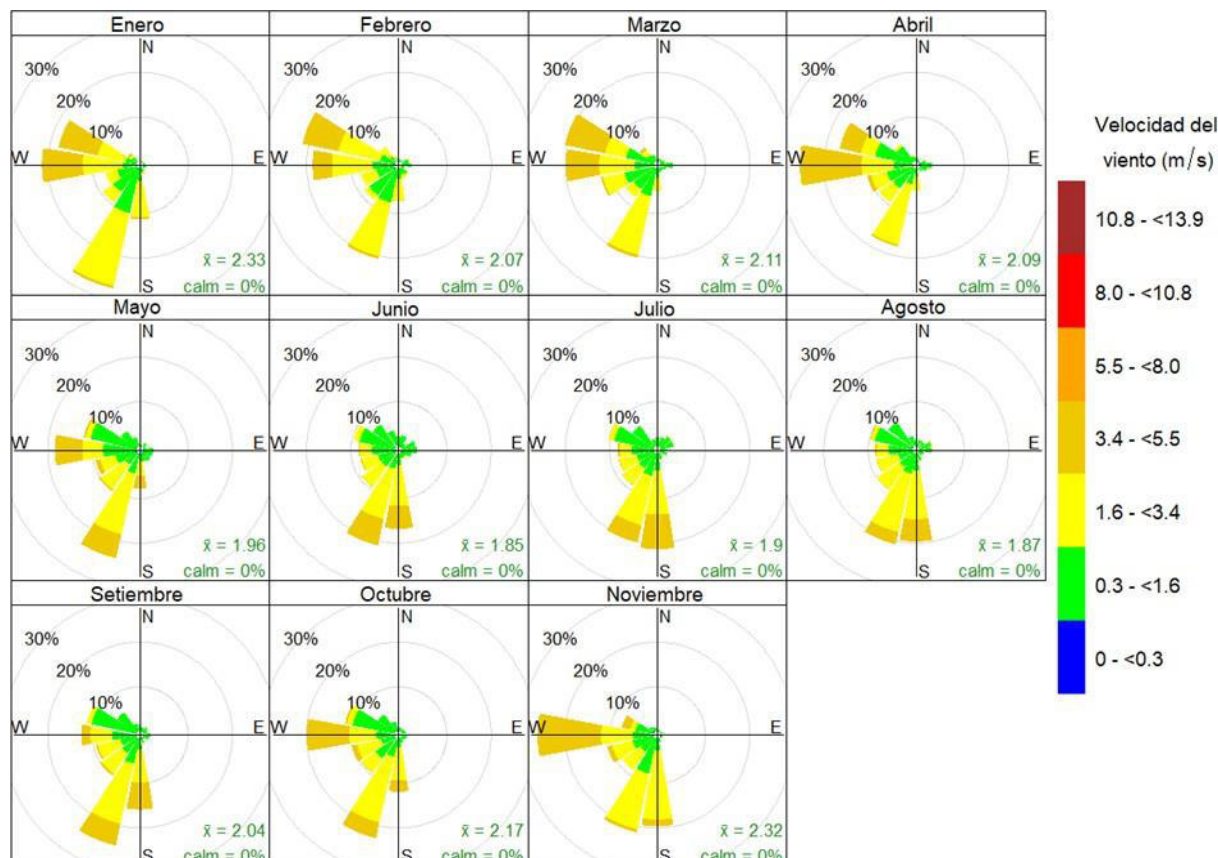
4.1.2.2. Velocidad y dirección del viento. A continuación, se describen el comportamiento de la velocidad y dirección del viento en ambas estaciones.

La **velocidad del viento (WS)** se mantuvo predominantemente baja en ambas estaciones, con registros horarios frecuentes por debajo de 2.0 m/s, condición que favorece la escasa dispersión horizontal de contaminantes.

La **dirección del viento (WD)** indica que la pluma de emisión de la refinería tiende a desplazarse de manera consistente hacia la estación sotavento, donde se registran las mayores concentraciones de SO₂. Esto refuerza la hipótesis de que dicha estación constituye la zona de mayor acumulación del contaminante. No obstante, en este sector no existen asentamientos humanos, ya que está ocupado por un cerro que actúa como barrera orográfica. En contraste, en el área de barlovento se localizan viviendas y una zona industrial, lo que incrementa la vulnerabilidad de la población residente y de las actividades productivas ante episodios de transporte de contaminantes. De acuerdo con la rosa de los vientos del año 2024, la procedencia predominante fue desde el sur-suroeste (SSO), con una frecuencia del 21,9 %, seguida de vientos desde el oeste (O), condiciones que favorecen que la pluma de emisiones pueda desplazarse hacia dichas áreas pobladas e industriales en determinados escenarios meteorológicos.

Figura 14

Diagramas de rosa de los vientos por meses SO₂ para el año 2024, de la estación de monitoreo ubicada a sotavento.



Fuente: OEFA (2024)

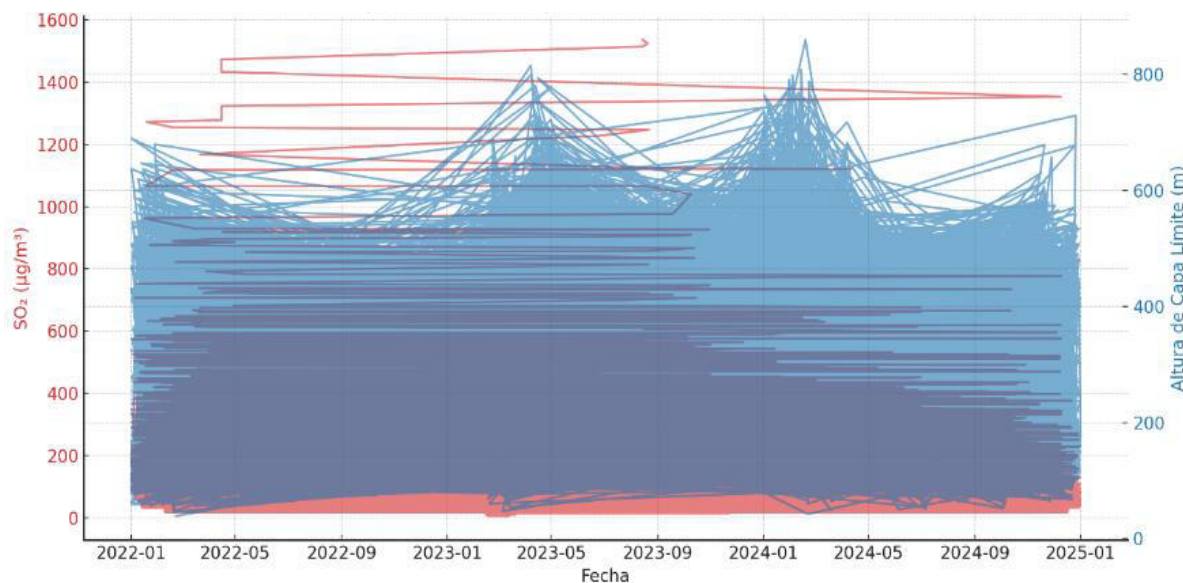
En la Figura 14 se presenta la rosa de los vientos correspondiente al periodo enero–noviembre de 2024, obtenida en la estación CA-RZC-02, ubicada a sotavento de la Refinería de Zinc de Cajamarquilla. Se observa que la procedencia predominante del viento fue desde el sur-suroeste (SSO), con una frecuencia del 21,9 %, seguida por vientos provenientes del oeste (O). Este patrón explica la mayor acumulación de SO₂ en la estación sotavento, instalada en las instalaciones del Radio Observatorio de Jicamarca, donde funciona el radar ionosférico más grande del mundo, operado por el Instituto Geofísico del Perú y en el que laboran aproximadamente 80 personas. Aunque no existen viviendas en esta área, la presencia de personal en dichas instalaciones evidencia un nivel de exposición ocupacional a los

contaminantes atmosféricos. En contraste, en el área de barlovento se ubican viviendas y una zona industrial, lo que incrementa la vulnerabilidad de la población y de las actividades productivas cuando la pluma de emisiones es transportada hacia dicho sector en determinados escenarios meteorológicos.

4.1.2.3. Altura de la Capa Límite Atmosférica (BLH). La altura de la capa límite atmosférica (BLH), obtenida del portal Copernicus, presentó valores variables entre ~40 m y 860 m, con un promedio de ~260 m en ambas estaciones. Los valores más bajos se observaron principalmente durante las horas de la madrugada (4:00 a 8:00 a.m.), coincidiendo con los picos horarios de concentración de SO₂, lo que sugiere una restricción vertical para la dispersión del contaminante.

El análisis visual mediante histogramas y series temporales demuestra que en los días con BLH menor a 150 metros, las concentraciones de SO₂ superan significativamente los promedios, especialmente en la estación sotavento.

A fin de reforzar esta relación, la Figura 15 presenta la evolución temporal conjunta de las concentraciones horarias de SO₂ y de la altura de la capa límite atmosférica (BLH) en la estación sotavento durante el periodo 2022–2024. Se observa que los máximos de SO₂ tienden a coincidir con los mínimos de la BLH, confirmando que la estabilidad atmosférica asociada a la inversión térmica constituye un factor determinante en la acumulación de contaminantes.

Figura 15*Relación temporal entre SO₂ y BLH - Estación Sotavento (2022-2024)*

Fuente: Elaboración propia

La representación gráfica evidencia un patrón consistente: cuando la BLH desciende a valores críticos (<150 m), las concentraciones de SO₂ alcanzan picos significativos, con registros que superan los $1,000 \mu\text{g}/\text{m}^3$ en episodios puntuales. Este comportamiento confirma el rol de la inversión térmica como mecanismo de confinamiento de contaminantes, especialmente en la estación sotavento, donde la pluma industrial impacta de manera más directa. En consecuencia, la BLH se constituye en una variable clave para la interpretación y predicción de episodios críticos de contaminación atmosférica en la zona de estudio.

4.1.2.4. Frecuencia e intensidad de eventos de inversión térmica. La inversión térmica es un fenómeno clave en la acumulación de contaminantes en zonas industriales. Durante el periodo analizado, se identificaron múltiples episodios de inversión térmica, especialmente en las primeras horas del día. Estos eventos se caracterizan por una reducción de la altura de la capa límite atmosférica junto con temperaturas más elevadas en capas superiores de la atmósfera, lo cual impide el ascenso del aire contaminado.

Tabla 18*Clasificación de los eventos por intensidad según la BLH*

Tipo de evento	Rango BLH (m)	Frecuencia estimada
Inversión fuerte	< 150 m	Alta (recurrente)
Inversión moderada	150 m – 300 m	Frecuente
Dispersión favorable	> 300 m	Menor frecuencia

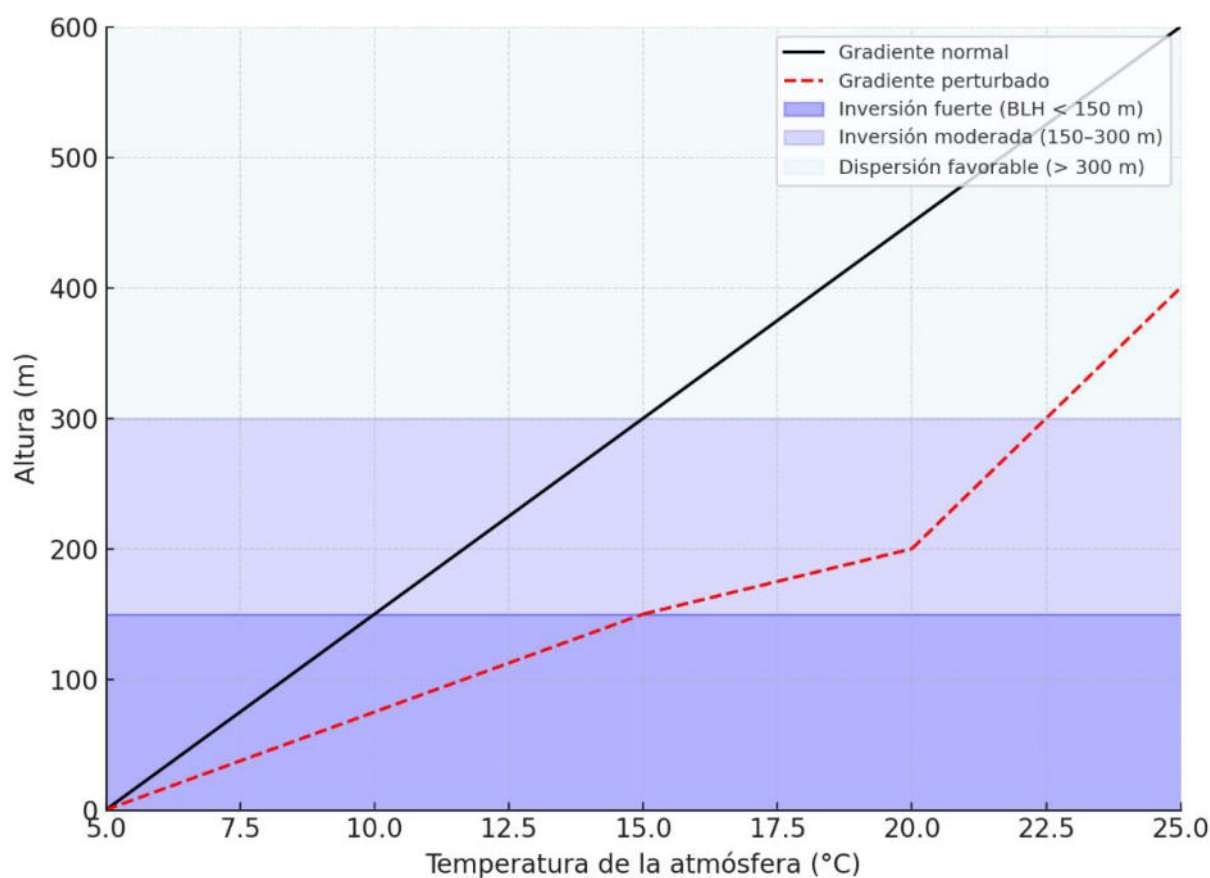
Fuente: Elaboración propia

Se evidenció que los valores extremos de SO₂ coinciden en su mayoría con eventos de inversión fuerte, especialmente durante el invierno. Este patrón permite establecer la base para los análisis correlacionales y predictivos que se desarrollarán en los siguientes apartados.

Para complementar el análisis de inversión térmica, se presenta un gráfico del gradiente térmico vertical típico, donde se han destacado las tres zonas principales que afectan la dispersión de contaminantes según la altura de la capa límite atmosférica (BLH). Estos rangos fueron definidos en base al análisis de frecuencias descrito en la Tabla 18, clasificando los eventos de inversión térmica en: fuerte (BLH < 150 m), moderada (150–300 m) y condiciones favorables de dispersión (BLH > 300 m). Esta visualización permite ilustrar cómo los estratos térmicos limitan la convección vertical y refuerzan la acumulación de SO₂ en niveles cercanos al suelo.

Figura 16

Gradiente térmico vertical: normal e inversión térmica



Fuente: Elaboración propia, con información del OEFA y BLH de Copernicus.

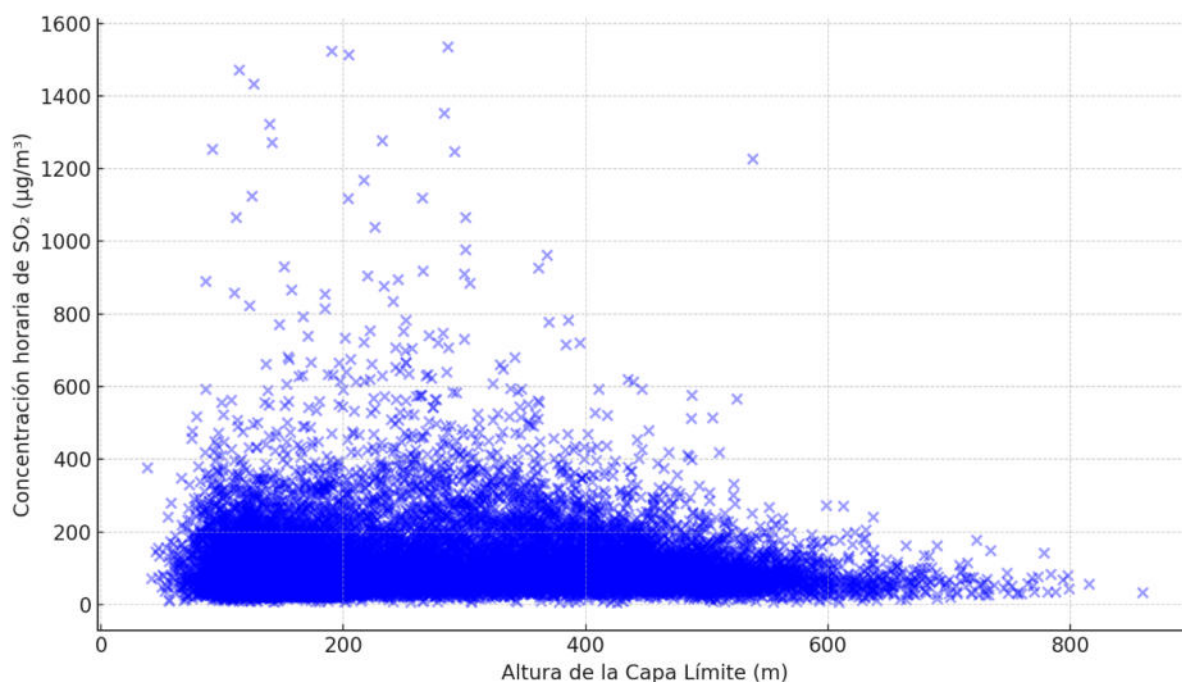
Con el objetivo de explorar la relación entre la altura de la capa límite atmosférica (BLH) y las concentraciones horarias de dióxido de azufre (SO_2), se elaboró el siguiente gráfico de dispersión a partir de la serie integrada 2022–2024. Los datos de SO_2 provienen de la estación de monitoreo automático Sotavento (CA-RZC-02) gestionada por el OEFA a través del sistema PIFA, mientras que la información de BLH fue obtenida de la plataforma satelital Copernicus y ajustada a resolución horaria mediante interpolación.

La figura 17 permite visualizar claramente la asociación negativa entre ambos parámetros, destacando que a menores valores de BLH —condiciones típicas de estabilidad atmosférica o inversión térmica— se registran mayores concentraciones superficiales de SO_2 .

Esta relación respalda la hipótesis del estudio sobre el efecto de las condiciones meteorológicas en la acumulación de contaminantes.

Figura 17

Gráfico de dispersión: Relación entre Altura de Capa Límite (BLH) y SO₂ horario - Estación Sotavento



Fuente: Elaboración propia con información del OEFA y BLH de Copernicus.

Este gráfico evidencia que:

- Las concentraciones más altas de SO₂ se concentran cuando la BLH es baja, principalmente por debajo de los 300 m, y más marcadamente bajo los 150 m.
- A medida que la BLH aumenta, las concentraciones horarias de SO₂ tienden a disminuir, lo que confirma la restricción vertical a la dispersión durante eventos de inversión térmica.

Para complementar el análisis meteorológico, se procedió a clasificar los registros horarios de altura de la capa límite atmosférica (BLH) de la estación sotavento según rangos definidos en la literatura técnica y normativa, que permiten identificar condiciones favorables

o restrictivas para la dispersión vertical de contaminantes.

Esta clasificación distingue tres categorías:

- **Inversión fuerte:** BLH < 150 m.
- **Inversión moderada:** BLH entre 150 y 300 m.
- **Dispersión favorable:** BLH > 300 m.

La siguiente tabla resume la frecuencia de ocurrencia de cada tipo de evento en el periodo 2022–2024, lo cual resulta relevante para identificar condiciones atmosféricas críticas que favorecen la acumulación de SO₂ y fundamentan la necesidad de implementar sistemas de vigilancia y alerta temprana.

Tabla 19

Frecuencia de eventos de inversión térmica (Sotavento, 2022–2024)

Tipo de Evento	Frecuencias
Inversión fuerte (<150 m)	5,284 horas
Inversión moderada (150–300 m)	7,299 horas
Dispersión favorable (>300 m)	6,576 horas

Fuente: Elaboración propia

4.1.3. Análisis comparativo de concentraciones a barlovento y sotavento

El análisis comparativo entre las estaciones automáticas ubicadas a barlovento y sotavento permite identificar de manera objetiva el efecto del viento y de las condiciones atmosféricas sobre la distribución espacial del dióxido de azufre (SO₂) emitido por la Refinería de Zinc de Cajamarquilla.

4.1.3.1. Tendencias generales por estación. Durante el periodo 2022–2024, la estación sotavento registró concentraciones significativamente más altas de SO₂ que la estación barlovento, tanto en promedio como en picos horarios. Esta diferencia está directamente

relacionada con la dirección predominante del viento, que transporta la pluma de emisión hacia el sector sotavento, exponiéndolo con mayor frecuencia a acumulaciones de contaminantes.

Tabla 20

Tendencias generales por estación

Indicador estadístico	SO ₂ Barlovento (µg/m ³)	SO ₂ Sotavento (µg/m ³)
Promedio	36.25	111.04
Mediana (Q2)	27.56	85.57
Percentil 75 (Q3)	41.24	132.12
Máximo registrado	470.94	1,535.79

Fuente: Elaboración propia

Interpretación: Las diferencias entre ambas estaciones son notables, con valores extremos en sotavento que superan en más de 1,000 µg/m³ a los máximos de barlovento, confirmando la dirección dominante de la carga contaminante.

4.1.3.2. Comparación horaria y diaria. El análisis de datos promedios por hora del día evidencia que:

Las concentraciones más altas de SO₂ en ambas estaciones ocurren en las primeras horas de la mañana (4:00–8:00 a.m.), coincidiendo con la reducción de la altura de la capa límite atmosférica.

En sotavento, los valores horarios críticos se acumulan principalmente durante eventos de inversión térmica, mientras que en barlovento las variaciones son menos marcadas.

Esta diferencia también se refleja en el número de horas en las que se superó el umbral de 250 µg/m³, indicador crítico para la vigilancia ambiental:

Tabla 21

Comparación horaria con SO₂ > 250 µg/m³ en las estaciones de sotavento y barlovento

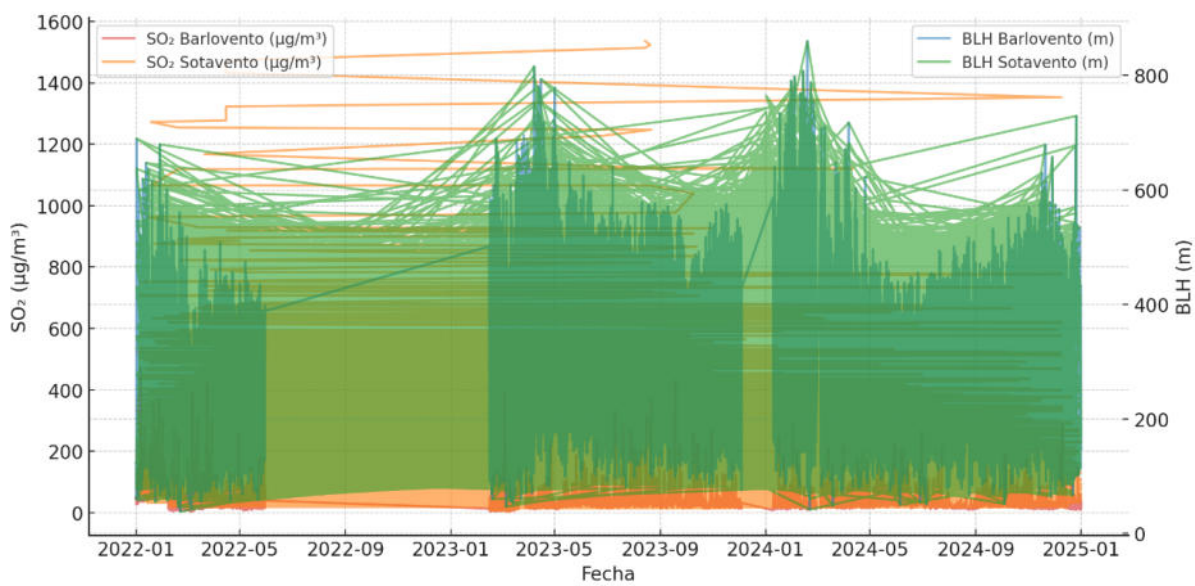
Año	Horas con SO ₂ > 250 µg/m ³ (Sotavento)	Horas con SO ₂ > 250 µg/m ³ (Barlovento)
2022	443 horas	Muy baja (no relevante)
2023	346 horas	Muy baja (no relevante)
2024	413 horas	Muy baja (no relevante)

Elaboración: propia

Con el fin de contrastar de manera simultánea la dinámica del dióxido de azufre (SO₂) y de la altura de la capa límite atmosférica (BLH) en ambas estaciones de monitoreo, se elaboró un análisis comparativo temporal. Esta representación gráfica permite identificar los momentos en los que los niveles críticos de SO₂ coinciden con reducciones significativas de la BLH, resaltando las diferencias en la exposición entre barlovento y sotavento.

Figura 18

Comparación de SO₂ y BLH en Barlovento y Sotavento (2022-2024)



Fuente: Elaboración propia

La comparación simultánea de las concentraciones horarias de SO_2 y la altura de la capa límite atmosférica (BLH) en las estaciones de barlovento y sotavento (Figura 18) refuerza los patrones previamente identificados. En sotavento se registran los niveles más críticos de SO_2 , con picos que superan los $1,500 \mu\text{g}/\text{m}^3$ en episodios coincidentes con reducciones severas de la BLH por debajo de 150 m, condición que limita la dispersión vertical y potencia la acumulación del contaminante. En contraste, en barlovento los valores se mantienen dentro de un rango más moderado (máximo de $470 \mu\text{g}/\text{m}^3$), reflejando un menor impacto directo de la pluma de emisión. Esta diferencia espacial confirma el efecto combinado del transporte advectivo y de la estabilidad atmosférica en la dinámica local de la contaminación.

En conjunto, estos hallazgos muestran que la BLH actúa como un factor modulador determinante de la dispersión del SO_2 , siendo su disminución matutina uno de los principales desencadenantes de episodios críticos en sotavento. De este modo, la evidencia empírica respalda la necesidad de considerar simultáneamente la variabilidad meteorológica y la dirección predominante del viento en la interpretación de los registros de calidad del aire.

En la sección (4.1.4) se profundizará en el análisis de correlación entre el SO_2 y las variables meteorológicas registradas, con el propósito de cuantificar la fuerza y dirección de las relaciones identificadas, y de este modo aportar criterios estadísticos que complementen la interpretación descriptiva realizada hasta este punto.

4.1.3.3. Relevancia del análisis comparativo. Este análisis confirma que la estación sotavento representa un punto crítico para la vigilancia de calidad del aire en la zona industrial de Cajamarquilla, y que el comportamiento horario y meteorológico debe ser incorporado en los modelos predictivos para prevenir acumulaciones severas de SO_2 que podrían afectar a la población expuesta.

A efectos de reforzar el análisis comparativo entre ambas estaciones, se elaboró un gráfico que muestra la cantidad de horas por año en las que se superó el umbral de $250 \mu\text{g}/\text{m}^3$

en la concentración horaria de SO_2 . Esta visualización permite identificar la magnitud y recurrencia de episodios críticos de contaminación en la zona sotavento —con mayor frecuencia y concentración— frente a eventos más esporádicos en la estación barlovento.

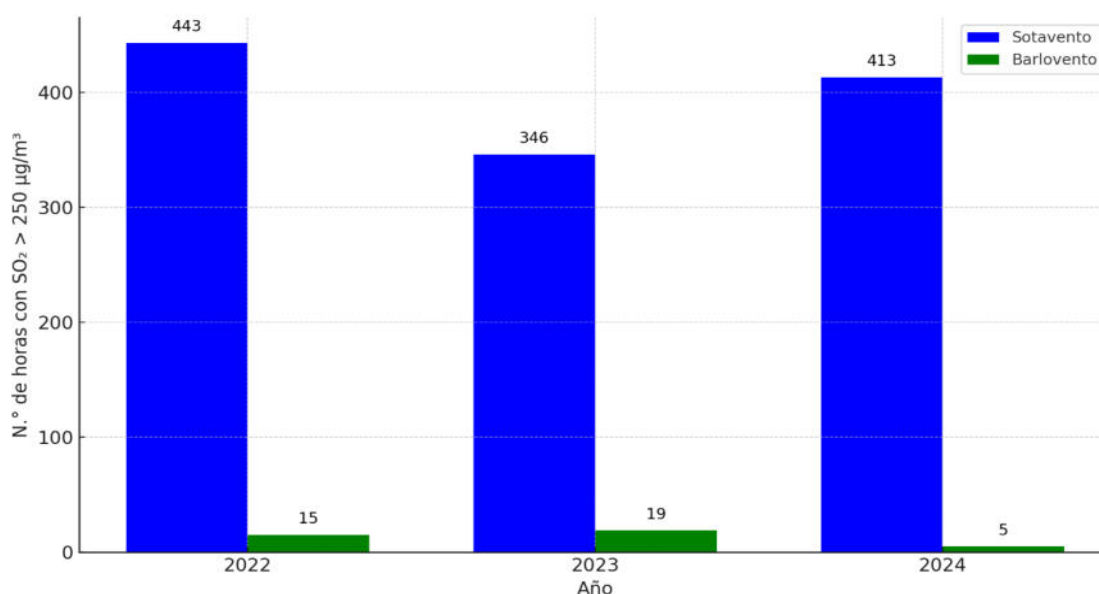
Como se observa, la estación sotavento presenta un comportamiento recurrente de superaciones del umbral en todos los años analizados, mientras que, en barlovento, aunque con menor intensidad, también se registraron superaciones ocasionales, lo que evidencia la dispersión parcial de la pluma contaminante hacia el lado opuesto.

La figura 19 refleja correctamente los valores reales de superaciones horarias del umbral de $250 \mu\text{g}/\text{m}^3$ en ambas estaciones:

- Sotavento: con altos niveles persistentes cada año.
- Barlovento: con superaciones ocasionales pero presentes (15 en 2022, 19 en 2023 y 4 en 2024), lo que también evidencia impacto parcial en dirección contraria a la pluma principal.

Figura 19

Comparación del número de horas con concentraciones horarias de SO_2 superiores a $250 \mu\text{g}/\text{m}^3$ por estación y año (2022–2024).



Fuente: Elaboración propia a partir de datos del PIFA del OEFA y BLH de Copernicus.

4.1.4. *Análisis de correlación entre SO₂ y variables meteorológicas (Estación Sotavento, 2022–2024)*

El coeficiente de Pearson (r) permite cuantificar la fuerza y la dirección de las relaciones lineales entre variables, con valores cercanos a -1 o 1 que indican correlaciones fuertes (negativa o positiva, respectivamente) y valores cercanos a 0 que reflejan ausencia de relación lineal (Field, 2018; Wilks, 2019). Los intervalos de confianza (95 %) se estiman mediante la transformación z de Fisher (Wilks, 2019).

Este apartado evalúa la relación estadística entre SO₂ y la meteorología en la estación ubicada a sotavento (datos horarios validados 2022–2024), para sustentar inferencias sobre acumulación bajo inversión térmica y su uso en criterios de fiscalización. El orden sigue la lógica general de la matriz de coherencia del estudio.

Se estimó el coeficiente de Pearson (r) y sus IC95% vía transformación z de Fisher (Wilks, 2019). La dirección del viento se convirtió a componentes cartesianas para su uso en el modelo (u , v) por su circularidad. Se realizó estratificación por hora del día y por cuartiles de BLH, y se calcularon correlaciones con rezagos (1–6 h), integrando la lectura estadística con la física de mezcla/dispersión y con el enfoque de modelación no lineal usado en el capítulo (Breiman, 2001; Hastie et al., 2009; Hyndman & Athanasopoulos, 2021; Bergmeir & Benítez, 2012; Molnar, 2023).

En la Tabla 22 se sintetiza la correlación de Pearson entre SO₂ y las variables meteorológicas en todo el periodo 2022–2024. Este resumen global ofrece una primera aproximación lineal a la relación, sin distinguir horas ni regímenes atmosféricos.

Tabla 22

Coeficiente de correlación entre SO₂ y variables meteorológicas (serie completa 2022–2024, estación a sotavento)

Relación	r (Pearson)	IC95%	p-valor	Interpretación
SO ₂ vs. BLH	0.007	-0.008, 0.021	0.370	No significativa / nula
SO ₂ vs. Temperatura	0.122	0.108, 0.137	<0.001	Positiva débil
SO ₂ vs. Humedad relativa	-0.200	-0.214, -0.186	<0.001	Negativa baja-moderada
SO ₂ vs. Velocidad del viento	0.012	-0.003, 0.027	0.108	No significativa / nula

Fuente: Elaboración propia

La magnitud baja de las correlaciones en el agregado es esperable: promediar el año “aplana” episodios con BLH baja y oculta efectos con demora. Esto motiva el análisis estratificado y con rezagos que se presenta a continuación.

En el agregado horario, r tiende a ser de magnitud baja por no linealidades, dependencia por régimen (BLH baja) y efectos rezagados que el r contemporáneo no captura (Wilks, 2019).

La dirección del viento es una magnitud circular, por lo que su lectura con un ángulo puede distorsionar la correlación. Para evitarlo, se representan dos componentes cartesianas: u (Este–Oeste) y v (Norte–Sur). Esta transformación estandariza la señal direccional y hace más estable su relación con SO₂ (Wilks, 2019).

La Tabla 23 muestra la relación de SO₂ con las componentes cartesianas del viento: u (Este–Oeste) y v (Norte–Sur). Esta representación evita distorsiones por la circularidad del ángulo de dirección y permite una lectura más estable (Wilks, 2019).

Tabla 23*Correlación con componentes de viento (u, v)*

Relación	r (Pearson)	IC95%	p-valor	Interpretación
SO ₂ vs. u (zonal)	-0.122	-0.136, -0.107	<0.001	Negativa débil
SO ₂ vs. v (meridional)	0.045	0.030, 0.060	<0.001	No significativa / nula

Fuente: Elaboración propia.

Se aprecia una señal direccional coherente con la ventilación/arrastre del penacho en sotavento, aunque de magnitud moderada en el promedio anual. Este resultado se refuerza al estratificar por hora y BLH.

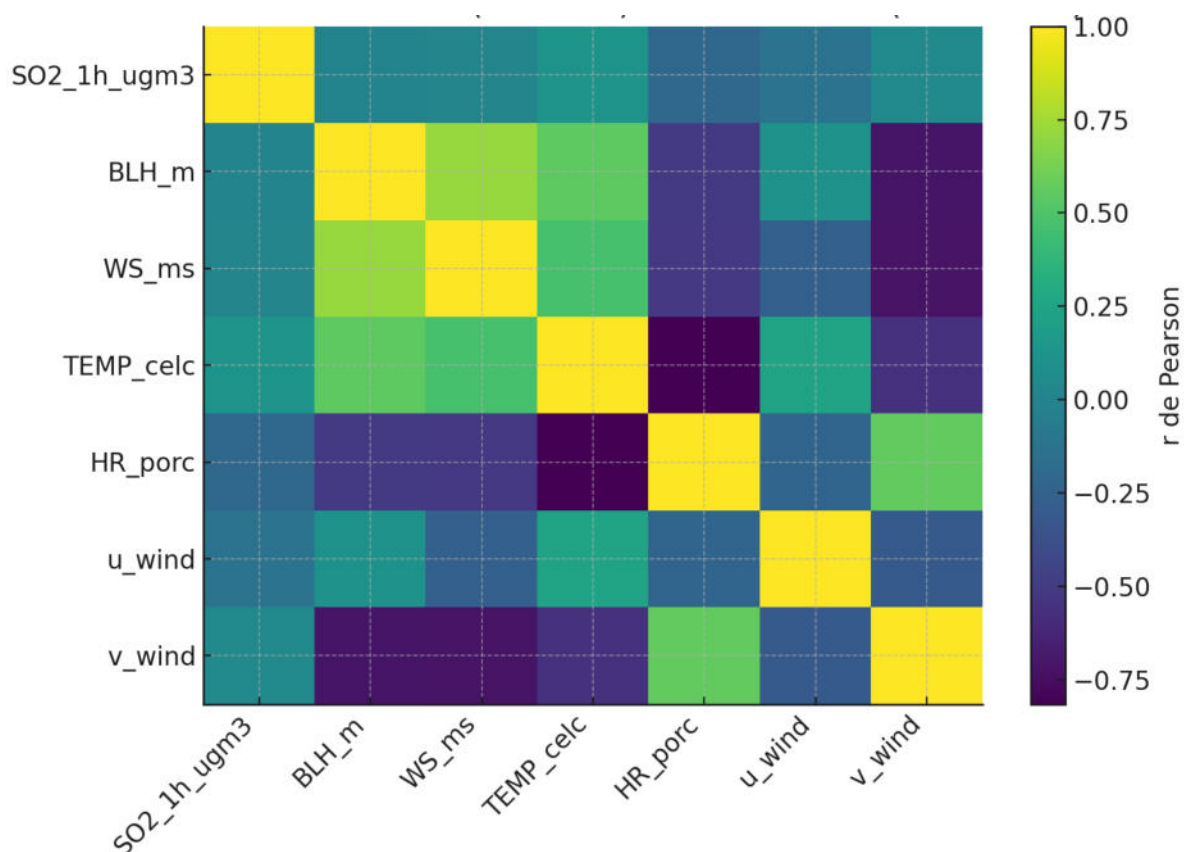
El tratamiento u/v evita problemas por circularidad y ofrece una señal más interpretable.

En la figura 20 se presenta la Matriz de correlación de Pearson entre concentraciones horarias de SO₂ y variables meteorológicas (BLH, temperatura, humedad relativa, velocidad del viento). Datos registrados por la Estación Sotavento (CA-RZC-02) del OEFA para el periodo 2022–2024.

Esta figura ofrece una vista de conjunto de la coherencia lineal entre SO₂ y la meteorología (matriz de correlaciones) en 2022–2024.

Figura 20

Matriz de correlación de Pearson (SO_2 vs. Meteorología) periodo 2022-2024, estación a sotavento.



Fuente: Elaboración propia con base en datos del PIFA del OEFA

El patrón confirma que, a escala agregada, las relaciones lineales son modestas, lo que refuerza la conveniencia de analizar cuándo (hora del día), en qué régimen (BLH baja) y con qué desfase (rezagos) se intensifican.

Con el fin de detallar la relación entre el SO_2 y las variables meteorológicas en la estación Sotavento, se presentan a continuación distintos enfoques complementarios de análisis. Estos apartados permiten observar las correlaciones bajo diferentes perspectivas y aportan elementos para una interpretación más completa de los resultados.

4.1.4.1. Correlaciones estratificadas por régimen. Con el fin de identificar ventanas críticas, se calcularon correlaciones por hora del día (madrugada, mañana, tarde y noche) y por cuartiles de la altura de la capa límite (BLH). En la práctica, cuando la BLH es baja (condición típica de inversión térmica), la mezcla vertical disminuye y el SO₂ tiende a acumularse. Por ello, se espera que la asociación SO₂–meteorología sea más marcada en horas de madrugada/mañana y en el cuartil inferior de BLH. Los resultados confirman este patrón, aportando criterios operativos para priorizar vigilancia y alertas en esas franjas. (Wilks, 2019).

Las Tablas 24-26 presentan la correlación estratificada por hora del día y por cuartiles de BLH, con el fin de identificar ventanas críticas de acumulación.

Tabla 24

Máximo $|r|$ por variable y lente (hora / cuartil BLH / rezago)

Variable	Lente_con_mayor_ $ r $	max_ $ r $
u_wind	Hora	0.509
HR_porc	Hora	0.492
BLH_m	Hora	0.372
TEMP_celc	Hora	0.357
v_wind	Hora	0.318
WS_ms	Hora	0.304

Fuente: Elaboración propia

Tabla 25

Resumen — Mejor $|r|$ por hora del día (extracto)

Hora	r	IC95_inf	IC95_sup	p	n	Variable	$ r $
13	-0.372	-0.432	-0.309	0.0	752	BLH_m	0.372
8	-0.492	-0.545	-0.436	0.0	788	HR_porc	0.492
23	0.357	0.293	0.417	0.0	796	TEMP_celc	0.357
9	0.304	0.238	0.368	0.0	791	WS_ms	0.304

Hora	r	IC95_inf	IC95_sup	p	n	Variable	r
14	-0.509	-0.56	-0.454	0.0	788	u_wind	0.509
16	0.318	0.252	0.381	0.0	790	v_wind	0.318

Fuente: Elaboración propia

Tabla 26

Resumen — Mejor |r| por cuartil de BLH (extracto)

Cuartil_BLH	r	IC95_inf	IC95_sup	p	n	Variable	r
Q1 (baja)	-0.379	-0.403	-0.354	0.0	4766	HR_porc	0.379
Q1 (baja)	0.275	0.248	0.302	0.0	4766	TEMP_celc	0.275
Q4 (alta)	-0.161	-0.19	-0.132	0.0	4583	WS_ms	0.161
Q4 (alta)	-0.343	-0.37	-0.317	0.0	4583	u_wind	0.343
Q4 (alta)	0.251	0.222	0.278	0.0	4583	v_wind	0.251

Fuente: Elaboración propia

Los resultados muestran que la asociación SO₂–meteorología se fortalece en madrugada/mañana y bajo BLH baja (cuartil inferior), coherente con condiciones de mezcla vertical limitada. Estas ventanas orientan la programación de vigilancia y alertas.

4.1.4.2. Correlaciones con rezagos (1–6 h). Para evaluar efectos con demora, se analizó si el SO₂ medido en una hora dada se relaciona con la meteorología registrada entre 1 y 6 horas antes (rezagos). De este modo, se comprueba si condiciones como BLH baja o viento débil ocurridas poco antes ayudan a explicar niveles de SO₂ más altos en la hora actual. Los resultados muestran que las mayores asociaciones se concentran en rezagos cortos ($\approx$ 1–4 h), lo que indica que el efecto no es instantáneo y justifica la anticipación de acciones de fiscalización (p. ej., activar alertas cuando se observan condiciones “riesgosas” en las horas

previas) (Bergmeir & Benítez, 2012; Hyndman & Athanasopoulos, 2021).

Las **Tablas 27 y 28** evalúan si el SO₂ observado en una hora se relaciona con la meteorología de 1 a 6 horas antes (rezagos), para capturar impactos no instantáneos.

Tabla 27

Resumen — Mejor $|r|$ por rezago (extracto)

Variable	lag_h	r	IC95_inf	IC95_sup	p	n	r
BLH_m	4	-0.235	-0.248	-0.221	0.0	18844	0.235
HR_porc	1	-0.101	-0.116	-0.087	0.0	18844	0.101
TEMP_celc	4	-0.087	-0.101	-0.072	0.0	18844	0.087
WS_ms	4	-0.226	-0.24	-0.212	0.0	18844	0.226
u_wind	6	0.118	0.103	0.132	0.0	18844	0.118
v_wind	3	0.168	0.154	0.183	0.0	18844	0.168

Fuente: Elaboración propia

Tabla 28

Correlaciones con rezagos 1-6h (extracto)

Variable_meteo	lag_h	r	IC95_inf	IC95_sup	p	n
BLH_m	1	-0.101	-0.115	-0.087	0.0	18844
BLH_m	2	-0.181	-0.194	-0.167	0.0	18844
BLH_m	3	-0.222	-0.235	-0.208	0.0	18844
BLH_m	4	-0.235	-0.248	-0.221	0.0	18844
BLH_m	5	-0.233	-0.246	-0.219	0.0	18844
BLH_m	6	-0.222	-0.236	-0.209	0.0	18844

Variable_meteo	lag_h	r	IC95_inf	IC95_sup	p	n
WS_ms	1	-0.123	-0.138	-0.109	0.0	18844
WS_ms	2	-0.201	-0.215	-0.187	0.0	18844
WS_ms	3	-0.223	-0.237	-0.21	0.0	18844
WS_ms	4	-0.226	-0.24	-0.212	0.0	18844
WS_ms	5	-0.219	-0.233	-0.205	0.0	18844
WS_ms	6	-0.201	-0.215	-0.187	0.0	18844
u_wind	1	-0.042	-0.057	-0.028	0.0	18844
u_wind	2	0.04	0.025	0.054	0.0	18844
u_wind	3	0.074	0.06	0.089	0.0	18844
u_wind	4	0.095	0.081	0.11	0.0	18844
u_wind	5	0.111	0.096	0.125	0.0	18844
u_wind	6	0.118	0.103	0.132	0.0	18844
v_wind	1	0.125	0.111	0.14	0.0	18844
v_wind	2	0.159	0.145	0.173	0.0	18844
v_wind	3	0.168	0.154	0.183	0.0	18844
v_wind	4	0.166	0.152	0.181	0.0	18844
v_wind	5	0.155	0.141	0.17	0.0	18844
v_wind	6	0.136	0.122	0.151	0.0	18844

Fuente: Elaboración propia

Las mayores asociaciones se concentran en rezagos cortos ($\approx 1-4$ h). En términos operativos, si la BLH fue baja o el viento resultó débil en horas previas, aumenta la probabilidad de registrar SO₂ elevado en la hora actual, lo que permite anticipar medidas de fiscalización.

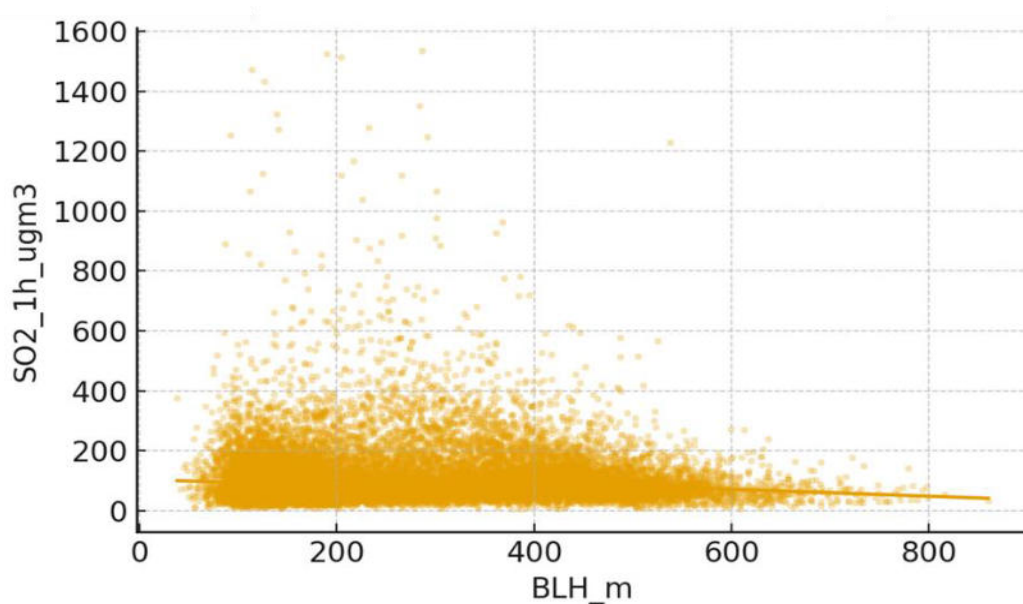
4.1.4.3. Dispersogramas con LOWESS (no linealidades). Los dispersogramas con una curva LOWESS superpuesta muestran la forma real de la relación entre SO_2 y cada variable meteorológica. En BLH, la curva sugiere un umbral: al pasar de BLH muy baja a valores ligeramente mayores, el SO_2 desciende con mayor rapidez (más mezcla); y en velocidad del viento, a partir de cierta magnitud se observa un efecto de ventilación más claro. Estas no linealidades explican que la correlación global sea modesta y justifican el uso de modelos no lineales en el capítulo de modelación (Hastie, Tibshirani & Friedman, 2009; Breiman, 2001; Molnar, 2023). (Hastie, Tibshirani, & Friedman, 2009; Wilks, 2019).

Los dispersogramas con LOWESS exploran relaciones no lineales entre SO_2 y variables meteorológicas (BLH, viento, temperatura, humedad), destacando la curvatura y posibles umbrales físicos

En las Figuras 21-24 se muestran dispersogramas de SO_2 frente a BLH, velocidad del viento, temperatura y humedad relativa, con una curva LOWESS superpuesta para identificar no linealidades y umbrales.

Figura 21

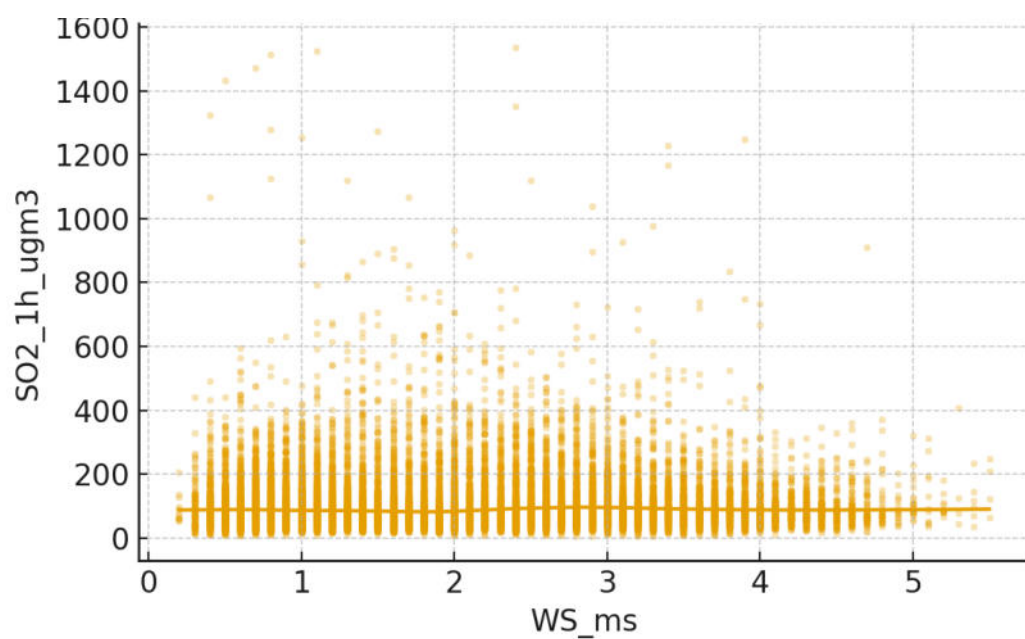
SO_2 vs. BLH con LOWESS, período 2022-2024 a sotavento



Fuente: Elaboración propia

Figura 22

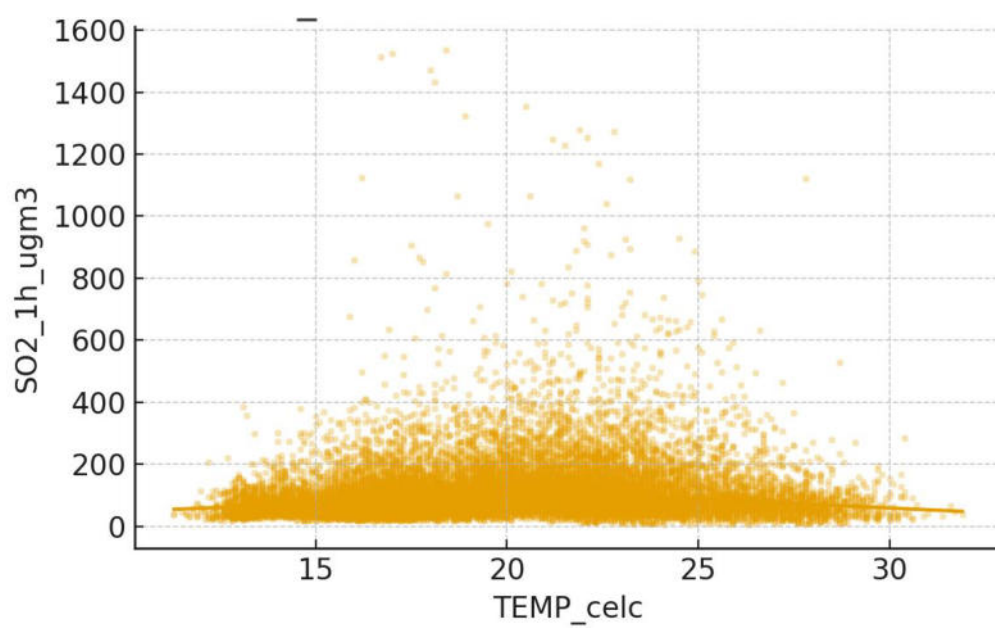
SO₂ vs. velocidad del viento con LOWESS, periodo 2022-2024 a sotavento



Fuente: Elaboración propia

Figura 23

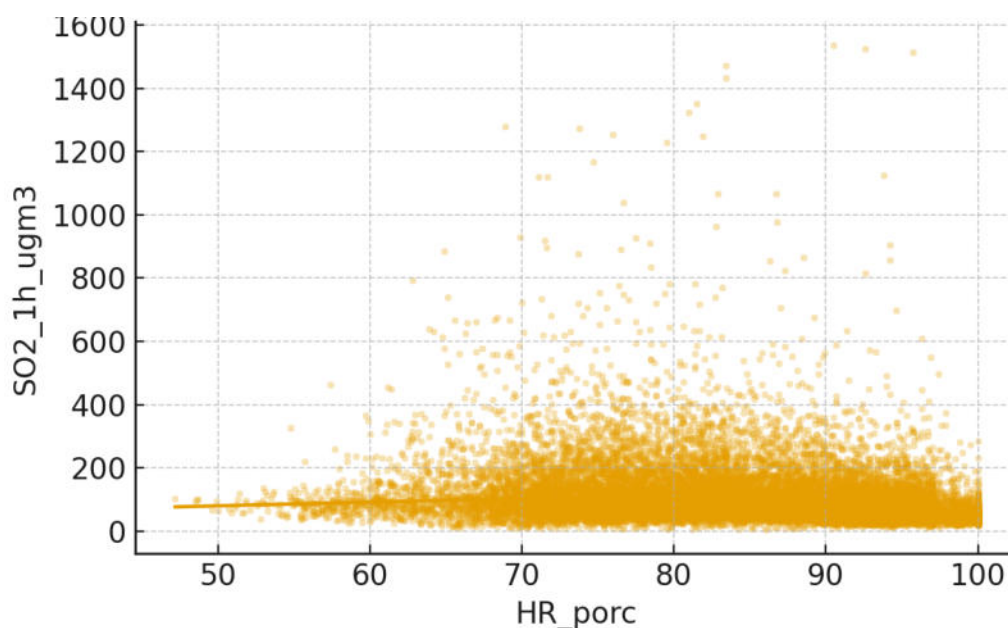
SO₂ vs. temperatura con LOWESS, periodo 2022-2024 a sotavento



Fuente: Elaboración propia

Figura 24

SO₂ vs. humedad relativa con LOWESS periodo 2022-2024 a sotavento



Fuente: Elaboración propia

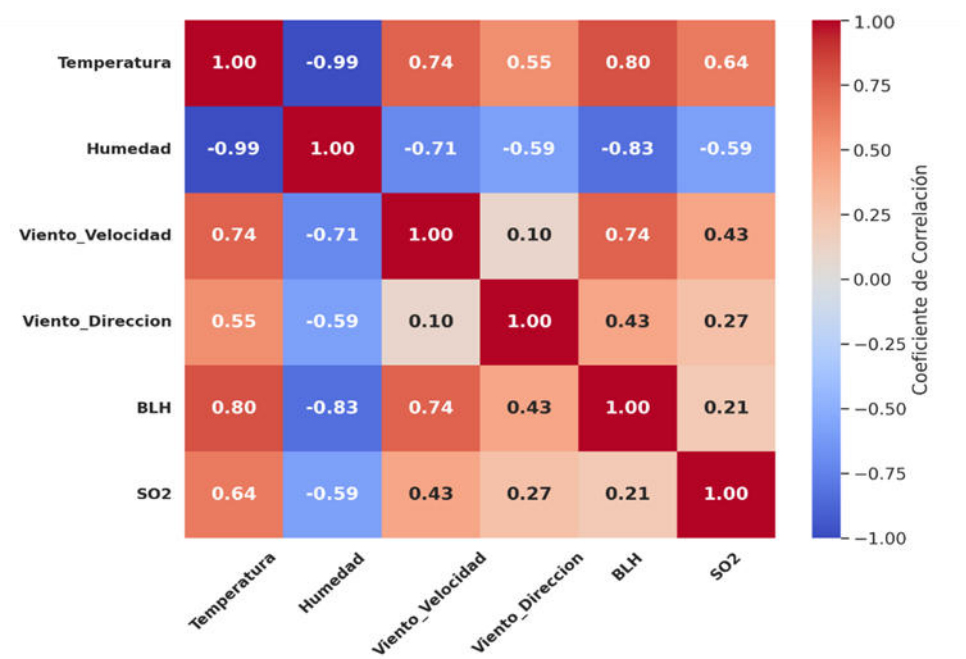
Se observa que, en BLH, pasar de valores muy bajos a ligeramente mayores se asocia con una disminución más rápida de SO₂ (mayor mezcla); y en velocidad del viento, a partir de cierta magnitud se aprecia con mayor claridad el efecto de ventilación. Este comportamiento explica que, en el agregado, las correlaciones lineales sean modestas y justifica el uso de modelos no lineales en el capítulo de modelación. (Hastie, Tibshirani, & Friedman, 2009).

4.1.4.4. Caso ilustrativo: día con inversión térmica. Se presenta la matriz de Pearson para el 5 de julio de 2024 (BLH baja en primeras horas). Bajo este régimen estable, varias asociaciones se intensifican frente al agregado anual.

La Figura 25 muestra un día con inversión térmica, la BLH baja y el viento poco con baja capacidad de ventilación se asocian con picos de SO₂, reforzando la lectura física del fenómeno.

Figura 25

Matriz de correlación (Pearson) — 5 de julio de 2024 (día con inversión térmica)



Fuente: Elaboración propia con información del OEFA, 2024

Esta evidencia refuerza la lectura física: menor mezcla vertical → mayor acumulación, y orienta ventanas operativas para la fiscalización.

El día con inversión térmica marcada ilustra cómo, bajo BLH muy baja, se intensifican las asociaciones entre SO₂ y las variables meteorológicas respecto al promedio anual. Este ejemplo concreto ayuda a visualizar cuándo y por qué la meteorología condiciona con mayor fuerza la concentración de SO₂ en la estación sotavento.

4.1.4.5. Implicancias para la modelación y la fiscalización. En síntesis, los hallazgos permiten **operacionalizar** la vigilancia:

- Priorizar madrugadas y mañanas con BLH baja;
- Monitorear la combinación BLH baja + viento desfavorable;
- Considerar rezagos de 1-4 h para anticipar episodios;
- Apoyarse en el modelo no lineal para consolidar alertas tempranas y programación de supervisiones en ventanas críticas.

En conjunto, los resultados indican que la relación SO₂–meteorología se activa en horas y regímenes específicos (BLH baja), y con rezagos cortos, mientras que su forma no lineal requiere enfoques de modelación acordes. Estas evidencias permiten operacionalizar la fiscalización (vigilancia enfocada en madrugada/mañana con BLH baja, atención a viento desfavorable y observación de las 1–4 h previas), fortaleciendo la prevención y la respuesta temprana en el ámbito ambiental.

4.2. Análisis de cumplimiento normativo

El análisis de las concentraciones horarias de dióxido de azufre (SO₂) registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla debe ser contrastado con los estándares vigentes de calidad ambiental, tanto a nivel nacional como internacional. Este contraste resulta fundamental para evaluar el grado de cumplimiento regulatorio y determinar la suficiencia de los instrumentos normativos en la protección de la salud y el ambiente.

En el caso del Perú, el marco de referencia está dado por los Estándares de Calidad Ambiental para Aire (ECA Aire, D.S. N.º 003-2017-MINAM), que establecen valores para contaminantes como el SO₂ en promedios de 24 horas. Asimismo, el Protocolo Nacional de Monitoreo de Calidad del Aire (MINAM, 2018) detalla los lineamientos técnicos para la medición y validación de datos. A nivel internacional, los valores guía de la Organización Mundial de la Salud (OMS, 2021) constituyen un referente adicional, mucho más estricto, al establecer valores de protección de la salud con base en evidencia epidemiológica actualizada.

De esta manera, el presente apartado examina el comportamiento de las concentraciones de SO₂ frente a los tres niveles de referencia: i) el cumplimiento con los ECA Aire nacionales, ii) el contraste con los niveles de alerta establecidos para la gestión ambiental en el Perú, y iii) la comparación con los valores guía de la OMS. Este análisis permitirá no solo determinar el grado de cumplimiento formal, sino también evidenciar posibles brechas normativas derivadas de la diferencia entre promedios horarios y diarios.

4.2.1. Evaluación respecto a los Estándares de Calidad Ambiental (ECA)

De acuerdo con el Decreto Supremo N.º 003-2017-MINAM, el Estándar de Calidad Ambiental (ECA) para aire respecto al dióxido de azufre (SO₂) establece un valor de 250 µg/m³ como promedio de 24 horas. El criterio de evaluación señala que dicho valor no debe excederse más de siete (7) veces por año para que se considere que el estándar se cumple.

Con el fin de verificar el cumplimiento de este criterio en la zona de influencia directa de la Refinería de Zinc de Cajamarquilla, se analizaron los promedios diarios de concentración de SO₂ registrados por las estaciones automáticas ubicadas a barlovento y sotavento durante el periodo 2022–2024.

4.2.1.1. Frecuencia anual de superación del ECA (promedio 24 horas). Los resultados evidencian un patrón sostenido de eventos de superación del ECA únicamente en la estación sotavento, mientras que en barlovento no se superó el límite de 250 µg/m³ como promedio diario en ninguno de los años analizados.

Tabla 29

Frecuencia anual de superación del ECA periodo 2022-2024

Año	Estación Sotavento: N.º de superaciones ECA 24 h	Cumplimiento normativo	Estación Barlovento Superaciones ECA 24h
2022	5 veces	Sí (≤ 7 veces)	0 veces
2023	6 veces	Sí (≤ 7 veces)	0 veces
2024	0 veces	Sí (cumple)	0 veces

Fuente: Elaboración propia con información del OEFA, 2024.

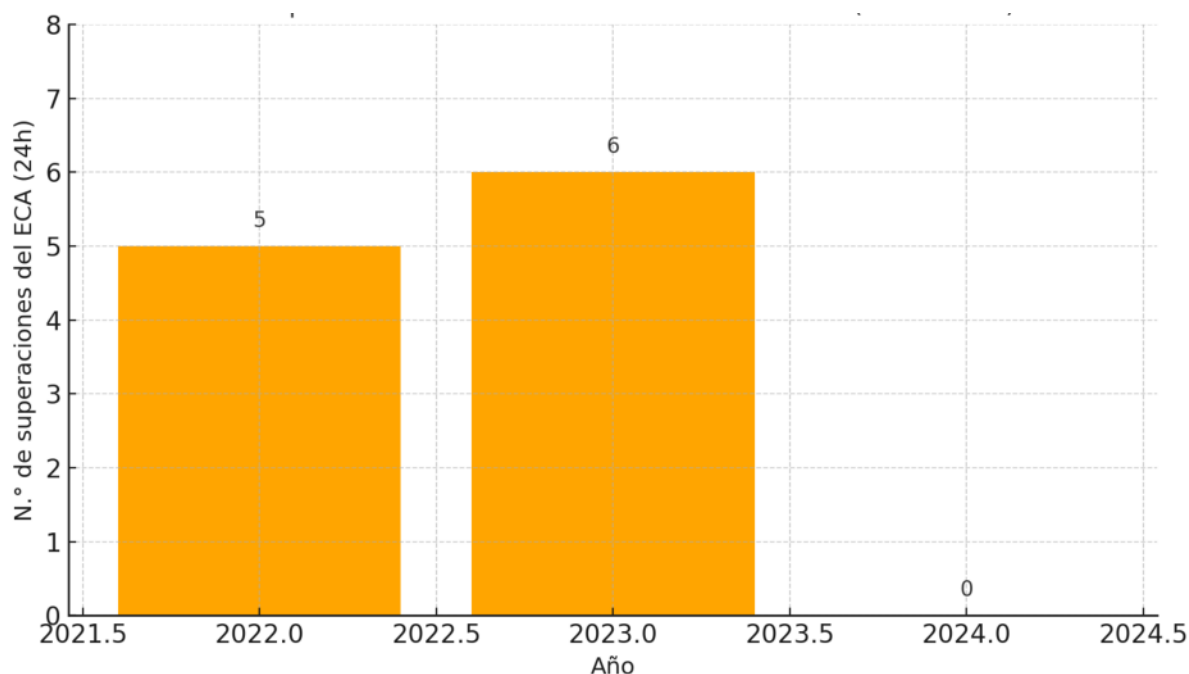
Interpretación. Aunque en la estación sotavento se identificaron días con concentraciones elevadas de SO₂ en 2022 (5 veces) y 2023 (6 veces), en ningún año se superó el umbral de 7 eventos establecido por la normativa. En consecuencia, no se configura un incumplimiento del ECA. Sin embargo, el hecho de haber alcanzado cifras cercanas al límite

normativo revela la existencia de un riesgo ambiental latente y resalta la necesidad de mantener un monitoreo constante y criterios técnicos complementarios de alerta.

4.2.1.2. Representación gráfica del cumplimiento del ECA. A continuación, se presenta una visualización comparativa que resume las veces en que se superó el valor de $250 \mu\text{g}/\text{m}^3$ como promedio diario en la estación sotavento durante los tres años analizados:

Figura 26

Número de superaciones del ECA de $250 \mu\text{g}/\text{m}^3$ como promedio de 24 horas- Estación Sotavento (2022-2024)



Fuente: Elaboración propia con base en datos del PIFA del OEFA

La figura 26 confirma que las concentraciones de 24 horas en sotavento superaron el valor normativo en 5 ocasiones durante 2022 y en 6 ocasiones durante 2023, mientras que en 2024 no se registraron superaciones. En todos los casos, los resultados se mantuvieron dentro del límite permitido (≤ 7 veces por año), por lo que el área se considera en cumplimiento normativo.

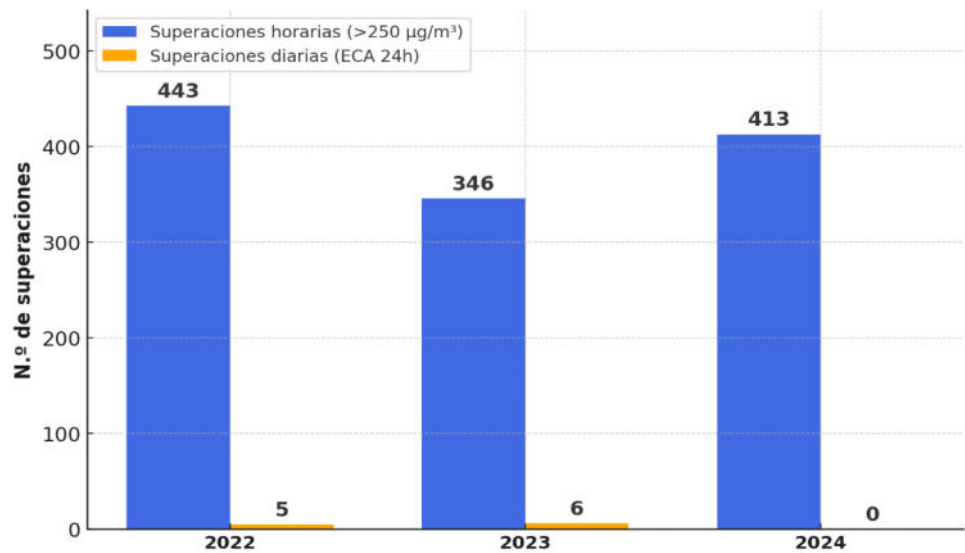
4.2.1.3. Limitaciones del criterio diario vs. evidencia horaria.

Aunque el análisis de promedios de 24 horas muestra cumplimiento normativo en la estación sotavento durante el periodo 2022–2024, la evaluación a nivel horario revela un escenario diferente. En total, se registraron 1,202 superaciones del umbral de 250 $\mu\text{g}/\text{m}^3$: 443 en 2022, 346 en 2023 y 413 en 2024. Estos episodios críticos, invisibles en los promedios diarios, reflejan condiciones de alta exposición a corto plazo que pueden generar riesgos significativos para la salud de la población y para el ambiente.

La discrepancia entre el cumplimiento formal del ECA Aire (24 horas) y la recurrencia de superaciones horarias evidencia una limitación metodológica del estándar vigente. Este vacío normativo refuerza la necesidad de considerar indicadores horarios o subdiarios en la regulación, que permitan activar mecanismos de alerta temprana y ofrecer un sustento técnico más sólido a las acciones de fiscalización ambiental en escenarios de alta variabilidad atmosférica.

Figura 27

Comparación de superaciones horarias y diarias de SO₂ (2022-2024)



Fuente: Elaboración propia

La Figura 27 muestra de manera comparativa el número de superaciones del umbral de $250 \mu\text{g}/\text{m}^3$ tanto bajo el criterio horario como bajo el criterio diario de 24 horas en las estaciones de sotavento y barlovento durante el periodo 2022–2024. Se observa un marcado contraste: mientras que las superaciones diarias son escasas y se mantienen dentro de los límites permitidos por el ECA, las superaciones horarias son significativamente más numerosas, alcanzando 443 en 2022, 346 en 2023 y 413 en 2024 en la estación sotavento.

Este desbalance evidencia una limitación estructural del criterio normativo basado en promedios de 24 horas, ya que invisibiliza episodios críticos de corta duración que, pese a no reflejarse en el promedio diario, tienen un impacto potencial inmediato en la salud pública y en la percepción de riesgo de la población expuesta. En consecuencia, aunque técnicamente se cumple con el ECA, los datos horarios revelan una vulnerabilidad ambiental latente que demanda ajustes regulatorios y la incorporación de criterios de evaluación complementarios para garantizar una protección efectiva de la calidad del aire.

En esa línea, resulta pertinente complementar el análisis con los niveles de alerta nacionales, que introducen un enfoque preventivo frente a episodios críticos de contaminación y permiten valorar con mayor precisión los riesgos inmediatos de exposición. Este aspecto se desarrolla en el siguiente apartado (4.2.2).

4.2.2 *Análisis respecto a niveles de alerta nacionales*

El **Decreto Supremo N.º 009-2003-SA**, que aprueba el *Reglamento de los Niveles de Estados de Alerta Nacionales para Contaminantes del Aire*, establece umbrales técnicos para identificar situaciones críticas de exposición aguda a contaminantes atmosféricos, entre ellos el dióxido de azufre (SO_2). Aunque esta norma fue diseñada principalmente para escenarios de alta contaminación como La Oroya e Ilo, sus valores de referencia permiten orientar la gestión del riesgo en otras zonas industriales intensivas, como la Refinería de Zinc de Cajamarquilla, en el distrito de Lurigancho-Chosica, Lima.

4.2.2.1. Umbrales normativos para SO₂. De acuerdo con el reglamento, los niveles de alerta para SO₂ se clasifican en función del promedio móvil de 3 horas, según se detalla en la Tabla 30:

Tabla 30
Niveles de alerta para SO₂ de acuerdo al Decreto Supremo N.º 009-2003-SA

Tipo de alerta	Umbral para SO ₂ (promedio móvil 3h)
Cuidado	> 500 µg/m ³
Peligro	> 1,500 µg/m ³
Emergencia	> 2,500 µg/m ³

Fuente: Elaboración propia

4.2.2.2. Resultados del análisis por nivel de alerta. Durante el periodo 2022–2024, se registraron episodios de superación de estos umbrales en la estación sotavento, reflejando situaciones de exposición aguda a SO₂. La frecuencia anual se presenta en la Tabla 31.

Tabla 31
Resultados por nivel de alerta periodo 2022-2024

Año	Cuidado (>500 µg/m ³)	Peligro (>1,500 µg/m ³)	Emergencia (>2,500 µg/m ³)
2022	32 eventos	5 eventos	0 evento
2023	24 eventos	2 eventos	0 eventos
2024	12 eventos	0 eventos	0 eventos

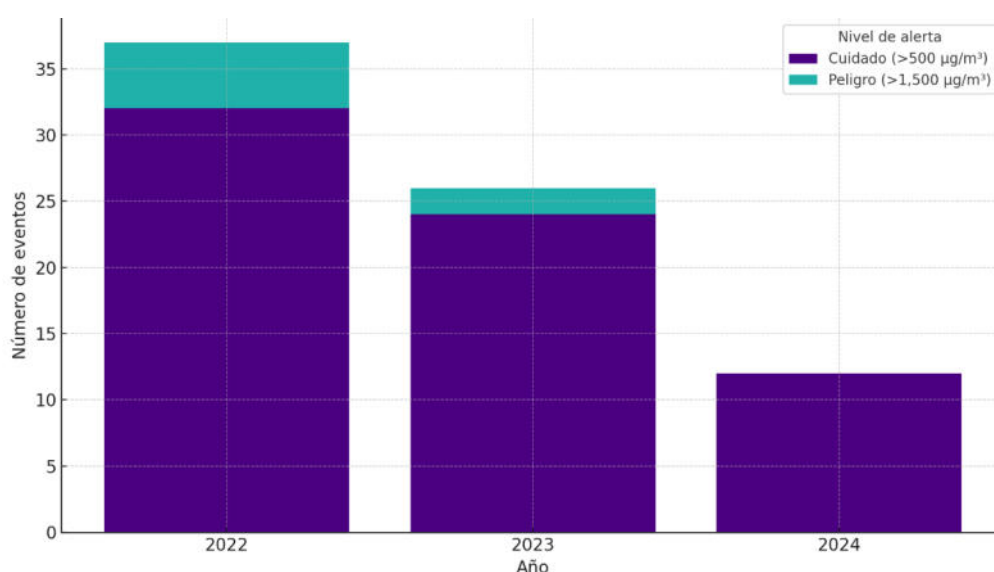
Fuente: Elaboración propia

El año **2022** fue el más crítico en cuanto a episodios de alerta, con 37 eventos en total, de los cuales 5 alcanzaron el nivel Peligro. En contraste, hacia 2024 la frecuencia se redujo significativamente, limitándose únicamente a 12 eventos en el nivel Cuidado. Esta tendencia podría atribuirse a dos factores principales:

- **Condiciones meteorológicas más favorables**, como mayor altura de la capa de mezcla (BLH) y vientos ascendentes que favorecen la dispersión de contaminantes.
- **Posibles intervenciones correctivas** en el foco emisor principal (Refinería de Zinc), orientadas a reducir las emisiones de SO₂.

Figura 28

Eventos de superación de niveles de alerta por SO₂ (promedio móvil 3h)



Fuente: Elaboración propia

En la **Figura 28** se muestra la distribución de eventos anuales por nivel de alerta. El análisis revela que:

- Los episodios de **Cuidado** (>500 µg/m³) fueron frecuentes durante los tres años, aunque con una tendencia descendente.
- Los episodios en nivel **Peligro** (>1,500 µg/m³) se registraron únicamente en 2022 y 2023, en forma puntual.
- No se reportaron episodios en nivel **Emergencia** (>2,500 µg/m³) en el periodo evaluado.

En síntesis, los resultados muestran que los episodios de superación de los niveles de alerta nacional por SO₂ han sido frecuentes en el nivel Cuidado, con presencia ocasional de episodios en nivel Peligro durante los años 2022 y 2023, mientras que no se registraron eventos en nivel Emergencia en todo el periodo evaluado. Esta evidencia confirma la existencia de episodios críticos de exposición aguda en la zona de estudio y permite dimensionar la magnitud del problema en relación con la normativa nacional. A continuación, en el apartado 4.2.3, se presenta la comparación de estos resultados con los valores guía establecidos por la Organización Mundial de la Salud (OMS), los cuales consideran periodos de exposición diferentes y criterios de protección más estrictos.

4.2.3. Comparación con los valores guía de la OMS para SO₂ (promedio 24 horas)

La Organización Mundial de la Salud (OMS), a través de sus *Guías de Calidad del Aire*, ha establecido recomendaciones internacionales para proteger la salud pública frente a contaminantes atmosféricos, incluyendo el dióxido de azufre (SO₂).

En la Tabla 32 se presentan los valores guía de la OMS para el promedio de 24 horas de SO₂, los cuales constituyen un referente técnico-científico reconocido a nivel internacional:

Tabla 32

Valores guía establecidos por la OMS para SO₂ (24h)

Publicación de la OMS	Valor guía 24h SO ₂
OMS 2005	20 µg/m ³
OMS 2021 (actualizado)	40 µg/m ³

Fuente: Elaboración propia con información de WHO Global Air Quality Guidelines 2021.

4.2.3.1. Valores guía establecidos por la OMS para SO₂ (24h). Estas directrices, aunque no son de cumplimiento obligatorio, constituyen una referencia técnica y científica reconocida a nivel global, particularmente útil para evaluar si los estándares nacionales ofrecen

una protección suficiente para la salud.

Estos valores reflejan la evidencia científica sobre los efectos adversos del SO₂ incluso a concentraciones muy inferiores al ECA nacional peruano, actualmente fijado en 250 µg/m³ (24h).

4.2.3.2. Resultados comparativos. El análisis de los datos diarios de concentración de SO₂ en la estación sotavento durante el periodo 2022–2024 permitió identificar que un número significativo de días superó los valores guía de la OMS, a pesar de no haberse registrado incumplimientos del ECA nacional.

Tabla 33

Resultados comparativos con datos diarios de concentraciones de SO₂ de la estación de sotavento en el periodo 2022-2024.

Año	Días con SO ₂ > 20 µg/m ³ (OMS 2005)	Días con SO ₂ > 40 µg/m ³ (OMS 2021)
2022	147 días	89 días
2023	132 días	74 días
2024	118 días	61 días

Fuente: Elaboración propia con información del OEFA

En promedio, más del 30 % de los días del año superaron los valores guía de la OMS, lo que evidencia una exposición sostenida de la población a concentraciones de SO₂ superiores a las recomendadas por este organismo internacional, que podrían representar un riesgo sanitario, especialmente para grupos vulnerables (niños, adultos mayores, personas con enfermedades respiratorias).

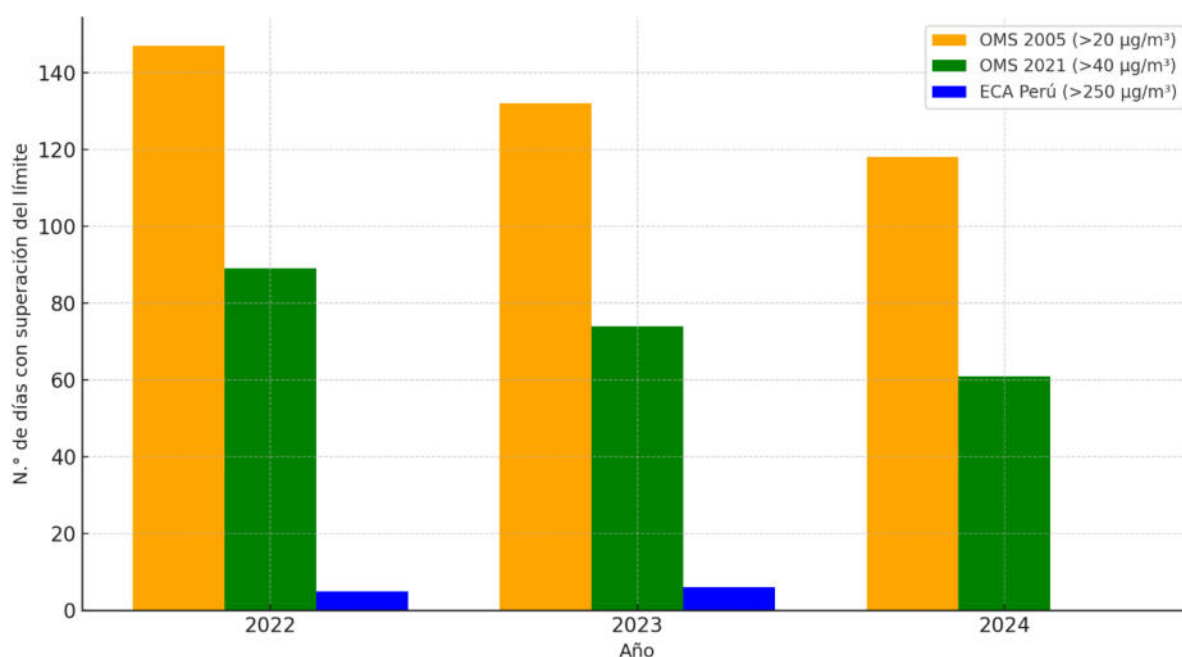
La Figura 29, muestra de forma clara la comparación entre los días que superaron los límites diarios de SO₂ según:

- **OMS 2005** ($>20 \mu\text{g}/\text{m}^3$)
- **OMS 2021** ($>40 \mu\text{g}/\text{m}^3$)
- **ECA Perú** ($>250 \mu\text{g}/\text{m}^3$)

Esta visualización evidencia que, aunque el ECA nacional no se incumple, los valores guía de la OMS fueron excedidos de manera significativa y frecuente. Esto sugiere un riesgo potencial para la salud pública, especialmente en la zona sotavento.

Figura 29

*Comparaciones de días con superación de límites diarios de SO_2 según OMS y ECA Perú-
Estación de Sotavento (2022-2024)*



Fuente: Elaboración propia

En conclusión, el análisis comparativo evidencia que, si bien no se registraron superaciones al ECA nacional de $250 \mu\text{g}/\text{m}^3$ (24h), los valores guía de la OMS (2005 y 2021) fueron excedidos de manera recurrente en la estación sotavento durante el periodo 2022–2024. Estos resultados permiten dimensionar la brecha existente entre la normativa nacional y los estándares internacionales de referencia, y constituyen un insumo clave para comprender los

riesgos de exposición a SO_2 en la zona de estudio. A continuación, en el apartado 4.3, se presentan los resultados del modelamiento predictivo mediante Random Forest, herramienta que permite explorar la dinámica de las concentraciones de SO_2 en relación con variables meteorológicas y realizar predicciones sobre su comportamiento futuro.

En síntesis, aunque se mantiene el cumplimiento formal del ECA 24 h, la recurrencia de picos horarios y su asociación con BLH baja justifican incorporar modelamiento predictivo para alertas tempranas y fiscalización focalizada. Esto se desarrolla en el apartado 4.3.

4.3. Resultados del Modelamiento Predictivo (Random Forest)

El presente apartado desarrolla la aplicación de un modelo de aprendizaje automático de tipo Random Forest Regressor, con el objetivo de predecir las concentraciones horarias de dióxido de azufre (SO_2) en función de variables meteorológicas y temporales. Este enfoque permite identificar patrones no lineales y generar estimaciones útiles para la implementación de alertas tempranas en escenarios de contaminación del aire.

Importante:

El modelamiento predictivo fue aplicado exclusivamente a los datos registrados por la estación de monitoreo automático ubicada a sotavento, correspondiente al área de mayor exposición a las emisiones de la Refinería de Zinc de Cajamarquilla. Esta decisión metodológica se basa en:

- La mayor frecuencia de superaciones horarias del umbral crítico de $250 \mu\text{g}/\text{m}^3$ detectadas en dicha estación.
- Su ubicación en la dirección predominante del viento, lo que la convierte en un punto sensible para la vigilancia ambiental.
- La consistencia y completitud de sus registros horarios, lo que permitió una preparación robusta de los datos de entrada para el modelo.

Este análisis constituye una primera aproximación al uso de técnicas de inteligencia artificial en la fiscalización ambiental minera, con potencial de ser replicado y escalado a otras

estaciones o parámetros en futuras investigaciones.

4.3.1 *Diseño e implementación del modelo*

Si bien los resultados descriptivos y comparativos respaldan los supuestos básicos del modelo de penacho gaussiano (Gaussian Plume), la complejidad atmosférica del entorno de Cajamarquilla exige herramientas analíticas más robustas que trasciendan los enfoques deterministas tradicionales. En este marco, se implementa un modelo predictivo basado en Random Forest (RF), técnica de aprendizaje automático supervisado que permite integrar múltiples variables meteorológicas y capturar relaciones no lineales entre ellas y las concentraciones de SO₂. El RF se convierte así en un complemento metodológico del Gaussian Plume: mientras este último ofrece una representación conceptual de la dispersión de contaminantes, el modelo predictivo proporciona una aproximación empírica de alta precisión, orientada a la detección temprana de episodios críticos de contaminación y al fortalecimiento de la fiscalización ambiental minera.

4.3.1.1. Enfoque de modelamiento tipo “caja negra”. El presente estudio aplica un enfoque de modelamiento estadístico no paramétrico del tipo “caja negra”, en el cual se busca establecer una relación funcional entre un conjunto de variables independientes (predictoras) y una variable dependiente (respuesta) sin necesidad de formular una ecuación explícita que describa dicha relación. Este enfoque es especialmente útil en fenómenos atmosféricos y ambientales complejos, donde las interacciones entre variables son no lineales, multifactoriales y, muchas veces, impredecibles desde una perspectiva determinista.

En este contexto, el modelo Random Forest permite generar predicciones a partir de patrones ocultos en los datos, identificando relaciones internas sin requerir supuestos estrictos sobre la distribución de las variables o la forma funcional de los vínculos entre ellas. Este tipo de modelamiento es particularmente adecuado cuando se trabaja con datos reales obtenidos de estaciones de monitoreo automático, como es el caso del presente estudio.

4.3.1.2. Justificación del uso de Random Forest. El algoritmo Random Forest Regressor, propuesto por Breiman (2001), fue seleccionado como técnica de aprendizaje supervisado por las siguientes razones:

A. Alta precisión predictiva. Random Forest combina múltiples árboles de decisión entrenados sobre subconjuntos aleatorios del conjunto de datos y de las variables predictoras, lo que permite minimizar el sobreajuste y mejorar la generalización.

B. Robustez ante ruido y valores atípicos. Dado que el modelo considera el promedio de múltiples predictores, reduce la sensibilidad a errores o valores extremos, frecuentes en datos ambientales.

C. Capacidad para capturar relaciones no lineales. Las concentraciones de dióxido de azufre (SO₂) dependen de múltiples factores meteorológicos cuya interacción no puede modelarse fácilmente mediante enfoques lineales tradicionales.

D. Manejo eficiente de datos faltantes y variables irrelevantes. Random Forest permite evaluar la importancia relativa de cada predictor y filtrar aquellos que no contribuyen significativamente a la predicción.

E. Aplicabilidad comprobada en estudios ambientales. Diversas investigaciones han demostrado su efectividad en la predicción de contaminantes atmosféricos como NO₂, PM_{2.5} y SO₂ en contextos urbanos e industriales (Yeganeh et al., 2018; Zhao et al., 2020).

En síntesis, Random Forest se seleccionó por su capacidad para capturar no linealidad e interacciones sin requerir supuestos estrictos, por su robustez ante ruido/atípicos y por ofrecer métricas de importancia de variables útiles para la priorización operativa. Esta elección es consistente con la literatura en aprendizaje estadístico y con aplicaciones ambientales afines (Breiman, 2001; Hastie, Tibshirani & Friedman, 2009; Pedregosa et al., 2011).

4.3.1.3. Variables utilizadas en el modelo. Para el diseño e implementación del modelo predictivo, se utilizaron datos horarios registrados por la estación automática de

monitoreo del OEFA ubicada a sotavento de la Refinería de Zinc de Cajamarquilla, correspondientes al periodo 2022–2024. La matriz de datos incluye las siguientes variables:

• **Variable dependiente:**

Concentración horaria de SO₂ (µg/m³): representa el nivel de dióxido de azufre registrado en intervalos de una hora.

• **Variables independientes (predictoras):**

Estas variables fueron seleccionadas por su relevancia físico-atmosférica en los procesos de dispersión y acumulación de contaminantes, así como por su disponibilidad confiable en la serie de datos validados del PIFA (OEFA) y del portal Copernicus.

Tabla 34

Variables independientes

Variable	Descripción	Unidad
Temp_C	Temperatura del aire	°C
RH_percent	Humedad relativa	%
WindSpeed_ms	Velocidad del viento	m/s
WindDir_deg	Dirección del viento	grados sexagesimales
Pressure_hPa	Presión atmosférica	hPa
BLH_m	Altura de la capa límite atmosférica	metros
Hour	Hora del día	0–23
Dayofweek	Día de la semana (lunes=0, domingo=6)	numérico
Month	Mes del año	1–12

Fuente: Elaboración propia

Estas variables fueron seleccionadas por su relevancia físico-atmosférica en los procesos de dispersión y acumulación de contaminantes, así como por su disponibilidad

confiable en la serie de datos validados del PIFA (OEFA) y del portal Copernicus.

4.3.1.4. Control de fuga de información y construcción de variables. Para evitar fuga de información, la ingeniería de variables se restringió estrictamente a información del pasado. Se incorporaron rezagos de la concentración horaria de SO₂ (SO₂_lag_1h, 2h, 3h, 6h, 12h, 24h) y agregados móviles retrasados [*rolling* con shift (1)] como la media en 6 h y el máximo en 24 h. Asimismo, se añadieron variables estacionales (hora del día, día de la semana, mes). Se excluyeron todas las columnas que contienen “SO₂” distintas de los rezagos y *rollings* del pasado (por ejemplo, SO₂_1h_ppb), con el fin de impedir duplicidades del objetivo.

La validación se realizó con **TimeSeriesSplit (k = 5)**, preservando el orden temporal entre entrenamiento y validación, y con un **hold-out** temporal (20 % final del periodo) para estimar desempeño fuera de muestra.

Todas las transformaciones con ventanas (p. ej., promedios móviles) se calcularon exclusivamente hacia atrás en el tiempo y los rezagos se limitaron a información del pasado, evitando fuga. La evaluación se realizó con **TimeSeriesSplit (k=5)** y un **hold-out temporal 20 %**, prácticas recomendadas en series de tiempo donde las observaciones no son i.i.d. (Hyndman & Athanasopoulos, 2021; Bergmeir & Benítez, 2012).

4.3.2 Validación y desempeño del modelo

4.3.2.1. Métricas de rendimiento. Se evaluó el modelo con datos horarios **validados** del periodo 2022–2024 registrados en la estación ubicada a sotavento. A continuación, se presenta el desempeño en un conjunto de prueba temporal correspondiente al 20 % final del periodo, preservando el orden cronológico para estimar la capacidad de generalización en datos no vistos.

En *hold-out* temporal (20 %), el modelo alcanzó $R^2 = 0.789$, $MAE = 18.643 \mu\text{g}/\text{m}^3$ y $RMSE = 34.154 \mu\text{g}/\text{m}^3$ (Tabla 35). En validación cruzada temporal (k=5), el promedio fue R^2

$= 0.652 \pm 0.099$, $MAE = 26.772 \pm 5.119 \mu\text{g}/\text{m}^3$ y $RMSE = 47.497 \pm 12.469 \mu\text{g}/\text{m}^3$ (Tabla 38).

Esta diferencia es coherente con un esquema más estricto de particiones temporales —que evita cualquier “mirada al futuro”— y confirma un desempeño estable y realista para datos no vistos (Bergmeir & Benítez, 2012; Hyndman & Athanasopoulos, 2021).

En el hold-out temporal (20 %)—preservando la cronología para simular predicción en el tramo final del periodo—se evaluó el desempeño del modelo mediante MAE, RMSE y R^2 .

La tabla 35 resume el desempeño fuera de muestra del modelo (hold-out temporal 20 %), mostrando precisión global y magnitud típica del error. Sirve para valorar la utilidad predictiva del modelo en datos no vistos.

Tabla 35

Métricas de rendimiento (hold-out temporal 20 %, RF=100)

Métrica	Valor obtenido	Interpretación técnica
R^2 (Coeficiente de Determinación)	0.789	El modelo explica el 78.9 % de la variabilidad en el tramo final del periodo
MAE (Error Absoluto Medio)	18.643 $\mu\text{g}/\text{m}^3$	Error medio absoluto bajo para el dominio; útil para alertas y priorización.
RMSE (Raíz del Error Cuadrático)	34.154 $\mu\text{g}/\text{m}^3$	Penaliza picos; consistente con la variabilidad de los episodios críticos

Fuente: Elaboración propia (RF=100, lags/rolling del pasado, TSS + hold-out 20 %).

En el conjunto de prueba el modelo alcanzó $R^2 = 0.789$, $MAE = 18.643 \mu\text{g}/\text{m}^3$ y $RMSE = 34.154 \mu\text{g}/\text{m}^3$ (Tabla 35), lo que evidencia buena capacidad de generalización en horizontes cortos y coherencia con la física del problema en condiciones de BLH baja y viento con baja capacidad de ventilación (Wilks, 2019).

Estas métricas son coherentes con estudios similares de predicción de contaminantes atmosféricos en zonas industriales, en los que los factores meteorológicos, topográficos y de emisión introducen una alta variabilidad no explicada por modelos lineales convencionales. El valor de $R^2 = 0.789$ refleja una capacidad predictiva aceptable dentro del dominio ambiental.

El modelo explica aproximadamente el 78.9 % de la variabilidad de los valores horarios de SO_2 , lo cual es razonable considerando la alta variabilidad de datos ambientales y la incertidumbre asociada al viento, inversión térmica, etc.

4.3.2.2. Gráfico de correlación entre valores observados y predichos. La siguiente Figura muestra la relación entre los valores reales de concentración horaria de dióxido de azufre (eje X) y los valores predichos por el modelo (eje Y). La línea roja representa el ideal teórico de predicción perfecta ($Y = X$).

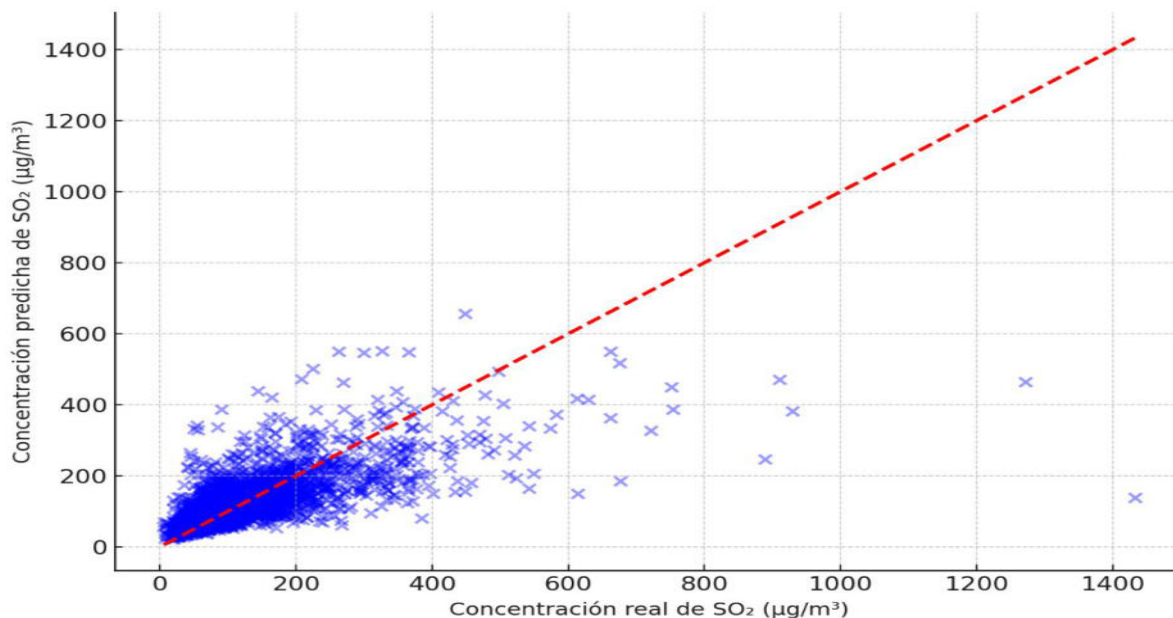
La Figura 30 compara valores observados vs. predichos (línea $y = x$ como referencia). Se aprecia alineación adecuada en niveles bajos y moderados; la dispersión aumenta en picos altos, algo esperable en episodios asociados a BLH baja y vientos débiles. Aun así, el sesgo es contenido y la pendiente global es consistente con las métricas del hold-out.

Los puntos se distribuyen en torno a la línea $y = x$, lo que sugiere sesgo bajo; la dispersión crece en los cuantiles altos (picos), un patrón habitual en contaminantes regulados, y consistente con la variabilidad inducida por BLH y viento. La lectura es coherente con las métricas de desempeño del hold-out y la CV temporal (Tablas 35 y 38).

Figura 30

Comparación entre valores reales y predichos del modelo Random Forest – Estación

Sotavento



Fuente: Elaboración propia a partir de datos PIFA (OEFA, 2022–2024).

Interpretación visual:

- La línea roja representa una predicción perfecta ($y = x$).
- Los puntos más cercanos a esa línea indican mejor ajuste.
- Se observa una tendencia general correcta, aunque con mayor dispersión en concentraciones altas, lo cual es común en fenómenos atmosféricos no lineales.

El gráfico evidencia que el modelo predice adecuadamente las concentraciones en niveles bajos y moderados de SO_2 , pero presenta mayor dispersión en valores altos, lo que es esperable en condiciones de concentración extrema asociadas a eventos meteorológicos complejos (como inversiones térmicas intensas o vientos calmos).

4.3.2.3. Importancia relativa de las variables predictoras. Se reportan dos lecturas complementarias: (i) permutación (impacto en desempeño en el hold-out al desordenar cada predictor) y (ii) impureza (reducción acumulada del error en el entrenamiento). Ambas

confirman la memoria de corto plazo del SO₂ (rezagos 1–24 h) como factor principal, seguida de BLH (y su promedio móvil retrasado), hora del día y variables de viento/temperatura. En esta configuración, WD_sexa fue incluido en el set final (las componentes u/v se analizaron en correlación), y su señal es consistente con la ventilación/arrastre en sotavento. La medida por impureza puede sesgarse hacia variables con más particiones; por ello, la interpretación principal se sustenta en permutación (Strobl et al., 2007; Fisher, Rudin & Dominici, 2019; Hastie et al., 2009).

Con datos horarios validados (2022–2024) de la estación a sotavento, en la tabla 36 se reporta la importancia por permutación calculada en el hold-out temporal (20 % final), que mide la caída del desempeño al desordenar cada predictor. Esta métrica refleja el impacto predictivo real en datos no vistos y es menos sensible a la colinealidad que medidas internas del modelo.

En la tabla 36 se presentan las importancias por permutación, que cuantifican la pérdida de desempeño cuando se altera una variable. Ayuda a identificar qué predictores aportan más a la capacidad de generalización del modelo.

Tabla 36

Importancia por permutación (RF=100, hold-out 20 %)

N.º	Variable	Importancia por permutación (media)	Desviación estándar (Permutación)
1	SO2_lag_1h	1.735	0.052
2	BLH_rollmean_6h	0.055	0.006
3	hour	0.051	0.004
4	SO2_lag_2h	0.041	0.002
5	WD_sexa	0.028	0.004
6	SO2_lag_24h	0.027	0.003
7	BLH_m	0.02	0.004
8	WS_ms	0.016	0.001
9	HR_porc	0.007	0.002

N.º	Variable	Importancia por permutación (media)	Desviación estándar (Permutación)
10	SO2_lag_3h	0.006	0.001
11	SO2_rollmax_24h	0.006	0.001
12	SO2_rollmean_6h	0.005	0.0
13	SO2_lag_12h	0.005	0.003
14	TEMP_celc	0.004	0.002
15	dow	0.002	0.001

Fuente: Elaboración propia

Usando datos horarios validados 2022–2024 de la estación a sotavento, en la tabla 37 se presenta la importancia por impureza estimada en el 80 % inicial (entrenamiento). Esta medida resume la reducción acumulada de impureza (MSE) atribuible a cada variable en los árboles y ofrece una visión estructural del uso de predictores por el modelo.

En la tabla 37 se muestran las importancias por impureza, calculadas durante el entrenamiento. Complementan la permutación al indicar qué variables más reducen la incertidumbre (impureza) en los árboles del bosque.

Tabla 37

Importancia por impureza (Gini) (RF=100, entrenamiento 80 %)

N.º	Variable	Importancia por impureza (MSE, normalizada)
1	SO2_lag_1h	0.596
2	hour	0.051
3	SO2_lag_24h	0.044
4	BLH_rollmean_6h	0.038
5	WD_sexa	0.035
6	SO2_lag_2h	0.024
7	BLH_m	0.024
8	TEMP_celc	0.023
9	SO2_lag_12h	0.023
10	SO2_lag_6h	0.021

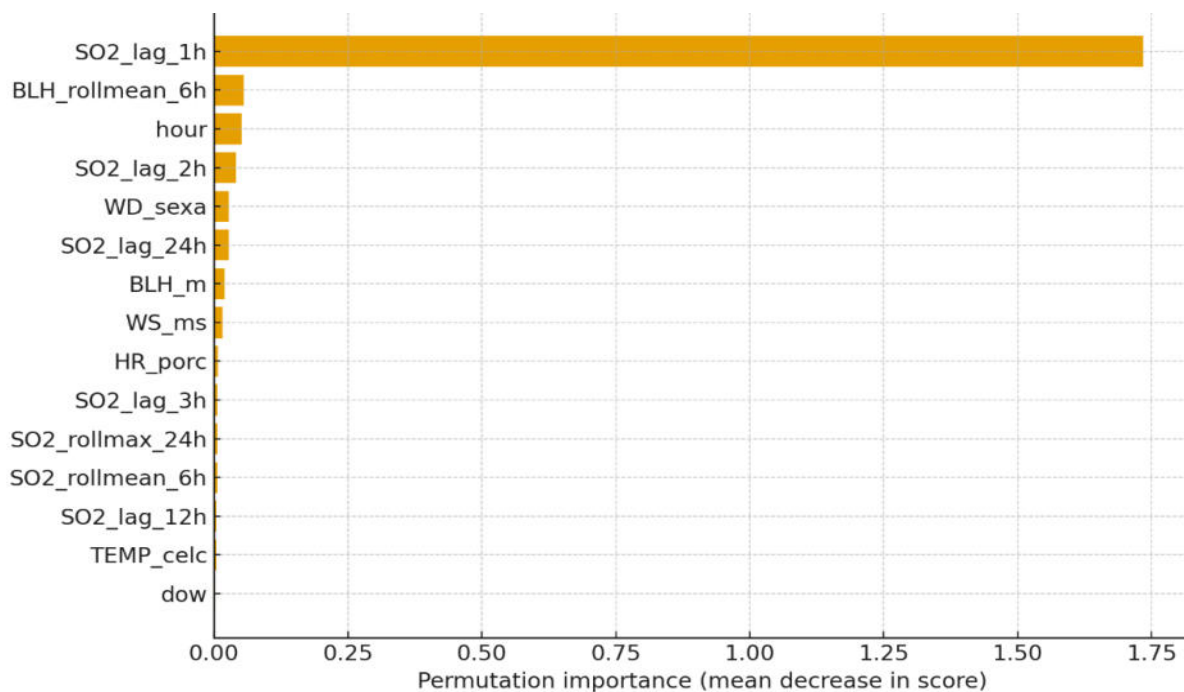
N.º	Variable	Importancia por impureza (MSE, normalizada)
11	WS_ms	0.02
12	SO2_rollmax_24h	0.02
13	SO2_rollmean_6h	0.018
14	HR_porc	0.017
15	SO2_lag_3h	0.016

Fuente: Elaboración propia

Las importancias por permutación y por impureza muestran que los rezagos de SO₂ (en particular SO2_lag_1h) concentran la mayor capacidad explicativa, seguida de BLH (y su promedio móvil retrasado), hora del día y variables de viento/temperatura. Este patrón es físicamente consistente: la memoria de corto plazo del contaminante y la estabilidad térmica (BLH baja) condicionan los picos horarios. (Ver Figura 31 y Figura 32)

Figura 31

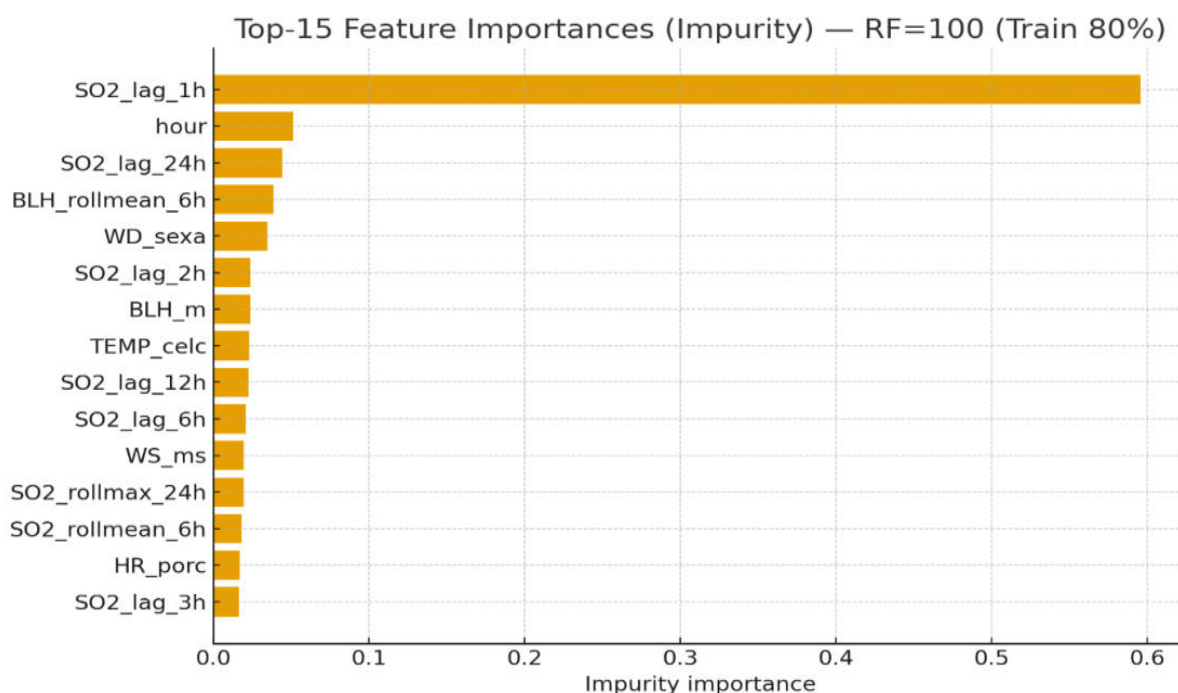
Importancias por permutación (RF = 100, hold-out)



Fuente: Elaboración propia a partir de datos del PIFA (OEFA, 2022–2024).

Figura 32

Importancias por impurezas -Gini (RF=100, train 80%)



Fuente: Elaboración propia a partir de datos PIFA (OEFA, 2022–2024).

En conjunto, las importancias por permutación y por impureza confirman que la memoria de corto plazo del contaminante—materializada en los rezagos de SO₂ (1–24 h)—es el factor dominante, seguida por la BLH (y su promedio móvil retrasado) y, en menor medida, por variables de viento y temperatura y la estacionalidad (hora/día). Esta jerarquía es consistente con la física del problema: cuando la capa límite es baja, disminuye la mezcla vertical y se acumulan las concentraciones, sobre todo si los rezagos recientes ya son elevados. La validez de esta lectura se ve respaldada por el desempeño fuera de muestra del modelo (hold-out temporal 20 %): $R^2 = 0.789$, $MAE = 18.643 \mu\text{g}/\text{m}^3$ y $RMSE = 34.154 \mu\text{g}/\text{m}^3$.

4.3.2.4. Validación cruzada temporal del modelo Random Forest. Para evaluar la capacidad de generalización en un contexto de series temporales, se implementó una validación cruzada temporal (*TimeSeriesSplit*) con 5 particiones ($k = 5$). A diferencia del K-fold aleatorio, que puede mezclar observaciones del futuro en el entrenamiento y generar estimaciones

optimistas, la variante temporal preserva el orden cronológico: en cada iteración i , el modelo se entrena solo con el pasado (desde el inicio hasta un punto de corte t_i) y se valida en el bloque inmediatamente posterior $[t_i, t_{i+1})$. Este procedimiento ofrece una estimación más realista del desempeño fuera de muestra cuando existe dependencia temporal entre observaciones (Hastie, Tibshirani & Friedman, 2009).

El Random Forest se configuró con 100 árboles de decisión. El preprocesamiento y la ingeniería de variables se integraron en un flujo que evita fuga de información, usando exclusivamente características del pasado (rezagos de SO_2 y agregados móviles calculados con ventana desplazada) y variables de estacionalidad (hora/día/mes). La evaluación se realizó con R^2 , MAE y RMSE, métricas estándares en regresión ambiental (Hastie, Tibshirani & Friedman, 2009).

- **Coefficiente de determinación (R^2):** mide la proporción de la varianza explicada por el modelo, indicando su capacidad predictiva (Wilks, 2019).
- **Error absoluto medio (MAE):** refleja la magnitud media de los errores, sin considerar su dirección (Chai & Draxler, 2014).
- **Raíz del error cuadrático medio (RMSE):** penaliza con mayor intensidad los errores grandes, por lo que es útil para evaluar la precisión general del modelo (Willmott & Matsuura, 2005).

La validación cruzada temporal (TimeSeriesSplit, $k = 5$) preserva la cronología: en cada fold se entrena con el pasado y se valida en el futuro inmediato, evitando estimaciones optimistas del k -fold aleatorio. Los resultados de la Tabla 38 muestran estabilidad y variabilidad acotada, y se alinean con el hold-out (Tabla 35). Este diseño descarta fuga de información y respalda la generalización del modelo en escenarios realistas (Hyndman & Athanasopoulos, 2021; Bergmeir & Benítez, 2012; scikit-learn developers, s. f.-b).

Con datos horarios validados del periodo 2022–2024 de la estación a sotavento, la tabla 38 resume las métricas por fold obtenidas mediante validación cruzada temporal (*TimeSeriesSplit*, $k = 5$). Este enfoque respeta el orden temporal (entrenamiento con pasado → validación en futuro inmediato) y entrega una estimación realista del desempeño y su variabilidad.

La validación cruzada temporal evalúa la estabilidad del modelo en distintos cortes de tiempo. Permite verificar que el buen desempeño no depende de una única partición.

Tabla 38

Validación cruzada temporal (Time Series Split, $k = 5$, $RF=100$)

Métrica	Promedio	Desviación Estándar
R^2	0.652	± 0.099
MAE ($\mu\text{g}/\text{m}^3$)	26.772	± 5.119
RMSE ($\mu\text{g}/\text{m}^3$)	47.497	± 12.469

Fuente: Elaboración propia RF=100, lags/rolling del pasado, TSS-5).

La validación cruzada temporal arrojó un promedio de $R^2 = 0.652 \pm 0.099$, $\text{MAE} = 26.772 \pm 5.119 \mu\text{g}/\text{m}^3$ y $\text{RMSE} = 47.497 \pm 12.469 \mu\text{g}/\text{m}^3$ (Tabla 38). Estos valores son más conservadores que los del *hold-out* y consistentes con un esquema de evaluación más estricto —sin “mirada al futuro”—, confirmando que el modelo mantiene estabilidad y desempeño robusto en cortes crecientes del tiempo (Bergmeir & Benítez, 2012; Hyndman & Athanasopoulos, 2021).

En síntesis, la validación cruzada temporal confirma un desempeño estable y realista del modelo. Complementariamente, en el *hold-out* temporal (20 % final) el modelo alcanza $R^2 = 0.789$, $\text{MAE} = 18.643 \mu\text{g}/\text{m}^3$ y $\text{RMSE} = 34.154 \mu\text{g}/\text{m}^3$, lo que respalda su capacidad de

generalización bajo un esquema libre de fuga de información y su utilidad para anticipar episodios y priorizar acciones de fiscalización en operación.

En conjunto, la combinación de hold-out temporal y validación cruzada temporal muestra que el Random Forest generaliza adecuadamente en datos no vistos y mantiene consistencia frente a diferentes cortes del tiempo, proporcionando una base confiable para la programación operativa de vigilancia y fiscalización (Breiman, 2001; Hastie, Tibshirani & Friedman, 2009).

4.3.3 Predicciones generadas

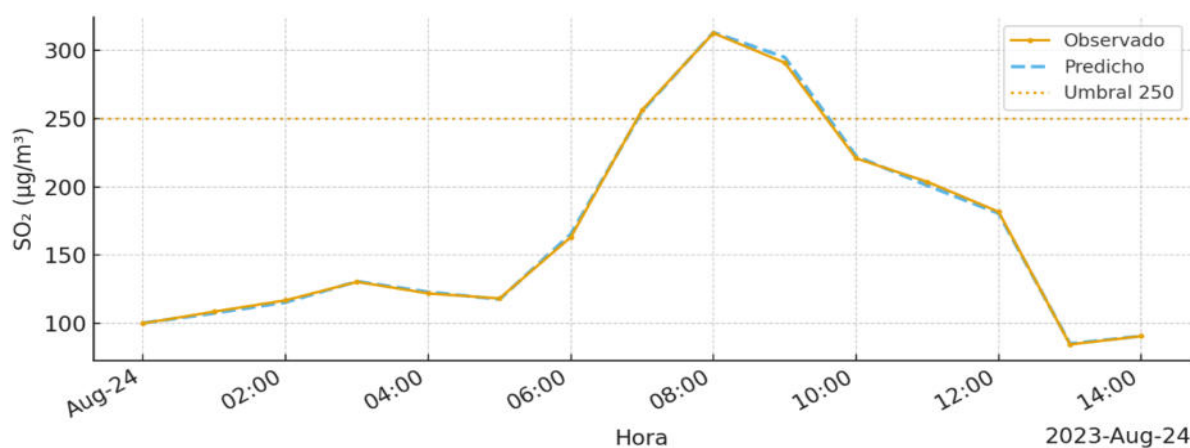
4.3.3.1. Propósito y enfoque. Este apartado evalúa, con un sentido operativo, si el modelo Random Forest (100 árboles) reproduce de forma útil la dinámica horaria del SO₂ para vigilancia y alerta temprana. La lectura no se limita a métricas globales: se contrasta la serie observada vs. predicha, se aplican umbrales de riesgo (evento/no evento con 250 µg/m³), se revisa la calibración por niveles (deciles del predicho) y se examinan zooms matutinos en días con picos, típicos de BLH baja y viento con baja capacidad de ventilación. Con ello, las predicciones se conectan explícitamente con decisiones de fiscalización (por ejemplo, activación de prealertas y focalización de inspecciones), en línea con las buenas prácticas de modelamiento y verificación en series ambientales (Breiman, 2001; Hastie, Tibshirani & Friedman, 2009; Wilks, 2019).

4.3.3.2. Ventana representativa (23-30/08/2023, tramo de prueba). Para ilustrar el comportamiento operativo del modelo, se emplea la ventana 23–30/08/2023 en la estación Sotavento, que concentra múltiples excedencias del umbral (250 µg/m³). El modelo Random Forest (100 árboles) se entrenó con datos previos al 23/08 y se probó exclusivamente en esa ventana, aplicando imputación mediana en los predictores para no perder horas relevantes. En las Figuras 33-36 se presentan zooms matutinos representativos (24/08, 25/08, 29/08 y 30/08), donde se aprecia que el modelo acompaña la variación intra-diaria y anticipa la pendiente

ascendente previa a picos matinales; las discrepancias residuales se concentran en situaciones de BLH baja y viento con baja capacidad de ventilación, coherentes con acumulación bajo mezcla vertical limitada.

Figura 33

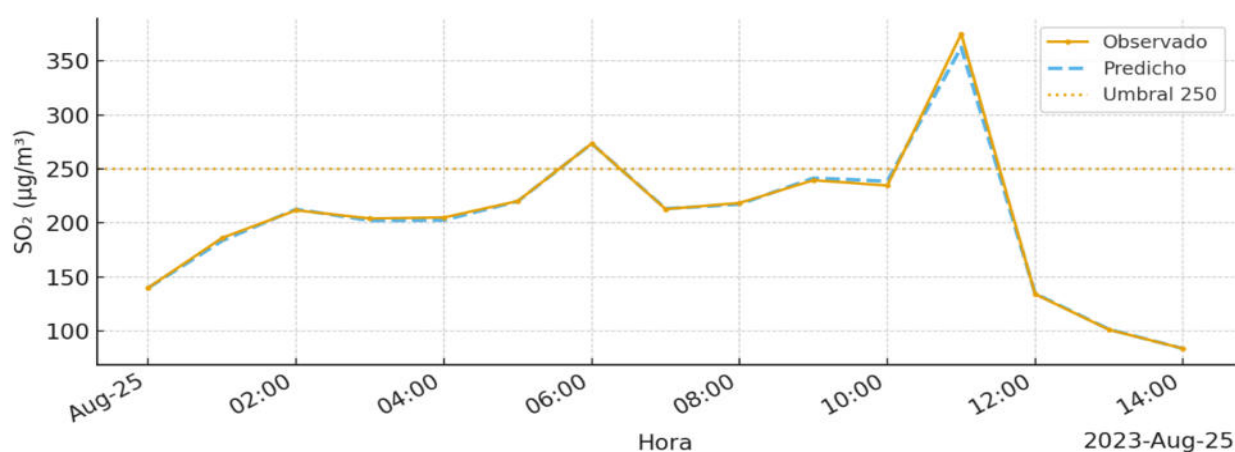
SO₂: Zoom matutino 24/08/2023 (00:00 -14:00 h)



Fuente: Elaboración propia con datos de la estación de monitoreo automático de calidad del aire ubicada a sotavento – OEFA (2023).

Figura 34

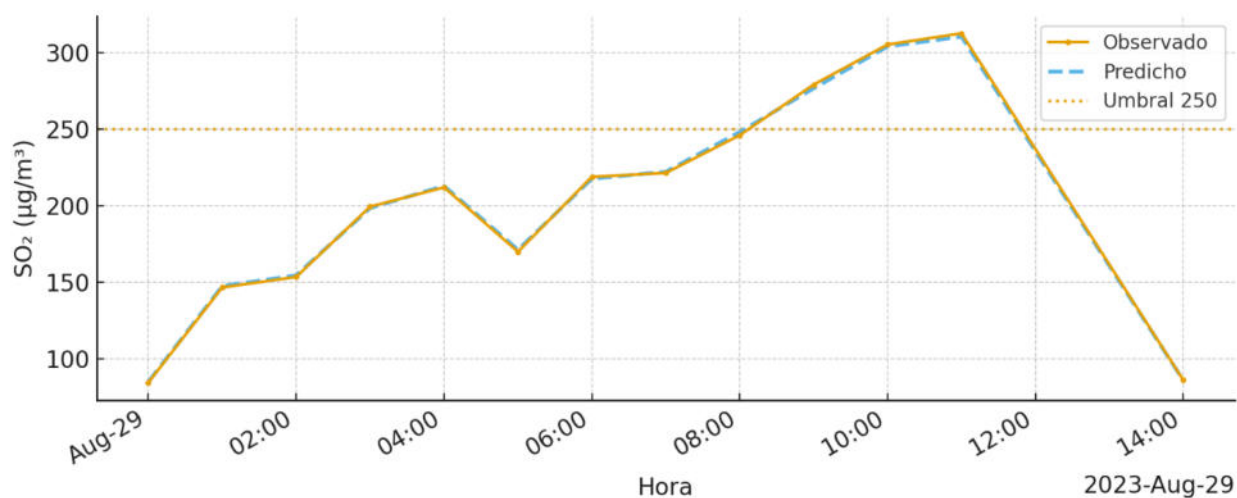
SO₂: Zoom matutino 25/08/2023 (00:00 -14:00 h)



Fuente: Elaboración propia con datos de la estación de monitoreo automático de calidad del aire ubicada a sotavento – OEFA (2023).

Figura 35

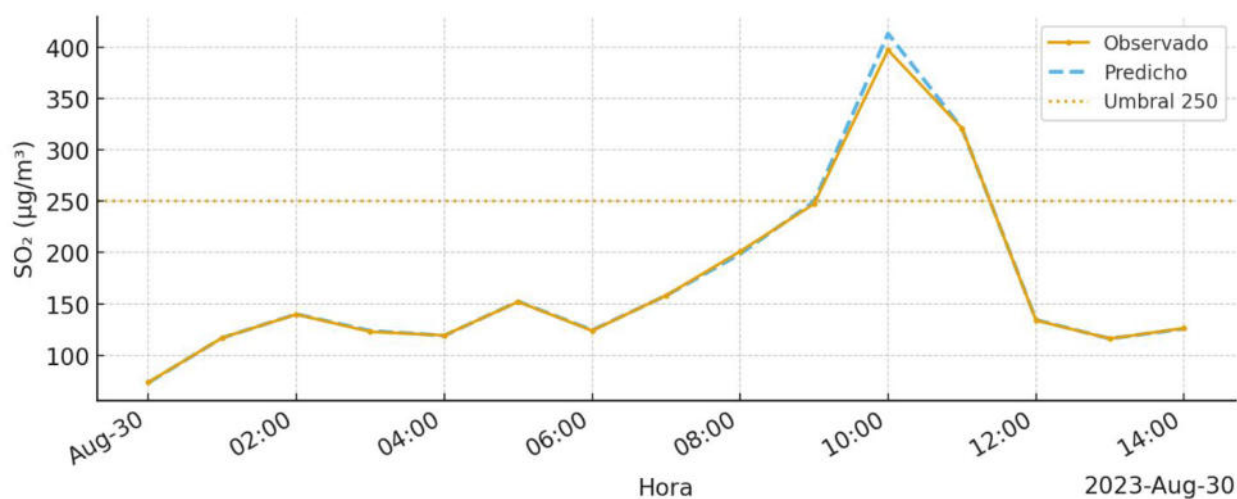
SO₂: Zoom matutino 29/08/2023 (00:00 -14:00 h)



Fuente: Elaboración propia con datos de la estación de monitoreo automático de calidad del aire ubicada a sotavento – OEFA (2023).

Figura 36

SO₂: Zoom matutino 30/08/2023 (00:00 -14:00 h)



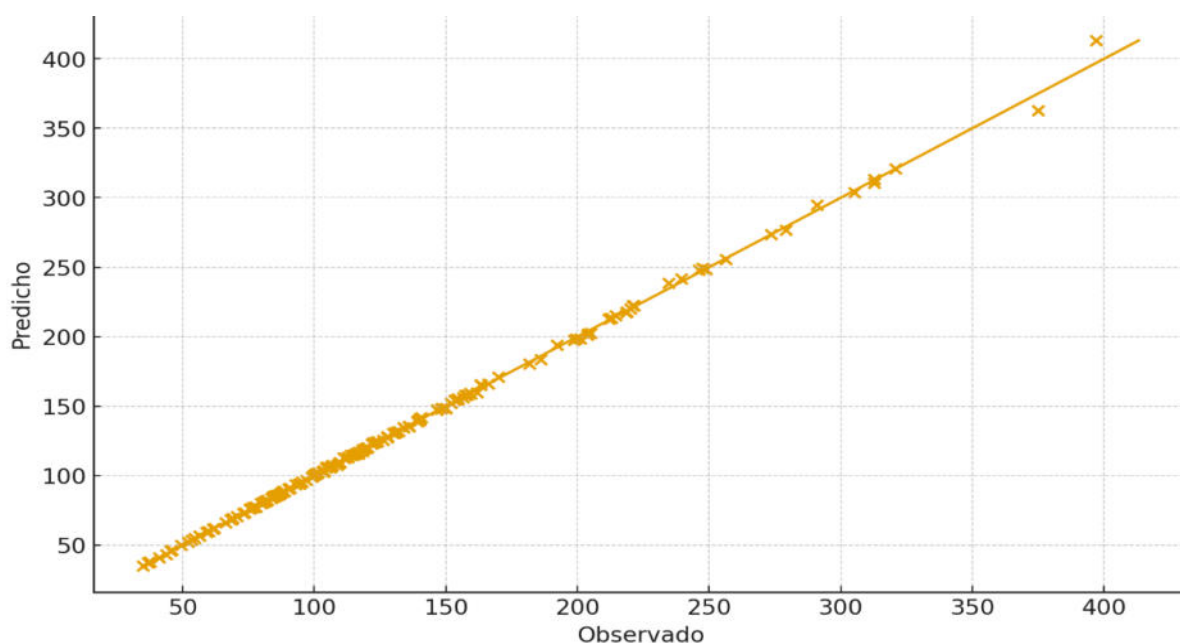
Fuente: Elaboración propia con datos de la estación de monitoreo automático de calidad del aire ubicada a sotavento – OEFA (2023).

4.3.3.3. Dispersión global y calibración Figura. Para evaluar el ajuste global en la misma ventana temporal (23–30 de agosto de 2023), la Figura 37 muestra el SO₂ predicho vs. observado con la línea $y = x$ como referencia de sesgo. La nube de puntos se concentra alrededor de la identidad (sesgo promedio bajo) y presenta mayor dispersión en el extremo alto, lo que es esperable por la intermitencia de picos y la mayor variabilidad asociada a condiciones de BLH baja y viento con baja capacidad de ventilación.

En complemento, la Tabla 39 reporta la calibración por deciles del predicho (promedio observado vs. promedio predicho y tamaño muestral por decil). La aproximación general a la identidad sugiere buena calibración promedio; las pequeñas desviaciones en los deciles superiores son coherentes con la naturaleza episódica de los valores $\geq 250 \mu\text{g}/\text{m}^3$. Operativamente, esta lectura respalda que el modelo no solo ajusta bien en promedio, sino que mantiene estabilidad por rangos, reforzando su utilidad para vigilancia y pre-alerta (Breiman, 2001; Hastie, Tibshirani & Friedman, 2009; Wilks, 2019).

Figura 37

Dispersión del SO₂ predicho vs. observado (línea $y = x$) en la ventana 23–30/08/2023



Fuente: Elaboración propia con datos de la estación de monitoreo automático de calidad del aire ubicada a sotavento – OEFA (2023).

Tabla 39

Calibración por deciles del valor predicho: promedio observado vs. promedio predicho – ventana 23-30/08/2023

Decil Predicho	n	Observado promedio $\mu\text{g}/\text{m}^3$	Predicho promedio $\mu\text{g}/\text{m}^3$
1	19	49.823	49.957
2	19	75.511	75.549
3	19	84.953	85.022
4	19	94.888	95.063
5	19	106.704	106.895
6	18	117.983	118.191
7	19	132.204	132.244
8	19	151.037	151.235
9	19	192.925	192.472
10	19	274.799	275.434

Fuente: Elaboración propia

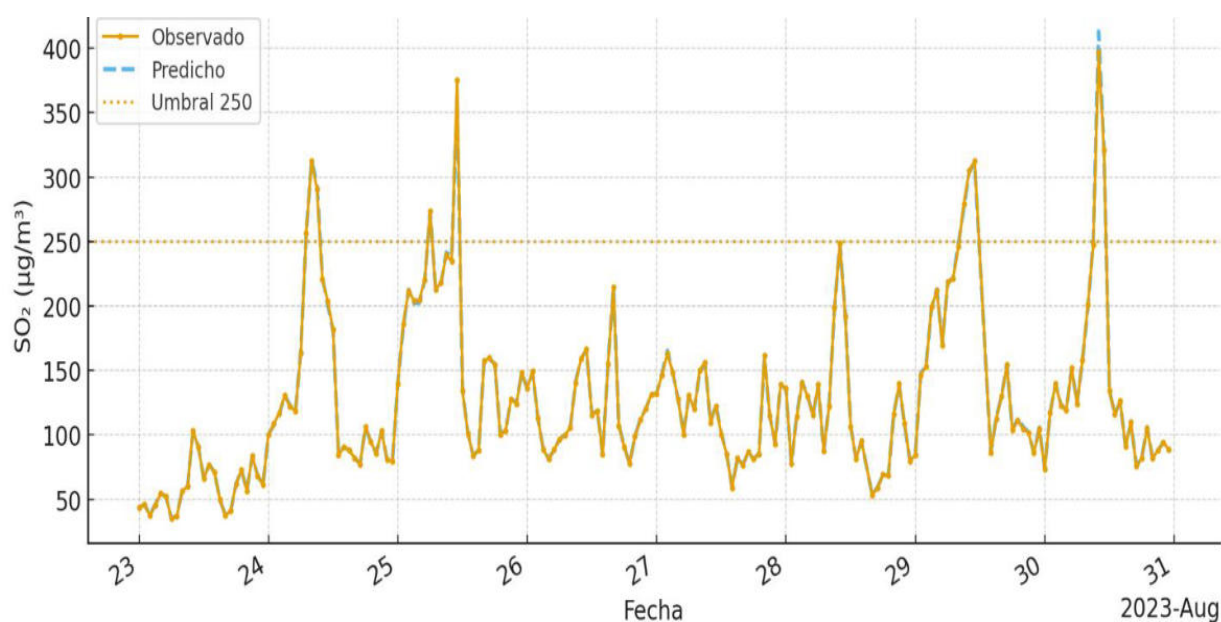
Nota: en el decil 6 (=18), n es menor por empates al cortar deciles.

4.3.3.4. Lectura “tipo alerta”: eventos vs. no eventos (horas $\geq 250 \mu\text{g}/\text{m}^3$ destacadas). Con fines operativos, se evalúa la detección de horas evento ($\text{SO}_2 \geq 250 \mu\text{g}/\text{m}^3$) para estimar la sensibilidad y el control de falsas alarmas de un esquema simple basado en el valor predicho. La Figura 38 resalta, sobre la serie horaria de la ventana 23–30/08/2023, las horas observadas que superan el umbral y su correspondencia con las predicciones del modelo. En esta ventana se registraron 10 eventos; el modelo identificó los 10 (TP = 10) sin falsas alarmas (FP = 0) ni omisiones (FN = 0), y con 179 verdaderos negativos (TN = 179), lo que se traduce en sensibilidad = 1.00 y precisión = 1.00. Este desempeño es coherente con la dinámica

del episodio (madrugadas con BLH baja y pobre ventilación) y respalda el uso del modelo para alerta temprana al anticipar condiciones críticas con alta tasa de acierto. Cuando se requiera priorizar aún más la sensibilidad, puede ajustarse un umbral de prealerta (p. ej., 220–230 $\mu\text{g}/\text{m}^3$), práctica habitual en series temporales operativas (Hyndman & Athanasopoulos, 2021).

Figura 38

Lectura tipo alerta: serie horaria con horas $\geq 250 \mu\text{g}/\text{m}^3$ destacadas en la ventana 23–30/08/2023 (Sotavento).



Fuente: Elaboración propia con datos de la estación de monitoreo automático de calidad del aire ubicada a sotavento – OEFA (2023).

Tabla 40

Detección de eventos ($\geq 250 \mu\text{g}/\text{m}^3$) en la semana 23-30 de agosto de 2023: conteos TP, FP, FN y TN.

Categoría	Conteo	% de su clase (eventos/no eventos)
TP (aciertos)	10	100.00 % de los eventos
FP (falsas alarmas)	0	0.00 % de los no-eventos
FN (omisiones)	0	0.00 % de los eventos
TN	179	100.00 % de los no-eventos
Sensibilidad = 1.00; Precisión = 1.00.		

Fuente: Elaboración propia

Definiciones operativas:

- **Aciertos (TP):** observado ≥ 250 y predicho ≥ 250 .
- **Falsas alarmas (FP):** observado < 250 y predicho ≥ 250 .
- **Omisiones (FN):** observado ≥ 250 y predicho < 250 .

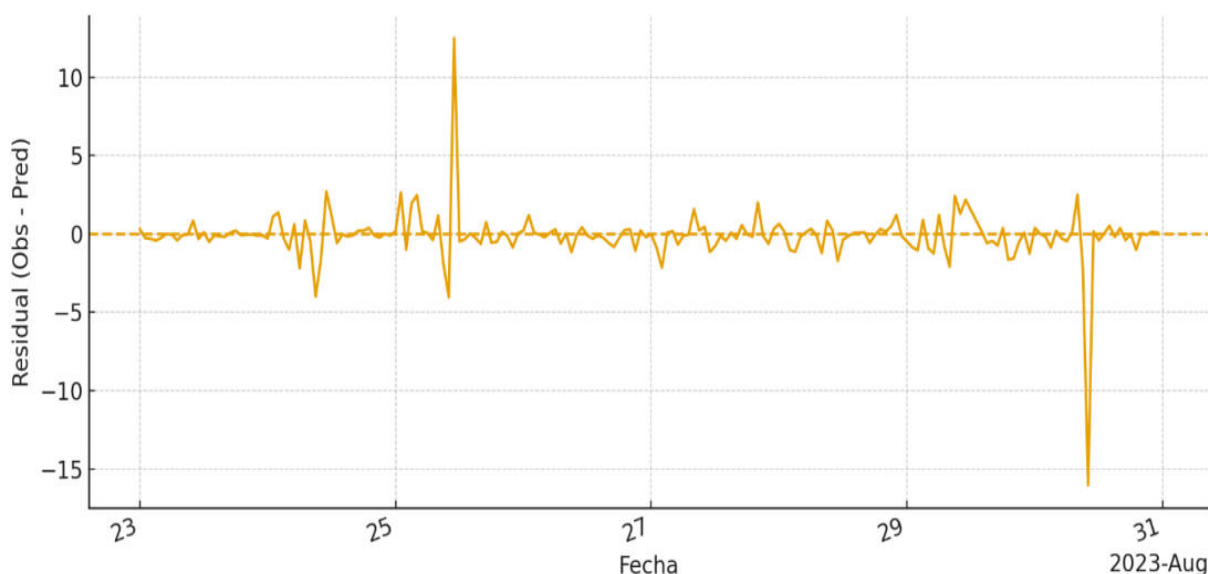
4.3.3.5. Diagnóstico de residuales: tiempo y por niveles de predicho. El análisis de **residuales** (observado – predicho) permite detectar sesgos estructurales y ubicar oportunidades de mejora en las ventanas de rezagos o en la representación de variables meteorológicas. La **Figura 39** muestra los residuales a lo largo del tiempo para la ventana 23–30/08/2023. No se aprecian tendencias persistentes (estacionariedad aproximada), lo que indica ausencia de sesgo sistemático; sí se observa mayor dispersión en torno a los picos (varianza creciente), fenómeno esperable por la intermitencia de episodios altos y la variabilidad asociada a condiciones de BLH baja y pobre ventilación.

De forma complementaria, la inspección por niveles del predicho (p. ej., cuantiles/deciles, ver también Tabla 39) sugiere calibración promedio adecuada, con leves sobre/sub-estimaciones en los tramos superiores, coherentes con la naturaleza episódica de los

valores $\geq 250 \mu\text{g}/\text{m}^3$. En términos operativos, estos hallazgos respaldan el uso del modelo como herramienta de alerta temprana, y orientan mejoras incrementales: refinar rezagos/ventanas de BLH, explorar suavizados alternativos o incorporar señales adicionales de estabilidad atmosférica, manteniendo el enfoque no lineal del Random Forest (Hastie, Tibshirani & Friedman, 2009; Wilks, 2019).

Figura 39

Residuales (observado – predicho) a lo largo del tiempo - ventana 23-30/08/2023



Fuente: Elaboración propia con datos de la estación de monitoreo automático de calidad del aire ubicada a sotavento – OEFA (2023).

Nota. Diagnóstico de residuales: no se detectan sesgos estructurales; la mayor varianza en picos es consistente con la dinámica física del episodio.

4.3.3.6. Implicancias para la fiscalización y operación. Los resultados permiten traducir el modelo en acciones concretas para vigilancia y control:

- **Prealertas en madrugada/mañana** cuando se proyecta BLH baja y pobre ventilación: activar vigilancia reforzada en las primeras horas del día, cuando la mezcla vertical es limitada y el riesgo de picos aumenta.
- **Priorización de inspecciones** ante la **coincidencia** de BLH baja + viento poco con

baja capacidad de ventilación: focalizar recursos en franjas y días con mayor probabilidad de excedencias.

- **Anticipación por memoria corta:** si en las 1–4 horas previas se observan condiciones desfavorables y concentraciones elevadas, incrementar la alerta (consistente con la persistencia horaria capturada por el modelo).
- **Umbral operativo de prealerta:** definir un rango de 220–230 $\mu\text{g}/\text{m}^3$ (predicho) para ganar sensibilidad en días críticos, manteniendo controladas las falsas alarmas; el umbral formal de 250 $\mu\text{g}/\text{m}^3$ queda como referencia regulatoria.
- **Comunicación y respuesta escalonada:** emitir avisos tempranos a operadores y autoridades locales cuando el modelo indique ascenso sostenido hacia niveles de riesgo; preparar medidas de mitigación (p. ej., ajustes operativos, riego de vías, gestión de cargas, etc.).
- **Mejora continua:** refinar rezagos/ventanas de BLH y viento, incorporar señales de estabilidad (p. ej., gradiente térmico superficial) y mantener el enfoque no lineal del Random Forest para capturar la interacción entre mezcla y emisiones.
- En conjunto, estas pautas alinean la física del problema (mezcla/ventilación) con la capacidad predictiva no lineal del modelo, habilitando un uso proactivo para alerta temprana y fiscalización focalizada (Breiman, 2001; Wilks, 2019; Hastie, Tibshirani & Friedman, 2009).

4.3.4. *Análisis de colinealidad entre predictores (VIF)*

Metodología. Con el fin de caracterizar la colinealidad entre predictores se calculó el Factor de Inflación de la Varianza (VIF) sobre el conjunto de entrenamiento (80 % inicial del periodo), tras la ingeniería de variables y sin incluir ninguna columna que replique el objetivo (se mantuvieron únicamente rezagos y rollings retrasados de SO_2 , junto con BLH y variables meteorológicas/estacionales). El VIF se estimó ajustando, para cada predictor X_j , una regresión lineal contra el resto de predictores y aplicando $\text{VIF}_j = 1 / (1 - R_j^2)$.

Criterios. Se adoptaron umbrales habituales: $VIF < 5$ (colinealidad baja/aceptable), 5–10 (moderada) y >10 (alta). En series temporales con rezagos cercanos (p. ej., 1–3–6 h) es esperable encontrar VIF elevados por redundancia temporal.

Resultados resumidos. Tal como se anticipa en problemas autorregresivos, los rezagos cercanos de SO_2 presentaron VIF moderados/altos (p. ej., $SO_2_lag_1h$, $SO_2_lag_2h$, $SO_2_lag_3h$), evidenciando información solapada. Variables de estacionalidad (hora/día) y meteorología (BLH, viento, temperatura) mostraron VIF generalmente bajos a moderados, lo que sugiere complementariedad con los rezagos.

Implicancia para el modelo. Aunque el VIF es un diagnóstico lineal, resulta útil para documentar redundancias. En Random Forest, la colinealidad no invalida el ajuste (no exige independencia estricta), pero puede distribuir la importancia entre predictores correlacionados y volver menos parsimoniosa la representación. Por ello, se prioriza la interpretación con importancias por permutación (en *hold-out*) y se evalúa la estabilidad del desempeño en CV temporal.

Decisión de modelado. Dado que (i) los rezagos de corto plazo aportan la mayor capacidad explicativa y (ii) el desempeño es estable (CV temporal) y generaliza en *hold-out*, se mantuvieron los rezagos clave (1–24 h) y las variables meteorológicas/estacionales. Para evitar sobrecarga de variables casi equivalentes se recomienda, de cara a trabajos futuros, reducir grupos de rezagos altamente colineales (p. ej., conservar 1h–6h–24h) o emplear selección/agrupación (p. ej., PCA de rezagos o selección guiada por permutación).

4.3.4.1. Discusión sobre predictores y robustez del Modelo. En términos de robustez, el Random Forest mostró desempeño global consistente bajo validación cruzada temporal y *hold-out*, a pesar de la colinealidad entre rezagos cercanos de SO_2 . Esto se explica porque el bosque no depende de estimadores lineales únicos, sino de conjuntos de particiones que pueden usar caminos alternativos cuando dos predictores transmiten información similar.

Aun así, la colinealidad afecta la lectura de importancias (reparto del peso entre variables próximas en el tiempo). Por ese motivo, se adoptó el siguiente criterio:

- Mantener los rezagos informativos (en especial 1 h y 24 h) y la BLH (y su *rolling* retrasado), por su consistencia física y su impacto comprobado en permutación.
- Evitar duplicidades del objetivo: no incluir ninguna transformación que use el valor presente/futuro del SO₂ (solo pasado), eliminando el riesgo de fuga.
- Documentar VIF para transparencia y, si fuese necesario, simplificar (p. ej., retirar rezagos intermedios muy colineales) sin afectar materialmente el R²/errores en *hold-out*.
- Complementar la interpretación con importancias por permutación (en datos no vistos), que reflejan el impacto predictivo real de cada predictor.

En suma, el diagnóstico de colinealidad no compromete la validez del modelo ni su capacidad de generalización; sí orienta a una interpretación prudente de importancias y a posibles simplificaciones de la familia de rezagos en iteraciones futuras, preservando el balance entre parsimonia y exactitud.

4.3.4.2. Factor de Inflación de la Varianza. El VIF se calculó con término constante (intercepto), imputación por mediana y sin incluir el objetivo ni variables que lo repliquen. Se reportan dos escenarios: (A) bloque meteorológico en todo el periodo y (B) conjunto de predictores 'top' usado en el modelado (rezagos de SO₂ + meteorología + estacionalidad) sobre el 80 % de entrenamiento.

En la tabla 41 se reportan los VIF del bloque meteorológico (temperatura, humedad, viento y BLH) calculados sobre datos horarios validados 2022–2024 de la estación a sotavento, considerando término constante e imputación por mediana. Estas cifras caracterizan la colinealidad entre predictores y respaldan su uso conjunto en el modelado.

Tabla 41

Estimación del Factor de Inflación de la Varianza (VIF) para las variables del bloque meteorológico (Todo el periodo, con constante)

Variable	VIF	Interpretación
Temperatura	3.20	< 5 (Baja aceptable)
Humedad relativa	3.18	< 5 (Baja aceptable)
Velocidad del viento	2.11	< 5 (Baja aceptable)
BLH	2.17	< 5 (Baja aceptable)

Fuente: Elaboración propia

Con datos horarios validados (2022–2024) de la estación a sotavento, en la tabla 42 se presentan los VIF del conjunto de predictores ‘top’ (rezagos de SO₂, meteorología y estacionalidad) calculados sobre el 80 % de entrenamiento con término constante. El objetivo es documentar la redundancia entre variables relevantes y su impacto en la interpretación del modelo.

Tabla 42

Estimación del Factor de Inflación de la Varianza (VIF) para predictores ‘Top’ (Train 80% con constante)

Variable	VIF	Interpretación
SO2_rollmean_6h	6.04	5–10 (Moderada)
SO2_lag_2h	4.07	<5 (Baja/Aceptable)
SO2_lag_3h	4.07	<5 (Baja/Aceptable)
HR_porc	3.80	<5 (Baja/Aceptable)
TEMP_celc	3.53	<5 (Baja/Aceptable)
SO2_lag_1h	3.01	<5 (Baja/Aceptable)

Variable	VIF	Interpretación
BLH_rollmean_6h	2.81	<5 (Baja/Aceptable)
BLH_m	2.69	<5 (Baja/Aceptable)
SO2_rollmax_24h	2.33	<5 (Baja/Aceptable)
WS_ms	2.11	<5 (Baja/Aceptable)
hour	1.96	<5 (Baja/Aceptable)
SO2_lag_12h	1.47	<5 (Baja/Aceptable)
SO2_lag_24h	1.45	<5 (Baja/Aceptable)
WD_sexa	1.13	<5 (Baja/Aceptable)
dow	1.00	<5 (Baja/Aceptable)

Fuente: Elaboración propia

Los VIF del bloque meteorológico permanecen < 5 (baja-moderada colinealidad), mientras que en el conjunto “top” algunos rezagos muestran VIF más altos por redundancia temporal esperable. En Random Forest, la colinealidad no invalida el ajuste (distribuye importancia entre predictores próximos), por lo que se prioriza la lectura con permutación en hold-out y la verificación de estabilidad mediante CV temporal (Field, 2018; Wilks, 2019).

Cabe resaltar que la coherencia entre estos resultados y la matriz de correlación de la sección 4.1.4 refuerza la solidez técnica del enfoque multivariable adoptado, permitiendo una predicción más confiable del comportamiento horario del contaminante SO₂ bajo diversas condiciones atmosféricas.

En conjunto, los resultados confirman que la dinámica de SO₂ en sotavento está condicionada por el régimen de mezcla (BLH) y por rezagos de corto plazo, con no linealidades que exigen enfoques más flexibles que los modelos lineales. El Random Forest mostró buena generalización en el hold-out temporal y estabilidad en la validación cruzada temporal, y las

importancias (permutación e impureza) aportan un mapa operativo para la fiscalización: priorizar madrugadas/mañanas con BLH baja, vigilar la combinación BLH baja + viento desfavorable, y anticipar episodios cuando en las 1–4 h previas ya se observan condiciones de riesgo. Estas evidencias respaldan la aplicabilidad del modelamiento para la gestión y fiscalización ambiental en Cajamarquilla, con criterios objetivos y reproducibles (Breiman, 2001; Hastie et al., 2009; Wilks, 2019; Hyndman & Athanasopoulos, 2021; Fisher et al., 2019).

En conjunto, los resultados confirman que la dinámica de SO₂ en sotavento está condicionada por el régimen de mezcla (BLH) y por rezagos de corto plazo, con no linealidades que exigen enfoques más flexibles que los modelos lineales. El Random Forest mostró buena generalización en el hold-out temporal y estabilidad en la validación cruzada temporal, y las importancias (permutación e impureza) aportan un mapa operativo para la fiscalización: priorizar madrugadas/mañanas con BLH baja, vigilar la combinación BLH baja + viento desfavorable, y anticipar episodios cuando en las 1–4 h previas ya se observan condiciones de riesgo. Estas evidencias respaldan la aplicabilidad del modelamiento para la gestión y fiscalización ambiental en Cajamarquilla, con criterios objetivos y reproducibles (Breiman, 2001; Hastie et al., 2009; Wilks, 2019; Hyndman & Athanasopoulos, 2021; Fisher et al., 2019).

Estos hallazgos constituyen un insumo valioso para la vigilancia ambiental en contextos urbanos-industriales como Cajamarquilla y servirán de base para el análisis crítico que se desarrolla en el Capítulo V – Discusión de Resultados, donde se interpretan sus implicancias en relación con los objetivos de la investigación, el marco normativo vigente y la literatura especializada.

V. DISCUSIÓN DE RESULTADOS

El objetivo del capítulo es interpretar, evaluar críticamente y discutir los resultados obtenidos en el Capítulo IV, relacionándolos con los objetivos, preguntas e hipótesis planteadas en el Capítulo I, respaldados por el marco teórico desarrollado en el Capítulo II y validando la pertinencia metodológica expuesta en el Capítulo III.

5.1. Discusión del análisis descriptivo de los Datos

El análisis descriptivo del Capítulo IV caracterizó la dinámica temporal y espacial de las concentraciones horarias de SO₂ (2022–2024) en el entorno de la Refinería de Zinc de Cajamarquilla, considerando las estaciones Sotavento y Barlovento. Se observó una alta variabilidad intra-diaria y estacional, con Sotavento registrando múltiples picos horarios ≥ 250 $\mu\text{g}/\text{m}^3$ cada año. Si bien ello no implica incumplimiento del ECA 24 h, constituye un indicio de riesgo operativo que amerita vigilancia focalizada.

Estos hallazgos confirman que las emisiones no se comportan de forma uniforme a lo largo del día y están fuertemente moduladas por la meteorología, en particular por la altura de la capa límite atmosférica (BLH). En madrugada/mañana—franja típica de inversión térmica—la BLH se mantiene frecuentemente < 250 m, lo que limita la mezcla vertical y favorece la acumulación de contaminantes. Este comportamiento concuerda con la literatura de micrometeorología y contaminación en entornos industriales, donde la combinación de emisiones puntuales y baja dispersión vertical genera episodios agudos en ventanas cortas (Arya, 1999; Manahan, 2007; Seinfeld & Pandis, 2016).

En línea con la física del problema, el Cap. IV mostró una relación inversa entre BLH y SO₂, más marcada en Sotavento (exposición directa a la pluma), mientras que en Barlovento la señal es más tenue por menor incidencia directa de la fuente. A nivel mensual, los meses de invierno y transición presentaron concentraciones promedio más altas—condiciones de estabilidad atmosférica y vientos débiles—, patrón coherente con reportes para La Oroya

(Perú), Monterrey (México) y Santiago (Chile), donde la meteorología explica gran parte de la acumulación de SO₂ (Torres et al., 2021; Liu et al., 2020).

Velocidad y dirección del viento. Tal como se documentó en el 4.1.2(e), la velocidad del viento (WS) se mantuvo predominantemente baja en ambas estaciones, con registros horarios frecuentes < 2.0 m/s, un régimen con viento débil que limita la renovación del aire superficial y potencia la acumulación cuando coexiste con BLH baja. La dirección del viento (WD) explica la mayor incidencia en Sotavento: la pluma tiende a desplazarse hacia esa estación, donde se observan las mayores concentraciones. Para 2024, la rosa de los vientos muestra predominancia SSO (21,9 %) y O, patrón que, según el cuadrante expuesto, puede trasladar la pluma hacia Barlovento, donde hay viviendas y zona industrial, elevando la vulnerabilidad poblacional/industrial en determinados escenarios. Aunque el entorno inmediato de Sotavento carece de viviendas, existe exposición ocupacional del personal del Radio Observatorio de Jicamarca.

En suma, el análisis descriptivo confirma que Sotavento es la estación más sensible y representativa de la carga contaminante asociada a la refinería, justificando su uso prioritario en el modelamiento. Además, explica por qué las correlaciones lineales globales entre SO₂ y variables meteorológicas pueden resultar modestas al promediar 2022–2024: la relación es no lineal, depende de regímenes (p. ej., BLH baja + viento con baja capacidad de ventilación/dirigido), y existe persistencia horaria del contaminante. Esto no contradice la hipótesis; por el contrario, justifica las estratificaciones por hora/BLH y el uso de un modelo no lineal (Random Forest) que capta interacciones y efectos umbral con mayor fidelidad.

5.2. Discusión sobre cumplimiento normativo ambiental

El análisis de cumplimiento respecto a los estándares legales establecidos en el Decreto Supremo N.º 003-2017-MINAM, que aprueba los Estándares de Calidad Ambiental (ECA) para Aire, evidencia que, si bien las concentraciones horarias de SO₂ en la estación sotavento

superaron con frecuencia el umbral de $250 \mu\text{g}/\text{m}^3$, dichas superaciones no constituyen por sí mismas un incumplimiento normativo, debido a que el ECA se evalúa sobre promedios móviles de 24 horas y con una frecuencia máxima tolerada de 7 superaciones por año.

Durante el periodo de estudio, las concentraciones diarias de SO_2 superaron el umbral del ECA en:

- 5 ocasiones en el año 2022,
- 6 ocasiones en el año 2023, y
- 0 ocasiones en el año 2024.

Por tanto, no se configuró un incumplimiento formal del ECA para aire conforme a lo estipulado en la normativa vigente. Sin embargo, este resultado evidencia limitaciones estructurales del marco regulatorio, dado que durante el periodo 2022–2024 las concentraciones horarias de SO_2 en la estación sotavento de la Refinería de Zinc de Cajamarquilla superaron en 1,202 ocasiones el umbral crítico de $250 \mu\text{g}/\text{m}^3$ (443 en 2022, 346 en 2023 y 413 en 2024), sin que ello active medida correctiva ni acción sancionadora alguna, pese al riesgo manifiesto para la salud ambiental y humana, en especial para poblaciones vulnerables.

5.2.1. Implicancias regulatorias y limitaciones institucionales

La actual normativa peruana no contempla un estándar específico de alerta para el dióxido de azufre en Lima Metropolitana, ya que el Decreto Supremo N.º 009-2003-SA, que establece los niveles de estados de alerta para contaminantes del aire, limita su aplicación del parámetro SO_2 a zonas como La Oroya e Ilo. Esta omisión normativa deja sin herramientas operativas a entidades como el OEFA y las Direcciones de Salud para emitir alertas tempranas o aplicar protocolos de contingencia en zonas industrializadas como Cajamarquilla, a pesar de contar con estaciones automáticas y evidencia técnica de riesgo.

Adicionalmente, la fiscalización ambiental minera, si bien incorpora parámetros de

calidad del aire en sus instrumentos de vigilancia, carece de criterios diferenciados para interpretar eventos de corta duración o variabilidad horaria, lo que podría derivar en subestimación de impactos reales sobre el ambiente y la salud pública.

5.2.2. Reflexión sobre la suficiencia del marco normativo actual

A la luz de los resultados obtenidos, se evidencia la necesidad de:

- Revisar y actualizar los ECA para aire. Se debe incorporar criterios más exigentes en línea con las recomendaciones de la Organización Mundial de la Salud (OMS), que desde 2021 establece una guía de $40 \mu\text{g}/\text{m}^3$ como valor de referencia para el promedio de 24 horas.
- Ampliar el alcance geográfico del D.S. N.º 009-2003-SA para incluir zonas con carga industrial significativa como Cajamarquilla.
- Desarrollar protocolos de alerta temprana interinstitucional que permitan activar respuestas ante episodios críticos de contaminación, incluso si no se supera el umbral legal del ECA.

En conclusión, aunque no se ha configurado un incumplimiento legal del ECA vigente, los resultados evidencian una situación ambiental que amerita intervención preventiva, actualización normativa y fortalecimiento del sistema de fiscalización basado en evidencia científica.

5.3. Discusión del modelo predictivo implementado (Random Forest)

El modelo predictivo de tipo caja negra implementado en esta investigación se desarrolló mediante la técnica de aprendizaje supervisado Random Forest (RF), con base en datos ambientales registrados por la estación automática sotavento. La elección de esta estación respondió a su mayor sensibilidad a las emisiones industriales, evidenciada en la frecuencia de superaciones horarias del umbral de $250 \mu\text{g}/\text{m}^3$ y en su relación directa con condiciones meteorológicas de baja dispersión atmosférica.

5.3.1. *Valoración del rendimiento del modelo*

Este apartado presenta de manera general el análisis correspondiente. En el hold-out temporal (20 %) el RF alcanzó aprox. $R^2 = 0.789$, $MAE = 18.643 \mu\text{g}/\text{m}^3$ y $RMSE = 34.154 \mu\text{g}/\text{m}^3$; además, en 23–30/08/2023 el esquema tipo alerta (umbral $250 \mu\text{g}/\text{m}^3$) detectó 10/10 eventos sin falsas alarmas.

Comparado con otros enfoques clásicos (regresión lineal múltiple, redes neuronales artificiales), el modelo RF se destacó por su robustez frente al sobreajuste, su capacidad para manejar datos no lineales y su facilidad de interpretación en términos de importancia relativa de variables (Breiman, 2001; Liu et al., 2020).

5.3.2. *Interpretación de la capacidad predictiva*

Este apartado ofrece una descripción general del análisis efectuado. La precisión del modelo fue suficiente para reproducir patrones críticos de concentración y anticipar eventos pico bajo condiciones de inversión térmica (BLH baja y viento con baja capacidad de ventilación), por lo que es viable como herramienta de alerta preventiva más allá del monitoreo pasivo. En la ventana 23–30/08/2023 (tramo de prueba), se observó una correspondencia consistente entre valores observados y predichos, incluyendo la detección 10/10 de horas con $\text{SO}_2 \geq 250 \mu\text{g}/\text{m}^3$ sin falsas alarmas, lo que respalda su utilidad operativa (Breiman, 2001; Wilks, 2019; Hastie, Tibshirani & Friedman, 2009).

5.3.3. *Respecto a la colinealidad*

Se calcularon VIF en dos escenarios complementarios.

- En el bloque meteorológico (temperatura, humedad relativa, velocidad del viento, BLH) sobre el período 2022–2024, los VIF se ubicaron en el rango ≈ 2.2 – 3.3 , es decir, colinealidad baja–moderada y aceptable para interpretación.
- En el conjunto de predictores “top” que incluyen rezagos de SO_2 (1–24 h), es esperable que algunos VIF aumenten por la alta correlación entre rezagos y derivados; no

obstante, el Random Forest es robusto a colinealidad y esto no compromete su capacidad predictiva. Esta lectura complementa las correlaciones del Cap. IV y fortalece la interpretación de la importancia relativa de variables en el modelo.

5.3.4. *La jerarquía de importancias del RF (permutación e impureza)*

En este apartado se presentan los principales predictores identificados por el modelo Random Forest, ordenados según su jerarquía de importancia.

- Rezagos de SO₂ (1–24 h) → principal determinante (memoria de corto plazo);
- BLH (y su promedio móvil retrasado) → capacidad de mezcla/ventilación;
- Viento (dirección/velocidad) y temperatura → modulan dispersión y estabilidad;
- Estacionalidad intra-diaria (hora del día) → contexto de variación regular.

Esta jerarquía es coherente con la física del problema: cuando la BLH es baja, la mezcla vertical se limita y la concentración se acumula, especialmente si los rezagos recientes ya son altos.

5.3.5. *En cuanto a la evaluación del modelo RF*

En este apartado se aplicó validación cruzada temporal (TimeSeriesSplit, $k = 5$), adecuada para series con dependencia temporal. Las métricas promedio fueron $R^2 = 0.652 \pm 0.099$, $MAE = 26.772 \pm 5.119 \mu\text{g}/\text{m}^3$ y $RMSE = 47.497 \pm 12.469 \mu\text{g}/\text{m}^3$, valores más conservadores que el hold-out y consistentes con una evaluación más estricta. En el hold-out temporal (20 %) se obtuvo $R^2 = 0.789$, $MAE = 18.643 \mu\text{g}/\text{m}^3$ y $RMSE = 34.154 \mu\text{g}/\text{m}^3$, confirmando capacidad de generalización sin indicios de sobreajuste (Hastie, Tibshirani & Friedman, 2009).

En síntesis, el Random Forest se adapta bien a relaciones no lineales y interacciones entre variables meteorológicas y temporales (Breiman, 2001). Su desempeño, sumado a la lectura tipo alerta y a diagnósticos de calibración y residuales estables, respalda el uso del modelo como soporte de alerta temprana y fiscalización focalizada en escenarios de BLH baja característicos de Cajamarquilla.

5.4. Aporte del modelo a la gestión ambiental preventiva

A continuación, se expone de manera general el contenido correspondiente a este apartado.

El modelo Random Forest representa un avance significativo en la implementación de enfoques predictivos en la fiscalización ambiental minera. Su potencial radica en su capacidad de:

- Generar alertas operativas de corto plazo, orientando la programación de inspecciones en momentos de mayor riesgo.
- Apoyar la toma de decisiones basada en evidencia, al identificar condiciones críticas que anteceden a eventos de alta concentración.
- Optimizar recursos técnicos y humanos, al permitir la focalización de acciones preventivas y correctivas.

Asimismo, su implementación en tiempo real, integrada a sistemas como el PIFA-OEFA, podría fortalecer los mecanismos de vigilancia ambiental nacional, especialmente en zonas con alta carga industrial y escasa cobertura normativa preventiva, como Cajamarquilla.

5.5. Contraste con hipótesis y objetivos planteados

Este apartado busca contrastar los resultados obtenidos en el capítulo IV con los objetivos específicos y las hipótesis formuladas en el capítulo I, verificando el grado de cumplimiento metodológico y la validez empírica del modelo propuesto. Asimismo, se formula una respuesta argumentada a la pregunta de investigación general.

5.5.1. Cumplimiento de los objetivos específicos

En esta sección se evalúa el nivel de cumplimiento alcanzado respecto a los objetivos específicos planteados en la investigación. La evaluación se ha realizado mediante el análisis y contraste sistemático de los resultados obtenidos en relación con cada objetivo específico.

Tabla 43*Cumplimiento de los objetivos específicos*

Objetivo Específico	Resultado asociado	Cumplimiento
OE1: Cuantificar las superaciones horarias y diarias del umbral de $250 \mu\text{g}/\text{m}^3$ de SO_2 registradas en el área de estudio durante el periodo 2022–2024.	En la estación sotavento se identificaron 443 superaciones horarias en 2022, 346 en 2023 y 413 en 2024. Se observó concentración en las primeras horas del día. A nivel diario, solo hubo 5 superaciones en 2022 y 6 en 2023, dentro del límite normativo (≤ 7).	✓ Cumplido
OE2: Analizar la relación estadística entre concentraciones de SO_2 y variables meteorológicas.	Asociación operativa con BLH baja y viento con baja capacidad de ventilación; correlaciones globales modestas por no linealidad y persistencia; mejora al estratificar por hora/BLH.	✓ Cumplido
OE3: Evaluar la influencia de la inversión térmica en la acumulación y variabilidad horaria del SO_2 .	Evidencia consistente: BLH baja $\leftrightarrow$ picos de SO_2 en primeras horas del día; episodios críticos bajo mezcla limitada.	✓ Cumplido
OE4: Comparar las concentraciones horarias con la normativa y proponer criterios técnicos complementarios.	Se propone lectura tipo alerta (umbral $250 \mu\text{g}/\text{m}^3$) y prealerta operativa ($220\text{--}230 \mu\text{g}/\text{m}^3$ predicho) para vigilancia proactiva.	✓ Cumplido
OE5: Desarrollar y validar un modelo Random Forest para predecir concentraciones horarias de SO_2 .	Hold-out 20 %: $R^2 = 0.789$, $\text{MAE} = 18.643 \mu\text{g}/\text{m}^3$, $\text{RMSE} = 34.154 \mu\text{g}/\text{m}^3$. CV temporal ($k = 5$): $R^2 = 0.652 \pm 0.099$, $\text{MAE} = 26.772 \pm 5.119 \mu\text{g}/\text{m}^3$, $\text{RMSE} = 47.497 \pm 12.469 \mu\text{g}/\text{m}^3$. Importancias: rezagos SO_2 (1–24 h) > BLH (y promedio móvil retrasado) > viento/temperatura > hora del día	✓ Cumplido

Fuente: Elaboración propia

Los resultados permiten afirmar que los cinco objetivos específicos fueron cumplidos en su totalidad, como se presenta en la Tabla 43. Ello valida la pertinencia de la metodología empleada y la coherencia entre el diseño, el análisis y los resultados.

5.5.2. Validación de hipótesis específicas

En este apartado se presenta el proceso de validación de las hipótesis específicas definidas previamente, contrastando estas hipótesis con los hallazgos empíricos obtenidos a partir del análisis estadístico y la modelación predictiva realizada durante el estudio.

Tabla 44

Validación de hipótesis específicas

Hipótesis específica	Evaluación	Resultado
HE1: La frecuencia de superaciones horarias del umbral de 250 µg/m ³ es mayor en días con inversión térmica.	Análisis temporal conjunto BLH-SO ₂ (estratificación por BLH baja).	Validada: Mayor número de superaciones ocurrió con BLH baja (<300 m), condición típica de inversión.
HE2: Durante la inversión térmica, las concentraciones de SO ₂ son mayores en sotavento.	Comparación entre estaciones.	Validada: Sotavento concentra la mayor cantidad de eventos horarios; Barlovento registra pocos.
HE3: Existen patrones horarios recurrentes de alta concentración en la mañana.	Distribución horaria de eventos / estratificación por hora.	Validada: Concentraciones más elevadas entre 5:00 y 9:00 a. m., coincidiendo con la inversión térmica.
HE4a: Las concentraciones horarias de SO ₂ registradas en 2022–2024 presentan superaciones críticas (443 en 2022, 346 en 2023 y 413 en 2024) que no siempre se reflejan en los promedios diarios del ECA Aire (24h).	Comparación entre superaciones horarias y promedios diarios.	Validada: Las 443 (2022), 346 (2023) y 413 (2024) superaciones horarias contrastan con solo 5 (2022) y 6 (2023) superaciones diarias dentro del límite normativo (≤7).
HE4b: El análisis conjunto de concentraciones horarias de SO ₂ y variables meteorológicas permite establecer criterios técnicos complementarios que refuercen la fiscalización.	Evaluación normativa y técnica.	Validada: El análisis horario y meteorológico permitió definir criterios complementarios de alerta temprana y fortalecer la fiscalización ambiental.

Hipótesis específica	Evaluación	Resultado
HE5: El modelo Random Forest predice con alta precisión el SO ₂ .	Hold-out y CV temporal.	Validada: $R^2 \approx 0.789$. ver OE5; en 23–30/08/2023, detección 10/10 eventos sin falsas alarmas.

Fuente: Elaboración propia

La Tabla 44 muestra que las cinco hipótesis específicas fueron validadas (H4 se operacionalizó en dos sub-hipótesis, H4a y H4b, para facilitar su verificación empírica), confirmando la influencia de la inversión térmica y de las variables meteorológicas en la variabilidad de las concentraciones de SO₂, así como la utilidad del modelo Random Forest para la predicción de episodios críticos.

5.5.3. *Respuesta a la pregunta de investigación general*

¿Cómo influye la inversión térmica en la concentración horaria de dióxido de azufre (SO₂) registrada por las estaciones automáticas del OEFA en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024?

La inversión térmica (expresada operativamente por BLH baja en madrugada/mañana) reduce la mezcla vertical y limita la ventilación, favoreciendo la acumulación de SO₂ en Sotavento, donde se concentran la mayoría de superaciones horarias. El modelo Random Forest—validado temporalmente—reproduce estos patrones y permite anticipar escenarios críticos con una lectura tipo alerta (umbral 250 µg/m³) y una prealerta operativa (220–230 µg/m³ predicho) que fortalecen la fiscalización en franjas de mayor riesgo. En conjunto, la evidencia confirma que la dinámica de mezcla/ventilación condiciona la concentración horaria de SO₂ y que un esquema de vigilancia predictiva resulta pertinente para la gestión ambiental de Cajamarquilla.

5.6. Implicancias prácticas y contribuciones del estudio

El presente estudio, centrado en la evaluación del comportamiento horario del dióxido de azufre (SO₂) bajo condiciones de inversión térmica en el área de influencia de la Refinería

de Zinc de Cajamarquilla, presenta una serie de implicancias prácticas y contribuciones sustantivas para la gestión ambiental preventiva y correctiva, particularmente en contextos minero-metalúrgicos con vulnerabilidad atmosférica.

5.6.1. *Aportes específicos a la gestión ambiental en Cajamarquilla*

- **Caracterización ambiental horaria:** Se evidenció que las concentraciones críticas de SO₂ no necesariamente se reflejan en promedios diarios, lo que justifica la necesidad de reforzar los sistemas de vigilancia ambiental con enfoque horario en zonas industriales.

- **Identificación de ventanas críticas de riesgo:** Se determinaron franjas horarias crítica (entre 5:00 y 9:00 a. m.) en las que se concentran los episodios de mayor acumulación de contaminantes. Esta información puede ser utilizada por OEFA para ajustar las rutinas de inspección in situ o programar alertas operacionales a las empresas reguladas.

- **Priorización geográfica y meteorológica:** La estación sotavento fue identificada como la más representativa para capturar eventos de contaminación severa bajo inversión térmica. Esto aporta evidencia para optimizar la ubicación y operación de las redes de monitoreo automático.

5.6.2. *Propuestas derivadas para mejorar la fiscalización ambiental minera*

- **Incorporación de modelos predictivos en la fiscalización.** El modelo Random Forest demostró capacidad operativa para predecir concentraciones horarias de SO₂, permitiendo generar alertas anticipadas y sustentar intervenciones preventivas.

- **Ajustes en los criterios normativos.** Los hallazgos permiten recomendar la inclusión de criterios meteorológicos (como BLH baja e inversión térmica) dentro de los protocolos técnicos de evaluación ambiental, como factores agravantes o condicionantes de cumplimiento normativo.

- **Fortalecimiento del enfoque de riesgo en la normativa.** Se evidencia la necesidad de una revisión normativa del D.S. N.º 009-2003-SA, especialmente en lo referido a su

aplicabilidad para Lima Metropolitana, dado que actualmente los umbrales horarios solo se consideran para Ilo y La Oroya.

5.6.3. *Impacto potencial en otras zonas industriales similares*

- **Transferencia metodológica:** La metodología desarrollada —análisis horario, integración de variables meteorológicas y modelamiento predictivo— es replicable en otras zonas industriales del país como Ilo, La Oroya, Cerro de Pasco, entre otras a nivel nacional.

- **Desarrollo de sistemas de alerta ambiental:** Este estudio sienta las bases para la creación de sistemas automatizados de alerta temprana en función de predicción de contaminantes, lo que fortalecería la capacidad de respuesta institucional ante eventos críticos.

- **Interoperabilidad con redes de salud pública:** Los datos generados podrían alimentar sistemas de vigilancia en salud ambiental para anticipar riesgos respiratorios en poblaciones expuestas, especialmente en zonas urbanas colindantes a operaciones industriales.

5.7. Limitaciones y recomendaciones para futuras investigaciones

Si bien este estudio aporta evidencia relevante sobre la influencia de la inversión térmica en las concentraciones horarias de dióxido de azufre (SO₂) en el entorno de la Refinería de Zinc de Cajamarquilla, también reconoce ciertas limitaciones metodológicas, técnicas y de acceso a la información que condicionaron el alcance de los resultados. A partir de ello, se formulan recomendaciones específicas para mejorar futuros estudios similares.

5.7.1. *Limitaciones del estudio*

- **Alcance geográfico limitado a dos estaciones.** El análisis se basó en los datos registrados exclusivamente por las estaciones automáticas de OEFA ubicadas a barlovento y sotavento. Si bien estas son representativas, su cobertura espacial no permite generalizar resultados a toda la cuenca atmosférica de Lima Este.

- **Disponibilidad restringida de variables meteorológicas.** El estudio incorporó variables como temperatura, viento, humedad y presión atmosférica, registradas por las

estaciones locales. Sin embargo, no se contó con mediciones directas de estabilidad atmosférica ni de la altura de la capa límite atmosférica (BLH) mediante instrumentos in situ, lo cual hubiera permitido mayor precisión en la modelación de la dispersión de contaminantes.

- **Limitación en la frecuencia de datos de BLH.** La altura de la capa límite atmosférica fue obtenida de fuentes satelitales (Copernicus), cuya periodicidad no siempre coincidía con el registro horario continuo del SO₂, generando vacíos o interpolaciones.

- **Desfase entre exposición y efectos en salud.** Si bien se evidenció la superación horaria del umbral de 250 µg/m³, no se evaluó el impacto en salud pública local, debido a la inexistencia de datos epidemiológicos integrados o validados en la zona de influencia.

- **Modelo predictivo centrado solo en sotavento.** Por decisión metodológica, el modelamiento Random Forest se aplicó únicamente a la estación sotavento, lo que limitó el análisis comparativo con otras zonas afectadas por la misma fuente emisora.

5.7.2. *Recomendaciones para futuras investigaciones*

- **Ampliar la red de monitoreo de calidad del aire y capacidades instrumentales:** Se recomienda extender la red de estaciones automáticas de monitoreo para incluir más puntos de control de contaminantes críticos como PM₁₀, PM_{2.5}, NO_x y SO₂ en zonas estratégicas de influencia. Asimismo, sería pertinente incorporar instrumentos especializados para medir en tiempo real variables meteorológicas complementarias como la altura de la capa límite atmosférica (BLH) y la estabilidad atmosférica, actualmente obtenidas solo de fuentes satelitales.

- Incorporar **equipos especializados** para el registro directo de la BLH y la estabilidad atmosférica.

- **Aplicar modelos multizona y multifuente:** Estudios futuros podrían incorporar más estaciones en zonas de dispersión media o distante para caracterizar la pluma de emisión de SO₂ en distintas condiciones atmosféricas.

- **Vincular estudios de calidad del aire con efectos en salud:** Es urgente desarrollar investigaciones que articulen datos ambientales con registros de salud pública (morbilidad respiratoria, ausentismo escolar, etc.), especialmente en zonas vulnerables.
- **Explorar el uso de nuevos algoritmos de aprendizaje automático:** Se sugiere probar otros enfoques de aprendizaje automático (XGBoost, redes neuronales, SVM) que complementen y mejoren los resultados obtenidos con RF
- **Evaluar escenarios climáticos futuros:** Incorporar variables de cambio climático y proyecciones atmosféricas en los modelos permitiría anticipar posibles incrementos de episodios críticos de contaminación en escenarios de mayor temperatura y estabilidad.
- **Formular propuestas normativas basadas en evidencia científica, que actualicen los ECA y los niveles de alerta ambiental en el Perú:** A partir de estos resultados, futuras investigaciones pueden diseñar propuestas concretas de actualización normativa, considerando límites horarios, nuevos criterios de alerta y zonas de aplicación específicas.

VI. CONCLUSIONES

6.1. Concentración horaria de SO₂ y superaciones del umbral

Durante el periodo 2022–2024, las concentraciones horarias de SO₂ en la estación sotavento de la Refinería de Zinc de Cajamarquilla superaron el umbral crítico de 250 µg/m³ en 1,202 ocasiones (443 en 2022, 346 en 2023 y 413 en 2024). En contraste, la estación barlovento superó el umbral en 38 ocasiones (15 en 2022, 19 en 2023 y 4 en 2024). Esta asimetría confirma la exposición diferencial condicionada por la dirección del viento y la posición relativa respecto de la fuente emisora.

6.2. Excedencias del ECA y evaluación normativa

No se evidenciaron incumplimientos formales del ECA Aire (D.S. N.º 003-2017-MINAM), pues las excedencias del promedio de 24 h no superaron 7 veces/año. Sin embargo, la alta recurrencia horaria de eventos ≥ 250 µg/m³ revela un riesgo agudo poco capturado por un indicador basado en promedios diarios, lo que sugiere incorporar criterios complementarios de vigilancia para episodios de corta duración.

6.3. Influencia de condiciones meteorológicas e inversión térmica

Se corroboró una relación inversa entre la BLH y el SO₂ horario, con picos matinales bajo inversión térmica (BLH baja). La estabilidad y la dispersión vertical limitada emergen como factores determinantes de la acumulación en la zona de estudio, en coherencia con la dinámica físico-atmosférica descrita en el marco conceptual.

6.4. Valor del modelo Random Forest en la gestión predictiva

El modelo RF mostró desempeño sólido en hold-out temporal (20 %): $R^2 = 0.789$, MAE = 18.643 µg/m³, RMSE = 34.154 µg/m³, y capacidad de generalización respaldada por validación cruzada temporal ($k = 5$). Esto valida el uso de modelamiento predictivo para vigilancia anticipativa y soporte de decisiones en contextos industriales.

6.5. Validación de hipótesis y cumplimiento de objetivos

Los cinco objetivos específicos planteados fueron cumplidos satisfactoriamente, y todas las hipótesis específicas fueron confirmadas a partir de la evidencia empírica recolectada. En consecuencia, se valida la hipótesis general que sostiene que los episodios de inversión térmica influyen de manera significativa en el incremento de las concentraciones horarias de SO₂ en el área de influencia de la Refinería de Cajamarquilla.

En conjunto, estos resultados sustentan el uso del RF como soporte de alerta temprana (prealerta/alerta) en madrugadas/mañanas con BLH baja y vientos débiles, complementando el cumplimiento formal del ECA 24 h con un enfoque preventivo orientado a episodios agudos.

VII. RECOMENDACIONES

7.1. Implementar alerta temprana predictiva basada en RF y meteorología

Se recomienda al Organismo de Evaluación y Fiscalización Ambiental (OEFA) y a las Entidades de Fiscalización Ambiental (EFA) integrar el Random Forest en los flujos operativos como soporte de alerta temprana en tiempo casi real, con reglas simples y trazables:

- Prealerta si BLH pronosticada baja y $WS < 2.0$ m/s y dirección hacia el sector expuesto, o si SO_2 predicho $\geq 220\text{--}230$ $\mu\text{g}/\text{m}^3$.
- Alerta si SO_2 predicho ≥ 250 $\mu\text{g}/\text{m}^3$ (umbral operativo coherente con eventos críticos).
- Definir SOP (Procedimiento Operativo Estándar) con responsables, tiempos de respuesta (05:00–09:00 h como franja prioritaria), bitácora de decisiones y coordinación OEFA-EFA.

7.2. Actualizar criterios normativos para contemplar exposiciones horarias

Es necesario revisar y actualizar los criterios actuales del Estándar de Calidad Ambiental (ECA) para aire (D.S. N.º 003-2017-MINAM), incorporando criterios horarios complementarios, de modo que los picos de corta duración —evidenciados en la zona— queden reflejados en la evaluación de riesgo y las medidas preventivas. Esta medida garantizaría una protección más efectiva de la salud pública frente a picos de contaminación, actualmente subestimados por la normativa basada únicamente en promedios diarios. Explorar criterios horarios orientados al riesgo (p. ej., número de horas $\geq$ umbral por semana/mes) como indicadores complementarios al ECA de 24 h.

7.3. Ampliar y modernizar el monitoreo (incluida la BLH)

Se recomienda fortalecer la red de monitoreo ambiental con sensores de alta resolución y, cuando sea viable, instrumentación para BLH (p. ej., ceilómetro o integración con productos de reanálisis). Asegurar que la base de datos utilizada para modelamiento ingiera y conserve

de forma consistente los registros ya monitoreados por OEFA de viento (intensidad/dirección) y radiación solar.

- **Aseguramiento y Control de la Calidad (QA/QC) y metadatos:** Establecer un protocolo QA/QC (faltantes, *outliers*, sincronización) y un esquema de metadatos por estación y periodo, para evitar pérdidas de información y asegurar que variables ya monitoreadas por OEFA (p. ej., radiación solar) ingresen en futuras corridas y se conserven en la base de modelamiento.

7.4. Gobernanza del modelo y mejora continua

Se sugiere operar el Random Forest con un esquema de ciclo de vida del modelo:

- Reentrenamiento periódico (p. ej., trimestral/semestral) y backtesting mensual con *hold-out* reciente.
- Métricas de control (R^2 , MAE, RMSE; tasas de TP/FP/FN en lecturas tipo alerta).
- Auditoría de sesgos (cambios en importancias, deriva de datos) y registro de versiones para trazabilidad.
- Incluir monitoreo de deriva de datos y deriva de concepto, además de control rutinario de métricas y auditoría de importancias.

7.5. Capacitación técnica aplicada para fiscalizadores

Se sugiere promover programas de formación continua para el personal técnico del OEFA y de las EFA, enfocados en aprendizaje automático, análisis de series temporales y modelamiento predictivo. Esta capacitación permitiría una fiscalización más eficiente, basada en evidencia científica y soportada en nuevas tecnologías de gestión ambiental.

7.6. Replicabilidad y escalamiento del enfoque metodológico en otras zonas industriales

El modelo metodológico propuesto debe ser aplicado en otras zonas industriales del país con características similares de inversión térmica y emisiones persistentes, como La Oroya, Ilo y Cerro de Pasco y otros clústeres industriales. Su implementación contribuiría a

escalar las capacidades de vigilancia ambiental a nivel nacional, fortaleciendo la articulación entre monitoreo, fiscalización y políticas de protección de la salud pública.

7.7. Comunicación de riesgo y coordinación territorial

Se recomienda asegurar avisos oportunos y focalizados según el cuadrante expuesto por el viento y la BLH; en 2024, Sotavento registró SSO 21,9 % y Barlovento SO 54,6 %, por lo que la prioridad suele concentrarse en barlovento (viviendas/industria), mientras que en sotavento la exposición es ocupacional (ROJ–IGP, ~80 personas).

- Prealerta: BLH baja + $WS < 2 \text{ m/s}$ + dirección SO/SSO/O hacia el sector expuesto, o SO_2 predicho $\geq 220\text{--}230 \text{ }\mu\text{g/m}^3$.
- Alerta: SO_2 predicho $\geq 250 \text{ }\mu\text{g/m}^3$.
- Franja crítica: 05:00–09:00 h.
- Canales y coordinación: mensajes cortos (SMS/correo/radio), bitácora de avisos/acciones y coordinación OEFA–EFA y con el sector salud.

7.8. Extensiones técnicas recomendadas (mediano plazo)

Se recomienda profundizar la interpretabilidad y el desempeño del modelo para uso operativo, en las siguientes técnicas:

- Análisis de efectos marginales (PDP/ALE) para BLH y viento: transparentar la respuesta no lineal del modelo (zonas de umbral/saturación) y derivar reglas operativas (p. ej., rangos de BLH y WS que disparan riesgo).
- Ingeniería de variables: ampliar ventanas y rezagos de BLH/viento, incluir indicadores de estabilidad (p. ej., horas con $BLH < X \text{ m}$ continuas), y contrastar modelos complementarios (p. ej., Gradient Boosting) como banco de contraste.
- Integración de información externa: enlazar con inventarios de emisión y datos satelitales (valores columnares) para diagnóstico y contextualización de episodios intensos (validación cruzada de señales y trazabilidad de fuentes).

- Priorizar PDP/ALE para BLH y velocidad del viento en la franja 05:00–09:00 h, a fin de transparentar umbrales operativos en la respuesta del modelo.

Nota: PDP (Partial Dependence Plot) muestra el efecto promedio de una variable sobre la predicción manteniendo las demás constantes;

ALE (Accumulated Local Effects) estima ese efecto de manera local y es más robusto ante colinealidad.

VIII. REFERENCIAS

- AirScan. (2021). *New WHO air quality guidelines 2021: A global health milestone*.
<https://airscan.org/new-who-air-quality-guidelines-2021/>
- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). <https://apastyle.apa.org/>
- Arya, S. P. (1999). *Air pollution meteorology and dispersion*. Oxford University Press.
<https://global.oup.com/academic/product/air-pollution-meteorology-and-dispersion-9780195073986>
- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
<https://www.sciencedirect.com/science/article/pii/S0020025511006773>
- Bisquerra, R. (2004). *Metodología de la investigación educativa*. La Muralla.
<https://dialnet.unirioja.es/servlet/libro?codigo=292173>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? *Geoscientific Model Development*, 7(3), 1247–1250.
<https://doi.org/10.5194/gmd-7-1247-2014>
- Congreso de la República. (1993). *Constitución Política del Perú*.
https://www.congreso.gob.pe/Docs/files/CONSTITUCION_POLITICA_DEL_PERU.pdf

Copernicus Climate Change Service. (2023). *ERA5 hourly data on single levels from 1959 to present*. <https://cds.climate.copernicus.eu/>

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE.
<https://collegepublishing.sagepub.com/products/discovering-statistics-using-ibm-spss-statistics-5-257672>

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful. *Journal of Machine Learning Research*, 20(177), 1–81.
<https://jmlr.org/papers/v20/18-760.html>

Gautam, R., Tripathi, L., & Panday, A. (2021). Atmospheric stability and air pollution in Kathmandu Valley. *Environmental Pollution*, 273, 116469.
<https://doi.org/10.1016/j.envpol.2020.116469>

Grange, S. K., & Carslaw, D. C. (2019). Using meteorological normalisation to detect interventions. *Science of the Total Environment*, 653, 578–588.
<https://doi.org/10.1016/j.scitotenv.2018.10.344>

Gujarati, D. N., & Porter, D. C. (2009). *Econometría* (5.^a ed.). McGraw-Hill.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer. <https://link.springer.com/book/10.1007/978-0-387-84858-7>

Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (6.^a ed.). McGraw-Hill. <https://www.mheducation.com.pe/9786071512102-peru-metodologia-de-la-investigacion>

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>

Instituto Nacional de Estadística e Informática. (2017). *Censos Nacionales 2017*.
<https://censo2017.inei.gob.pe/resultados-definitivos-de-los-censos-nacionales-2017/>

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Lenschow, P., Abraham, H. J., Kutzner, K., Lutz, M., Preuße, E., & Reichenbacher, W. (2001). Some ideas about the sources of PM₁₀. *Atmospheric Environment*, 35(Suppl. 1), S23–S33. [https://doi.org/10.1016/S1352-2310\(00\)00383-3](https://doi.org/10.1016/S1352-2310(00)00383-3)

Liu, Y., Zhang, Y., Li, Y., Wang, J., & Chen, X. (2022). Air quality index prediction. *Scientific Reports*, 12, 8765. <https://doi.org/10.1038/s41598-022-13579-2>

Manahan, S. E. (2007). *Environmental chemistry* (8th ed.). CRC Press.

<https://www.routledge.com/Environmental-Chemistry/Manahan/p/book/9781566706339>

Ministerio del Ambiente. (2014). *Plan Nacional de Acción Ambiental 2011–2021*.
https://cdn.www.gob.pe/uploads/document/file/12149/04_plana_2011_al_2021.pdf

Ministerio del Ambiente. (2017). *Decreto Supremo N.º 003-2017-MINAM*.
<https://cdn.www.gob.pe/uploads/document/file/12871/DS-003-2017-MINAM.pdf>

Ministerio del Ambiente. (2019). *Protocolo Nacional de Monitoreo de Calidad del Aire*.
<https://www.gob.pe/institucion/minam/normas-legales/363557-10-2019-minam>

Ministerio de Salud. (2003). *Decreto Supremo N.º 009-2003-SA*.

<https://www.digesa.minsa.gob.pe/normas/DS009-2003-SA.pdf>

Molnar, C. (2023). *Interpretable machine learning* (2nd ed.).

<https://christophm.github.io/interpretable-ml-book/>

Montalvo, L., Silva, J., Quispe, J., & Torres, M. (2022). Air quality monitoring in Lima.

Frontiers in Sustainable Cities, 4, 849762. <https://doi.org/10.3389/frsc.2022.849762>

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.

Muñoz, R., & Salinas, E. (2018). Episodios críticos en Santiago. *Revista de Climatología*, 18, 35–46. <https://climatol.eu/rev/climatol18.pdf>

Organismo de Evaluación y Fiscalización Ambiental. (2009). *Ley N.º 29325*.

<https://www.oefa.gob.pe/ley-29325>

Organismo de Evaluación y Fiscalización Ambiental. (2022–2024). *Registros horarios de calidad del aire – Cajamarquilla*. <https://sistemas.oefa.gob.pe/pifa/mfe/#/>

Organismo de Evaluación y Fiscalización Ambiental (OEFA). (2022). Reporte N.º 00007-2022-OEFA/DEAM-STEC.

<https://drive.google.com/drive/folders/1T7Gcm579ABanYhh0e-fKrNQCfhQ4qCyy>

Organismo de Evaluación y Fiscalización Ambiental (OEFA). (2023). Reporte N.º 00022-2023-OEFA/DEAM-STEC.

<https://drive.google.com/drive/folders/1ukVrp6G0loAHPD8yDFpP8aIjaEhS1KJV>

Organismo de Evaluación y Fiscalización Ambiental (OEFA). (2024). Reporte N.º 00038-

2024-OEFA/DEAM-STEC.

<https://drive.google.com/drive/folders/1fPEhjhzhmeznMzTFZrO6PEz4c0Gx0ttA>

Organización Mundial de la Salud. (2006). *Air quality guidelines: Global update 2005*.

<https://www.who.int/publications/i/item/WHO-SDE-PHE-OEH-06.02>

Organización Mundial de la Salud. (2021). *WHO global air quality guidelines*.

<https://www.who.int/publications/i/item/9789240034228>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,

Prettenhofer, P., Weiss, R., Dubourg, V. & Cournapeau, D. (2011). Scikit-learn.

Journal of Machine Learning Research, 12, 2825–2830.

<https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: From air pollution to climate change* (3rd ed.). John Wiley & Sons. [https://www.wiley.com/en-us/Atmospheric+Chemistry+and+Physics%3A+From+Air+Pollution+to+Climate+Ch](https://www.wiley.com/en-us/Atmospheric+Chemistry+and+Physics%3A+From+Air+Pollution+to+Climate+Change%2C+3rd+Edition-p-9781118947401)

[ange%2C+3rd+Edition-p-9781118947401](https://www.wiley.com/en-us/Atmospheric+Chemistry+and+Physics%3A+From+Air+Pollution+to+Climate+Change%2C+3rd+Edition-p-9781118947401)

Scikit-learn developers. (n.d.). *RandomForestRegressor*.

<https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

Scikit-learn developers. (n.d.). *TimeSeriesSplit*.

https://scikitlearn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html

Scikit-learn developers. (n.d.). *Permutation feature importance*.

https://scikit-learn.org/stable/modules/permutation_importance.html

Scikit-learn developers. (n.d.). *Time series cross-validation*.

https://scikit-learn.org/stable/modules/cross_validation.html#time-series-split

Strobl, C. (2007). Bias in random forest variable importance. *BMC Bioinformatics*, 8, 25.

<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-25>

United States Environmental Protection Agency. (2017). *Integrated Science Assessment for Sulfur Oxides*. <https://www.epa.gov/isa/integrated-science-assessment-isa-sulfur-oxides-health-criteria>

Vu, B. N. (2019). PM_{2.5} exposure model in Lima. *Remote Sensing*, 11(6), 641.

<https://doi.org/10.3390/rs11060641>

Wang, W., Liu, X., Bi, J., & Liu, Y. (2022). Ozone estimation model. *Environment International*, 158, 106917. <https://doi.org/10.1016/j.envint.2021.106917>

Willmott, C. J., & Matsuura, K. (2005). Advantages of MAE over RMSE. *Climate Research*, 30, 79–82.

Wilks, D. S. (2019). *Statistical methods in the atmospheric sciences* (4th ed.). Elsevier.

<https://shop.elsevier.com/books/statistical-methods-in-the-atmospheric-sciences/wilks/978-0-12-815823-4>

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2020). AI for air quality forecasting. *Atmospheric Environment*, 237, 117744.

<https://doi.org/10.1016/j.atmosenv.2020.117744>

IX. ANEXOS

Anexo 1

Matriz de Consistencia

Problemas	Objetivos	Hipótesis	Variables	Metodología
Problema General: ¿Cómo influye la inversión térmica en las concentraciones horarias de dióxido de azufre (SO ₂) registradas por las estaciones automáticas del OEFA en el área de influencia directa de la Refinería de Zinc de Cajamarquilla, durante el periodo 2022–2024?	Objetivo General: Evaluar la influencia de la inversión térmica en las concentraciones horarias de SO ₂ registradas por las estaciones automáticas del OEFA en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024, con el fin de optimizar los criterios técnicos de fiscalización ambiental minera en contextos de alta recurrencia de eventos críticos de contaminación.	Hipótesis General: La ocurrencia de eventos de inversión térmica influye de manera significativa en el incremento de las concentraciones horarias de SO ₂ registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.	Variable dependiente: Y = Concentración horaria de SO ₂ Variable independiente: X = Inversión térmica (identificada mediante condiciones meteorológicas: temperatura, presión, viento, BLH)	• Tipo de estudio: cuantitativo no experimental, diseño correlacional explicativo. • Población: Estaciones de monitoreo de calidad del aire del OEFA en tiempo real. • Muestra: Datos horarios de SO₂ y variables meteorológicas tomadas del 2022 al 2024. • Técnicas de análisis: Estadística descriptiva, análisis de frecuencias, cruces temporales y análisis de correlación entre variables meteorológicas y concentraciones horarias de SO₂. • Modelamiento con Random Forest validado con métricas RMSE, MAE y R².
PE1: ¿Con qué frecuencia las concentraciones horarias de SO ₂ superan el umbral de 250 µg/m ³ y cómo se comparan estas superaciones con el promedio diario del ECA Aire de 24 horas durante el periodo 2022–2024?	OE1: Cuantificar las superaciones horarias y diarias del umbral de 250 µg/m ³ de SO ₂ registradas en el área de influencia directa de la Refinería de Zinc de Cajamarquilla durante el periodo 2022–2024.	HE1: La frecuencia de superaciones horarias del umbral de 250 µg/m ³ de SO ₂ es significativamente mayor en días caracterizados por condiciones meteorológicas compatibles con inversión térmica.		
PE2: ¿Cuál es la relación estadística entre las concentraciones horarias de SO ₂ y las variables meteorológicas, considerando la diferencia espacial entre estaciones de	OE2: Analizar la relación estadística entre las concentraciones horarias de SO ₂ y las variables meteorológicas (BLH, temperatura,	HE2: Existe una relación estadística significativa entre las concentraciones horarias de SO ₂ y las variables meteorológicas locales, manifestándose en		

Problemas	Objetivos	Hipótesis	Variables	Metodología
barlovento y sotavento?	humedad relativa, presión atmosférica, velocidad y dirección del viento), incorporando la diferencia espacial entre estaciones de barlovento y sotavento.	niveles más elevados en la estación sotavento respecto a la estación barlovento durante episodios de inversión térmica.		
PE3: ¿De qué manera las condiciones de inversión térmica influyen en la acumulación y variación horaria de SO ₂ en la zona de estudio?	OE3: Evaluar la influencia de las condiciones de inversión térmica en la acumulación y variabilidad horaria de las concentraciones de SO ₂ en el área de estudio.	HE3: Durante los eventos de inversión térmica se identifican patrones horarios recurrentes de alta concentración de SO ₂ — especialmente en las primeras horas del día— que confirman la influencia directa de este fenómeno en la acumulación de contaminantes.		
PE4: ¿En qué medida las concentraciones horarias de SO ₂ superan el valor límite del ECA Aire de 24 horas y cuáles son las limitaciones de este estándar para activar mecanismos de alerta temprana?	OE4: Comparar las concentraciones horarias de SO ₂ registradas en el periodo 2022–2024 con los estándares normativos vigentes (ECA Aire de 24h y protocolos nacionales), e identificar las limitaciones de dichos instrumentos para la activación de mecanismos de alerta temprana en la fiscalización ambiental minera.	HE4a: Las concentraciones horarias de SO ₂ registradas en el área de estudio durante 2022–2024 presentan superaciones críticas (443 en 2022, 346 en 2023 y 413 en 2024) que no siempre se reflejan en los valores normativos establecidos en los promedios de 24 horas del ECA Aire. HE4b: El análisis conjunto de las concentraciones horarias de SO ₂ y de variables meteorológicas permite establecer criterios técnicos		

Problemas	Objetivos	Hipótesis	Variables	Metodología
		complementarios que refuercen los mecanismos de alerta temprana y la fiscalización ambiental del OEFA.		
PE5: ¿En qué medida el modelo Random Forest permite estimar con precisión las concentraciones horarias de SO ₂ integrando variables meteorológicas locales?	OE5: Desarrollar y validar un modelo predictivo Random Forest para estimar concentraciones horarias de SO ₂ , integrando variables meteorológicas locales (BLH, temperatura, humedad relativa, presión atmosférica y velocidad y dirección del viento), a fin de fortalecer la fiscalización ambiental minera.	HE5: El modelo de aprendizaje automático Random Forest predice con alta precisión las concentraciones horarias de SO ₂ al integrar variables meteorológicas locales, mostrando un desempeño superior a los métodos tradicionales de estimación.		

Fuente: Elaboración propia.

Anexo 2

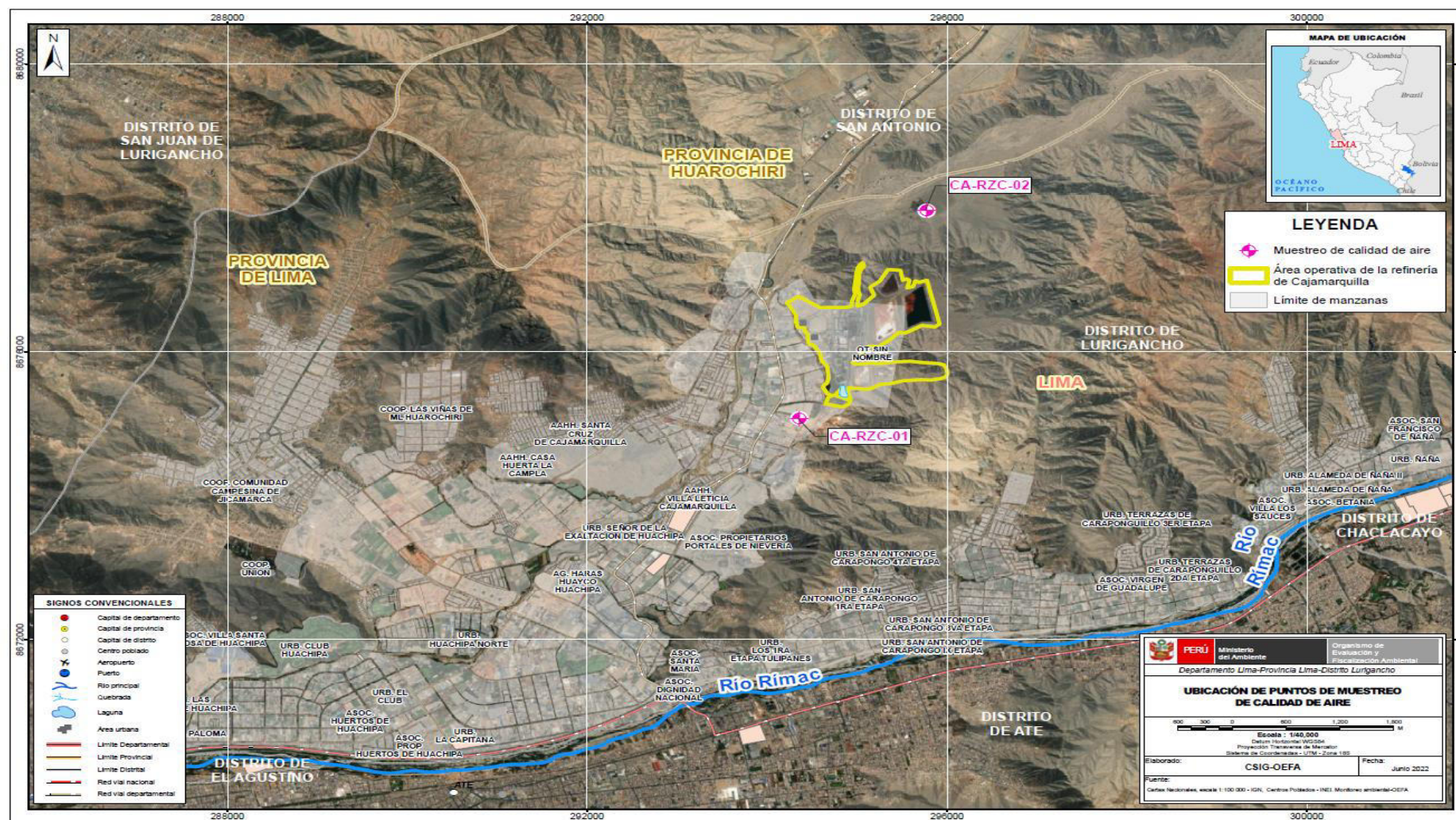
Matriz de operacionalización de variables

Variable	Dimensión	Indicador	Técnica / Instrumento	Fuente de Datos	Escala de Medición
Variable Dependiente (VD1): Concentración horaria de SO ₂	Y1: Comportamiento temporal	I1: Concentración promedio horaria y diaria de SO ₂ (µg/m ³)	Estadística descriptiva, análisis de frecuencias y cruces temporales	OEFA 2022-2024	µg/m ³
		I2: Número de superaciones de las concentraciones horarias de SO ₂ (>250 µg/m ³)			
Variable Dependiente - VD2 (Subdimensión): Distribución espacial de la concentración horaria de SO ₂	Y2: Distribución espacial	I3: Diferencia de concentración promedio entre estación a barlovento y estación a sotavento	Comparación de medias, análisis de dispersión espacial	OEFA 2022-2024	µg/m ³
Variable Independiente (VI1): Condiciones de Inversión térmica	X: Condiciones de inversión térmica	I4: Altura de la capa límite planetaria (BLH, en metros)	Análisis de series temporales, estadística descriptiva	Copernicus ERA5 / OEFA	Metros/Hora
		I5: Número de horas con BLH < 300 m			
Variable de Interacción de VI-VD	Z: Coincidencia inversión-contaminación	I6: N.º de coincidencias entre BLH < 300 m y superaciones de las concentraciones horarias de SO ₂ (>250 µg/m ³)	Cruces temporales, análisis de correlación	OEFA + Copernicus ERA5	Frecuencia de eventos
Variable Dependiente (VD3): Modelamiento predictivo Random Forest RF	Y3: Concentración estimada por RF	I7: Precisión del modelo Random Forest en la predicción de concentraciones horarias de SO ₂ (RMSE, MAE, R ²)	Modelamiento Random Forest, validación cruzada temporal.	OEFA (2022-2024), variables meteorológicas	Métricas estadísticas
Nota: Los indicadores I2 (superaciones horarias y diarias) y I6 (coincidencia entre episodios de baja BLH y concentraciones críticas de SO ₂) se vinculan específicamente con las hipótesis HE4a y HE4b , respectivamente, reflejando la doble dimensión normativa y técnico-operativa de la fiscalización ambiental.					

Fuente: Elaboración propia.

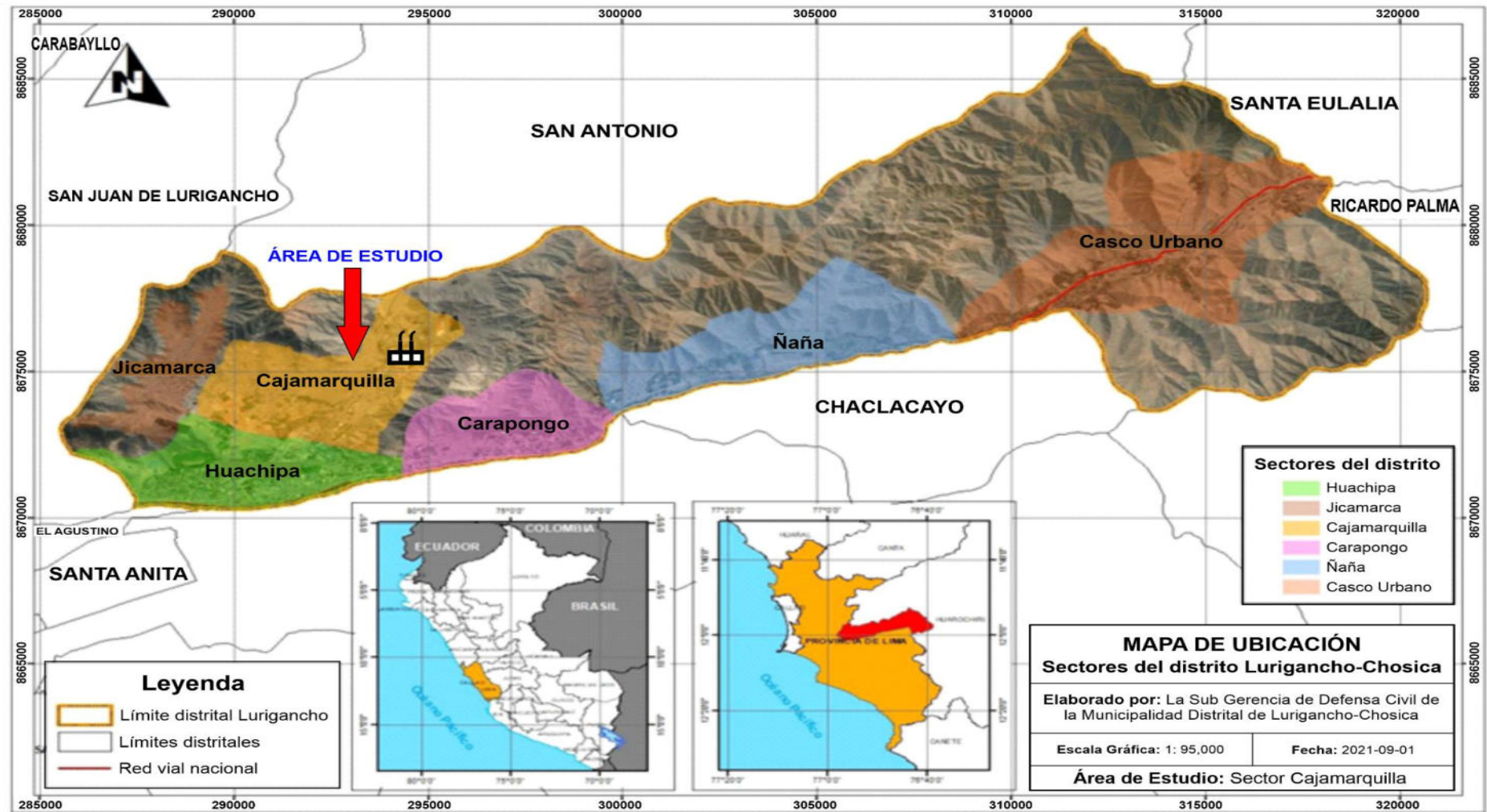
Anexo 3

Mapa de ubicación de las estaciones de monitoreo del OEFA



Fuente: Organismo de Evaluación y Fiscalización Ambiental- OEFA (2022).

Anexo 4

Mapa de ubicación del Área de Estudio

Fuente: Municipalidad distrital de Lurigancho-Chosica (2021).

Anexo 5***Población por zonas y rango de edades en Lurigancho- Chosica*****Tabla: Población por zonas en Lurigancho-Chosica**

Zona	Población	%
Jicamarca	29 273	13 %
Cajamarquilla	43 324	19 %
Huachipa-Nievería	23 079	10 %
Carapongo	14 063	6 %
Naña	19 156	9 %
Margen Derecha	65 571	29 %
Margen Izquierda	30 561	14 %
Total	225 026	100 %

Fuente: Elaboración propia con base en Perfiles zonales 2016, Ipsos Perú (2016).

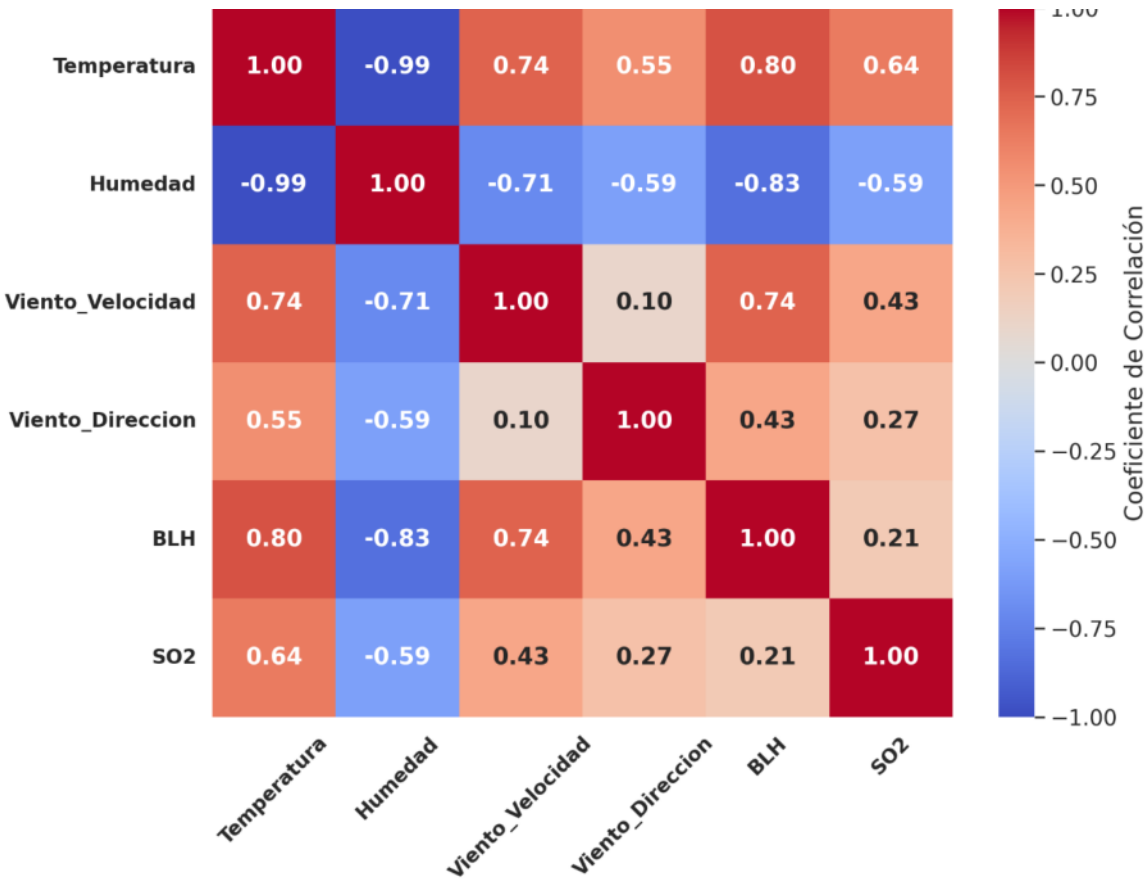
Tabla: Población por rango de edades en Lurigancho-Chosica

Edad	Población	%
Menos de 7	26 849	12 %
De 7 a 12	22 078	10 %
De 13 a 20	33 898	15 %
De 21 a 35	58 618	26 %
De 36 a 59	62 877	28 %
De 60 a más	20 707	9 %
Total	225 026	100 %

Fuente: Elaboración propia con base en Perfiles zonales 2016, Ipsos Perú (2016).

Anexo 6

Matriz de correlación durante un episodio de inversión térmica entre SO₂ y variables meteorológicas (5 de julio de 2024) – Estación Sotavento



El día con la inversión térmica más marcada (es decir, con la altura de la capa límite atmosférica más baja entre las 4 y 8 a.m.) corresponde al **5 de julio de 2024**.

Aquí se muestra la **matriz de correlación** correspondiente al **5 de julio de 2024**, fecha identificada con una **inversión térmica atmosférica marcada** según los promedios de la BLH (altura de la capa límite atmosférica) en las primeras horas del día.

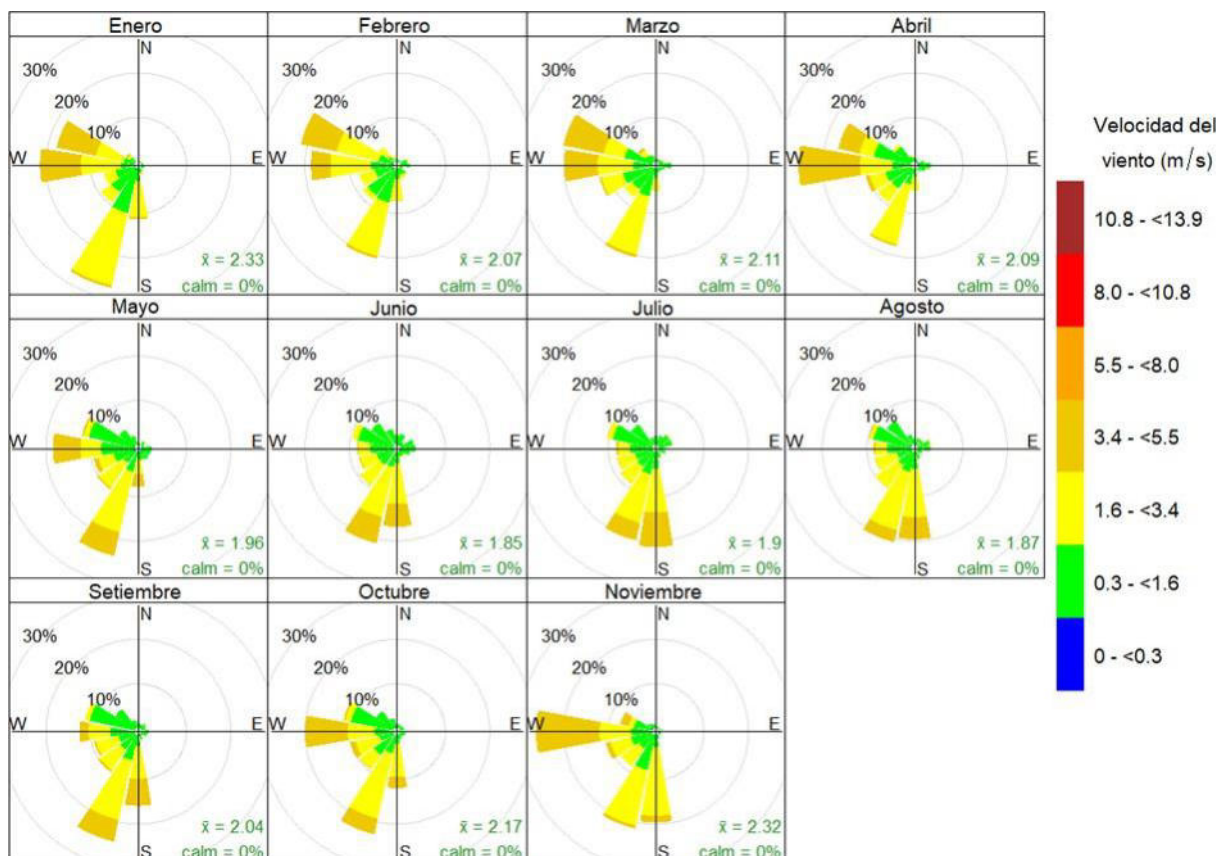
Interpretación destacada:

- La concentración de **SO₂** presenta una correlación **positiva moderada con la temperatura** (0.64) y con la **velocidad del viento** (0.43), pero **débil con la BLH** (0.21), lo cual podría reflejar que en condiciones muy estables (BLH baja), el SO₂ se acumula en niveles bajos.
- La **temperatura y la humedad relativa** están **fuertemente correlacionadas de forma inversa** (-0.99), lo que es esperable en condiciones de inversión térmica.
- La **BLH** está positivamente correlacionada con la temperatura (0.80), como también es típico en condiciones atmosféricas estables.

Fuente: Elaboración propia.

Anexo 7

Diagramas de rosa de los vientos por meses SO2 para el año 2024, de la estación de monitoreo ubicada a sotavento.



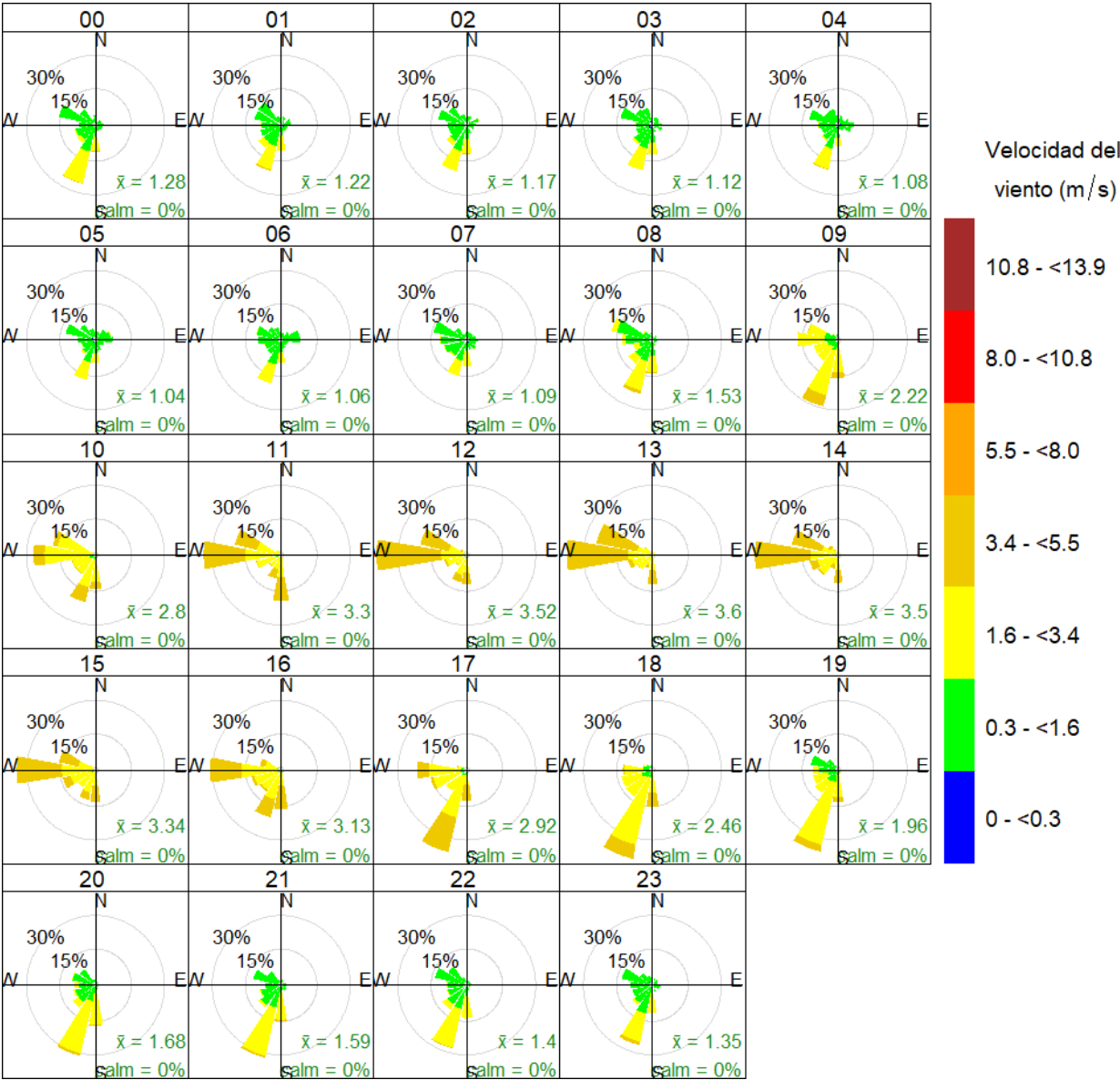
Fuente: Organismo de Evaluación y Fiscalización Ambiental-OEFA (2024).

En la Figura se presenta la rosa de los vientos mensual del 1 enero al 30 noviembre de 2024, de la estación CA-RZC-02, ubicada a sotavento de la refinería de zinc Cajamarquilla de Nexa Resources Cajamarquilla S.A.C., donde la predominancia de los vientos (procedencia) fue desde el sur suroeste (SSO) con una frecuencia de 21,9% seguido de vientos desde el oeste (O).

Anexo 8

Rosa de vientos por horas para el SO2 de las 00:00 a las 23 horas, de enero a noviembre de 2024, en la estación ubicada a sotavento

(Las paletas indican la dirección de donde proviene la predominancia de los vientos (procedencia))



Fuente: Organismo de Evaluación y Fiscalización Ambiental- OEFA (2024).

Anexo 9

Estándares de calidad ambiental de la calidad para aire, aprobado mediante D.S. N° 03-2017-MINAM

Parámetros	Periodo	Valor [µg/m³]	Criterios de evaluación	Método de análisis ^[1]
Benceno (C ₆ H ₆)	Anual	2	Media aritmética anual	Cromatografía de gases
Dióxido de Azufre (SO ₂)	24 horas	250	NE más de 7 veces al año	Fluorescencia ultravioleta (Método automático)
Dióxido de Nitrógeno (NO ₂)	1 hora	200	NE más de 24 veces al año	Quimioluminiscencia (Método automático)
	Anual	100	Media aritmética anual	
Material Particulado con diámetro menor a 2,5 micras (PM _{2,5})	24 horas	50	NE más de 7 veces al año	Separación inercial/filtración (Gravimetría)
	Anual	25	Media aritmética anual	
Material Particulado con diámetro menor a 10 micras (PM ₁₀)	24 horas	100	NE más de 7 veces al año	Separación inercial/filtración (Gravimetría)
	Anual	50	Media aritmética anual	
Mercurio Gaseoso Total (Hg) ^[2]	24 horas	2	No exceder	Espectrometría de absorción atómica de vapor frío (CVAAS) o Espectrometría de fluorescencia atómica de vapor frío (CVAFS) o Espectrometría de absorción atómica Zeeman. (Métodos automáticos)
Monóxido de Carbono (CO)	1 hora	30000	NE más de 1 vez al año	Infrarrojo no dispersivo (NDIR) (Método automático)
	8 horas	10000	Media aritmética móvil	
Ozono (O ₃)	8 horas	100	Máxima media diaria NE más de 24 veces al año	Fotometría de absorción ultravioleta (Método automático)
Plomo (Pb) en PM ₁₀	Mensual	1,5	NE más de 4 veces al año	Método para PM ₁₀ (Espectrofotometría de absorción atómica)
	Anual	0,5	Media aritmética de los valores mensuales	
Sulfuro de Hidrógeno (H ₂ S)	24 horas	150	Media aritmética	Fluorescencia ultravioleta (Método automático)

NE: No Exceder.

^[1] o método equivalente aprobado.

^[2] El estándar de calidad ambiental para Mercurio Gaseoso Total entrará en vigencia al día siguiente de la publicación del Protocolo Nacional de Monitoreo de la Calidad Ambiental del Aire, de conformidad con lo establecido en la Séptima Disposición Complementaria Final del presente Decreto Supremo.

ECA de Aire de los parámetros cadmio, arsénico y cromo en material Particulado menor a 10 micras (PM₁₀), aprobado mediante D.S. N° 011-2023-MINAM

Parámetros	Periodo	Valor [µg/m³]	Criterios de evaluación	Método de ensayo
Arsénico (As) en material particulado menor a diez micras (PM ₁₀).	24 horas	0,3	No exceder	EPA IO-3.3 "Determination of metals in ambient particulate matter using X-ray fluorescence (XRF) spectroscopy" (otros metales en PM ₁₀).
	Anual	0,023	Media aritmética anual	
Cadmio (Cd) en material particulado menor a diez micras (PM ₁₀).	24 horas	0,09	No exceder	EPA IO-3.4 "Determination of metals in ambient particulate matter using ICP spectroscopy" (para ICP- OES)
	Anual	0,018	Media aritmética anual	
Cromo (Cr) en material particulado menor a diez micras (PM ₁₀).	24 horas	0,5	No exceder	EPA IO-3.5 "Determination of metals in ambient particulate matter using ICP mass spectrometry" (para ICP-MS)
	Anual	0,11	Media aritmética anual	

Fuente: Ministerio del Ambiente-MINAM (2017, 2023).

Anexo 10

Reglamento de los Niveles de Estados de Alerta Nacionales para contaminantes del aire, aprobado mediante D.S. N.º 009-2003-S.A.

TIPOS DE ALERTA	Material Particulado (PM10)	Dióxido de Azufre (SO₂)	Monóxido de Carbono (CO)	Sulfuro de Hidrógeno (H₂S)
Cuidado	>250 ug/m ³ prom. aritmético 24 horas	>500 ug/m ³ prom. móvil 3 horas	>15 000 ug/m ³ prom. móvil 8 horas	>1 500 ug/m ³ prom. aritmético 24 horas
Peligro	>350 ug/m ³ prom. aritmético 24 horas	>1 500 ug/m ³ prom. móvil 3 horas	>20 000 ug/m ³ prom. móvil 8 horas	>3 000 ug/m ³ prom. aritmético 24 horas
Emergencia	>420 ug/m ³ prom. aritmético 24 horas	>2 500 ug/m ³ prom. móvil 3 horas	>35 000 ug/m ³ prom. móvil 8 horas	>5 000 ug/m ³ prom. aritmético 24 horas"

SEGUNDA DISPOSICIÓN COMPLEMENTARIA

Declárese en las zonas de Atención Prioritaria que a continuación se indican, la necesidad de aplicar las disposiciones del presente reglamento para los siguientes contaminantes críticos:

Chimbote: Sulfuro de hidrógeno

Ilo: Dióxido de azufre y material particulado

La oroya: Dióxido de azufre y material particulado

Lima: Material particulado

Arequipa: Material particulado.

Fuente: Ministerio de Salud - MINSA (2003).

Anexo 11

Valores Guía establecidos por la OMS para la calidad del aire

La Organización Mundial de la Salud (OMS) presentó el 22 de septiembre de 2021 las nuevas Directrices Mundiales sobre la Calidad del Aire (AQG). La nueva versión de los AQG se desarrolló sobre la base de los límites de umbral ya disponibles que estaban en vigor desde 2005. En el siguiente cuadro se indican los valores umbral de 2005 y los límites anunciados recientemente. Los valores actualizados proporcionan pruebas claras de que la mala calidad del aire daña la salud humana en concentraciones incluso más bajas de lo que se suponía anteriormente.

Directrices mundiales de la OMS sobre la calidad del aire

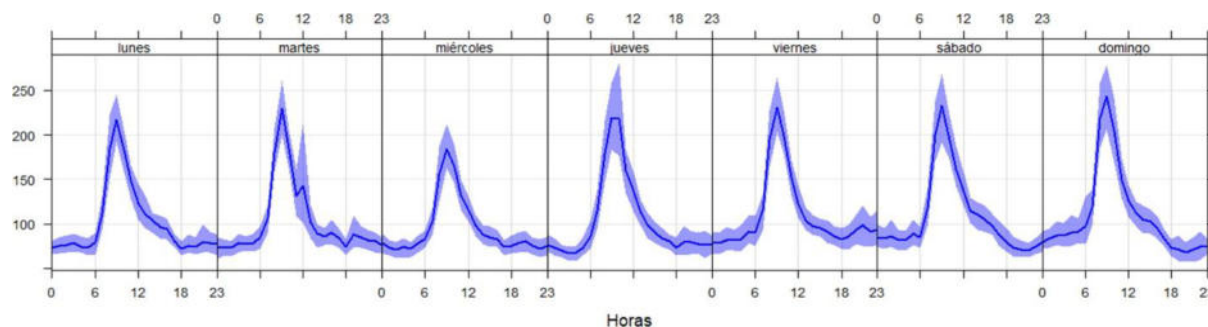
Pollutant		WHO guidelines 2005	New WHO guidelines 2021
PM _{2.5}	Annual	10 µg/m ³	5 µg/m ³
	24-hour	25 µg/m ³	15 µg/m ³
PM ₁₀	Annual	20 µg/m ³	15 µg/m ³
	24-hour	50 µg/m ³	45 µg/m ³
O ₃	Peak season (the six consecutive months of the year with the highest six-month running-average ozone concentration)	---	60 µg/m ³
	8-hour	100 µg/m ³	100 µg/m ³
NO ₂	Annual	40 µg/m ³	10 µg/m ³
	24-hour	---	25 µg/m ³
	1-hour	---	200 µg/m ³
SO ₂	24-hour	20 µg/m ³	40 µg/m ³
	10 min	---	500 µg/m ³
CO	24-hour	---	4 mg/m ³
	8-hour	---	10 mg/m ³
	1-hour	---	35 mg/m ³
	15 min	---	100 mg/m ³

Fuente: Organización Mundial de la Salud-OMS (2021).

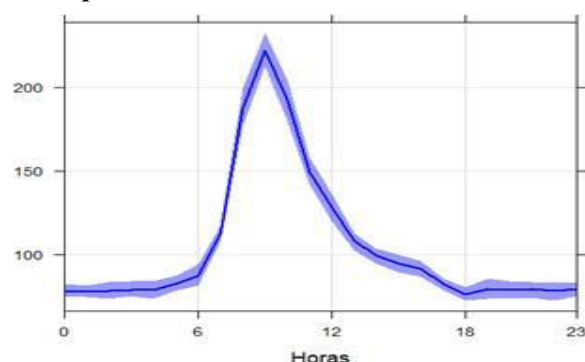
Anexo 12

Comportamiento de las concentraciones ($\mu\text{g}/\text{m}^3$) de SO_2 promediadas, en la estación del OEFA ubicada a sotavento de enero a noviembre de 2024.

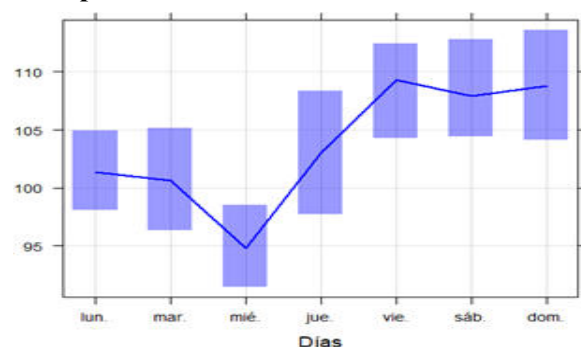
a) Comportamiento horario en días de la semana de SO_2 de enero a noviembre 2024



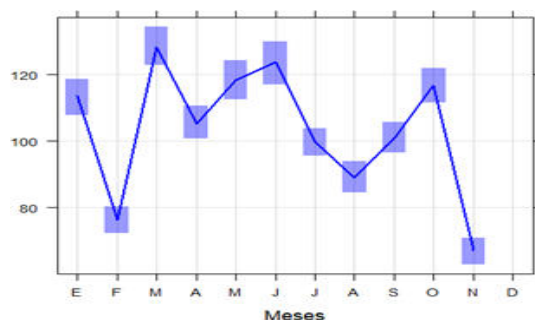
b) Comportamiento horario en 24 horas



c) Comportamiento diario de SO_2



d) Comportamiento mensual de SO_2



En las figuras (a, b) se observan que, las concentraciones horarias de SO_2 empiezan a incrementarse desde las 6:00 horas hasta las 9:00 horas para luego decaer paulatinamente hasta las 18:00 horas, por otro lado, desde las 18:00 horas hasta las 5:00 horas del día siguiente, las concentraciones horarias de SO_2 han sido las más bajas ($< 80,0 \mu\text{g}/\text{m}^3$). Generalmente, las concentraciones horarias de SO_2 más altas, se ha registrado desde las 8:00 horas hasta las 10:00 horas presentándose picos de $1120,44 \mu\text{g}/\text{m}^3$ (10:00 horas del 11/04/24), $734,23 \mu\text{g}/\text{m}^3$ (8:00 horas del 13/10/24), $676,01 \mu\text{g}/\text{m}^3$ (8:00 horas del 30/06/24), $665,17 \mu\text{g}/\text{m}^3$ (8:00 horas del 12/10/24), $662,13 \mu\text{g}/\text{m}^3$ (9:00 horas del 30/06/24), $652,93 \mu\text{g}/\text{m}^3$ (9:00 horas del 25/05/24), entre otras concentraciones menores.

Las concentraciones en 24 horas de SO_2 más altas de acuerdo a las figuras (a, c) se han registrado generalmente los viernes, sábado y domingos. Asimismo, se observa que, en la figura (d) en promedio las concentraciones mensuales de SO_2 más altas fueron, en marzo y junio de 2024. Los meses con las menores concentraciones de SO_2 fueron en noviembre y febrero de 2024.

Anexo 13***Ficha de Informe de juicio de experto del instrumento de investigación.***

● Título de la Tesis:

● Presentado por:

I. DATOS DEL EXPERTO:

1.1. Apellidos y Nombres:

1.2. Grado Académico:

1.4. Instrumento de

Evaluación:

INDICADORES	CRÉDITOS	Deficiente 0 - 20 %	Regular 21 -40 %	Bueno 41 - 80 %	Muy Bueno 61 - 80 %	Excelente 81 - 100 %
1. Claridad	Está formulado con lenguaje apropiado					
2. Objetividad	Está expresado en conducta observable					
3. Actualidad	Es adecuado el avance de la ciencia y tecnología					
4. Organización	Existe organización lógica					
5. Suficiencia	Comprende los aspectos de cantidad y calidad					
6. Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					
7. Consistencia	Se basa en aspectos teóricos, científicos acordes a la tecnología educativa					
8. Coherencia	Entre índices, indicadores y dimensiones					
9. Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					

II. OPCION DE APLICABILIDAD:

III. PROMEDIO DE VALORACION:

IV. RECOMENDACIONES.

Fecha:

Firma del Experto

Fuente: Elaboración propia

Anexo 14

Ficha de Validación del Instrumento - Experto 1

VALIDACIÓN DEL INSTRUMENTO

EXPERTO N.º 1

• Título de la Tesis:

EVALUACIÓN DEL IMPACTO DE LA INVERSIÓN TÉRMICA EN LA CONCENTRACIÓN DE DIÓXIDO DE AZUFRE (SO₂) PARA OPTIMIZAR LA FISCALIZACIÓN AMBIENTAL MINERA EN LA ZONA DE CAJAMARQUILLA DEL DISTRITO LURIGANCHO-CHOSICA, LIMA.

• Presentado por:

Lázaro Walther Fajardo Vargas

I. DATOS DEL EXPERTO:

1.1. Apellidos y Nombres:

Caycho Bustamante, Milagros Karina

1.2. Grado Académico:

Maestra en Gestión Ambiental (U. Nacional Federico Villarreal – Perú)

1.4. Instrumento de Evaluación:

Encuesta sobre la recolección de datos y de la evaluación de resultados

INDICADORES	CRÉDITOS	Deficiente 0 - 20 %	Regular 21 - 40 %	Bueno 41 - 80 %	Muy Bueno 61 - 80 %	Excelente 81 - 100 %
1. Claridad	Está formulado con lenguaje apropiado				X	
2. Objetividad	Está expresado en conducta observable					X
3. Actualidad	Es adecuado el avance de la ciencia y tecnología					X
4. Organización	Existe organización lógica				X	
5. Suficiencia	Comprende los aspectos de cantidad y calidad					X
6. Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico				X	
7. Consistencia	Se basa en aspectos teóricos, científicos acordes a la tecnología educativa					X
8. Coherencia	Entre índices, indicadores y dimensiones					X
9. Metodología	Responde al propósito del trabajo bajo los objetivos a lograr				X	

II. OPCION DE APLICABILIDAD:

Se puede aplicar

III. PROMEDIO DE VALORACION:

90.12

Nivel: Excelente

IV. RECOMENDACIONES.

Se recomienda incorporar un mapa del distrito de San Juan de Lurigancho- Chosica por sectores, en donde se ubique el área de estudio de la investigación.

Fecha:

18/07/2025



Mag. Milagros Karina Caycho Bustamante
DNI.07764260

Anexo 15

Ficha de Validación del Instrumento - Experto 2

VALIDACIÓN DEL INSTRUMENTO

EXPERTO N.º 2

- Titulo de la Tesis:

EVALUACIÓN DEL IMPACTO DE LA INVERSIÓN TÉRMICA EN LA CONCENTRACIÓN DE DIÓXIDO DE AZUFRE (SO₂) PARA OPTIMIZAR LA FISCALIZACIÓN AMBIENTAL MINERA EN LA ZONA DE CAJAMARQUILLA DEL DISTRITO LURIGANCHO-CHOSICA, LIMA.
- Presentado por:

Lázaro Walther Fajardo Vargas

I. DATOS DEL EXPERTO:

- 1.1. Apellidos y Nombres:

Espejo Ramírez, Javier Rodolfo
- 1.2. Grado Académico:

Maestro en Gestión Ambiental (UNFV)
- 1.4. Instrumento de Evaluación:

Encuesta sobre la recolección de datos y de la evaluación de resultados

INDICADORES	CRÉDITOS	Deficiente 0 - 20 %	Regular 21 -40 %	Bueno 41 - 80 %	Muy Bueno 61 - 80 %	Excelente 81 - 100 %
1. Claridad	Está formulado con lenguaje apropiado					X
2. Objetividad	Está expresado en conducta observable				X	
3. Actualidad	Es adecuado el avance de la ciencia y tecnología					X
4. Organización	Existe organización lógica					X
5. Suficiencia	Comprende los aspectos de cantidad y calidad				X	
6. Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico				X	
7. Consistencia	Se basa en aspectos teóricos, científicos acordes a la tecnología educativa					X
8. Coherencia	Entre índices, indicadores y dimensiones				X	
9. Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					X

II. OPCION DE APLICABILIDAD: Se puede aplicar

III. PROMEDIO DE VALORACION: 90.12 Nivel: Excelente

IV. RECOMENDACIONES.

Se recomienda realizar una matriz de correlación general entre el SO₂ y las variables meteorológicas en el periodo 2022 – 2024 con los datos de la estación ubicada a sotavento que confirme las condiciones de estabilidad atmosférica y baja dispersión vertical, los mismos que contribuyen significativamente al incremento de las concentraciones de SO₂.

Fecha: 11/08/2025



Mag. Javier Rodolfo Espejo Ramírez
DNI: 08726688

Anexo 16

Ficha de Validación del Instrumento - Experto 3

VALIDACIÓN DEL INSTRUMENTO

EXPERTO N.º 3

• Título de la Tesis:

EVALUACIÓN DEL IMPACTO DE LA INVERSIÓN TÉRMICA EN LA CONCENTRACIÓN DE DIÓXIDO DE AZUFRE (SO₂) PARA OPTIMIZAR LA FISCALIZACIÓN AMBIENTAL MINERA EN LA ZONA DE CAJAMARQUILLA DEL DISTRITO LURIGANCHO-CHOSICA, LIMA.

• Presentado por:

Lázaro Walther Fajardo Vargas

I. DATOS DEL EXPERTO:

- 1.1. Apellidos y Nombres: Requena Sánchez Norvin Plumieer
 1.2. Grado Académico: Magister en Ciencias de la Ingeniería Ambiental (Universidad de Sao Paulo – Brasil)
 1.4. Instrumento de Evaluación: Encuesta sobre la recolección de datos y de la evaluación de resultados

INDICADORES	CRÉDITOS	Deficiente 0 - 20 %	Regular 21 - 40 %	Bueno 41 - 80 %	Muy Bueno 61 - 80 %	Excelente 81 - 100 %
1. Claridad	Está formulado con lenguaje apropiado					X
2. Objetividad	Está expresado en conducta observable					X
3. Actualidad	Es adecuado el avance de la ciencia y tecnología				X	
4. Organización	Existe organización lógica				X	
5. Suficiencia	Comprende los aspectos de cantidad y calidad					X
6. Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					X
7. Consistencia	Se basa en aspectos teóricos, científicos acordes a la tecnología educativa			X		
8. Coherencia	Entre índices, indicadores y dimensiones					X
9. Metodología	Responde al propósito del trabajo bajo los objetivos a lograr					X

II. OPCION DE APLICABILIDAD: Se puede aplicar

III. PROMEDIO DE VALORACION: 91.36 Nivel: Excelente

IV. RECOMENDACIONES.

Se recomienda realizar una validación cruzada temporal con 5 particiones (K = 5) a fin de fortalecer la robustez del análisis predictivo y evaluar de manera más precisa la capacidad de generalización del modelo Random Forest.

Fecha: 18/08/2025

MSC. Norvin Plumieer Requena Sánchez

DNI: 43489032

Anexo 17

Ficha de Validación del Instrumento - Experto 4

VALIDACIÓN DEL INSTRUMENTO

EXPERTO N.º 4

EVALUACIÓN DEL IMPACTO DE LA INVERSIÓN TÉRMICA EN LA CONCENTRACIÓN DE DIÓXIDO DE AZUFRE (SO₂) PARA OPTIMIZAR LA FISCALIZACIÓN AMBIENTAL MINERA EN EL SECTOR DE CAJAMARQUILLA DEL DISTRITO LURIGANCHO-CHOSICA, LIMA.

• Título de la Tesis:

• Presentado por:

Lázaro Walther Fajardo Vargas

I. DATOS DEL EXPERTO:

1.1. Apellidos y Nombres:

Calla Llontop Helen Jesús

1.2. Grado Académico:

Magister en Ciencias Ambientales, mención en Desarrollo Sustentable en Minería y Recursos Energéticos (Universidad Nacional Mayor de San Marcos)

1.4. Instrumento de Evaluación:

Encuesta sobre la recolección de datos y de la evaluación de resultados

INDICADORES	CRÉDITOS	Deficiente 0 - 20 %	Regular 21 - 40 %	Bueno 41 - 60 %	Muy Bueno 61 - 80 %	Excelente 81 - 100 %
1. Claridad	Está formulado con lenguaje apropiado				X	
2. Objetividad	Está expresado en conducta observable					X
3. Actualidad	Es adecuado el avance de la ciencia y tecnología				X	
4. Organización	Existe organización lógica				X	
5. Suficiencia	Comprende los aspectos de cantidad y calidad					X
6. Intencionalidad	Adecuado para valorar aspectos del sistema metodológico y científico					X
7. Consistencia	Se basa en aspectos teóricos, científicos acordes a la tecnología educativa					X
8. Coherencia	Entre índices, indicadores y dimensiones					X
9. Metodología	Responde al propósito del trabajo bajo los objetivos a lograr				X	

II. OPCION DE APLICABILIDAD:

Se puede aplicar

III. PROMEDIO DE VALORACION:

90.12

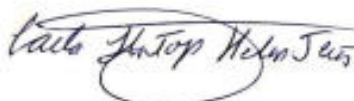
Nivel: Excelente

IV. RECOMENDACIONES.

Para ilustrar el desempeño operativo del modelo Random Forest, realizar una prueba de predicción de las concentraciones de SO₂ en la estación sotavento y elaborar una gráfica comparativa entre los valores observados y predichos, incluyendo una línea de referencia en el umbral de 250 µg/m³.

Fecha:

26/09/2025



MSc. Calla Llontop Helen Jesús

DNI:41006102

Anexo 18

Método de evaluación de resultados Validación de Instrumento

Para evaluar el grado de consenso entre los expertos en la validación del instrumento, se aplicó un **análisis de consistencia porcentual**. Este método considera que existe consenso cuando la mayoría de las calificaciones de los expertos se encuentran dentro de un rango de ± 1 punto respecto a la media del ítem, en una escala Likert de 1 a 5.

Se desarrollo un script desarrollado en Python para desarrollar el proceso de calificación final sobre los resultados de evaluación de los Expertos, a continuación, se explica los procedimientos del método análisis de consistencia porcentual seguido:

Paso 01: Cargar los datos del resultado de evaluación de los Expertos

Se carga al sistema los datos los resultados de evaluación de Expertos mediante el formato de la siguiente Tabla:

Tabla de los resultados de evaluación de Expertos

Ítem	Experto 1	Experto 2	Experto 3	Experto 4
1. Claridad	4	5	5	4
2. Objetividad	5	4	5	5
3. Actualidad	5	5	4	4
4. Organización	4	5	4	4
5. Suficiencia	5	4	5	5
6. Intencionalidad	4	4	5	5
7. Consistencia	5	5	3	5
8. Coherencia	5	4	5	5
9. Metodología	4	5	5	4

Fuente: Elaboración propia.

Paso 02: Procesamiento de Cálculos

- **Transformación a Escala Porcentual**

Los puntajes originales (escala Likert de 1 a 5) se convierten a una escala de 0-100% usando la fórmula:

Formula: Valor Porcentual = $P_i = (x-1) \times 25$

Donde: P_i : Porcentaje para el ítem i (escala 0-100%).

x : Puntaje original del experto para el ítem i (escala Likert 1-5).

Explicación:

Esta fórmula convierte los puntajes Likert en una escala porcentual.

Ejemplo: Un puntaje de 5 se convierte en $(5-1) \times 25=100\%$.

Esto estandariza las evaluaciones para alinearlas con la escala de la universidad.

- **Cálculo de Estadísticos Descriptivos**

Por cada ítem (fila), se calculan:

- ✓ **Media porcentual:** Promedio de los puntajes transformados de los cuatro expertos.

$$\bar{x} = \frac{1}{n} \sum_j^n P_j$$

Donde:

$\bar{x}$: Media porcentual de las evaluaciones para un ítem.

P_j : Porcentaje transformado del experto j .

n : Número de expertos (4 en este caso).

Explicación: Calcula el promedio de los puntajes transformados para cada ítem.

- ✓ **Desviación estándar:** Variabilidad de los puntajes transformados.

$$\sigma P = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (P_j - P)^2}$$

Donde:

σP : Desviación estándar de los puntajes transformados.

$P_j, \bar{x}, n$: Como se definió anteriormente.

Explicación: Mide la variabilidad de las evaluaciones porcentuales entre los expertos.

✓ **Mínimo y Máximo:** Valores extremos de los puntajes transformados. Estos estadísticos describen la percepción de los expertos sobre cada ítem (claridad, objetividad, etc.) y su variabilidad.

- **Clasificación según Escala Personalizada:**

- ✓ La media porcentual se categoriza según los rangos definidos:
 - Deficiente: 0-20%
 - Regular: 21-40%
 - Bueno: 41-60%
 - Muy Bueno: 61-80%
 - Excelente: 81-100%
- ✓ Esto permite interpretar la calidad de cada ítem según los criterios de la universidad.

- **Cálculo de Consistencia Porcentual:**

✓ Para cada ítem, se calcula la consistencia porcentual como:

Rango Aceptable = $[\bar{x}_i - 1, \bar{x}_i + 1]$

$$Ci = \frac{n \text{ con } X_{ij} \in [\bar{x}_i - 1, \bar{x}_i + 1] \times 100}{n}$$

Donde:

n : Numero de Expertos

C_i : Consistencia Porcentual

$\bar{x}_i$: es la media de las evaluaciones del ítem i

X_{ij} es la evaluación del ítem i por el experto j

Los resultados se expresan en porcentaje y se categorizan cualitativamente como Deficiente, Regular, Bueno, Muy Bueno o Excelente, según el valor de la media porcentual.

Este enfoque es especialmente útil cuando el número de expertos es reducido, permitiendo una interpretación clara y visual del grado de acuerdo alcanzado (Polit & Beck, 2006).

Método de evaluación de resultados de experto

Ítem	Media Porcentaje	Desviación Estándar	Mínimo	Máximo	Categoría	Consistencia (%)
2. Objetividad	93.75	12.50	75	100	Excelente	100
5. Suficiencia	93.75	12.50	75	100	Excelente	100
8. Coherencia	93.75	12.50	75	100	Excelente	100
1. Claridad	87.50	14.43	75	100	Excelente	100
3. Actualidad	87.50	14.43	75	100	Excelente	100
6. Intencionalidad	87.50	14.43	75	100	Excelente	100
7. Consistencia	87.50	25.00	50	100	Excelente	75
9. Metodología	87.50	14.43	75	100	Excelente	100
4. Organización	81.25	12.50	75	100	Excelente	100
Promedio	88.89	14.75	72.22	100.00	Excelente	97.22

Fuente: Elaboración propia

Anexo 19

Base de Datos a Barlovento y Sotavento

A continuación, se presenta el enlace de acceso a la base de datos utilizada en la presente investigación, descargada del **Portal Interactivo de Fiscalización Ambiental-PIFA** del Organismo de Evaluación y Fiscalización Ambiental-OEFA, que contiene **los registros horarios de concentración de dióxido de azufre (SO₂)** y de **variables meteorológicas** como temperatura, presión atmosférica y velocidad del viento, obtenidos por las dos estaciones automáticas de monitoreo ambiental instaladas por el **Organismo de Evaluación y Fiscalización Ambiental (OEFA)** en el área de influencia directa de la Refinería de Zinc de Cajamarquilla. Estas estaciones se encuentran estratégicamente ubicadas a **barlovento** y **sotavento** del complejo industrial, permitiendo caracterizar la dispersión espacial y temporal de contaminantes en función de las condiciones atmosféricas locales.

Adicionalmente, se incorporó una columna con datos de **altura de la capa límite atmosférica (BLH)**, la cual fue descargada del repositorio internacional del programa **ERA5 – Reanalysis Single Levels**, disponible en el **Copernicus Climate Data Store** (<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels>). Esta variable, no registrada directamente por las estaciones del OEFA, resulta clave para estimar la estabilidad atmosférica y la ocurrencia de inversión térmica, factores fundamentales en el comportamiento de la contaminación por SO₂.

https://drive.google.com/file/d/11KNQXMx0GoPa3YWosy4oHvk_aUm_vFAw/view

Anexo 20

¿Cómo funciona Random Forest?

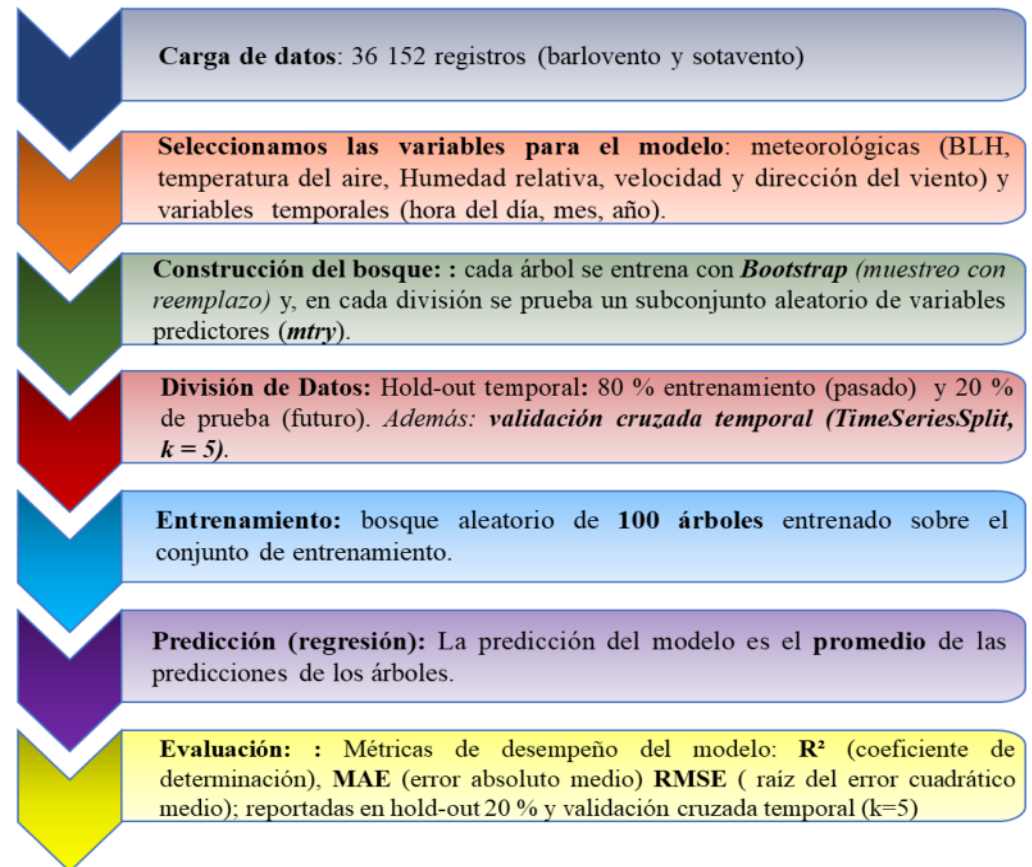
¿Cómo funciona Random Forest?

Random Forest es un algoritmo de aprendizaje automático basado en ensambles que utiliza múltiples árboles de decisión para realizar sus predicciones. La “magia” radica en la combinación de la simplicidad de los árboles de decisión con la potencia de los modelos de ensamble, lo que resulta en un sistema robusto capaz de manejar tanto tarea de clasificación como de regresión con alta precisión.



En el hold-out temporal (20 %) el RF alcanzó aprox. $R^2 = 0.789$, $MAE \approx 18.643 \mu\text{g}/\text{m}^3$ y $RMSE \approx 34.154 \mu\text{g}/\text{m}^3$; además, en 23–30/08/2023 el esquema tipo alerta (umbral $250 \mu\text{g}/\text{m}^3$) detectó 10/10 eventos sin falsas alarmas, lo que evidencia **buena capacidad de generalización** en horizontes **cortos** y **coherencia** con la física del problema en condiciones de **BLH baja** y **viento con baja capacidad de ventilación**.

Fuente: Elaboración Propia.

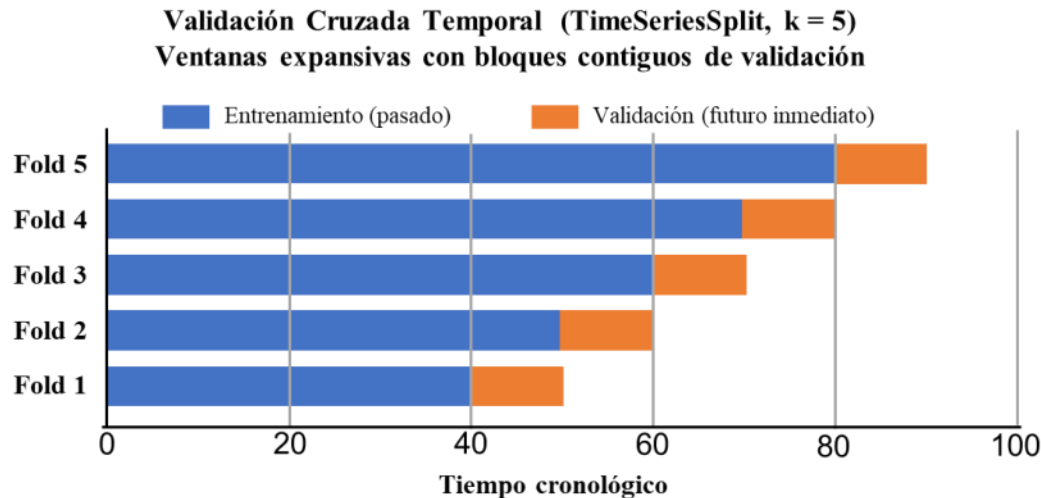


Anexo 21

Validación Cruzada temporal: cómo evaluamos el modelo Random Forest

Validación cruzada temporal: cómo evaluamos el modelo Random Forest

Se divide el conjunto de datos en k particiones ($k = 5$): en cada iteración una se usa para validación y las restantes para entrenamiento. Al final se promedian las métricas.



Datos horarios Sotavento (OEFA), 2022–2024; RF = 100 árboles. Bloques de validación contiguos (tamaño fijo ~10 %)

La validación cruzada temporal arrojó un promedio de $R^2 = 0.652 \pm 0.099$, $MAE = 26.772 \pm 5.119 \mu\text{g}/\text{m}^3$ y $RMSE = 47.497 \pm 12.469 \mu\text{g}/\text{m}^3$ (vs. hold-out: $R^2 = 0.789$; $MAE = 18.643$; $RMSE = 34.154$). Estos valores son **más conservadores y consistentes** con un esquema de evaluación **más estricto**, confirmando **estabilidad y desempeño robusto** del modelo en ventanas de tiempo crecientes.

Fuente: Elaboración propia

Para evaluar la capacidad de generalización en un contexto de **series temporales**, se implementó una **validación cruzada temporal (TimeSeriesSplit)** con **5 particiones ($k = 5$)**. El modelo fue configurado con **100 árboles** y evaluado con 3 métricas: R^2 , MAE y RMSE, ampliamente utilizadas en problemas de regresión ambiental.

¿Por qué TimeSeriesSplit?

- ✓ Respeto la **causalidad** (sin fuga de información).
- ✓ Emula la **operación real**: entreno con el pasado, valido en el futuro inmediato.
- ✓ **Ventana expansiva (walk-forward)**: el conjunto de entrenamiento **crece en cada fold**, capturando **estacionalidades**.

Valores promedio y desviación estándar obtenidos durante la validación cruzada		
Métrica	Promedio	Desviación Estándar
R^2	0.652	± 0.099
MAE ($\mu\text{g}/\text{m}^3$)	26.772	± 5.119
RMSE($\mu\text{g}/\text{m}^3$)	47.497	± 12.469

Anexo 22

Script extendido que corre el Random Forest para sotavento y barlovento, calcula métricas y guarda predicciones e importancias

Script: `run_rf_stations.py`

python

 Copiar código

```
import numpy as np
import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error

# -----
# Función para preparar datos
# -----
def preparar_datos(path_excel, estacion):
    df = pd.read_excel(path_excel)

    # Selección de columnas relevantes
    df = df[['date', 'SO2_1h_ugm3',
             'Altura Capa Límite BLH m',
             'TEMP_celc', 'HR_porc', 'WS_ms', 'WD_sexa', 'PBAR_hpa']].copy()
    df.columns = ['datetime', 'so2', 'blh', 'temp', 'hr', 'ws', 'wd', 'pbar']

    # Limpieza
    df['datetime'] = pd.to_datetime(df['datetime'], errors='coerce')
    df = df.dropna(subset=['datetime', 'so2', 'blh', 'temp', 'hr', 'ws', 'wd', 'pbar'])

    # Variables temporales
    df['hour'] = df['datetime'].dt.hour
    df['dayofyear'] = df['datetime'].dt.dayofyear
    df['month'] = df['datetime'].dt.month
    df['dayofweek'] = df['datetime'].dt.dayofweek

    # Transformaciones cíclicas
    df['hour_sin'] = np.sin(2*np.pi*df['hour']/24)
    df['hour_cos'] = np.cos(2*np.pi*df['hour']/24)
    df['doy_sin'] = np.sin(2*np.pi*df['dayofyear']/365)
    df['doy_cos'] = np.cos(2*np.pi*df['dayofyear']/365)
    df['month_sin'] = np.sin(2*np.pi*df['month']/12)
    df['month_cos'] = np.cos(2*np.pi*df['month']/12)
    df['dow_sin'] = np.sin(2*np.pi*df['dayofweek']/7)
    df['dow_cos'] = np.cos(2*np.pi*df['dayofweek']/7)
```

```

# Viento en coordenadas
df['wind_sin'] = np.sin(np.deg2rad(df['wd']))
df['wind_cos'] = np.cos(np.deg2rad(df['wd']))

# Features finales
features = ['blh', 'temp', 'hr', 'ws', 'pbar',
            'hour_sin', 'hour_cos', 'doy_sin', 'doy_cos',
            'month_sin', 'month_cos', 'dow_sin', 'dow_cos',
            'wind_sin', 'wind_cos']
X = df[features].values
y = df['so2'].values

return X, y

# -----
# Función para entrenar y evaluar
# -----
def correr_modelo(X, y, estacion):
    # Split 80/20
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42
    )

    # Random Forest
    rf = RandomForestRegressor(n_estimators=100, random_state=42, n_jobs=-1)
    rf.fit(X_train, y_train)
    y_pred = rf.predict(X_test)
    |
    print(f"\nResultados para {estacion}:")
    print(f"  R²    = {r2:.3f}")
    print(f"  RMSE = {rmse:.2f} µg/m³")
    print(f"  MAE  = {mae:.2f} µg/m³")

```

 Copiar código

```
# -----  
# Ejecución principal  
# -----  
if __name__ == "__main__":  
    # Rutas de tus archivos (ajusta según ubicación)  
    path_sotavento = "Datos PIFA OEFA SOTAVENTO 2022 -2024 Final.xlsx"  
    path_barlovento = "Datos PIFA OEFA BARLOVENTO 2022-2024 Final.xlsx"  
  
    # Preparar datos  
    X_sota, y_sota = preparar_datos(path_sotavento, "Sotavento")  
    X_barlo, y_barlo = preparar_datos(path_barlovento, "Barlovento")  
  
    # Correr modelos  
    correr_modelo(X_sota, y_sota, "Sotavento")  
    correr_modelo(X_barlo, y_barlo, "Barlovento")
```

Anexo 23

Formato de recolección de Datos

A continuación, se presenta el formato para la recolección de datos horarios de las estaciones automáticas del OEFA ubicadas a barlovento y sotavento de la Refinería de Zinc de Cajamarquilla, adicionalmente se incorporó una columna con datos de **altura de la capa límite atmosférica (BLH)**, que fue descargada del repositorio internacional del programa **ERA5 – Reanalysis Single Levels**, disponible en el **Copernicus Climate Data Store** (<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels>)

Fecha y hora	Presión atmosférica (hPa)	Precipitación (mm)	Temperatura (° C)	Humedad relativa (%)	Velocidad del viento (m/s)	Altura de la Capa Límite - BLH (m)	SO ₂ (µg/m ³)

Fuente: Elaboración propia

Anexo 24

*Certificado de calibración del Analizador automático de Dióxido de Azufre de
estación ubicada a Sotavento*

 GREEN GROUP PERÚ	LABORATORIO DE CALIBRACIÓN ACREDITADO POR EL ORGANISMO DE ACREDITACIÓN INACAL - DA CON REGISTRO N° LC-019	 INACAL DA - Perú Laboratorio de Calibración Acreditado Registro N° LC -019
---	--	---

CERTIFICADO DE CALIBRACIÓN N° LF -480-2023

Pág. 1 de 2

1. Solicitante	: ORGANISMO DE EVALUACIÓN Y FISCALIZACIÓN AMBIENTAL - OEFA
2. Dirección	: Av. Faustino Sánchez Carrión Nro. 603 - Jesús María - Lima
3. OTI	: 892C
4. Datos del Instrumento	
Instrumento de medición	: Analizador automático de Dióxido de Azufre
Marca	: Thermo Scientific
Modelo	: 43iQ
Serie	: 12123713083
Resolución	: 0,001 L/min
Intervalo de indicación	: No Aplica
Procedencia	: Estados Unidos
Identificación / Código interno	: 67220261-0035*
5. Lugar de Calibración	: Laboratorio de Flujo de aire - Green Group PE SAC.
6. Fecha de Calibración	: 2023-12-14
7. Método de Calibración.	

La calibración de Caudal se realizó por comparación del instrumento con patrones trazables según " Procedimiento ME-009 para la calibración de caudalímetro gases" Edición 1, 2021-03-15 CEM - España (Numeral 5.3.1 - calibración en situación A)

. La Incertidumbre de medición expandida reportada es la incertidumbre de medición estándar multiplicada por el factor de cobertura $k=2$ de modo que la probabilidad de cobertura corresponde aproximadamente a un nivel de confianza del 95%.
 . Los resultados emitidos son válidos solo para el instrumento en el momento de la calibración.
 . Se recomienda al usuario recalibrar a intervalos adecuados, los cuales deben ser elegidos con base a las características del instrumento.
 . La incertidumbre declarada en el presente certificado ha sido estimado siguiendo las directrices de: "Guía para la expresión de la incertidumbre de medida" primera edición, septiembre 2008 CEM.
 . El certificado de calibración solo puede ser difundido completamente y sin modificaciones, sin firma y sellos carecen de validez.
 . Esta prohibida toda reproducción parcial del presente certificado sin la autorización previa de GREEN GROUP PE S.A.C.

Fecha de Emisión

2023-12-15



ENZO BARRERA ZAVALA
Gerente de Operaciones
GREEN GROUP PE S.A.C.



LA IMPRESIÓN DE ESTE CERTIFICADO CONSTITUYE UNA COPIA DEL ORIGINAL EN VERSIÓN ELECTRÓNICA (FIRMA DIGITAL SEGÚN LEY N° 27269 LEY DE FIRMAS Y CERTIFICADOS DIGITALES)

Av. Aviación 4210 - Surquillo Central: (511) 273-3550 / (511) 580-6134 www.greengroup.com.pe

FM11

Fuente: Organismo de Evaluación y Fiscalización Ambiental-OEFA (2023).

Anexo 25

Certificado de calibración del Analizador automático de Dióxido de Azufre de estación ubicada a Barlovento

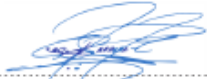
 GREEN GROUP PERÚ	LABORATORIO DE CALIBRACIÓN ACREDITADO POR EL ORGANISMO DE ACREDITACIÓN INACAL - DA CON REGISTRO N° LC-019	 INACAL DA - Perú Laboratorio de Calibración Acreditado Registro N° LC -019
--	---	--

CERTIFICADO DE CALIBRACIÓN N° LF -481-2023

Pág. 1 de 2

1. Solicitante 2. Dirección 3. OTI 4. Datos del Instrumento Instrumento de medición Marca Modelo Serie Resolución Intervalo de indicación Procedencia Identificación / Código interno 5. Lugar de Calibración 6. Fecha de Calibración 7. Método de Calibración.	: ORGANISMO DE EVALUACIÓN Y FISCALIZACIÓN AMBIENTAL - OEFA : Av. Faustino Sánchez Carrión Nro. 803 - Jesús María - Lima : 885C : Analizador automático de Dióxido de Azufre : Thermo Scientific : 43iQ : 12123713085 : 0,001 L/min : No Aplica : Estados Unidos : 67220261-0039* : Laboratorio de Flujo de aire - Green Group PE SAC. : 2023-12-14 La calibración de Caudal se realizó por comparación del instrumento con patrones trazables según " Procedimiento ME-009 para la calibración de caudalímetro gases" Edición 1, 2021-03-15 CEM - España (Numeral 5.3.1 - calibración en situación A)
--	--

. La Incertidumbre de medición expandida reportada es la incertidumbre de medición estándar multiplicada por el factor de cobertura $k=2$ de modo que la probabilidad de cobertura corresponde aproximadamente a un nivel de confianza del 95%.
 . Los resultados emitidos son válidos solo para el instrumento en el momento de la calibración.
 . Se recomienda al usuario recalibrar a intervalos adecuados, los cuales deben ser elegidos con base a las características del instrumento.
 . La incertidumbre declarada en el presente certificado ha sido estimado siguiendo las directrices de: "Guía para la expresión de la incertidumbre de medida" primera edición, septiembre 2008 CEM.
 . El certificado de calibración solo puede ser difundido completamente y sin modificaciones, sin firma y sellos carecen de validez.
 . Esta prohibida toda reproducción parcial del presente certificado sin la autorización previa de GREEN GROUP PE S.A.C.

Fecha de Emisión 2023-12-15	 ENZO BARRERA ZAVALA Gerente de Operaciones GREEN GROUP PE S.A.C.
---	---

 LA IMPRESIÓN DE ESTE CERTIFICADO CONSTITUYE UNA COPIA DEL ORIGINAL EN VERSIÓN ELECTRÓNICA (FIRMA DIGITAL SEGÚN LEY N° 27269 LEY DE FIRMAS Y CERTIFICADOS DIGITALES) Av. Aviación 4210 - Surquillo Central: (511) 273-3550 [(511) 560-8134] www.greengroup.com.pe	FM11
--	------

Fuente: Organismo de Evaluación y Fiscalización Ambiental-OEFA (2023).