

UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI

POSGRADO



MAESTRÍA EN ESTADÍSTICA APLICADA

“Incidencia de la meteorología en la calidad del aire antes, durante y después de la pandemia COVID-19 en Quito – Ecuador”.

Trabajo de titulación previa la obtención del

Título de Magister en Estadística Aplicada

Autor: Pablo Neptalí Escobar Tibán.

Tutora: Ph.D. Mercy Lucila Ilbay Yupa

Tulcán, diciembre 2023

CERTIFICADO DEL TUTOR

Certifico que el estudiante Pablo Neptalí Escobar Tibán con el número de cédula 1600526378 ha elaborado el Trabajo de Titulación: “Incidencia de la meteorología en la calidad del aire antes, durante y después de la pandemia COVID-19 en Quito – Ecuador”.

Este trabajo se sujeta a las normas y metodología dispuestas en la Codificación del Reglamento de Régimen Académico y de Estudiantes de la Universidad Politécnica Estatal del Carchi con RESOLUCIÓN No. 171-CSUP-2023, por lo tanto, autorizo su presentación para la sustentación respectiva.

f.....


Ph.D. Mercy Lucila Ilbay Yupa

Tulcán, diciembre 2023

AUTORÍA DE TRABAJO

El presente trabajo de titulación constituye un requisito previo para la obtención del título de Magister en Estadística Aplicada.

Yo, Pablo Neptalí Escobar Tibán, ciudadano ecuatoriano con cédula de identidad número 1600526378 declaro: que la investigación es absolutamente original, autentica, personal y los resultados y conclusiones a los que he llegado son de mi absoluta responsabilidad.

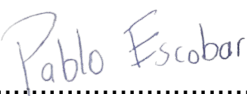
f. 
f.....

Pablo Neptalí Escobar Tibán

Tulcán, diciembre 2023

ACTA DE CESIÓN DE DERECHOS DEL TRABAJO DE TITULACIÓN

Yo, Pablo Neptalí Escobar Tibán declaro ser autor de los criterios emitidos en el trabajo de titulación: “Incidencia de la meteorología en la calidad del aire antes, durante y después de la pandemia COVID-19 en Quito – Ecuador” y eximo expresamente a la Universidad Politécnica Estatal del Carchi y a sus representantes legales de posibles reclamos o acciones legales.

f. 

Pablo Neptalí Escobar Tibán

Tulcán, diciembre 2023

DEDICATORIA

Dedico mi trabajo a mis hermanos Andrés y Aracely quienes han sabido estar apoyándome de una u otra manera para que siga adelante, y por tal motivo les invito a que sigan triunfando como lo han venido haciendo hasta el momento.

También dedico este trabajo a Pepa y Niki quienes han traído una gran calma a mi ser cuando más he necesitado, al saber estar por sobre cualquier incidente que haya sucedido. Espero que se encuentren por muchos años más compartiendo conmigo sus vivencias y travesuras.

AGRADECIMIENTO

Agradezco inmensamente a mis padres quienes han sido un pilar fundamental para poder graduarme de la Maestría, por su incondicional ayuda tanto económica como con mensajes de aliento para no decaer y completar la meta.

A su vez agradezco a Barbara Villacís, quien ha sido un gran apoyo moral, y que se ha preocupado por mí para que pudiera avanzar a cumplir con las actividades; como mi compañera de trabajo durante la maestría y como consejera cuando lo he necesitado.

De igual manera quiero poner un agradecimiento formal a la doctora Mercy Ilbay quien ha sido mi tutora en este proceso y ha sabido guiarme de la mejor manera. Espero siga teniendo mucho éxito en su vida, tanto de manera laboral como personal, porque ha demostrado ser una persona honorable, y de gran favor para ir subsanando dudas.

ÍNDICE

RESUMEN	vii
ABSTRACT	viii
CAPÍTULO I	9
PROBLEMA	9
1.1. Planteamiento del problema.....	9
1.2. Preguntas de investigación	12
1.3. Objetivos de investigación.....	12
1.3.1. <i>Objetivo General</i>	12
1.3.2. <i>Objetivos Específicos</i>	13
1.4. Justificación.....	13
CAPÍTULO II	17
FUNDAMENTACIÓN TEÓRICA.....	17
2.1. Antecedentes de investigación	17
2.2. Marco teórico	24
2.2.1. <i>Contaminación del Aire</i>	24
2.2.2. <i>Gases Contaminantes</i>	24
2.2.3. <i>Índice de la Calidad del Aire (ICA)</i>	27
2.2.4. <i>Clima</i>	30
2.2.5. <i>Modelos para predecir la calidad del aire</i>	31
2.3. Marco legal.....	33
CAPÍTULO III	36
METODOLOGÍA.....	36
3.1. Descripción del área de estudio/grupo de estudio	36
3.2. Enfoque y tipo de investigación.....	38
3.2.1. <i>Enfoque</i>	38

3.2.2. Tipo de Investigación.....	38
3.3. Definición y operacionalización de variables.....	39
3.3.1. Variables.....	39
3.4. Procedimiento	42
3.4.1. Comportamiento del índice de la calidad del aire antes, durante y después de la pandemia del COVID-19 en Quito – Ecuador mediante técnicas estadísticas	42
3.4.2. Contaminación del aire con respecto a la variabilidad climática .	499
3.4.3 Modelos estadísticos de la Calidad del Aire teniendo en cuenta las variables meteorológicas para la predicción de la contaminación.....	56
3.5. Consideraciones bioéticas	61
CAPÍTULO IV.....	63
RESULTADOS Y DISCUSIÓN.....	63
4.1 Índice de la calidad del aire antes, durante y después de la pandemia del COVID-19 en Quito.....	63
4.1.1 Índice de la calidad del aire antes de la pandemia	64
4.1.2 Índice de la calidad del aire durante la pandemia.....	67
4.1.3 Índice de la calidad del aire después la pandemia	70
4.1.4 Análisis de tendencia.....	73
4.2 Contaminación del aire con respecto a la variabilidad climática	75
4.2.1 Análisis de la base de datos	75
4.2.2 Gráficamente	75
4.2.3 Correlación de Pearson y Spearman entre calidad de Aire y Clima	80
4.2.4 Teoría Cópula.....	83
4.3 Modelos estadísticos.....	84
4.3.1. Modelo de Regresión Lineal Múltiple (RLM).....	85
4.3.2. Modelo ARIMAX.....	87
4.3.3. Validación.....	88

CONCLUSIONES Y RECOMENDACIONES	92
Conclusiones	92
Recomendaciones	95
REFERENCIAS	96
ANEXOS	1088

ÍNDICE DE TABLAS

Tabla 1 Índice Quiteño de la Calidad del Aire	29
Tabla 2 Estaciones Meteorológicas REMMAQ.....	37
Tabla 3 Operacionalización de Variables Estudio	40
Tabla 4 Expresiones matemáticas para cada rango de concentración	45
Tabla 5 Valores Descriptivos IQGCDA - Antes de la Pandemia.....	65
Tabla 6 Valores Descriptivos IQGCDA - Durante la Pandemia.....	688
Tabla 7 Valores Descriptivos IQGCDA – Después de la Pandemia.....	71
Tabla 8 Prueba de Mann - Kendall para la Tendencia	7474
Tabla 9 Prueba de Correlación de Pearson y Spearman	82
Tabla 10 Teoría Cópula – Relación entre Clima y IQGCDA.....	83
Tabla 11 Valores Regresión Lineal Múltiple	86
Tabla 12 Valores ARIMAX	87
Tabla 13 Validación de la Regresión Lineal Múltiple y ARIMAX - Primer y Segundo Momento Tendencia, Etapa de calibración	899
Tabla 14 Evaluación del modelo de Regresión Lineal Múltiple y ARIMAX - Primer y Segundo Momento Tendencia, Etapa validación	90
Tabla 15 Pruebas de Normalidad.....	11414
Tabla 16 Simetría Sin Transformación	11616
Tabla 17 Simetría después de la Transformación.....	116
Tabla 18 Revisión de Supuestos - Regresión Lineal Múltiple.....	12222
Tabla 19 Revisión de Supuestos - ARIMAX.....	122

ÍNDICE DE FIGURAS

Figura 1 Área de Estudio	37
Figura 2 Librerías R.....	44
Figura 3 IQGCDA - Todas las Estaciones - 01/01/2019 - 31/01/2023.....	64
Figura 4 IQGCDA - Todas las Estaciones – Antes de la Pandemia.....	67
Figura 5 IQGCDA - Todas las Estaciones – Durante la Pandemia	699
Figura 6 IQGCDA - Todas las Estaciones – Después de la Pandemia.....	72
Figura 7 Diagramas Dispersión, Elipse e Índices de Correlación - Belisario.....	76
Figura 8 Diagramas Dispersión, Elipse e Índices de Correlación - Carapungo	77
Figura 9 Diagramas Dispersión, Elipse e Índices de Correlación - Centro	788
Figura 10 Diagramas Dispersión, Elipse e Índices de Correlación - Cotocollao	79
Figura 11 Diagramas Dispersión, Elipse e Índices de Correlación - Guamaní..	80
Figura 12 Simulación de los modelos de RLM y ARIMAX en comparación con lo observado	91
Figura 13 Certificado Abstract	10908
Figura 14 Datos Perdidos dentro de la Base - Belisario.....	10909
Figura 15 Imputación con la media en Belisario.....	11010
Figura 16 Imputación mediante regresión lineal en Belisario.....	11111
Figura 17 Imputación mediante regresión estocástica en Belisario	11212
Figura 18 Imputación Múltiple en Belisario.....	11313
Figura 19 Belisario - Relación Mediante Funciones Copula.....	11717
Figura 20 Carapungo - Relación Mediante Funciones Copula	1188
Figura 21 Centro - Relación Mediante Funciones Copula	11919
Figura 22 Cotocollao - Relación Mediante Funciones Copula	120
Figura 23 Guamaní - Relación Mediante Funciones Copula.....	121

ÍNDICE DE ANEXOS

Anexo A. Certificado del abstract por parte de idiomas	108
Anexo B. Resultados Imputación De Datos - Belisario	109
Anexo C. Normalidad	114
Anexo D. Simetría	116
Anexo E. Teoría Copula	117
Anexo F. Supuestos Regresiones	122

RESUMEN

Quito está situada en un valle interandino rodeado de montañas, ello la convierte en una ciudad vulnerable a la contaminación atmosférica; su elevada altitud afecta la calidad del aire volviéndolo más fino disminuyendo la dispersión de contaminantes. La presente investigación evalúa la relación entre la calidad del aire y las variables meteorológicas, para la comprensión de la situación ambiental antes, durante y después del COVID-19 en la ciudad de Quito – Ecuador. Se utilizaron bases de datos de las estaciones de control meteorológico repartidas por la ciudad, que incluyen parámetros de la calidad del aire, como material particulado, ozono, dióxido de nitrógeno, monóxido de carbono y dióxido de azufre. Los datos recogidos desde enero del 2019 hasta enero de 2023 fueron procesados mediante el software estadístico R-Studio. Se estableció la correspondencia entre variables utilizando: la Correlación de Pearson, de Spearman y la Teoría Cópula. El estudio se realizó en tres momentos específico. El análisis de tendencia determinó que únicamente hay dos momentos definidos; antes y desde el inicio de la situación de pandemia en Ecuador. El descenso de la contaminación al inicio de las paralizaciones fue efímero debido al auge de empresas de plásticos, transporte y materiales del hogar, entre otros, que han empeorado la situación ambiental. De ello se destaca la influencia estadísticamente significativa que tiene la radiación solar “positiva” y la velocidad del viento “negativa” en la calidad del aire. La Regresión Lineal Múltiple fue el método más eficaz en la predicción de la calidad del aire y se validó mediante RMSE, MSE, MAE y R-cuadrado basándose en datos meteorológicos.

Palabras Clave: Índice de Calidad del Aire, Meteorología, Incidencia.

ABSTRACT

Quito is located in an inter-Andean valley surrounded by mountains, which makes it a city vulnerable to air pollution; Its high altitude affects the quality of the air, making it thinner, reducing the dispersion of pollutants. This research evaluates the relationship between air quality and meteorological variables, to understand the environmental situation before, during and after COVID-19 in the city of Quito - Ecuador. Databases from meteorological control stations distributed throughout the city were used, which include air quality parameters, particulate matter, ozone, nitrogen dioxide, carbon monoxide and sulfur dioxide. The data collected from January 2019 to January 2023 were processed using the R-Studio statistical software. Correspondence between variables was established using Pearson's correlation, Spearman's correlation, and Copula Theory. The study was carried out at three specific moments. The trend analysis determines that there are only two defined moments: before and since the beginning of the pandemic situation in Ecuador. The decrease in pollution at the beginning of the shutdowns was ephemeral due to the rise of plastics, transportation, and household materials companies, among others, which have worsened the environmental situation. This highlights the statistically significant influence that "positive" solar radiation and "negative" wind speed have on air quality. Multiple Linear Regression was the most effective method in predicting air quality and was validated using RMSE, MSE, MAE and R-Square based on meteorological data.

Keywords: Air Quality Index, Meteorology, incidence.

CAPÍTULO I

PROBLEMA

1.1. Planteamiento del problema

A lo largo de la historia, la humanidad, ha pasado de la era donde únicamente se preocupaba en cazar, cosechar, a encontrarse en la actualidad con avances en todos los campos, tanto como especie individual, y como sociedad (Veenhoven, 2010). Esto tiene particularidades positivas como negativas; por una parte, se han visto progresos en ámbitos muy variados como el económico, industrial, tecnológico y, por otro lado, se han descuidado rasgos importantes como la propia salud humana, o el conservar los espacios naturales preservando las especies que tan importantes son en la cadena de la vida. Principalmente el auge de las industrias, la expansión masiva de las ciudades y la globalización ha causado el deterioro de los ambientes habitados, por ende, en la calidad de la salud humana (Álvarez, 2011).

UNICEF Ecuador (2020) la contaminación del aire ha provocado un alarmante número de muertes de aproximadamente 7 millones anuales. Los niños, catalogados como uno de los grupos más vulnerables, han sufrido más de 600,000 muertes relacionadas directamente con la contaminación de su entorno. Esto refleja una urgente necesidad de estudiar la calidad del aire y adoptar medidas que salvaguarden la salud y el bienestar de la población en general, quienes se ven afectados directamente. El estudio del comportamiento de gases contaminantes, tanto primarios como secundarios, ayuda a comprender la propagación, el impacto que pueden causar, además de buscar estrategias de mitigación.

La situación al salirse de los límites controlados produce problemas de Estado, ya no solo en el ámbito ambiental, sino también se tiene también problemas de salud, pero con el apoyo adecuado de instituciones gubernamentales como de algunas no gubernamentales se han presentado planes que buscan frenar la expansión de la contaminación ambiental. De este modo, las mediciones automáticas de la concentración de los contaminantes proporcionan un registro instantáneo de la contaminación nociva para informar o alertar a los habitantes de la zona del peligro existente (García *et al.*, 2013).

En países latinoamericanos no se ha tenido una buena estrategia, ya sea por desconocimiento o por desinterés de los entes superiores encargados de la regularización de la contaminación, como nos lo menciona Almeida *et al.* (2020):

Se debe destacar que la calidad del aire en países de primer mundo ha incrementado en los últimos años gracias a los esfuerzos académicos y legislativos. No ocurre lo mismo en Latinoamérica, donde todavía se evidencia una baja calidad de aire, que se relaciona fuertemente a problemas de salud y a políticas implementadas por los estados (p.152).

En Ecuador, la normativa actual respaldada por la Constitución vigente ha impulsado medidas que han contribuido de manera significativa a reducir el impacto de la contaminación ambiental. Esto se enmarca en el esfuerzo por alcanzar un estado de bienestar que promueva el “Buen Vivir”, buscando así mejorar la calidad de vida de la población y, de manera concomitante, la calidad del aire que todos respiramos. Estudios como el de Mila y Yanez (2019) han demostrado los efectos positivos de estas acciones en la disminución de la contaminación ambiental. Se han establecido normativas con el objetivo de proteger y preservar el medio ambiente, asegurando su salud y equilibrio (Asamblea Constituyente, 2008).

Se ha trabajado en la promoción de un desarrollo sostenible, como lo destaca el Ministerio del Ambiente de Ecuador (Ministerio del Ambiente, 2010). Este enfoque ha llevado a la implementación de regulaciones específicas para mejorar la calidad del aire. A pesar de estos avances, aún existen desafíos pendientes en nuestro camino hacia una mejora significativa de la salud pública (Moreira, 2018). Para lograr un impacto verdaderamente relevante que se

traduzca en una mejora constante en la calidad del aire en los próximos años, se requiere un esfuerzo continuo y un enfoque en la investigación de las interacciones entre la calidad del aire y las variables meteorológicas, como lo ha estado haciendo Almeida *et al.*, (2020) en sus investigaciones además de otros investigadores.

La ciudad de Quito situada en la sierra de Ecuador se ha enfrentado a importantes problemas de calidad del aire, debido a su singular topografía y altitud, así como a actividades humanas como el transporte y la industria. Como resultado, hay una creciente necesidad de estrategias eficaces para mitigar y gestionar la contaminación del aire en la ciudad. Desde el estudio realizado por Hayas Barrú (1991), que mostró por primera vez la relación entre las variables meteorológicas con la contaminación atmosférica, se han llevado a cabo estudios como: (Di y Li, 2019; Liu *et al.*, 2022; Wang y Li, 2022; Zhu *et al.*, 2021), que han descubierto que algunos factores meteorológicos como la temperatura, la humedad, la velocidad del viento son predictores significativos de la calidad del aire en un sector, sin embargo, aún queda mucho por aprender y comprender sobre cómo interactúan estas variables entre sí, además de su impacto sobre la concentración de contaminantes en el aire (Almeida *et al.*, 2020).

A pesar de existir varias estaciones de control en la ciudad, éstas aún no garantizan una correcta toma de datos, ya que suelen hallarse períodos donde las tomas que se realizan llevan valores atípicos o de igual manera valores negativos que no tienen ninguna explicación físico-química. Esto dificulta el desarrollo de intervenciones eficaces en respuesta a los cambios en la calidad del aire y limita la capacidad de los investigadores y responsables políticos para estudiar la relación entre la meteorología y la calidad del aire en tiempo real. Por tanto, es importante la identificación de las variables meteorológicas que inciden en la alteración de la calidad de aire en las diferentes etapas planteadas de pandemia en el Distrito Metropolitano de Quito, examinando su relación estadística y desarrollando modelos para predecir los cambios en la calidad del aire basándose en datos meteorológicos.

Se requiere la recopilación, imputación, análisis de numerosos datos meteorológicos y de calidad del aire procedentes de estaciones de vigilancia y

otras fuentes. Es necesario utilizar métodos estadísticos básicos, como la estadística descriptiva, obtención de algunos coeficientes de la correlación que guardan, además de métodos más avanzados como uso de técnicas de imputación multivariante, modelos de regresión lineal múltiple, análisis de series temporales.

El estudio identifica a las variables meteorológicas que influyen en la alteración de la calidad de aire de la ciudad de Quito, con la finalidad de contribuir al desarrollo de estrategias eficaces para reducir la contaminación atmosférica en Quito. Se desarrollan modelos estadísticos avanzados para predecir los cambios en la calidad del aire a partir de los datos meteorológicos, lo cual permitirá contar con investigaciones que contribuyan a la construcción de políticas públicas encaminadas a garantizar un ambiente sano y una salud adecuada.

1.2. Preguntas de investigación

El estudio plantea las siguientes preguntas de investigación:

¿Existe relación entre las variables meteorológicas y la calidad del aire antes, durante y después de la pandemia COVID-19 en Quito Ecuador?

¿Durante qué periodo en torno a la pandemia producida por el COVID-19, la calidad de aire de la ciudad de Quito mostró variaciones considerables?

¿Qué variables meteorológicas inciden en la calidad de aire dentro de Quito?

¿Qué modelos estadísticos son más efectivos para predecir la contaminación atmosférica en Quito, considerando las variables meteorológicas?

1.3. Objetivos de investigación

1.3.1. Objetivo General

Evaluar la relación entre la calidad de aire y las variables meteorológicas, para la comprensión de la situación ambiental antes, durante y después del COVID-19 en la ciudad de Quito – Ecuador.

1.3.2. Objetivos Específicos

- Examinar el comportamiento del índice de la calidad del aire antes, durante y después de la pandemia del COVID-19 en Quito – Ecuador mediante técnicas estadísticas.
- Comparar los valores del Índice de Calidad de Aire a las variables meteorológicas mediante pruebas de correlación, y otras técnicas estadísticas, para la confirmación de la relación entre las variables del estudio.
- Comparar modelos estadísticos de los gases contaminantes del aire teniendo en cuenta las variables meteorológicas como control de la contaminación atmosférica.

1.4. Justificación

La contaminación ha venido incrementándose por parte del accionar de los humanos, la densidad poblacional que existe en las metrópolis dentro de países como China, Japón, Estados Unidos, India y Europa ponen en claro ejemplo el cómo a pesar de existir buen desarrollo como sociedad, se tienen los niveles más altos de contaminación (Marlier *et al.*, 2016). Es más, la situación presente que se vive ha puesto en evidencia la íntima relación que guarda tanto el estado de salud y el medio ambiente en el que se desarrolla la vida (UNICEF Ecuador, 2020).

Ecuador se enfrenta a un importante problema de contaminación desde hace varias décadas. La contaminación en Ecuador se debe principalmente a la explotación de los recursos naturales, que ha provocado la contaminación de los cursos de agua, el suelo y en especial el aire. La contaminación en Ecuador ha dejado una huella significativa en la salud y el bienestar de la población del país (Moreira, 2018). Muchas personas que viven en zonas contaminadas sufren problemas respiratorios, afecciones cutáneas y otros problemas de salud. El gobierno de Ecuador ha tomado medidas para hacer frente a la contaminación en el país, como el establecimiento de normativas para controlar la polución y la creación de zonas protegidas para preservar el medio ambiente (Ministerio del

Ambiente, 2010). Sin embargo, aún queda mucho por hacer para garantizar la protección de la salud y el bienestar de las comunidades locales.

La ciudad de Quito situada en la sierra andina de Ecuador, es conocida por su rico patrimonio cultural y sus impresionantes paisajes naturales, sin embargo, como muchas ciudades de América Latina, Quito también se enfrenta a importantes problemas de calidad del aire, debido en parte a su elevada altitud y topografía como a la industrialización que tiene de manera vigente (Emperatriz y Murillo, 2018). El gobierno de la ciudad ha puesto en marcha una serie de intervenciones destinadas a reducir la contaminación atmosférica y mejorar la salud pública. Se ha llegado a determinar que los factores meteorológicos desempeñan un papel crucial en la calidad del aire. Algunos factores meteorológicos, como la velocidad y dirección del viento, precipitaciones y estabilidad atmosférica, pueden influir en la calidad del aire(Almeida *et al.*, 2020).

La incidencia de la meteorología en la calidad del aire de Quito (Ecuador) se ha vuelto una necesidad que requiere un estudio y un análisis minucioso. De esto surge el interés de realizar la investigación sobre la relación que plantea tener gases contaminantes y las variables meteorológicas antes, durante y después de la pandemia de COVID-19. Al comprender la compleja relación entre los factores meteorológicos y la calidad del aire, los investigadores y los responsables políticos pueden desarrollar intervenciones específicas para proteger la salud pública y el medio ambiente.

La investigación realiza un análisis del comportamiento del índice de calidad de aire en el período comprendido entre 2019 a 2022, haciendo mayor énfasis en 3 etapas, que se las puede indicar como el antes, durante y después de la pandemia producida por el COVID-19 en el centro de la ciudad de Quito Ecuador. Esto ayuda a comprender cómo es que la contaminación se empezó a distribuir una vez existieron paralizaciones dentro del país, como de sus efectos colaterales.

En la investigación se realizará la comparación entre el índice de calidad de aire (ICA) con respecto a cómo el clima ha ido variando, teniendo en cuenta la variabilidad climática de cómo la precipitación, la velocidad del viento, la dirección del viento, la humedad relativa, la temperatura, presión atmosférica,

radiación solar cambian alrededor del tiempo. Lo que brinda un sustento para saber cuáles están más íntimamente relacionadas con la calidad del aire.

Es necesario investigar esta relación, para determinar los factores que contribuyeron a los cambios observados en los niveles de contaminación atmosférica. Además, a medida que se reanudan las actividades económicas después de la pandemia, el impacto potencial de los factores meteorológicos en la calidad del aire se convierte en una preocupación, por lo que se vuelve necesario comprender la relación entre la meteorología y la calidad del aire en Quito antes, durante y después de la pandemia.

Se puede desarrollar modelos de predicción, como aquellos mencionados en OMM (2018), sin embargo, debe tenerse en cuenta que para obtener validez en estas predicciones se sugiere tener al menos un aserio de 10 años para confirmar la variabilidad climática en el lugar de estudio. Para periodos cortos de estudio, lo que se obtiene es una comparación de variables para analizar su causa y efecto, mismos que se enfocan en comprender qué medidas tomar una vez suceda nuevamente una paralización como la acontecida durante la pandemia de COVID-19. Dentro de los modelos utilizados existen los modelos de regresión lineal múltiple, modelos basados en series de tiempo y modelos econométricos espaciales. Cada uno de ellos se pueden utilizar para predecir variaciones diarias, semanales o estacionales de la calidad del aire, y pueden actualizarse en tiempo real a medida que se dispone de nuevos datos meteorológicos.

Este trabajo es importante y lleva consigo grandes beneficios para que en un futuro se pueda tener una idea de cómo puede variar la contaminación ambiental, en especial durante una pandemia global. La investigación establece medidas para mejorar la salud de las personas para que las muertes por esta causa se reduzcan, ya que dentro de un país en vía de desarrollo como lo es Ecuador, las personas perjudicadas son quienes pertenecen a sectores de mayor pobreza, determinándose por ende como beneficiarios directos de la investigación.

Los aportes al Plan Nacional de Desarrollo 2021-2025 y Línea de Investigación del Programa de Postgrado: “Aplicación de la Estadística en la solución de problemas del entorno”; a la que se adscribe la investigación, son para mejorar

la calidad de vida de las personas dentro de Ecuador. Esto enfocándose en comprender la relación entre la variabilidad meteorológica con el índice de calidad de aire, mismo que toma en cuenta los principales contaminantes del aire que existen dentro de Quito Ecuador. Se hace énfasis en el objetivo 12 que busca fomentar modelos de desarrollo sostenible para mitigar el cambio climático con la implementación de distintas medidas, esto nos ayuda a proteger cada una de las familias quienes deben tener una salud de calidad, para lo cual es importante que el medio ambiente se encuentre en sus niveles estables (Consejo Nacional de Planificación, 2021).

CAPÍTULO II

FUNDAMENTACIÓN TEÓRICA

2.1. Antecedentes de investigación

La contaminación afecta desde hace mucho tiempo al medio ambiente y a la salud humana. En el siglo XIX se empezó a comprender que las condiciones de vida insalubres y la contaminación del agua contribuyen a las epidemias (Walker *et al.*, 2015). La revolución industrial aceleró tanto la magnitud de las emisiones de contaminantes primarios como la extensión geográfica de los factores contribuyentes. La contaminación de la Tierra ofrece una historia social y política de la contaminación industrial, trazando sus trayectorias a lo largo de los siglos. A mediados del siglo XX, la contaminación atmosférica se había convertido en una dura consecuencia del crecimiento industrial en todo el país y el mundo (Fowler *et al.*, 2020). Las grandes catástrofes medioambientales también han influido en muchos países para establecer nuevos límites a los contaminantes industriales (Manisalidis *et al.*, 2020).

Todo ello ha conllevado al impulso de estudios a nivel internacional para en primer lugar mostrar cuales son los contaminantes. Según Guo y Feng, (2019) se determina que los principales contaminantes en Xi'an China fueron las $PM_{2.5}$, PM_{10} , sin embargo, la cantidad de contaminante ha ido incrementándose, siendo notable ahora las concentraciones de NO_2 y O_3 ; todo esto debido a posibles incrementos en vehículos a más de industrias que producen estos desechos tóxicos al ambiente. Tomando como eje fundamental el hecho que las variables indiquen directamente en la persona, inclusive el clima puede ser un gran indicador para tratar de contener la expansión de la contaminación ambiental. Con la meteorología se puede comprender en cierto punto, que fenómenos producen que la calidad del aire regional esté en buenas o malas condiciones, a más de promover la optimización en profundidad de los procedimientos de las

empresas industriales para de esta manera tratar de reducir el conflicto entre la protección del medio ambiente y el desarrollo económico, aspectos importantes en la vida del humano (Liu *et al.*, 2022).

La determinación de los contaminantes no es lo único relevante a estudiar sino el buscar soluciones o de preguntarse si algún factor extra puede suponer un cambio en la contaminación. Se logró comprobar que un buen indicador para conocer la distribución que seguirán los gases contaminantes es el estudio de las variables meteorológicas. Por ejemplo, en China en medio de un estudio realizado en 221 ciudades se encontró que “las diferencias espaciales y temporales en los valores umbrales de los distintos factores meteorológicos afectaron a las concentraciones de $PM_{2.5}$, PM_{10} , SO_2 , CO , NO_2 y O_3 ” (Ji *et al.*, 2020, p.2).

También en un estudio realizado en Xi'an en China muestra la relación entre el índice de calidad de aire y la precipitación, que es una de las variables típicas que se miden en meteorología. Indican que la relación entre las variables es importante siendo que influye no sólo en las concentraciones de los contaminantes, sino también en la correlación entre los contaminantes. Por tal motivo los gases expedidos quienes contaban principalmente con $PM_{2.5}$ y PM_{10} , eran quienes tenían una fuerte relación entre ellos y la relación con el clima muestra que las concentraciones presentaban las menores cantidades cuando la cantidad de precipitación era alta, por lo que los contaminantes son fáciles de sedimentar y la calidad del aire mejora. De esta manera el autor plantea que, en las épocas de mayor contaminación atmosférica, la calidad del aire puede mejorarse aumentando la cantidad de precipitaciones (Di y Li, 2019).

Respecto al impacto que tuvo en la calidad del aire la pandemia producida por el COVID-19 existen algunos indicativos de su relación. Se encuentra en varias investigaciones, que las paralizaciones a nivel de ciudades de Europa y Estados Unidos, varios contaminantes mostraron reducciones marcadas y sostenidas. El más notorio fue el NO_2 mismo que en el 93% de las investigaciones se redujo, seguido del CO que pasó el 88% de las investigaciones. Todos los estudios sobre NO_x y benceno informaron de reducciones significativas, aunque en números estas investigaciones no eran tan numerosas para afirmar la reducción total de

los contaminantes debido a las paralizaciones por el COVID-19 (Bakola *et al.*, 2022). Del mismo modo en lugares como Wuhan, Italia, España, USA, y Brasil, se observó por medio de investigación que la contaminación se redujo en un 30% (Urrutia *et al.*, 2020).

La contaminación del aire es un problema importante en América Latina y el Caribe, donde sus impactos no sólo contribuyen al cambio climático, sino que también crean importantes riesgos para la salud de quienes viven en la región. Según algunos organismos especializados como la Organización Mundial de la Salud, la contaminación del aire es el mayor riesgo ambiental para la salud en América Latina (Riojas *et al.*, 2016). La población expuesta a niveles de contaminantes atmosféricos que superan las directrices de la OMS representa un gran porcentaje del total de la población urbana de América Latina. La contaminación del aire es una de las tres principales causas de muerte en la región, y el acceso limitado a aire limpio aumenta el riesgo de enfermedades cardiovasculares, enfermedad pulmonar obstructiva crónica, infecciones agudas de las vías respiratorias inferiores y cáncer de pulmón (Husaini *et al.*, 2022).

Los países que tienen más contaminado el aire en América Latina son Chile, Perú, Argentina y México. En Chile, se registra una concentración anualmente media de $PM_{2.5}$ de $20,10 \mu g/m^3$, siendo el país con mayor contaminación de la región, aunque los siguientes países como Perú o México no se encuentran tan alejados en contaminación, teniendo una media anual de concentración de $14,60 \mu g/m^3$ y $14,50 \mu g/m^3$ de $PM_{2.5}$ respectivamente (Gouveia *et al.*, 2021). Las fuentes de contaminación que asolan el aire en latinoamericana son principalmente el aumento y el uso indiscriminado de productos de cuidado personal, productos alimenticios, fertilizantes, pesticidas, productos sanitarios, quema de derivados de fósiles como el combustible, entre otros. La mayoría de estos productos contienen trazas de sustancias químicas muy extendidas, normalmente conocidas como contaminantes emergentes que han provocado y están provocando una gran contaminación ambiental (Souza *et al.*, 2022).

La contaminación atmosférica aumenta el riesgo de enfermedades cardiovasculares, enfermedad pulmonar obstructiva crónica, infecciones agudas de las vías respiratorias inferiores y cáncer de pulmón. Las comunidades rurales

e indígenas son más vulnerables a estos riesgos para la salud debido a su limitado acceso a aire limpio (Manisalidis *et al.*, 2020). América Latina y el Caribe ha hecho algunos progresos en el desarrollo de políticas, marcos de control de la contaminación atmosférica del aire, la adopción de normas de calidad del aire y reducir las emisiones de distintas fuentes clave. Sin embargo, todavía hay países, ciudades en expansión constante con un crecimiento urbano que aún no dispone de un marco para prevenir, reducir la contaminación atmosférica o la capacidad de hacer cumplir planes y normativas existentes (UNEP, 2020).

Se han realizado algunos estudios en estos países que buscan revertir esta situación y que se han enfocado en tomar como referencia la contaminación del aire, para analizarlo con relación a las variables meteorológicas. Dentro de las investigaciones más importantes se debe indicar que en Chile se realizó la caracterización de la contaminación por $PM_{2.5}$ en Osorno, una ciudad fuertemente afectada por la contaminación. Tras realizarse un análisis de datos las mencionadas $PM_{2.5}$, que se generan a partir de la combustión de biomas, muestra altos niveles de concentración, alcanzando un nivel de emergencia de $1124 \mu g/m^3$, debido mayormente a la variabilidad climática provocada por las características topográficas de la ciudad de Osorno (Nakamura *et al.*, 2022).

A su vez en el país de Perú se realizó un estudio de evaluación de la calidad del aire y previsión de la contaminación mediante redes neuronales artificiales en Lima Metropolitana. Este estudio aborda el problema de la predicción de la concentración de PM_{10} a escala horaria basándose en los indicadores de calidad del aire de cinco estaciones de monitoreo en Lima, Perú en conjunto con datos de variabilidad climática del sector. Se obtuvieron resultados importantes para indicar que la predicción se la puede realizar tras tomar las variables meteorológicas adecuadas con un gran nivel de precisión (Cordova *et al.*, 2021).

Además, existen investigaciones que muestran cómo fue la afectación de la pandémica COVID-19 con relación a la contaminación ambiental. En Perú mismo se realizó un análisis de la Calidad del Aire en Lima, Perú, utilizando los niveles de NO_2 durante el Bloqueo Pandémico COVID-19. Los resultados muestran que el NO_2 en comparación con 2019 varía de entre el 60% al 40% en las dos primeras semanas y de entre el 50% al 25% en las dos semanas siguientes del

cierre de COVID-19. Sin embargo, este efecto no pudo observarse dos meses después del inicio del cierre (Velayarce *et al.*, 2022).

Situándonos en un contexto más enfocado en el ámbito nacional en los últimos años, la contaminación se ha vuelto un problema importante en Ecuador, sobre todo en la región amazónica, donde la explotación y producción de petróleo han provocado la contaminación de las fuentes de agua (García *et al.*, 2021). Se han producido varios derramamientos de petróleo que han contaminado ríos y afectado a comunidades indígenas (Coronel *et al.*, 2020). Un estudio que analizaba la calidad del agua de doce ríos de Ecuador reveló altos niveles de contaminación por *E. coli* (Vinueza *et al.*, 2021).

La contaminación ha provocado desastres medioambientales, como la contaminación del suelo, las aguas subterráneas, las corrientes superficiales y la contaminación del aire. A pesar de los esfuerzos realizados tanto por el Estado a través de la Constitución y a través de algunas normativas para tratar la contaminación de aire, se puede ver que la polución atmosférica sigue siendo aún un problema importante de tratar, produciendo efectos adversos para la salud humana y el medio ambiente (UNICEF Ecuador, 2020). Aunque se han aplicado varias medidas, como el uso de medios de transporte menos contaminantes y la restricción de las actividades industriales, aún no está clara la eficacia de estas medidas para reducir los niveles de contaminación (Rodríguez y Cuvi, 2019).

Se necesitan intervenciones específicas basadas en predicciones precisas y fiables de los cambios en la calidad del aire, lo que requiere el uso de modelos estadísticos avanzados. De esta manera se han venido llevando a cabo publicaciones científicas que traten este tema, siendo las más relevantes a nivel ecuatoriano aquellas llevadas a cabo en las principales ciudades del país. Por ejemplo, en Guayaquil una investigación muestra que los niveles de contaminación, en especial de material particulado ya sea grueso o fino se encuentra por debajo de los niveles admitidos, es decir se encuentra en una situación donde de no revertirse o aplicarse medidas lo más pronto posible, se puede llegar a tener graves problemas ambientales (Moran *et al.*, 2021).

En la ciudad de Cuenca, se han llevado a cabo publicaciones, tesis y trabajos relativos a la incidencia directa que tiene el clima sobre la calidad del aire. Se utilizan los modelos eulerianos Weather Research y Predicción con Química (WRF-ChemV3.2) para simular las variables meteorológicas y la calidad del aire (Parra, 2022). Además, se han realizado estudios sobre la calidad del aire durante el 2020 teniendo como resultado que ha habido un mejoramiento de esta, tanto por los primeros meses de las paralizaciones por COVID-19 como por la temprana operatividad del tranvía eléctrico y al futuro cambio de autobuses diésel a eléctricos en Cuenca (Parra y Espinoza, 2020).

Es decir, la pandemia COVID-19 dentro de Ecuador, a pesar de las preocupaciones sociales y económicas, se logró ver también consecuencias positivas inesperadas para la calidad del aire, ya que se redujeron las actividades humanas y las emisiones de la quema de combustible. Las paralizaciones se implementaron con el conjunto de las cuarentenas muy estrictas en el momento en que se produjo el crecimiento de los casos positivos respecto a la enfermedad, el 17 de marzo de 2020. Este gran descenso de las concentraciones de contaminación atmosférica se produjo en puntos estratégicos por ejemplo en las mediciones tomadas en Quito, como en Cuenca siendo fuente valiosa de la prueba del impacto antropogénico en la calidad del aire urbano (Zalakeviciute *et al.*, 2020).

La ciudad de Quito que está situada en la sierra centro de Ecuador, rodeada por la cordillera de los Andes, a más de 2 850 metros sobre el nivel del mar. Es una de las ciudades más pobladas del país, con más de 2 011 000 habitantes; es conocida por su arquitectura histórica y su importancia cultural. Sin embargo, como muchas zonas urbanas de todo el mundo, Quito se enfrenta a importantes problemas de contaminación, en especial del aire. Los problemas de calidad del aire de Quito se deben en gran medida a su geografía. La ciudad está situada en un valle rodeado de montañas, que pueden atrapar los contaminantes atmosféricos e impedir que se disipen. Además, la elevada altitud de Quito puede agravar los problemas de calidad del aire, ya que la menor presión atmosférica puede dificultar la dispersión de los contaminantes (Emperatriz y Murillo, 2018).

Como resultado de estos factores, previamente la calidad del aire en Quito ha sido una preocupación creciente, y la Organización Mundial de la Salud (OMS) informó de que la concentración media anual de partículas (PM) de la ciudad estaba por encima del valor guía recomendado en 2018. La exposición a altos niveles de contaminación del aire puede provocar una serie de problemas de salud, incluidas enfermedades respiratorias y cardiovasculares, así como impactos adversos en el medio ambiente (UNICEF Ecuador, 2020).

Otros factores que contribuyen a la mala calidad del aire en Quito son las emisiones del transporte, la actividad industrial y la quema de residuos al aire libre. Según una investigación realizada en la ciudad, existe una fuerte emisión de contaminantes por parte del tráfico vehicular, esto hablando de la zona urbana, pero, en sectores específicos donde se encuentra más industrializada la ciudad, la emisión de contaminantes se ve potenciada por el aporte que hacen las grandes industrias (Cornejo *et al.*, 2022).

Dada la importancia de la meteorología en la calidad del aire de Quito, es esencial estudiar estos factores en detalle. Los investigadores han utilizado diversos métodos para analizar los datos meteorológicos y su relación con la calidad del aire en la ciudad. Un modelo utilizado en la investigación de Almeida *et al.* (2020) es el análisis de regresión lineal mediante un software llamado AirQ2, que da muy buenos resultados a la hora de realizar predicciones. También han llevado a cabo modelos de Machine Learning para entender y analizar los comportamientos que van a tener la calidad del aire con relación a variables meteorológicas.

La pandemia COVID-19 y las medidas de cierre asociadas tuvieron un impacto significativo en la calidad del aire de Quito. Al ralentizarse las actividades económicas y disminuir el tráfico, la calidad del aire en la ciudad mejoró notablemente. Dentro de la zona centro de la ciudad durante la imposición de la parada de movilidad se obtienen los niveles más bajos de contaminación. Se lograron reducciones significativas de las concentraciones de NO_2 (-68%), SO_2 (-48%), CO (-8%) y $PM_{2.5}$ (-29%) en la capital, Quito, durante el primer mes de cuarentena (Zalakeviciute *et al.*, 2020). Los datos de calidad del aire muestran

que se puede mejorar la calidad del aire cuando se reducen las emisiones contaminantes producidas por la actividad humana.

2.2. Marco teórico

2.2.1. Contaminación atmosférica

La contaminación atmosférica se produce por la liberación de gases y sustancias químicas nocivas en el aire (Manisalidis *et al.*, 2020). Los seis contaminantes regulados por la EPA son el monóxido de carbono, plomo, los óxidos de nitrógeno, el ozono troposférico, la contaminación por partículas y los óxidos de azufre (EPA, 2023). El uso de gasolina contribuye a la contaminación atmosférica porque los vapores que desprende al evaporarse y las sustancias que produce al quemarse (monóxido de carbono, óxidos de nitrógeno, partículas e hidrocarburos no quemados) contribuyen a la contaminación atmosférica (Huang *et al.*, 2020).

La industria del petróleo y el gas natural también contribuye a la contaminación atmosférica. La normativa de la EPA para esta industria ayuda a combatir el cambio climático y a reducir la contaminación atmosférica que perjudica la salud pública (EPA, 2023). La contaminación atmosférica tiene muchas repercusiones negativas en la salud humana. Puede causar problemas respiratorios como asma, cáncer de pulmón, cardiopatías, infartos y otras enfermedades. La contaminación atmosférica también perjudica al medio ambiente al contribuir al cambio climático (Walker *et al.*, 2015).

2.2.2. Gases Contaminantes.

La calidad del aire se caracteriza por diferencias espaciales y temporales relacionadas con condiciones meteorológicas locales. Las emisiones contaminantes de alta intensidad y las condiciones meteorológicas desfavorables son los factores clave para el desarrollo de procesos de contaminación atmosférica, a continuación, se detalla los mismos:

PM significa material particulado (también llamado contaminación por partículas), se usa para definir una mezcla de partículas sólidas y gotas líquidas que se encuentran en el aire. Algunas partículas, como el polvo, la suciedad, el

hollín, o el humo, son lo suficientemente grandes y oscuras como para verlas a simple vista. Otras son tan pequeñas que solo pueden detectarse mediante el uso de un microscopio electrónico (Thangavel *et al.*, 2022).

Son partículas inhalables finas que tienen diámetros de, por lo general, 2,5 micrómetros y menores. En base a una serie de estudios epidemiológicos más recientes respaldados por pruebas toxicológicas, la revisión de la OMS (2006) concluyó que las partículas finas ($PM_{2.5}$) son más peligrosas en términos de mortalidad cardiovascular y respiratoria. Por lo tanto, se recomendó utilizar también las $PM_{2.5}$ como indicador de los efectos sobre la salud inducidos por la contaminación por partículas. Por otro lado, las partículas gruesas pueden tener impactos más visibles en la morbilidad respiratoria (Thangavel *et al.*, 2022).

PM_{10} son partículas inhalables que tienen diámetros de, por lo general, 10 micrómetros y menores. Los aerosoles también afectan al equilibrio radiactivo y, por tanto, contribuyen al cambio climático. Estas partículas pueden proceder de diversas fuentes, como el polvo, los gases de escape de los vehículos y las emisiones industriales. Las partículas PM_{10} pueden inhalarse y causar problemas respiratorios, sobre todo a las personas con enfermedades preexistentes, como asma o cardiopatías (Xing y Sun, 2022). Las PM_{10} son una de las diversas métricas utilizadas para medir la calidad del aire y están reguladas por muchos gobiernos y organizaciones de todo el mundo.

El ozono (O_3) es un gas compuesto por tres átomos de oxígeno. El ozono se encuentra tanto en la atmósfera superior de la Tierra como a nivel del suelo. El ozono puede ser bueno o malo, dependiendo de dónde se encuentre. El ozono bueno, llamado ozono estratosférico, se encuentra de forma natural en la atmósfera superior, donde forma una capa protectora que nos protege de los dañinos rayos ultravioleta del sol. Este ozono beneficioso ha sido parcialmente destruido por productos químicos de origen humano, causando lo que a veces se llama un "agujero en el ozono". El ozono a nivel del suelo es un contaminante atmosférico perjudicial, por sus efectos sobre las personas y el medio ambiente, y es el principal ingrediente del "smog" (US EPA, 2022a).

El Monóxido de Carbono (CO), es un gas incoloro e inodoro que puede resultar de la combustión incompleta de cualquier material carbonoso. Las principales

fuentes de monóxido de carbono en interiores son el humo del tabaco, las estufas de gas y las luces piloto, las estufas de leña y las chimeneas, los calentadores de gas y queroseno, los motores de gasolina, las linternas y estufas de camping y los garajes adosados y los respiraderos de entrada a nivel de la calle (Dey y Dhal, 2019).

El monóxido de carbono es peligroso porque se une a la hemoglobina en el torrente sanguíneo, reduciendo la cantidad de oxígeno que puede transportarse a órganos vitales como el cerebro y el corazón. La exposición a niveles elevados de monóxido de carbono puede provocar dolores de cabeza, mareos, náuseas e incluso la muerte (US EPA, 2022b).

El dióxido de azufre es el componente más preocupante y se utiliza como indicador del grupo más amplio de óxidos de azufre gaseosos (SO_x) (US EPA, 2023a). Las normas nacionales de calidad del aire ambiente de la Environmental Protection Agency (EPA) para el SO_2 están diseñadas para prevenir la exposición a todo el grupo de óxidos de azufre (SO_x). Otros SO_x gaseosos (como el SO_3) se encuentran en la atmósfera en concentraciones mucho menores que el SO_2 .

En general, cabe esperar que las medidas de control para el SO_2 reduzcan la exposición de las personas a todos los SO_x gaseosos. Esto puede tener el importante beneficio de reducir la formación de partículas contaminantes de azufre, como las partículas finas de sulfato.

La mayor fuente de SO_2 en la atmósfera es la quema de combustibles fósiles por parte de las centrales eléctricas y otras instalaciones industriales. Otras fuentes menores de emisiones son: los procesos industriales, como la extracción de metales del mineral; las fuentes naturales, como los volcanes; y las locomotoras, los barcos y otros vehículos y equipos pesados que queman combustible con un alto contenido de azufre (US EPA, 2023a).

Velayarce et al. (2022) explica que el dióxido de nitrógeno es el óxido de nitrógeno más importante y omnipresente en el gas ambiente. En altas concentraciones, el NO_2 es un gas de color marrón anaranjado. La inhalación de NO_2 produce una sensación acre y ofensiva similar a la que se experimenta cuando se inhala amoníaco o cloro. En concentraciones más bajas, el NO_2

parece amarillo o invisible y tiene un olor similar al de la lejía, pero es menos penetrante y huele algo más dulce que el cloro. Además, es menos soluble en agua que los gases que son inmediatamente irritantes, el cloro y el amoníaco. Por ello puede provocar una lesión retardada sin provocar síntomas de alarma inmediatos. El NO_2 es uno de los seis "criterios" de contaminantes del aire (US EPA, 2023b).

2.2.3. Índice de la Calidad del Aire (ICA)

El Índice de Calidad del Aire es una medida utilizada para informar de la calidad del aire en una zona específica. Es una herramienta que ayuda a la gente a comprender lo contaminado que está el aire y cómo puede afectar a su salud. El ICA se basa en mediciones de varios contaminantes atmosféricos, como las partículas, el ozono, el monóxido de carbono, el dióxido de azufre y el dióxido de nitrógeno (Suman, 2019).

La escala del ICA va de 0 a 500, donde 0 representa una calidad del aire excelente y 500 representa la calidad del aire más peligrosa. Cuanto más alto es el valor ICA, mayor es el nivel de contaminación atmosférica y mayor el riesgo para la salud pública (Suman, 2019). Los distintos países pueden tener escalas ICA ligeramente diferentes, pero el principio subyacente de medición de la calidad del aire es el mismo, difiriendo en muchos casos por la ubicación de cada lugar por lo que a distinta presión atmosférica que se tiene, se debe modificar el índice para que nos muestre información idéntica sobre la contaminación del aire.

El índice de la calidad de aire ha sido planteado exclusivamente tomando las características de la ciudad, en este caso el Municipio del Distrito Metropolitano de Quito (MDMQ), quienes por medio de la Red Metropolitana de Monitoreo Atmosférico (REMMAQ) de la Secretaría de Ambiente, plantearon un índice de calidad del aire que sirve para comunicar la calidad que este presenta diariamente (Tabla 1).

Utiliza categorías que proporcionan afirmaciones para cada categoría que le informan sobre la calidad del aire en su zona, se basa en las concentraciones de los gases contaminantes, a más de indicar lo peligros asociados con la salud que

pueden llevar a provocar. El índice se le toma entre un rango de valores, siendo el valor más alto el que indica la mayor contaminación del aire, por lo cual es necesario y siempre se busca tener valores pequeños (Secretaría de Ambiente de Quito, 2011).

Tabla 1*Índice Quiteño de la Calidad del Aire*

Rango	Categoría	CO	O₃	NO₂	SO₂	PM_{2.5}	PM₁₀
0-50	Nivel deseable u óptimo	0-5000	0-50	0-100	0-62.5	0-25	0-50
51-100	Nivel aceptable o bueno	5001-10000	51-100	101-200	63.5-125	26-50	51-100
101-200	Nivel de precaución	10001-15000	101-200	201-1000	126-200	51-150	101-250
201-300	Nivel de alerta	15001-30000	201-400	1001-2000	201-1000	151-250	251-400
301-400	Nivel de alarma	30001-40000	401-600	2001-3000	1001-1800	251-350	401-500
400-500	Nivel de emergencia	>40000	>600	>3000	>1800	>350	>500

Nota. Categorías del índice con sus respectivos valores límites (Secretaría de Ambiente de Quito, 2011)

2.2.4. Clima

La meteorología es la ciencia que se ocupa de la atmósfera y sus fenómenos. Al hablar de la tecnología de los instrumentos, los sensores y los sistemas de observación, se ha de seguir las categorías de la Organización Meteorológica Mundial, establecidas en el Manual de Instrumentos y Métodos de Observación (OMM, 2014).

Las condiciones meteorológicas tienen un impacto significativo en la calidad del aire al influir en la dispersión y el transporte de contaminantes en la atmósfera. Algunos de los factores meteorológicos críticos que afectan a la calidad del aire son la temperatura, la humedad, la velocidad del viento, la precipitación (Zhu *et al.*, 2021). Aunque aún existen variables como la radiación solar, dirección del viento, humedad que pueden estar causando diferencias en los niveles de contaminación.

La temperatura es una medida de la energía cinética de las moléculas, en cuanto a meteorología, se la define como el contenido de calor de la atmósfera o de la superficie terrestre. La temperatura puede tener un impacto significativo en la calidad del aire. Cuando la temperatura es alta, el aire puede retener más humedad, lo que puede aumentar la formación de ozono troposférico, un contaminante nocivo que puede causar problemas respiratorios. Además, las altas temperaturas también pueden exacerbar los efectos de otros contaminantes, como las partículas (PM), que pueden provocar problemas respiratorios y cardiovasculares. Las altas temperaturas pueden hacer que las PM se concentren más, facilitando su inhalación y aumentando su impacto en el organismo (Kalisa *et al.*, 2018).

La velocidad del viento es el aire en movimiento, un vector que tiene tanta magnitud (rapidez de movimiento) y dirección (30°, 60°, 80°). La velocidad del viento es un factor crítico en la dispersión y el transporte de contaminantes atmosféricos. Una velocidad del viento elevada puede contribuir a dispersar los contaminantes y reducir su concentración en la atmósfera, lo que mejora la calidad del aire. Por el contrario, una baja velocidad del viento puede hacer que

los contaminantes se acumulen en las zonas urbanas, provocando una mala calidad del aire (Zhu *et al.*, 2021).

La humedad es el contenido de humedad de la atmósfera o lo cerca que está de la saturación. La humedad afecta a la calidad del aire al influir en la formación y dispersión de partículas (PM) en la atmósfera. Los altos niveles de humedad pueden aumentar el tamaño de las PM, haciendo que sea menos probable que se dispersen y más probable que permanezcan en el aire durante períodos prolongados. Los niveles bajos de humedad pueden hacer que las PM sean más pequeñas, lo que aumenta su dispersión (*op.ci*

t.)

La precipitación es el suministro de sustancia líquida de la atmósfera en formas de lluvia-nieve-granizo. Es muy discontinua en el espacio. La precipitación es una variable que tiene un alto grado de incidencia en cómo se comporta la calidad del aire ya que el mismo puede ayudar a que las partículas que se encuentran contaminando la atmósfera, se pueda precipitar en conjunto con el agua para que el aire se limpie. Es decir, las condiciones donde existe gran incidencia de la lluvia en un sector pueden beneficiar a limpiar el aire contaminado. Por el contrario, unas condiciones de sequía pueden ayudar a que el aire siga contaminado, por lo tanto, empeora la calidad del aire (Di y Li, 2019).

La presión atmosférica es una fuerza por unidad de superficie o peso de una columna de aire por unidad de superficie. A su vez la radiación solar es uno de los parámetros meteorológicos más importantes. El sol es la fuente de energía que está detrás de los procesos meteorológicos. Las principales mediciones meteorológicas, realizadas en prácticamente todas las redes, son las de la radiación solar global. Para el rango ultravioleta, que por cierto es dañino para la piel humana, se tienen sensores especializados para lograr estimar sus mediciones espectrales (OMM, 2014).

2.2.5. Modelos para predecir la calidad del aire

Para predecir la calidad del aire se tienen un conjunto de procedimientos que han sido usados, sin embargo, estos pueden ser directamente teniendo en cuenta las variables que lo componen como los contaminantes del aire, o bien

pueden ser muy refinados algunos otros que utilizan variables que lo afectan como el clima.

Factores meteorológicos como la temperatura, la humedad, la velocidad del viento y las precipitaciones suelen influir considerablemente en la formación y dispersión de los contaminantes atmosféricos. Por ejemplo, las altas temperaturas y las bajas velocidades del viento pueden provocar la acumulación de contaminantes en la atmósfera, lo que se traduce en una mala calidad del aire. Del mismo modo, las bajas temperaturas y las altas velocidades del viento pueden facilitar la dispersión de los contaminantes, lo que se traduce en una mejor calidad del aire. Es por ello que suelen ser muy utilizados para la creación de modelos predictivos; a continuación, se enumeran algunos de los modelos más utilizados:

Los modelos estadísticos utilizan datos meteorológicos históricos para identificar patrones y hacer predicciones sobre las condiciones de la contaminación. Ejemplos de modelos estadísticos son los modelos autorregresivos, que utilizan datos meteorológicos pasados para predecir valores futuros, y los modelos de regresión, que identifican relaciones entre variables meteorológicas y utilizan esas relaciones para hacer predicciones (Dalal *et al.*, 2019). Los modelos estadísticos suelen ser más sencillos que los modelos basados en la física y pueden ser más fáciles de interpretar.

La regresión lineal múltiple (MLR) se trata de un modelo estadístico básico que ayuda a la predicción en muchos casos. Esto lo realiza basándose en relaciones lineales entre múltiples variables, para el caso de la calidad del aire lo hace tomando en cuenta variables predictoras como la temperatura, la humedad, la velocidad del viento y la concentración de diversos contaminantes atmosféricos. El MLR es fácil de aplicar, pero asume relaciones lineales entre variables y puede no ser exacto si hay interacciones complejas entre las variables (Vilá *et al.*, 2019).

ARIMA son las siglas de Autoregressive Integrated Moving Average (media móvil integrado autorregresiva), un modelo de análisis estadístico que se utiliza para comprender mejor un conjunto de datos de series temporales o para predecir tendencias futuras, siendo autorregresivos porque se basan en valores pasados. Se utiliza ampliamente en la previsión de series temporales y proporciona

enfoques complementarios al problema junto con el suavizado exponencial (Palomino *et al.*, 2020).

Las redes neuronales artificiales (RNA) se trata de un modelo de aprendizaje automático diseñado para simular la estructura y función del cerebro humano. Las RNA pueden aprender relaciones complejas entre variables y pueden captar interacciones no lineales entre variables. Sin embargo, requiere una gran cantidad de datos para su entrenamiento y el modelo resultante puede ser difícil de interpretar (Dastres y Soori, 2021).

La regresión de vectores de soporte (SVR) se trata de otro modelo de aprendizaje automático que se basa en la idea de encontrar un hiperplano que separe mejor los puntos de datos en un espacio de alta dimensión. SVR puede manejar relaciones no lineales entre variables y puede funcionar bien con conjuntos de datos de tamaño pequeño o mediano. Sin embargo, puede no funcionar bien con conjuntos de datos muy complejos (Evgeniou y Pontil, 2001).

El bosque aleatorio (RF) se trata de un modelo de aprendizaje conjunto que combina varios árboles de decisión para mejorar la precisión de las predicciones. Random Forest puede manejar relaciones no lineales entre variables y es relativamente fácil de interpretar. En algunas situaciones puede no funcionar bien con conjuntos de datos muy desequilibrados (Breiman, 2001).

Los modelos numéricos utilizan complejas ecuaciones matemáticas para simular la física de la atmósfera y predecir las condiciones meteorológicas futuras. Los modelos utilizan datos de satélites, estaciones meteorológicas y otras fuentes para crear una cuadrícula tridimensional de la atmósfera y resolver ecuaciones para el movimiento del aire, la humedad y otras variables (Hernández, 2018). Estos modelos son ampliamente utilizados por las agencias meteorológicas y medioambientales debido a que suelen ser el tipo más preciso de modelo meteorológico disponible.

2.3. Marco legal

Basado en lo promulgado por la Constitución vigente aprobada en el año 2008: El Art. 14 habla sobre el medio ambiente y el derecho que se tiene a vivir en un ambiente sano, se proclama que debe mantenerse un equilibrio ecológicamente

hablando para poder de esta manera llegar a una sostenibilidad, sustentabilidad, proclamadas en el SUMAK KAWSAY que significa buen vivir para el pueblo ecuatoriano. Por tal motivo es indispensable buscar que el ambiente que nos rodea se lo preserve; teniendo en cuenta la gran biodiversidad que forma parte del patrimonio del país (Asamblea Constituyente, 2008).

El Plan Nacional de la Calidad de Aire de Ecuador, se desarrolla por la necesidad de hacer cumplir a mayor cabalidad lo expuesto en la Constitución. Este plan se puso en práctica en el año 2010, haciendo que se prime por lograr preservar el medio ambiente a través de un cuidado extensivo de la naturaleza y con esto llegar a cumplir metas indicadas en el Plan Nacional para el Buen Vivir 2009 - 2013.

La investigación se centra en cumplir con el Proyecto 3 y Proyecto 5 que señala la calidad del aire y modelos predictivos respectivamente. Los objetivos principalmente son en lo que se pone énfasis dentro de esta investigación, buscando desarrollar un modelo que ayude a gestionar de una mejor manera la calidad del aire. Pudiendo en muchos casos definir acciones luego de conocer las predicciones sobre la calidad del aire. Se espera bajar los niveles de contaminación, además de proveer de un camino por el cual dirigirse cuando volviera a existir una pandemia como la azotada por el COVID-19. Esto logra que la salud sustentable de los individuos empiece a mejorar (Ministerio del Ambiente, 2010).

Según la Norma de Calidad de Aire Ambiente del Ecuador aprobada en 2011. Toda la información recabada de acuerdo con el punto 4.1.1.7, y que ha sido resultado de algún programa para el ambiente será de carácter público para quien desee utilizarla. El control y monitoreo de cada una de las concentraciones de gases perjudiciales para el medio ambiente se encuentran regulados por la legislación del ambiente expedida por los Estados Unidos de América (Code of Federal Regulations, Anexos 40 CFR 50), y por cada una de las normas y regularizaciones brindadas por el consejo del aire de Europa (EPA), a más de la American Society for Testing and Materials (ASTM) (Ministerio del Ambiente, 2011).

La Red Metropolitana de Monitoreo Atmosférico De Quito (REMMAQ) es el ente encargado en salvaguardar y hacer cumplir lo anteriormente expuesto. Conforme a ello, están encargados de la recolección, toma y repartición de datos tanto meteorológicos como de calidad de aire. Ellos forman parte de la secretaría del Ambiente de Quito por lo que sus objetivos como sus principios están encaminados a tratar de mejorar la calidad de aire para que la ciudadanía de Quito y sus alrededores se encuentre con una salud digna de un humano por medio de un aire limpio. Para lograrlo han creado un manual donde explica la manera de obtener el índice Quiteño de la calidad del aire (IQCA), mismo que busca proporcionar información que sea valedera para un mejor entendimiento de los contaminantes en la ciudad y que de la misma manera ya se tome en cuenta la presión atmosférica como otras variables que afectan al índice (Secretaría de Ambiente de Quito, 2011).

CAPÍTULO III

METODOLOGÍA

3.1. Descripción del área de estudio/grupo de estudio

Dado que existe una relación entre la densidad poblacional en un sector y la contaminación que puede llegar a ese lugar (Marlier *et al.*, 2016), fue necesario considerar el lugar de estudio. Este es el caso Quito, siendo una de las más grandes ciudades a nivel de Ecuador y por lo tanto el área de estudio. Cabe indicar que Quito se ha visto aquejada por la contaminación ambiental no solo proveniente de la gran cantidad de personas que allí habitan, sino también por el uso de suelos, las industrias emisoras de gases contaminantes y la congestión vehicular (Rodríguez y Cuvi, 2019). Todo ello lleva a mostrar a Quito como un sector altamente contaminado dentro del cual deben plantearse estrategias para tratar de mejorar el ambiente, por lo que se necesita estudiar previamente cómo se comportan los contaminantes.

Dentro del distrito metropolitano de Quito existen algunas estaciones que trabajan para tomar datos meteorológicos como de las concentraciones de los principales contaminantes ambientales. En este estudio se trabajó con 5 de las estaciones pertenecientes al REMMAQ, mismas que se detallan a continuación en la Tabla 2:

Tabla 2

Estaciones Meteorológicas REMMAQ

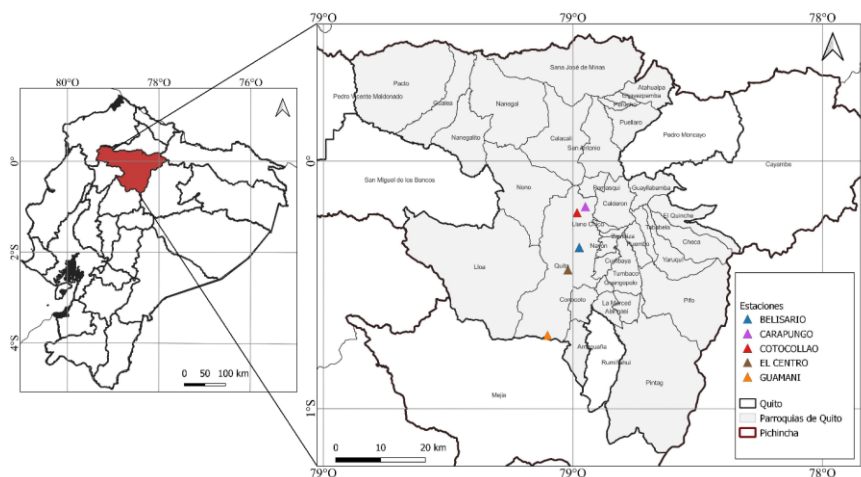
Estaciones Meteorológicas REMMAQ						
Nombre	Código	Provincia	Latitud	Longitud	Altitud (msnm)	Estado
Carapungo	M5036	Pichincha	-0,092	-78,475	2851	Operativa
Cotocollao	M5035	Pichincha	-0,104	-78,492	2807	Operativa
Belisario	M5037	Pichincha	-0,174	-78,487	2835	Operativa
El Centro	M5172	Pichincha	-0,220	-78,510	2820	Operativa
Guamaní	M5039	Pichincha	-0,352	-78,551	2887	Operativa

Nota: Detalle de la zona de estudio, indicando las características principales.

En la Figura 1 se detalla la zona de estudio, se indica cada una de las estaciones que son pertenecientes al REMMAQ y de las cuales se toman los datos para analizarlos. Las estaciones se distribuyen a lo largo del distrito Metropolitano de Quito; siendo Carapungo y Cotocollao la que se encuentra por la región norte, Belisario y El centro en la zona central y Guamaní al Sur de la capital.

Figura 1

Área de Estudio



Nota: Mapa Área de Estudio - Provincia de Pichincha.

3.2. Enfoque y tipo de investigación

3.2.1. Enfoque

El enfoque cuantitativo en la investigación se basa en la recopilación y análisis de datos numéricos con el propósito de identificar relaciones, patrones y tendencias de distintas variables de la población estudiada (Mohajan, 2020). En el contexto de esta investigación este enfoque se aplica, debido a su capacidad para proporcionar una comprensión precisa y objetiva de las variables involucradas. Empleando técnicas estadísticas como el análisis de regresión y predicciones, se descubre relaciones numéricas significativas entre las condiciones meteorológicas y la calidad del aire. Esto permite evaluar cuantitativamente la influencia de la meteorología en la calidad del aire, proporcionando información para la toma de decisiones y la implementación de medidas efectivas de gestión ambiental en la región.

En línea con los objetivos, el enfoque de investigación abarca aspectos descriptivos, correlativos y explicativos. Se diseña el análisis estadístico descriptivo con el fin de identificar relaciones e indicadores entre las variables dependientes e independientes. Asimismo, se busca establecer asociaciones entre estas variables para obtener resultados fiables mediante pruebas de hipótesis que ofrezcan un alto grado de precisión y confiabilidad, así como encontrar los fenómenos que los producen para una mejor predicción de que sucederá a futuro con el ambiente.

3.2.2. Tipo de Investigación

Dentro de este estudio se plantea el desarrollo de predicciones que servirán para que la gente esté preparada para el tiempo venidero y de ser posible se tomen medidas precautelares, que logren proporcionar una estabilidad en la ciudad, y en la calidad del aire existente. Se lleva a cabo una investigación descriptiva que busca determinar las características por medio del análisis estadístico de las variables estudia. Esta investigación es explicativa a su vez puesto que busca determinar cuáles han sido las causas que llevaron a tener los niveles de contaminación dentro de la ciudad, a más de comprobar que las variables tienen relación demostrada físicamente. Se desarrolla una investigación correlacional para observar si la relación es en ambos sentidos o de manera unidireccional.

Además, la investigación se centra en ser no experimental ya que directamente no se han tomado las medidas en campo abierto, sino que han sido recolectados de la base de datos que comparte el REMMAQ. Los mismos datos van a ser tratados por medio de estadística tanto descriptiva como inferencial. Cada una de las operaciones seguidas se detallan en la sección 3.4 Procedimiento y estos van diseñados para tratar de cumplir los objetivos de investigación planteados inicialmente.

3.3. Definición y operacionalización de variables

3.3.1. Variables

Los datos de las variables consideradas son proporcionados por el sistema que existe en la ciudad de Quito por parte del REMMAQ, quienes desde hace algunos años vienen tomando datos meteorológicos y las concentraciones de los principales contaminantes del aire. Para la toma de datos meteorológicos según la Secretaría del Ambiente MDMQ (2023) se dispone de nueve estaciones meteorológicas con sensores automáticos para medir rapidez y dirección de viento (MetOne), humedad relativa (Thies Clima), radiación solar (Kipp y Zonen y Hukseflux), temperatura ambiente (Thies Clima), presión atmosférica (Vaisala), y precipitación (Thies Clima y Metone). Estos sensores están instalados en las estaciones de monitoreo de calidad del aire previamente descritas con sus ubicaciones en la Figura 1 y Tabla 2.

Los datos presentados en este trabajo fueron tomados de la página web que dispone el REMMAQ (<http://www.quitoambiente.gob.ec/index.php/descarga-datos-historicos>), dentro de la cual han compartido datos desde 2006. Sin embargo, en este estudio se utilizó datos desde enero 2019 que se considera como el momento “Antes de la pandemia” hasta marzo de 2020 donde empieza una siguiente etapa; es decir, el inicio de las paralizaciones por la pandemia en Ecuador que queda identificado como el momento “Durante la pandemia”. Finalmente, desde el fin de estado de excepción se empezó un nuevo tiempo con semaforizaciones en septiembre 2019 que se lo toma como el periodo “Después de la pandemia”.

Se tiene principalmente como variables independientes a aquellas que muestran el clima que ha existido, por tal motivo se trabaja con temperatura ($^{\circ}C$), humedad relativa (%), velocidad (m/s) y dirección del viento ($^{\circ}$), precipitaciones (mm), radiación solar (W/m^2) y presión atmosférica (mb). Por otro lado, y dado a la relación probada en publicaciones científicas se toma como variable dependiente al índice de calidad de aire que viene dado por las concentraciones de distintos contaminantes del aire como el Monóxido de Carbono (CO), Partículas ($PM_{2.5}$), Dióxido de Azufre (SO_2), Dióxido de Nitrógeno (NO_2) y Ozono (O_3).

Tabla 3

Operacionalización de Variable Estudio

Variable	Indicadores	Unidades	Concepto	
Independiente	Clima	Temperatura	$^{\circ}C$	Contenido de calor de la atmósfera o de la superficie terrestre.
		Velocidad del Viento	m/s	Es el aire en movimiento, un vector que tiene tanta magnitud, ayuda a la dispersión y el transporte de contaminantes atmosféricos
		Dirección del Viento	$^{\circ}$	Indica la dirección que viene el viento
		Radiación Solar	W/m^2	Valor de la energía que recibimos del sol
		Precipitación	mm	Suministro de sustancia líquida de la atmósfera en formas de lluvia-nieve-granizo
		Humedad Relativa	%	Indica en el ambiente lo cerca que está de la saturación, influye en la formación y dispersión de partículas (PM)
		Presión Atmosférica	Mb	Peso de una columna de aire por unidad de superficie

Dependiente	Índice de Calidad del Aire	Monóxido de Carbono (CO)	mg/m^3	Gas incoloro e inodoro que puede resultar de la combustión incompleta de cualquier material carbonoso
		$PM_{2.5}$ ($PM_{2.5}$)	ug/m^3	Mezcla de partículas sólidas y gotas líquidas que se encuentran en el aire de diámetro aerodinámico menor a 2.5 micrones
		Dióxido de Azufre (SO_2)	ug/m^3	Gas perteneciente al grupo de óxidos de azufre (SO_x)
		Dióxido de Nitrógeno (NO_2)	ug/m^3	Gas más importante y omnipresente en el gas ambiente, la inhalación produce una sensación acre y malestar
		Ozono (O_3)	ug/m^3	Gas compuesto por tres átomos de oxígeno, dependiendo de dónde se encuentre puede ser bueno o malo

Nota: Detalle de las variables utilizadas (Secretaría del Ambiente MDMQ, 2023).

3.4. Procedimiento

El objetivo de este estudio fue investigar la relación entre la meteorología y la calidad del aire en Quito - Ecuador, entre 2019 y 2023; y desarrollar modelos estadísticos para predecir los cambios en la calidad del aire basados en datos meteorológicos. En esta sección se describen los procedimientos que se siguieron incluyendo el diseño de la investigación, recopilación, análisis de datos, y técnicas de modelización estadística.

3.4.1. Comportamiento del índice de la calidad del aire antes, durante y después de la pandemia del COVID-19 en Quito – Ecuador mediante técnicas estadísticas

3.4.1.1. Análisis exploratorio de datos:

Debido a que los datos iniciales tomados de la página web se encontraban horariamente, se realizó un previo análisis exploratorio en las 5 estaciones. Para esto primero se realizó un análisis de los datos para comprender si estos tienen valores perdidos, valores nulos; es decir, se realiza un estudio de sí todos los datos son correctos o si por el contrario ha existido algún problema (Komorowski *et al.*, 2016). Dentro de esta se pudo identificar que existen valores negativos de las concentraciones de contaminantes o simplemente datos que fueron definidos como valores perdidos (NA).

Los valores perdidos en los datos meteorológicos y de calidad del aire pueden tratarse mediante técnicas de imputación (Junninen *et al.*, 2004). Existen varios tipos de imputación como la media, regresión, regresión estocástica y la imputación múltiple fueron los procedimientos que se utilizaron en esta investigación. Se encontró que la imputación múltiple es el más exacto de los métodos planteados para el estudio ya que al realizar gráficas de la distribución de los datos, estos se ajustan de una manera más correcta a la distribución teórica que pudieran tener los datos; además, según Quinteros *et al.* (2019) el método de imputación múltiple ayuda a completar un 100% los datos que se están imputando y mantiene una diferencia de error pequeño con relación a más métodos de imputación.

El método utilizado para la imputación de datos se realizó en las bases (variables) que mostraban una pérdida de datos no mayor al 8%. Los resultados de los tratamientos propuestos se pueden visualizar en el Anexo 3 donde se muestran la distribución de los valores inicialmente. En este paso fue importante obtener los promedios diarios de cada una de las variables climatológicas para luego pronosticar el comportamiento de los contaminantes con relación al clima diariamente. Para el tratamiento de las concentraciones de contaminantes se explica más adelante los pasos seguidos dentro del apartado de Metodología.

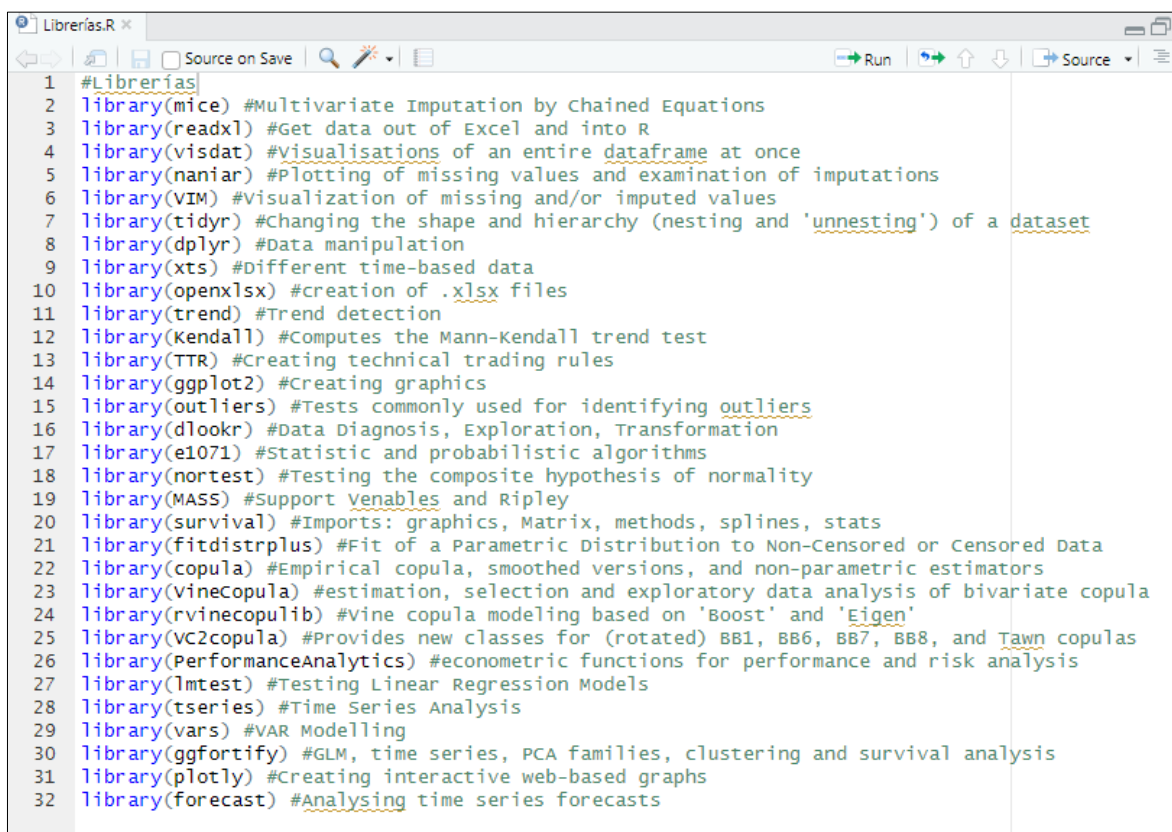
Es necesario hablar sobre el software que se usó para el estudio de la calidad del aire y el clima. R es un lenguaje de programación y un entorno fuertemente manejado para computación estadística y gráficos. Se utiliza ampliamente en el análisis de datos y la modelización estadística, incluido el análisis de datos meteorológicos y de calidad del aire (Setiawan, 2020). En el contexto del análisis meteorológico, R permite realizar diversas operaciones estadísticas. Sirve para manipular, limpiar y procesar grandes conjuntos de datos de parámetros meteorológicos como temperatura, humedad, precipitaciones, velocidad y dirección del viento. Ofrece una amplia gama de funciones y paquetes estadísticos para analizar tendencias, calcular estadísticas descriptivas, realizar pruebas de hipótesis y crear visualizaciones para comprender mejor los patrones meteorológicos y los cambios climáticos (Charalampopoulos, 2020).

En lo que respecta al análisis de la calidad del aire, permite explorar las concentraciones de contaminantes, identificar patrones espaciales y temporales y evaluar los niveles de contaminación atmosférica. R proporciona técnicas avanzadas de modelización estadística para evaluar las relaciones entre las variables meteorológicas y los indicadores de calidad del aire, ayudando a identificar el impacto de las condiciones meteorológicas en la dispersión y formación de contaminantes atmosféricos (Setiawan, 2020).

Se indica en la Figura 2 las librerías usadas.

Figura 2

Librerías R



```
1 #Librerías
2 library(mice) #Multivariate Imputation by Chained Equations
3 library(readxl) #Get data out of Excel and into R
4 library(visdat) #Visualisations of an entire dataframe at once
5 library(naniar) #Plotting of missing values and examination of imputations
6 library(VIM) #Visualization of missing and/or imputed values
7 library(tidyr) #Changing the shape and hierarchy (nesting and 'unnesting') of a dataset
8 library(dplyr) #Data manipulation
9 library(xts) #Different time-based data
10 library(openxlsx) #creation of .xlsx files
11 library(trend) #Trend detection
12 library(Kendall) #Computes the Mann-Kendall trend test
13 library(TTR) #Creating technical trading rules
14 library(ggplot2) #Creating graphics
15 library(outliers) #Tests commonly used for identifying outliers
16 library(dlookr) #Data Diagnosis, Exploration, Transformation
17 library(e1071) #Statistic and probabilistic algorithms
18 library(nortest) #Testing the composite hypothesis of normality
19 library(MASS) #Support Venables and Ripley
20 library(survival) #Imports: graphics, Matrix, methods, splines, stats
21 library(fitdistrplus) #Fit of a Parametric Distribution to Non-Censored or Censored Data
22 library(copula) #Empirical copula, smoothed versions, and non-parametric estimators
23 library(vinecopula) #estimation, selection and exploratory data analysis of bivariate copula
24 library(rvinecopulib) #vine copula modeling based on 'Boost' and 'Eigen'
25 library(VC2copula) #Provides new classes for (rotated) BB1, BB6, BB7, BB8, and Tawn copulas
26 library(PerformanceAnalytics) #econometric functions for performance and risk analysis
27 library(lmtest) #Testing Linear Regression Models
28 library(tseries) #Time Series Analysis
29 library(vars) #VAR Modelling
30 library(ggfortify) #GLM, time series, PCA families, clustering and survival analysis
31 library(plotly) #Creating interactive web-based graphs
32 library(forecast) #Analysing time series forecasts
```

Nota: Captura de pantalla de Librerías usadas en R.

3.4.1.3. Obtención de Índice Quiteño de Calidad del Aire

La base de datos original contiene valores horarios, los cuales previamente fueron imputados para que no existan valores perdidos o negativos, comprobando la validez de estos. Con la base verificada se obtuvo el índice de calidad de aire correspondiente a cada contaminante considerando su concentración. A continuación, se expone en la Tabla 4, el procedimiento seguido en la obtención del índice con relación al contaminante que se estudia con las respectivas consideraciones propuestas en el Distrito Metropolitano de Quito (Secretaría de Ambiente de Quito, 2011).

Tabla 4

Expresiones matemáticas para cada rango de concentración

Contaminante	Expresiones matemáticas para cada rango de concentración			
CO , concentración máxima de promedio de 8 horas, mg/m^3	$0 < C_i < 10$ $IQCA = 10C_i$	$10 < C_i < 15$ $IQCA = 20C_i - 100,00$	$15 < C_i < 30$ $IQCA = 6,67C_i + 100,00$	$30 < C_i$ $IQCA = 10C_i$
O_3 , concentración máxima de promedios de 8 horas, $\mu g/m^3$	$0 < C_i < 100$ $IQCA = C_i$	$100 < C_i < 200$ $IQCA = C_i$	$200 < C_i < 600$ $IQCA = 0,5C_i + 100,00$	$600 < C_i$ $IQCA = 0,5C_i + 100,00$
NO_2 , concentración máxima en 1 hora, $\mu g/m^3$	$0 < C_i < 200$ $IQCA = 0,50C_i$	$200 < C_i < 1000$ $IQCA = 0,125C_i + 75,00$	$1000 < C_i < 3000$ $IQCA = 0,1C_i + 100,00$	$3000 < C_i$ $IQCA = 0,1C_i + 100,00$
SO_2 , promedio en 24 horas, $\mu g/m^3$	$0 < C_i < 62.5$ $IQCA = 0,8C_i$	$62.5 < C_i < 125$ $IQCA = 1,333C_i - 66,667$	$125 < C_i < 200$ $IQCA = 0,125C_i + 175,00$	$200 < C_i$ $IQCA = 0,125C_i + 175,00$
$PM_{2.5}$, promedio en 24 horas, $\mu g/m^3$	$0 < C_i < 50$ $IQCA = 2,00C_i$	$50 < C_i < 250$ $IQCA = C_i + 50,00$	$250 < C_i$ $IQCA = C_i + 50,00$	
PM_{10} , promedio en 24 horas, $\mu g/m^3$	$0 < C_i < 100$ $IQCA = C_i$	$100 < C_i < 250$ $IQCA = 0,6667C_i + 33,33$	$250 < C_i < 400$ $IQCA = 0,6667C_i + 33,33$	$400 < C_i$ $IQCA = C_i - 100,00$

Nota. C_i : Concentración de un determinado contaminante (Secretaría de Ambiente de Quito, 2011)

Cuando los datos han sido procesados se tiene un valor de Índice Quiteño de Calidad del Aire (IQCA) correspondiente cada 8h o cada 24 h de acuerdo con el contaminante estudiado como se lo indicó en la Tabla 4.

Luego de esto fue necesario obtener los valores totales de un día que son determinados mediante el promedio de los valores de IQCA, a fin de obtener el Índice Quiteño De Calidad del Aire Diario (IQCAD).

$$IQCAD_i = \sum IQCA_i / n_i \quad (1)$$

donde:

$IQCAD_i$ son los Índices Quiteños de calidad del Aire Diario para cada i-ésimo contaminante.

$IQCA_i$ son los índices Quiteños de calidad del Aire dependiendo el i-ésimo contaminante.

n_i sería el número de los IQCA obtenidos ya sea cada 8h o cada 24 h.

Finalmente, con este índice diario, se procedió a obtener un índice global o general de todos los contaminantes que nos indica, el nivel de contaminación en ese día sin importar a que contaminante sea debido. Se elige el índice de valor mayor; es decir, el peor índice de calidad de aire (Central Pollution Control Board, 2015).

A este se lo denomina el Índice Quiteño General de Calidad del Aire Diario (IQGCAD), mismo que se toma como el indicativo diario de la calidad del aire en su respectiva fecha.

$$IQGCAD = \max(IQCAD_i) \quad (2)$$

donde:

$IQGCAD$ es el Índice Quiteño General de Calidad del Aire Diario.

$IQCAD_i$ son los índices Quiteños de calidad del Aire dependiendo el i -ésimo contaminante

3.4.1.4. *Variación de la Calidad del Aire*

Con los índices encontrados se procedió a realizar diagramas de dispersión, relacionando el tiempo de aparición con la calidad del aire, con esto se podrá determinar la variación temporal que ha seguido el índice de calidad de aire, tanto desde antes de la aparición de las restricciones vehiculares, pasando por las paralizaciones, hasta ahora que existe libre movilidad nuevamente. Para esto se propuso realizar el estudio en tres momentos específicos que se mencionan a continuación:

El primer período (Antes de la Pandemia), tomado desde el 1 de enero de 2019 hasta el 15 de marzo de 2020 que inició las paralizaciones por pandemia. A partir de ahí se tuvo el siguiente periodo (Durante la Pandemia) que está tomado desde el 16 de marzo de 2020 hasta el 11 de septiembre de 2020, donde se da el inicio del tercer y último periodo (Después de la pandemia) desde el 12 de septiembre de 2020 que es la fecha en la cual el COE y el Gobierno Nacional de Ecuador dieron ligeras aperturas, terminando el periodo de los estados de excepción. El último periodo termina con datos tomados hasta el 31 de enero de 2023. Estas fechas como se visualiza tiene alta relación con el transcurso por el COVID-19 en la ciudad de Quito - Ecuador. Para determinar las fechas se consideró los documentos oficiales del COE Nacional quienes en sesión permanente fueron determinando lo que se debía seguir durante la pandemia en Ecuador (Ocles, 2020; Salazar, 2021)

3.4.1.5. *Test No Paramétrico de Mann-Kendall*

Tras haber realizado la respectiva división de la base de datos en tres periodos se procedió aplicar la prueba de tendencia, en este caso se utilizó la prueba No Paramétrica de Mann-Kendall la cual es utilizada ampliamente para detectar tendencias en datos de series temporales. Al ser una prueba no paramétrica, no asume ninguna distribución específica para los datos, por lo cual se analizó globalmente cada uno de los períodos previamente indicados. La prueba Mann-Kendall funciona comparando los valores de una serie temporal con el valor mediano de la serie. Si los valores tienden a estar mayoritariamente por encima

o por debajo del valor mediano, esto indica una tendencia al alza o a la baja en los datos, respectivamente. La prueba calcula un valor denominado Tao de Kendall, que mide la fuerza de la tendencia. Si el Tao de Kendall es positivo, indica una tendencia al alza, y si es negativo, indica una tendencia a la baja (Cabral y Luna, 2020).

La prueba estadística de Mann Kendall se calcula por medio de la aplicación de la siguiente fórmula:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(y_j - y_i) \quad (3)$$

donde:

$\text{sign}(y_j - y_i)$ es igual a 1 cuando $y_j > y_i$, y es igual a 0 cuando $y_j = y_i$, además es igual a -1 cuando $y_j < y_i$

El valor de (Tao) de Kendall por medio de:

$$\tau = \frac{S}{n(n-1)} \quad (4)$$

Este valor oscila entre -1 y 1, siendo que el valor de $\tau=0$ muestra que no hay tendencia visible.

La prueba de Mann-Kendall se utiliza habitualmente en estudios medioambientales e hidrológicos para detectar tendencias en variables como el caudal de los ríos, las precipitaciones y la temperatura, aunque también se utiliza en finanzas y economía para detectar tendencias en los precios de las acciones, los tipos de cambio y otros indicadores económicos (Reshu *et al.*, 2014). Cabe señalar que la prueba de Mann-Kendall sólo detecta tendencias monótonas; es decir, tendencias que aumentan o disminuyen con el tiempo. No detecta tendencias que tengan una forma más compleja, como oscilaciones o ciclos. Además, como cualquier prueba estadística, la prueba de Mann-Kendall tiene limitaciones.

3.4.2. Contaminación del aire con respecto a la variabilidad climática

Para determinar si existe una dependencia entre las variables estudiadas, se pueden utilizar medidas estadísticas, aunque es importante tener en cuenta el contexto, la distribución y las posibles variables de confusión a la hora de interpretar los resultados. Aunque una relación estadísticamente significativa entre dos variables meteorológicas puede sugerir dependencia, también puede estar influenciada por otros factores que no se tienen en cuenta en el análisis.

3.4.2.1. Análisis de la base de datos

Para empezar a trabajar con cada una de las variables en los análisis siguientes, fue necesario comprender la distribución que siguen los datos para saber si podemos utilizar una prueba paramétrica o no. Para lograr esto se comprobó la normalidad de los datos, por medio de algunas pruebas. Se realizó un análisis de asimetría, y por medio de una transformación se buscó reducir la misma. Finalmente se realizó una estandarización de la base tomando en cuenta la mejor alternativa.

Se realiza la Comprobación de Normalidad, al conocer la desviación que toman los datos se debe comprobar el supuesto de normalidad. Este paso es muy importante porque permite conocer el comportamiento de las variables y la distribución de los datos de una manera más fiable se procedió con 4 métodos, mismos que se describen a continuación:

La prueba de Kolmogórov-Smirnov compara la función de distribución acumulativa (FDA) de los datos de la muestra con la FDA de la distribución normal esperada. Calcula la distancia vertical máxima (D) entre las dos FDA. A continuación, el estadístico de la prueba se compara con los valores críticos o p-valores para determinar si los datos se desvían significativamente de una distribución normal (Massey, 1951). Esta prueba es sensible a las diferencias tanto de localización como de forma.

El estadístico a utilizar en esta prueba viene dado por:

$$D = \sup_x |F_n(X) - F(X, \mu, \sigma)| \quad (5)$$

donde:

$F(X, \mu, \sigma)$ es la función de distribución acumulativa de la distribución normal y $F_n(X)$ es la función empírica de la distribución de los datos.

La prueba de Lilliefors es una modificación de la prueba de Kolmogórov-Smirnov diseñada específicamente para tamaños de muestra más pequeños. Estima los valores críticos de la estadística de la prueba en función del tamaño de la muestra, lo que mejora la potencia en comparación con la prueba estándar de Kolmogórov-Smirnov (Abdi y Molin, 2007). Sigue el mismo procedimiento que la prueba de Kolmogórov-Smirnov, comparando la FDA observada con la FDA esperada bajo el supuesto de normalidad.

El estadístico utilizado es:

$$T_1 = \sup_x |F_*(X) - S(X)| \quad (6)$$

donde:

$S(X)$ es la función de distribución empírica de los valores Z y $F_*(X)$ es la función normal estándar.

La prueba chi-cuadrado de normalidad compara las frecuencias observadas de los datos en intervalos predefinidos con las frecuencias esperadas suponiendo una distribución normal. Se logra calcular el estadístico de la prueba chi-cuadrado sumando las diferencias al cuadrado entre las frecuencias observadas y las esperadas, divididas por las frecuencias esperadas. El estadístico resultante sigue una distribución chi-cuadrado, y su significación se determina comparándolo con los valores críticos o p-valores (Lemeshko, 2015). La prueba chi-cuadrado es menos sensible a las diferencias de forma, pero puede detectar desviaciones en las colas de la distribución.

Es importante señalar que las pruebas mencionadas no son una prueba definitiva de normalidad, sino que proporcionan una evaluación de lo bien que los datos se ajustan a una distribución normal basándose en las pruebas disponibles. Es por ello por lo que como añadido fue importante también realizar una revisión

visual en un gráfico Q-Q y un histograma para complementar las pruebas estadísticas para una evaluación más completa de la normalidad (Rani, 2016).

La asimetría es una medida estadística que describe la falta de simetría en la distribución de probabilidad de una variable aleatoria. En términos más sencillos, podemos indicar cuánto se desvían de una distribución normal los datos (Orcan, 2020). Al obtener los coeficientes de asimetría se comprendió la forma de una distribución e identificó posibles valores atípicos o desviaciones de la simetría. Larasati *et al.* (2019), indica que la medida del coeficiente de asimetría puede tomar valores positivos, negativos o cero, lo que indica distintos tipos de asimetría:

Asimetría positiva: La distribución tiene una cola larga en el lado derecho y la mayoría de los valores se concentran en el lado izquierdo.

Asimetría negativa: La distribución tiene una cola larga en el lado izquierdo y la mayoría de los valores se concentran en el lado derecho.

Asimetría cero: La distribución es simétrica, es decir podemos inferir que la distribución va a acercarse a una normal.

Las medidas de tendencia central ayudan a conocer hacia que valor están tendiendo nuestras variables; sin embargo, no siempre es mejor elegir la media como medida indicativa, sino que en algunos casos la mediana es un mejor indicador. Por ello con el conocimiento previo de si las variables son simétricas y si son normales se procede a ir eligiendo que medida tomar como central.

De Nijs y Klausen (2013) nos indica que la media es adecuada cuando los datos tienen una distribución normal o aproximadamente simétrica; sin embargo, es una medida sensible a los valores extremos (valores atípicos), ya que tiene en cuenta todos los puntos de datos mientras que; la mediana es robusta frente a los valores atípicos porque sólo tiene en cuenta los valores medios y se ve menos afectada por los valores extremos es por ello adecuada para conjuntos de datos con distribuciones sesgadas o cuando hay valores atípicos que pueden influir en la media.

De no haberse comprobado la normalidad en todas las variables un paso importante dentro del análisis es el aplicar transformaciones. En este caso se aplicaron las principales transformaciones dependiendo si ayudaban a que los datos se volvieran normales o de igual manera de no hacerlo al menos reducir su coeficiente de asimetría (West, 2022). Como nos indica Watthanacheewakul (2021), alguna de las principales transformaciones y que se fueron probando en las variables de este estudio son:

Transformación logarítmica: Toma el logaritmo de los valores de los datos. Esta transformación es eficaz para reducir la asimetría a la derecha y hacer que la distribución sea más simétrica.

Transformación de la raíz cuadrada: Toma la raíz cuadrada de los valores de los datos. Esta transformación puede ayudar a reducir la asimetría a la derecha, pero es menos eficaz que la transformación logarítmica.

Transformación raíz cúbica: Al aplicar una transformación de raíz cúbica, los valores extremos se acercan más a la media, lo que reduce la influencia de los valores atípicos. Cuando los datos están sesgados a ambos lados de la distribución, la transformación de raíz cúbica puede reducir la asimetría y hacer que la distribución sea más simétrica.

Es importante tener en cuenta que las transformaciones alteran los valores originales de los datos, por lo que se llevó un registro con fines de análisis e interpretación posteriores. Después de aplicar la transformación, se calculó de nuevo la asimetría para evaluar la eficacia de la transformación.

Debido a que las variables pueden mostrar un comportamiento aún muy variado entre ellas, se procedió a estandarizar nuestra base de datos. Como método de estandarizar se utilizó el escalamiento robusto. Esta es una técnica de transformación de datos que se utiliza habitualmente en el preprocesamiento de datos para escalar características de forma robusta, sobre todo cuando se trata de datos que contienen valores atípicos o no están distribuidos normalmente (De Amorim *et al.*, 2022).

Para realizar el escalamiento robusto se debe aplicar la siguiente ecuación en cada uno de los datos:

$$x_i^{\wedge} = \frac{xi-Q_2(x)}{Q_3(x)-Q_1(x)} \quad (7)$$

donde:

x_i^{\wedge} es i-ésimo dato escalado,

xi es el i-ésimo dato al que se va a realizar el escalamiento, y

$Q_1(x), Q_2(x), Q_3(x)$ indicar el respectivo cuartil de la variable.

3.4.2.2 Coeficiente de Correlación de Pearson

El coeficiente de correlación de Pearson es una medida estadística que suele utilizarse para determinar la fuerza y la dirección de una relación entre dos variables continuas (Zinzenoff *et al.*, 2020). En este estudio se aplicó para evaluar la relación lineal entre la calidad del aire y variables meteorológicas; además, para conocer la relación que tienen entre sí las variables del clima pudiendo determinar si tienen alguna dependencia.

Un coeficiente de correlación positivo indica que a medida que aumenta una variable, la otra también tiende a aumentar, mientras que un coeficiente de correlación negativo indica que a medida que aumenta una variable, la otra tiende a disminuir. Un coeficiente de correlación cero indica que no existe relación lineal entre las dos variables (Sedgwick, 2012). El coeficiente de correlación de Pearson entre dos variables se calculó siguiendo la siguiente ecuación:

$$\rho = \frac{E[(X_t-\mu)(X_{t-\lambda}-\mu)]}{\sigma^2} \quad (8)$$

donde:

ρ representa el coeficiente de autocorrelación;

X_t representa los datos muestrales de la variable X en un momento t;

$X_{t-\lambda}$ representa los datos muestrales retrasados con respecto a X_t y con un tiempo de retardo λ ;

μ y σ representan la media y la desviación típica de la secuencia muestral, respectivamente.

Es importante señalar que, aunque el coeficiente de correlación de Pearson puede proporcionar información valiosa sobre la fuerza y la dirección de una relación lineal entre dos variables, no puede probar la causalidad. Otros factores pueden estar influyendo tanto en la calidad del aire como en las variables meteorológicas, por lo que sería necesario un análisis más detallado para determinar la relación causal entre ambas (Zinzenoff *et al.*, 2020).

En el presente estudio el coeficiente de correlación de Pearson fue únicamente utilizado cuando existía normalidad entre las dos variables relacionadas. En los casos en que una de las variables no mostró normalidad previamente encontrada se procedió a obtener el coeficiente de correlación de Spearman, que es un método más robusto frente a falta de normalidad.

3.4.2.3 Coeficiente de Correlación de Spearman

El test no paramétrico del coeficiente de correlación de Spearman ayuda en la obtención de la correlación de dos variables de las cuales se desconoce su distribución. La rho de Spearman (ρ) es una medida de la fuerza y la dirección de la relación monótona entre dos variables. Resulta especialmente útil cuando la relación entre variables es compleja, pero puede describirse mediante una función monótona (Zhang, 2023). Se utiliza la siguiente fórmula para calcular el coeficiente de Spearman:

$$\rho = 1 - \frac{6 \cdot \Sigma(d^2)}{n \cdot (n^2 - 1)} \quad (9)$$

donde:

ρ es el coeficiente de Spearman

$\Sigma(d^2)$ es la suma de las diferencias de rango al cuadrado

n es el número de observaciones

Sin embargo, al igual que ocurre con el coeficiente de correlación de Pearson, el coeficiente de correlación de Spearman por sí solo no puede demostrar la causalidad y muchas veces pueden malinterpretar los resultados, por lo que fue necesario unos análisis adicionales para determinar la relación entre la calidad del aire y las variables meteorológicas.

3.4.2.4 Teoría Cópula

La Teoría Cópula es un método estadístico que puede utilizarse para modelizar la dependencia entre dos o más variables, aunque sus distribuciones marginales no estén distribuidas normalmente. Es una potente herramienta para estudiar la relación entre la calidad del aire y la meteorología, que puede presentar dependencias complejas debido a las interacciones no lineales y complejas entre distintas variables meteorológicas (Wang y Li, 2022). En la teoría cópula, la estructura de dependencia entre dos variables se describe mediante una función de cópula, que transforma las distribuciones marginales de las variables en una distribución conjunta.

Según las variables utilizadas cuando $x = F(x)$, $y = G(y)$, y la función Cópula $C(u, v)$ existe, se tiene que $u = F(x)$, $v = G(y)$, $u, v \in [0,1]$, el coeficiente de correlación Kendall τ se puede expresar como:

$$\tau = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1 \quad (10)$$

La correlación entre los factores X y los contaminantes Y son calculados.

El coeficiente de correlación de cola puede expresarse mediante la función Cópula:

$$\lambda = \lim_{\alpha \rightarrow 1} P\{Y > G^{-1}(\alpha) | X > F^{-1}(\alpha)\} \quad (11)$$

$$\lambda = \lim_{\alpha \rightarrow 1} \frac{C(1-\alpha, 1-\alpha)}{1-\alpha} \quad (12)$$

Puede utilizarse una cópula Gaussiana para modelizar la dependencia entre la calidad del aire y la temperatura, mientras que una cópula T de Student puede ser más apropiada para modelizar la dependencia entre la calidad del aire y la

velocidad del viento. Eligiendo la función de cópula adecuada, es posible captar la compleja dependencia entre la calidad del aire y la meteorología, incluso cuando sus distribuciones marginales no son normales, y para esto ha ayudado el uso del software R. En el Anexo 5 se pueden encontrar más detalles de cada una de las gráficas simuladas por medio de la aplicación de distintas funciones Cópula.

La teoría de la cópula se ha aplicado ampliamente en estudios medioambientales, como modelizar la dependencia entre la contaminación atmosférica y diversas variables meteorológicas, como la temperatura, la humedad, la velocidad del viento y las precipitaciones, con el fin de desarrollar modelos de previsión de la calidad del aire más precisos para comprender mejor las complejas interacciones entre la calidad del aire y la meteorología (Wang y Li, 2022).

3.4.3 Modelos estadísticos de la Calidad del Aire teniendo en cuenta las variables meteorológicas para la predicción de la contaminación.

Dentro de la parte final de este estudio se logró conocer en cierta manera la participación de las distintas variables como a su vez por la incidencia que tuvo la paralización causada por la pandemia del COVID 19 en Quito - Ecuador.

A su vez para ayudar a la comprensión global de la contaminación dentro del Distrito Metropolitano de Quito, se sacaron promedios por día entre las 5 estaciones para obtener nuestros indicadores finales de cada una de las variables participes en el estudio para la modelización, análisis y validación de resultados.

Para el modelamiento y la comprobación de que tanto se acercan los datos predichos a los de la base de estudio, se dividió a la base de datos de cada uno de los momentos, en dos secciones. Se tomó para desarrollar y enseñar al modelo el 80% de datos, y por otra parte correspondiente al 20% de los datos para mirar y evaluar si la predicción fue correcta o no por medio de algunos indicadores detallados más adelante. Los porcentajes fueron tomados en cuenta la investigación realizada por Gholamy *et al.* (2018).

3.4.3.1 Predicción 1: Regresión Lineal Múltiple (RLM)

La RLM es un método estadístico utilizado para modelizar la relación entre una variable dependiente y dos o más variables independientes (Uyanık y Güler, 2013). En el contexto de la predicción de la calidad del aire, la regresión lineal múltiple se utilizó para modelizar la relación entre la calidad del aire y las variables meteorológicas.

Tras inicialmente haber identificado las relaciones, se construyó un modelo de regresión lineal múltiple. El modelo de regresión lineal múltiple adopta la forma de una ecuación del siguiente estilo:

$$ICA = b_0 + b_1 (X_1) + b_2 (X_2) + \dots + b_n (X_n) + e \quad (13)$$

donde:

b_0 es el intercepto,

b_n es el coeficiente de regresión para la n -ésima variable independiente

e es el término de error.

X_1, X_2, \dots, X_n son las variables meteorológicas

Los coeficientes de regresión se han estimado utilizando una técnica llamada regresión por mínimos cuadrados, que minimiza la suma de los errores al cuadrado entre los valores previstos y reales de la variable dependiente (Anusha *et al.*, 2019).

Una vez construido el modelo, se lo utilizó para hacer predicciones sobre la calidad del aire a partir de los valores de las variables independientes. Es importante señalar que los modelos de regresión lineal múltiple no son predictores perfectos, y es posible que haya otros factores que influyen en la calidad del aire que no se tuvieron en cuenta en el modelo aquí presentado. Sin embargo, la regresión lineal múltiple puede ser una herramienta útil para comprender la relación entre las variables (Wei, 2019), como en este estudio entre la meteorología y la calidad del aire.

3.4.3.2 Predicción 2: Modelo ARIMAX

ARIMA-X (AutoRegressive Integrated Moving Average with Exogenous variables) es un modelo de previsión de series temporales que combina el modelo ARIMAX con variables exógenas. Además de los componentes ARIMAX, de la modelación univariante, el modelo ARIMAX incluye variables exógenas, que son factores externos que pueden influir en la variable objetivo (Kongcharoen y Kruangpradit, 2013). En este caso, las variables meteorológicas como temperatura, humedad, velocidad del viento, etc., se consideran variables exógenas. La inclusión de estas variables en el modelo permitió captar el impacto de las condiciones meteorológicas en la calidad del aire.

La ecuación del modelo ARIMAX para modelizar la calidad del aire a partir de datos meteorológicos puede representarse de la siguiente manera:

$$Y(t) = c + \Phi_1 Y(t-1) + \dots + \Phi_p Y(t-p) + \theta_1 \varepsilon(t-1) + \dots + \theta_s \varepsilon(t-s) + \beta_1 X_1(t) + \dots + \beta_m X_m(t) + \varepsilon(t) \quad (14)$$

donde:

$Y(t)$ representa la medición de la calidad del aire en el momento t .

c es el intercepto o término constante.

Φ_1, \dots, Φ_p son los coeficientes autorregresivos correspondientes a los valores retardados de la variable de calidad del aire (Y).

$\theta_1, \dots, \theta_s$ son los coeficientes de media móvil correspondientes a los errores residuales retardados (ε).

β_1, \dots, β_m son coeficientes de las variables exógenas (variables meteorológicas) representadas como $X_1(t), \dots, X_m(t)$.

$\varepsilon(t)$ es el término de error, que representa el componente aleatorio del modelo en el momento t .

Los parámetros p, d y q representan los órdenes de los componentes autorregresivo, de diferenciación y de media móvil, respectivamente, de la

variable de calidad del aire (Liu *et al.*, 2018). El parámetro s representa el orden del componente de media móvil para los errores residuales.

3.4.3.3 Validación de Modelos

Los modelos se validaron utilizando medidas como el error medio absoluto (MAE) y la raíz cuadrada del error cuadrático medio (RMSE); además, de evaluarse que tan bien se ajustan los valores predichos a los valores usados para la validación por medio del coeficiente de determinación (R-cuadrado).

El error medio absoluto (MAE) es una medida de la magnitud media de los errores entre los valores previstos y los reales. En el contexto de la modelización de la calidad del aire, el MAE puede utilizarse como medida de la precisión predictiva de los modelos que utilizan variables meteorológicas para predecir la calidad del aire. El MAE mide la diferencia absoluta media entre los valores predichos y los valores reales, lo que da una indicación de la distancia entre las predicciones del modelo y los valores reales. Un MAE bajo indica que las predicciones del modelo se aproximan en general a los valores reales, mientras que un MAE alto indica que las predicciones del modelo se alejan mucho de los valores reales (Wang y Lu, 2018).

Sin embargo, como todas las medidas del rendimiento de un modelo, el MAE tiene algunas limitaciones. Por ejemplo, no indica si el modelo se ajusta en exceso o en defecto a los datos, y no proporciona ninguna información sobre la importancia relativa de las distintas variables del modelo (Cruz y Barrios, 2018).

El MAE es calculado de la siguiente manera:

$$MAE = \frac{1}{n} \sum_{i=1}^n |z_i - \hat{z}_i| \quad (15)$$

Al utilizar múltiples medidas del rendimiento del modelo y métodos de validación, es posible obtener una comprensión más completa de la precisión y fiabilidad del modelo para predecir la calidad del aire con la meteorología. Por lo tanto, es importante utilizar el MAE junto con otras medidas de rendimiento del modelo, como R-cuadrado, error cuadrático medio (RMSE) o error porcentual medio absoluto (MAPE), así como métodos de validación como la validación por

muestreo dividido, la validación cruzada y la validación de series temporales (Cruz y Barrios, 2018).

El RMSE (Root Mean Squared Error) es una métrica muy utilizada en el análisis de regresión para evaluar la precisión de las predicciones de un modelo sobre datos de validación (Wang y Lu, 2018). Ayuda a observar el rendimiento del modelo en términos de lo bien que las diferencias entre los valores previstos y los reales.

El error cuadrático medio (RMSE) se calcula como:

$$RMSE = \sqrt{(1/n) * \Sigma((valor\ previsto - valor\ real)^2)} \quad (16)$$

donde:

RMSE es la raíz cuadrada del error cuadrático medio

n es el número de observaciones del conjunto de datos de validación

Σ es la suma de las diferencias al cuadrado de todas las observaciones.

Un RMSE más bajo indica un mejor ajuste del modelo a los datos de validación, ya que significa menos error global entre los valores predichos y los reales. El RMSE se utiliza mucho porque penaliza más los errores grandes que los pequeños debido a la operación de elevar al cuadrado (Chai y Draxler, 2014).

El Coeficiente de Determinación o R-Cuadrado puede utilizarse para evaluar hasta qué punto el modelo se ajusta a los datos observados. El R-cuadrado oscila entre -1 y 1, y un valor de 0 indica que no se acoplan los datos predichos a los que se utilizaron para validar el modelo (Crocker, 1972). En esta investigación en específico un valor R-cuadrado cercano a 1 indica que una gran proporción de la variación de la calidad del aire puede explicarse por las variables meteorológicas incluidas en el modelo, mientras que un valor bajo de R-cuadrado (cercano a 0) indica que el modelo no es capaz de explicar gran parte de la variable dependiente (Chicco *et al.*, 2021).

El coeficiente se lo puede calcular de la siguiente manera:

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2} \quad (17)$$

Sin embargo, es importante señalar que R-cuadrado tiene algunas limitaciones cuando se utiliza para la validación de modelos. Por ejemplo, R-cuadrado no proporciona información sobre si el modelo se ajusta en exceso o en defecto a los datos, y no proporciona ninguna información sobre la exactitud predictiva del modelo para los nuevos datos que no se utilizaron para desarrollar el modelo.

3.5. Consideraciones bioéticas

En este estudio se pone en consideración los principios bioéticos de precautelar la vida por medio del mejoramiento e informe de la calidad del aire, el cual es el principal componente dentro del tratamiento de la contaminación ambiental. En el estudio del índice de calidad del aire en relación con la meteorología se plantea varias consideraciones bioéticas, enunciadas a continuación:

Justicia ambiental: Se ha demostrado que la contaminación atmosférica afecta de forma desproporcionada a las comunidades marginadas (Rodríguez y Cuvi, 2019). Por lo tanto, es importante tener en cuenta el posible impacto diferencial de la contaminación atmosférica en las distintas comunidades y poblaciones. Se han considerado medidas para garantizar que su investigación no perpetúe o agrave las disparidades existentes en materia de salud.

Intercambio de datos: Los datos sobre la calidad del aire en Quito están a disposición del público, siendo que se utilizan aquellos brindados por el REMMAQ y que están a disposición del público en general. Al pertenecer a un instituto gubernamental como el REMMAQ, la forma de compartirlos es abierta siempre y cuando se utilicen de la mejor manera como en esta investigación donde sirven para el mejoramiento y conocimiento de cómo se comporta la contaminación en una ciudad tan acelerada como lo es el Distrito Metropolitano de Quito. Dentro de los resultados que se muestran se trató de ser lo más transparente para cumplir la normativa y las directrices indicadas por los organismos reguladores.

Colaboración interdisciplinar: En el estudio de la calidad del aire y el cambio climático intervienen diversas disciplinas, como las ciencias medioambientales,

la física, la estadística, la matemática, la salud pública. Por lo tanto, la colaboración interdisciplinar es esencial para garantizar que la investigación se lleva a cabo de forma ética y que los resultados se interpretan y comunican adecuadamente.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

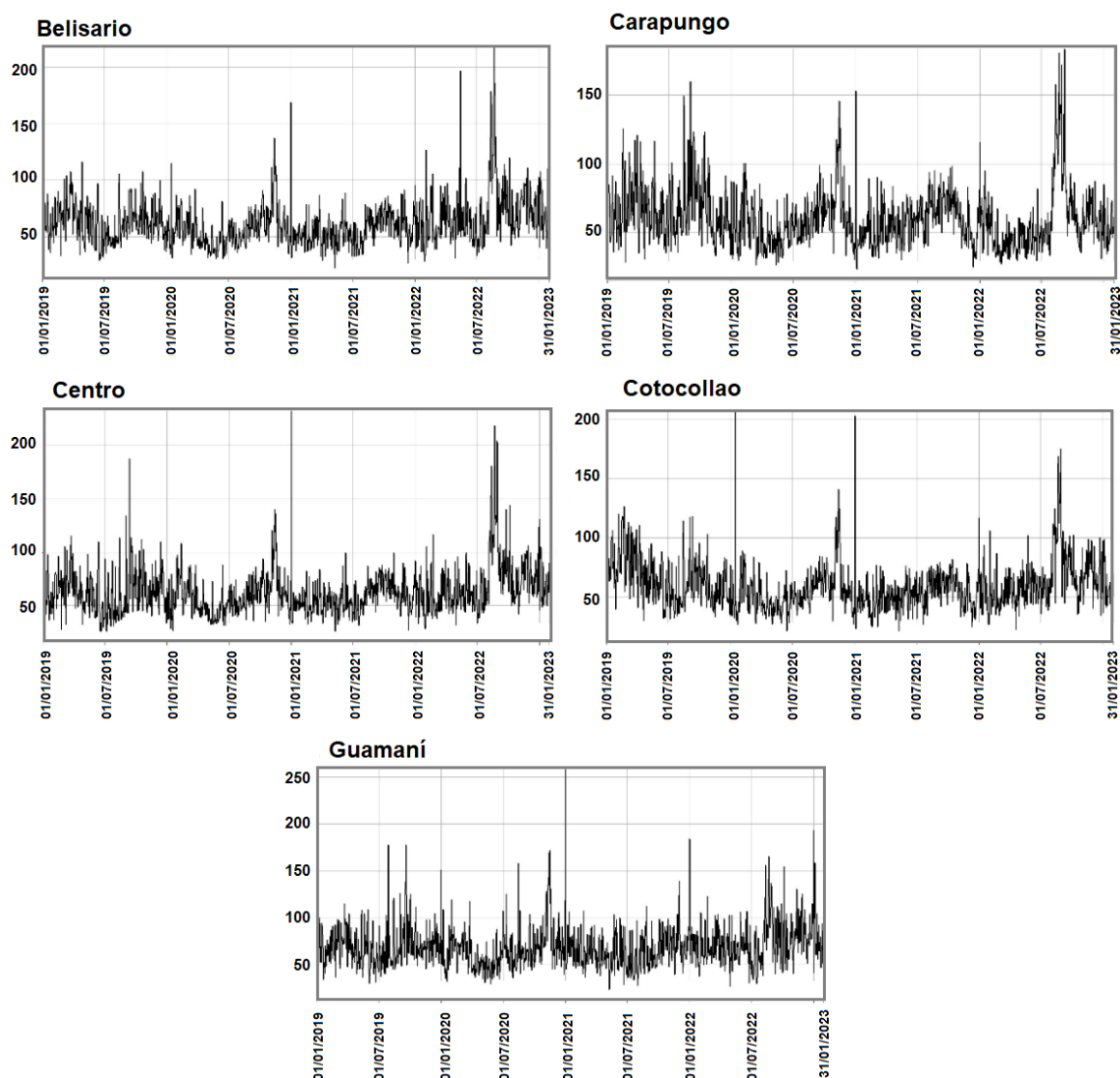
A continuación, se presentan los resultados obtenidos en este trabajo, haciendo una distinción respecto a cada uno de los objetivos secundarios planteados. Es decir, se analiza la calidad del aire en el período de estudio, luego se observa las relaciones existentes entre las variables climatológicas y la calidad del aire. Finalmente se busca realizar un repaso de algunos métodos para predecir la calidad del aire, sin embargo para predicciones o pronósticos de incidencia climática necesitaríamos por lo menos un aserio de datos de 10 años, para justificar su variabilidad climática esto según la OMM, (2018). Lo que se realiza entonces no es un pronóstico y más bien se lo toma como una comparación de causa y efecto de ambas variables durante los periodos de tiempo antes, durante y después de pandemia.

4.1 Índice de la calidad del aire antes, durante y después de la pandemia del COVID-19 en Quito

En la Figura 3 se visualiza el comportamiento el IQGCDA en cada una de las estaciones a lo largo de todo el periodo del estudio; es decir, son datos que van desde 01/01/2019 hasta el 31/01/2023. En ella se nota como en todas las estaciones existen picos más sobresalientes, como por ejemplo aquellos picos positivos que tenemos entre cada 31 de diciembre y 01 de enero de los años correspondientes debido a que en estas fechas se quema los monigotes y las festividades por fin de año que llevan a que los índices se disparen mostrando una mayor contaminación del aire (Municipio del Distrito Metropolitano de Quito, 2020). Más adelante se detalla los picos y su posible causa haciendo la distinción entre las tres etapas del estudio; es decir, antes de la pandemia, durante la pandemia y finalmente después de la pandemia.

Figura 3

IQGCDA - Todas las Estaciones - 01/01/2019 - 31/01/2023



Nota: Gráfica en series de Tiempo de IQGCDA, de todas las estaciones durante todo el tiempo de estudio.

4.1.1 Índice de la calidad del aire antes de la pandemia

En la Tabla 5 se observan algunos valores representativos de la variable IQGCDA como los valores mínimos, máximos, media, mediana, desviación estándar y el coeficiente de variación del IQGCDA para el primer período (Antes de la Pandemia). La estación de Belisario registró la menor contaminación del Distrito de Quito con valor de $61,80 \pm 16,07$; esto puede deberse a que se obtuvo, también el valor mínimo en la máxima calidad de aire y con una mediana

variabilidad diaria de calidad del aire ($CV < 0,30$). Por otro lado, en la estación del Centro presentó un valor menor en la mínima calidad del aire, por lo que la mediana en este lugar toma su menor valor. Guamaní en cambio mostró la peor contaminación de aire, siendo que refleja una media de $69,10 \pm 27,79$ mediana de 65,60 y teniendo sus respectivos valores mínimos y máximos como lo más altos de todas las estaciones. Las demás estaciones se encuentran intermedias con valores que no superan los máximos de las estaciones ya mencionadas ni con valores mínimos inferiores. Esto se puede visualizar de una manera gráfica también en la Figura 4.

Los coeficientes de variación de todas las estaciones Carapungo, Centro, Cotocollao y Guamaní son parecidos lo cual nos indica un nivel tendiendo a alto ya que tenemos mayor al 30% de variabilidad (Pélabon *et al.*, 2020).

Tabla 5

Valores Descriptivos IQGCDA - Antes de la Pandemia

	Mínimo	Máximo	Media	Mediana	SD	CV
Belisario	28,99	116,23	61,83	60,37	16,07	0,26
Carapungo	26,07	160,19	66,54	62,09	21,51	0,32
Centro	25,56	187,59	62,56	60,28	20,41	0,33
Cotocollao	26,65	207,45	64,16	61,82	21,27	0,33
Guamaní	29,52	246,73	69,09	66,50	21,79	0,31

Nota: Valores descriptivos en el período comprendido entre 01/01/2019 - 15/03/2020.

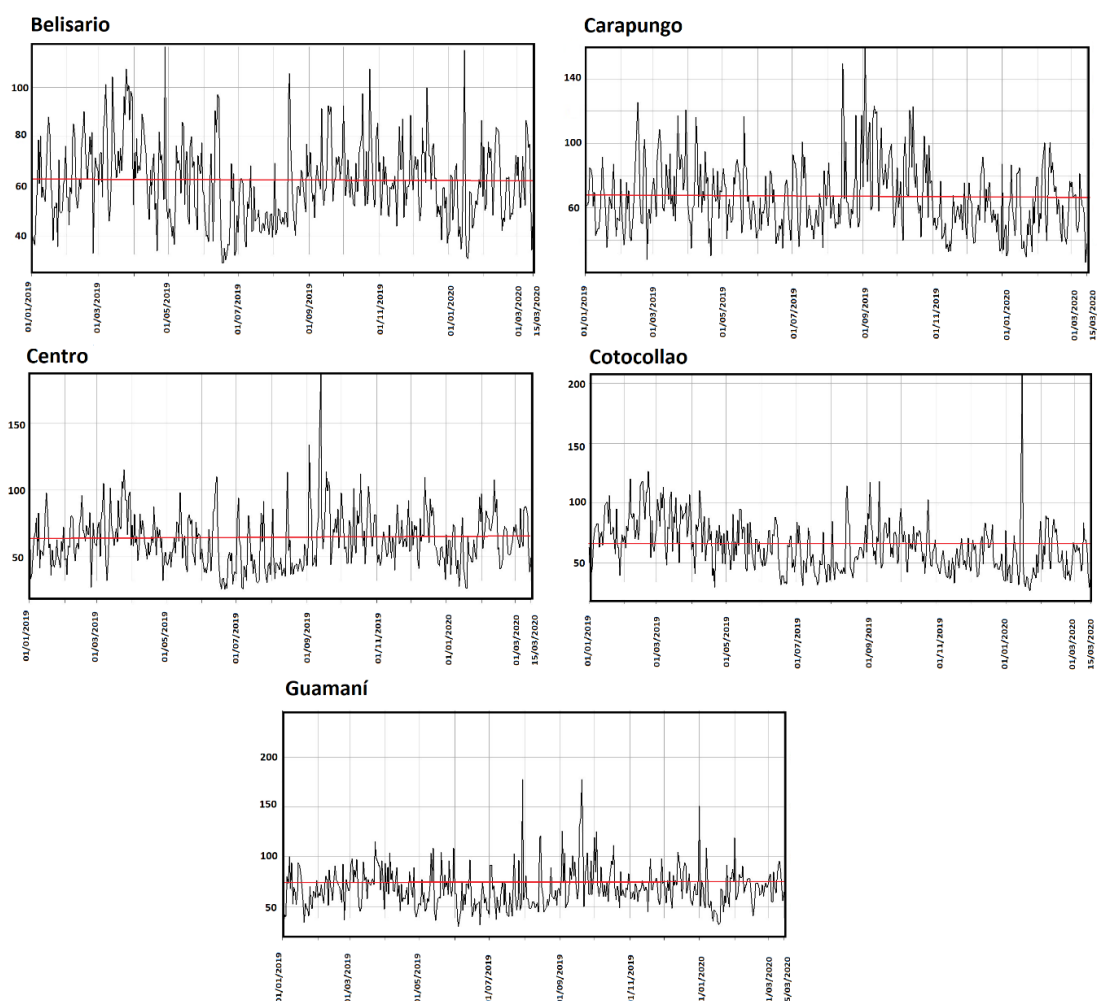
Dentro de la Figura 4 se observa una serie de tiempo correspondiente al período Antes de la Pandemia que va desde 01/01/2019 a 15/03/2020 que como se ha explicado anteriormente es cuando empezó las paralizaciones debido a la Pandemia COVID-19. Se ha trazado la media de cada estación para tener un indicativo de cómo se fue comportando la calidad del aire. En Guamaní se nota que el 01/01/2019 hay un pico debido a la quema de monigotes en fin de año; sin embargo, en los demás lugares se nota controlado el aspecto de la contaminación ambiental en esa fecha.

En todas las estaciones entre los meses de marzo y abril de 2019 se tiene la primera subida de contaminación progresiva, esto puede deberse a problemas por transporte urbano en el cual se movilizan los ciudadanos para realizar sus actividades diarias, como de igual manera al estar aún en periodo de año escolar se necesita mayor movimiento humano. Es bien sabido que el uso de transporte urbano tiene un gran impacto en la contaminación ambiental (Borck, 2019).

Al acabar las clases empezamos a ver un descenso en la contaminación en los meses de julio, agosto e inicios de septiembre, para proceder a tender a crecer nuevamente en contaminación durante septiembre y octubre que las industrias de útiles, materiales continúan trabajando para suplir la demanda debido al nuevo año escolar. En diciembre tenemos el mismo problema de las festividades que no ayudan a que la contaminación disminuya teniendo picos notables a fin de año como anteriormente se había mencionado. Finalmente, en este periodo se nota que existe ya una incidencia en la contaminación durante los meses de febrero y marzo que es cuando se escuchaba sobre la crisis de salud que se esparcía por el mundo.

Figura 4

IQGCDA - Todas las Estaciones – Antes de la Pandemia



Nota: Gráfica en series de Tiempo de IQGCDA, de todas las estaciones Antes de la Pandemia.

4.1.2 Índice de la calidad del aire durante la pandemia

En la Tabla 6 encontramos los valores representativos del IQGCDA en el Segundo Período (Durante la Pandemia), donde se nota un cambio en el comportamiento de los datos; siendo que en Cotocollao se encuentra el punto con menos contaminación esto indicado tanto por su media, como por su mediana de 47,60 y de 46,10 respectivamente. A su vez, en Guamaní presentó la peor contaminación con valores muy altos, en especial por el valor máximo que sobrepasa los 158,10. La estación del Centro puesto que ahora ya no es una de las de menor contaminación, sino que ha subido con relación a las demás

estaciones. El coeficiente de variación denota que ha descendido con relación al primer periodo con una variabilidad inferior al 30% con excepción de Guamaní que tiene muchos datos extremos que producen este cambio. Es necesario indicar que los valores tanto de media como de mediana decrecieron con respecto al periodo Antes de la Pandemia. Este es un primer indicador de que puede haber tendencias de disminución de la calidad del aire; sin embargo, se debe realizar alguna prueba estadística para validar lo observado (Angel y Casas, 2023).

Tabla 6

Valores Descriptivos IQGCDA - Durante la Pandemia

Estación	Mínimo	Máximo	Media	Mediana	SD	CV
Belisario	30,47	92,25	49,03	47,39	11,16	0,23
Carapungo	26,14	86,28	50,83	48,48	12,83	0,25
Centro	31,24	88,21	50,57	48,44	12,30	0,24
Cotocollao	20,94	78,97	47,59	46,14	11,31	0,24
Guamaní	29,20	158,15	57,42	55,30	17,64	0,31

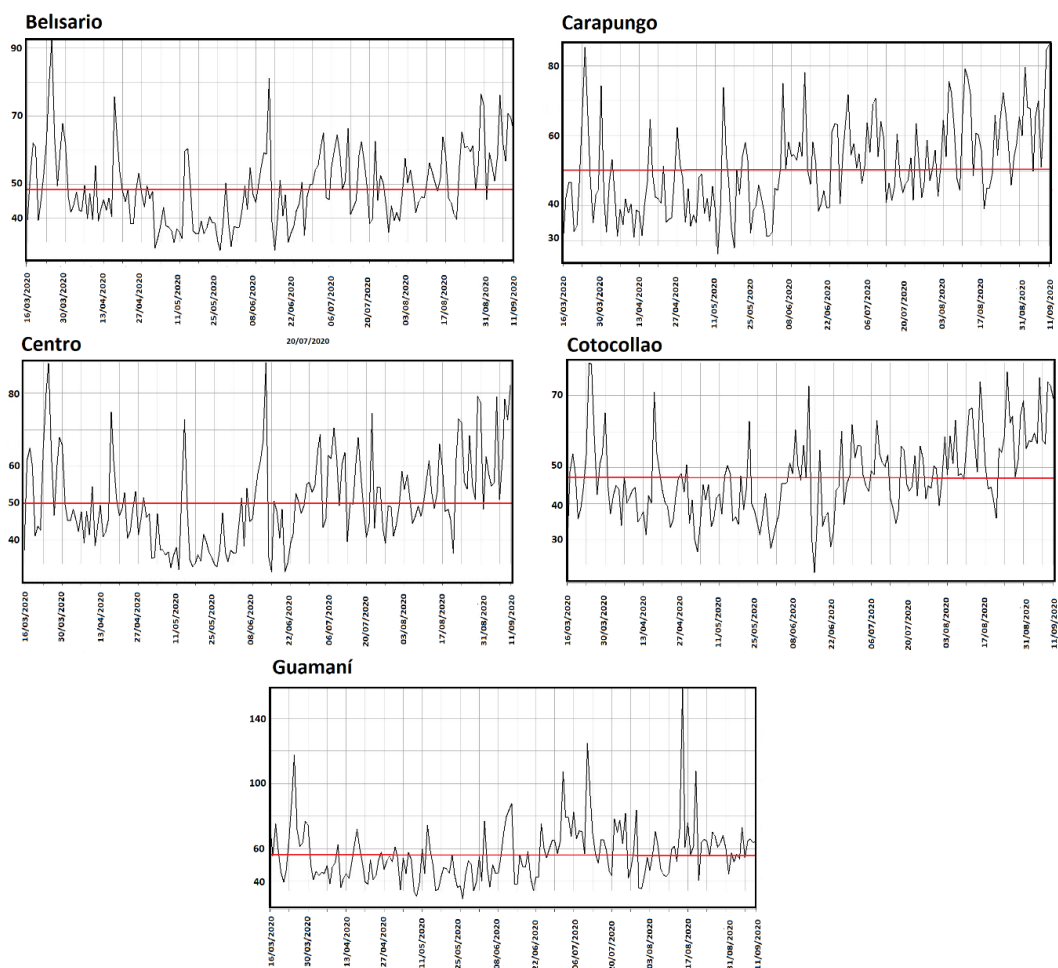
Nota: Valores descriptivos en el período comprendido entre 16/03/2020 - 11/09/2020.

En la Figura 5 se observa como la contaminación fue fluctuando, pudiendo destacar que la pandemia tiene un impacto al reducir los niveles de contaminación en el mes de abril y mayo, donde las paralizaciones fueron obligatorias en todo el territorio nacional (Zalakeviciute *et al.*, 2020). A pesar de ello luego de los primeros meses de pandemia; es decir, a partir del mes de junio la contaminación del aire vuelve a incrementarse pudiendo explicarse ya que al

empezar un periodo de prohibición de libre tránsito los mercados se vieron ahora ocupados por empresas que producen tecnología, alimentos, material del hogar y en especial desechos plásticos (Florian, 2021). A esto se suma el incremento en el número de las empresas que las distribuyen, pasando ahora a ser necesario el uso de motocicletas, que igual consumen restos fósiles para movilizarse produciendo contaminación del aire. Cabe añadir que en todas las estaciones se nota que hay incremento en el momento final de este periodo, cuando ya se estaba normalizando la situación y los estados de excepción llegaban a su fin, esto lo podemos equiparar tomando como base a la investigación realizada en Perú (Velayarce *et al.*, 2022).

Figura 5

IQGCDA - Todas las Estaciones – Durante la Pandemia.



Nota: Gráfica en series de Tiempo de IQGCDA, de todas las estaciones Durante la Pandemia.

4.1.3 Índice de la calidad del aire después la pandemia

La Tabla 7 muestra los principales indicadores del IQGCDA en el Tercer Período (Después de la Pandemia). En ella podemos ver que Guamaní sigue manteniéndose como el sector con mayor contaminación del aire, mostrando el valor máximo de todo el período con $258,40 \pm 23,65$; por esta razón, también su media y mediana es la máxima. La estación Centro juega un papel importante puesto que ocupa el segundo lugar más contaminado, donde la media y mediana es solo superada por Guamaní, lo cual es preocupante puesto que empezó siendo un lugar no tan contaminado a ser ya uno de los más contaminados. En las estaciones de Carapungo y Cotocollao se observó los valores mínimos siendo los lugares donde a pesar de volver a la normalidad tras la pandemia, se puede ver que mantienen niveles de contaminación controlados. La variación de los datos durante este periodo se ha visto afectada aumentando todos por encima del 30%. Es decir, los datos se alejan de la media volviéndose menos consistentes y que quizá deban ser transformados o estandarizados para tener un mejor desempeño en las pruebas a realizar.

En este período se visualiza que las medias y medianas se han elevado con respecto a lo visto en el segundo periodo. Sin embargo, como se ha indicado con anterioridad no se puede afirmar que estos tres periodos fueron como tal con tendencias de subida, luego bajada y finalmente subida respectivamente. Por tal razón más adelante se encuentran los datos obtenidos al realizar una prueba no paramétrica de Mann – Kendall sobre las tendencias.

Tabla 7*Valores Descriptivos IQGCDA - Después de la Pandemia*

	Mínimo	Máximo	Media	Mediana	SD	CV
Belisario	22,06	216,72	64,28	60,33	21,71	0,34
Carapungo	22,87	183,57	59,98	56,23	21,87	0,36
Centro	26,05	232,02	65,36	61,10	23,20	0,35
Cotacollao	21,11	203,00	59,59	56,32	20,47	0,34
Guamaní	23,17	258,41	71,21	67,38	23,65	0,33

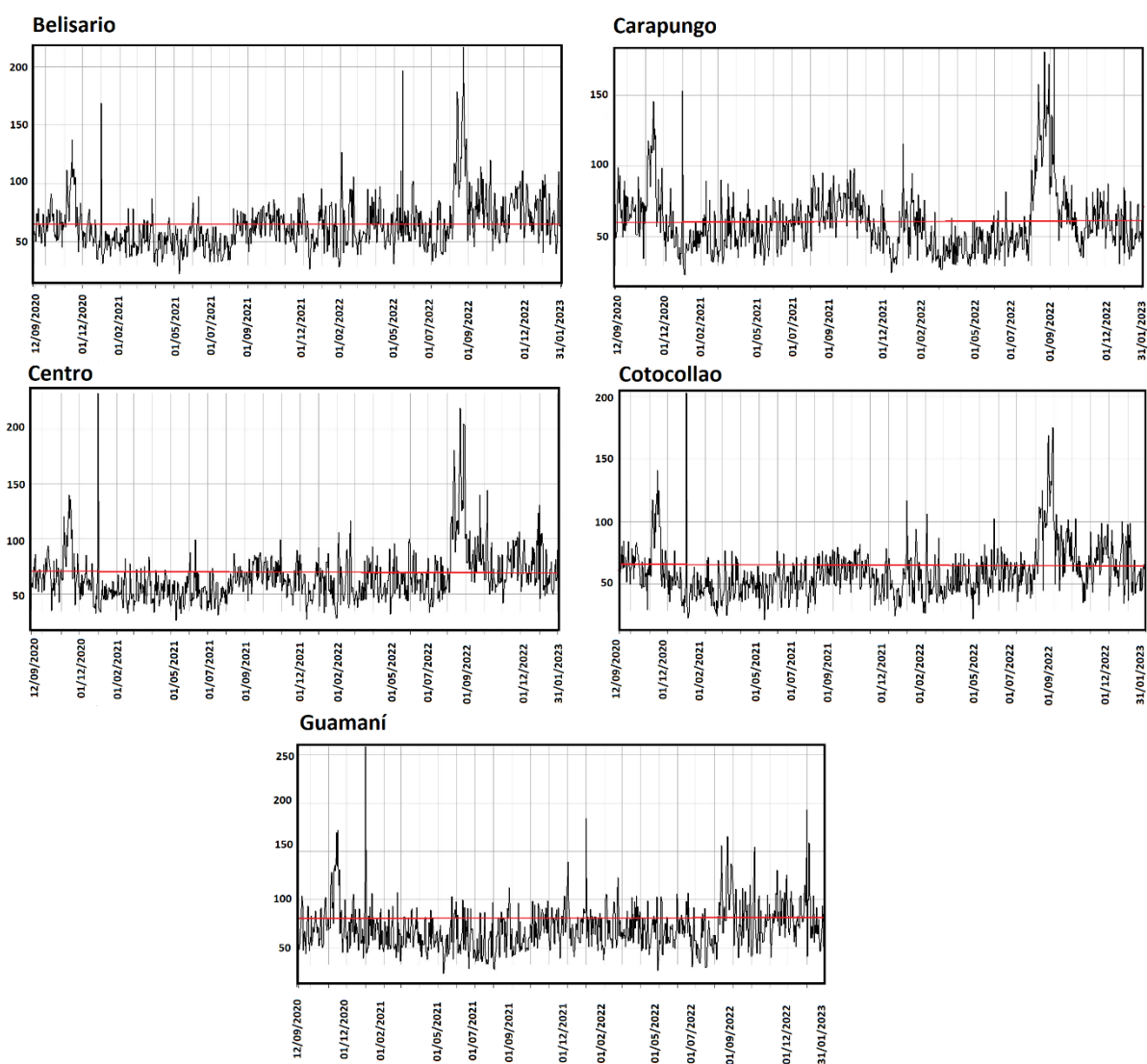
Nota: Valores descriptivos en el período comprendido entre 12/09/2020 – 31/01/2023.

De la Figura 6 se denota algunos aspectos, se visualiza una variabilidad muy fuerte las festividades de fin de año del 2020, quizá por las paralizaciones hubo un gran impacto para que ese año la contaminación se salga de control. Durante los siguientes meses de 2022 no se tiene mayor variabilidad, sin embargo, en todas las estaciones a excepción de Guamaní durante los meses de agosto y septiembre de 2022 se observa un incremento desmedido de contaminación que se puede explicar por el inicio de las clases del régimen sierra en Ecuador. Esto como ya se indicó produce un movimiento grande en empresas, agricultura, transporte que llevan a que la calidad del aire empeore. En adición a ello en la fecha primero de septiembre existe un feriado por la Independencia del País lo que produce movimiento por transporte mayoritario, es decir, gente que se moviliza a distintos lugares turísticos en la capital del Ecuador producen que se empeore y exista un pico de la contaminación ambiental. Es necesario resaltar que el año 2021 se presentó como el más estable en cuanto a contaminación

donde inclusive en festividades como fin de año no se elevaron los niveles de contaminación. Algunos feriados como el que se da el primero de mayo muestran una repercusión positiva, por lo que mejora la calidad del aire ya que el trabajo de empresas se ve detenido.

Figura 6

IQGCA - Todas las Estaciones – Después de la Pandemia



Nota: Gráfica en series de Tiempo de IQGCA, de todas las estaciones Después de la Pandemia.

4.1.4 Análisis de tendencia

Al realizar un primer análisis con medidas descriptivas y visualizando figuras de series de tiempo se pudo observar que las tres etapas previamente pensadas obtienen valores distintos de media. Por tal motivo procedió con el análisis de tendencia para observar si esto es evidencia suficiente para continuar con el análisis en estos tres períodos o si se debe realizar algún cambio. A continuación, se explican e indican los resultados obtenidos tras aplicar un análisis con la prueba no paramétrica de Mann-Kendall:

El análisis de tendencia que se llevó a cabo mediante la prueba de Mann - Kendall se muestra en la Tabla 8, donde se evidencia que algunas de las estaciones no muestran representatividad estadística altamente significativa ($p \leq 0,05$); esto sucede en Belisario, Centro, Guamaní en el periodo Antes de la Pandemia. Sin embargo, Carapungo y Cotocollao presentan una tendencia decreciente significativa, es decir, la calidad presentó mejoras antes de la pandemia. Para los períodos Durante y Después de la Pandemia se reflejan dentro de las estaciones una tendencia al incremento (deterioro de la calidad del aire), existiendo únicamente en Carapungo la tendencia de baja (Después de pandemia), aunque esta no es significativa. Es decir, estadísticamente la calidad del aire ha empeorado desde el inicio de la Pandemia COVID-19 en Quito-Ecuador.

Debido a la gran variedad de tendencias como de sus representatividades el primer periodo será tomado como un momento con tendencia mientras que el segundo y tercer periodo se los unen para formar un segundo momento con tendencia. Más adelante esta separación sirve para realizar las predicciones que se detallan en la sección 4.3.

Tabla 8*Prueba de Mann - Kendall para la Tendencia*

Periodo		Belisario	Carapungo	Centro	Cotocollao	Guamaní
Antes de pandemia	Tao	-0,0405	-0,0990	0,0272	-0,2752	0,0327
	P-val	0,2046	0,0019	0,3941	0,0000	0,3062
Durante la pandemia	Tao	0,1778	0,3656	0,2125	0,2967	0,1665
	P-val	0,0004	0,0000	0,0000	0,0000	0,0009
Después de pandemia	Tao	0,2138	-0,0267	0,1777	0,1574	0,1079
	P-Val	0,0000	0,2376	0,0000	0,0000	0,0000

Nota: Valores Tao y p-value para la prueba de Mann - Kendall en los tres períodos de estudio.

4.2 Contaminación del aire con respecto a la variabilidad climática

4.2.1 Análisis de la base de datos

Una vez se comprendió la distribución de los datos de IQGCDA alrededor del tiempo se procedió a realizar un análisis en conjunto con los datos de variación climática. Dentro de este análisis se realizaron los procedimientos detallados dentro del apartado 3.4.2 (Procedimiento). Se pueden visualizar los resultados tanto de Normalidad como de Simetría en el Anexo 4 y Anexo 5 respectivamente.

4.2.2 Gráficamente

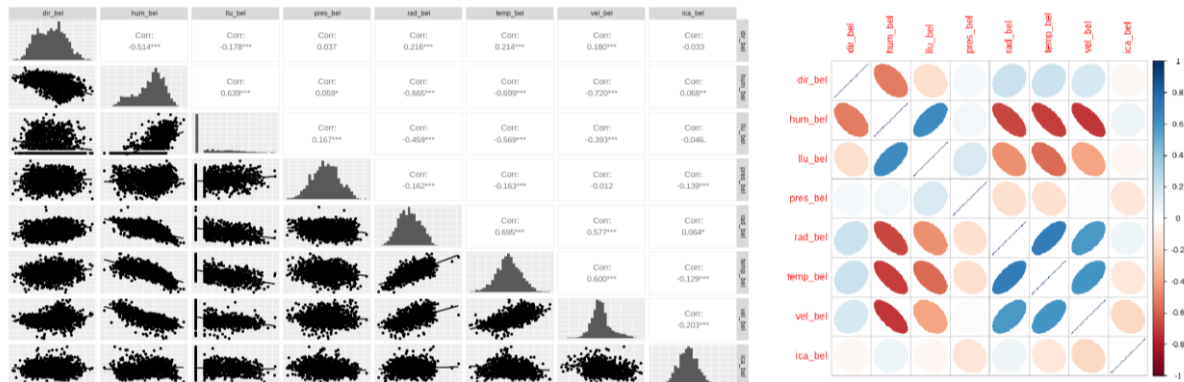
Dentro de este apartado se ilustran los resultados correspondientes al análisis de la relación que tuvieron las variables entre sí; además, de cada uno de los indicadores del clima con relación al Índice de Calidad de Aire. Como se indicó en la sección 3.4. (Procedimiento) se realiza primero un análisis gráfico que se describe a continuación por cada una de las estaciones.

4.2.2.1 Belisario

En la Figura 7, se observa algunas relaciones importantes como bien es la relación positiva fuerte que muestra tener la precipitación y la humedad; además, de la relación que existe entre la radiación, temperatura y velocidad del viento. Por otro lado, hay una relación negativa fuerte entre la humedad y la radiación, la temperatura y la velocidad del viento. Sin embargo, esta relación mostrada no evidencia que tengan colinealidad, esto se podrá analizar más adelante al momento de realizar las predicciones. Por parte del IQGCDA muestra leves relaciones, siendo las más importantes aquellas de la temperatura, la velocidad del viento y la presión atmosférica.

Figura 7

Diagramas Dispersión, Elipse e Índices de Correlación - Belisario



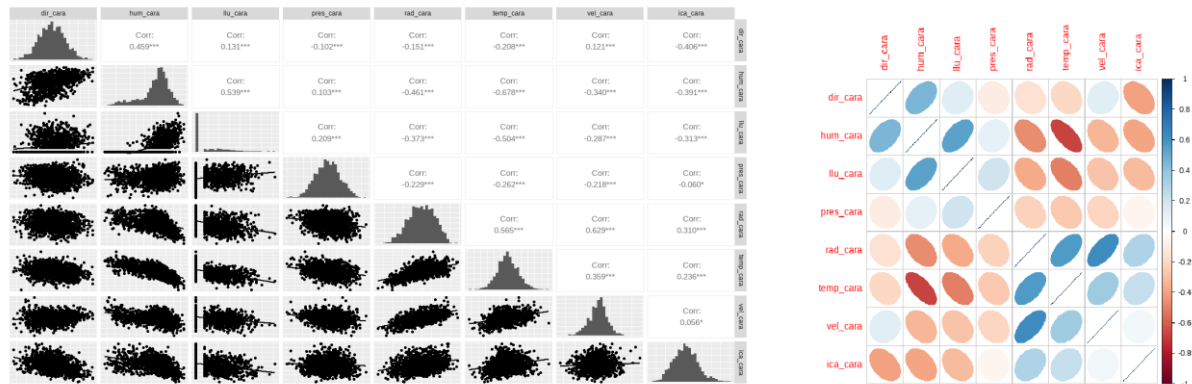
Nota: Visualización gráfica sobre la correlación que tienen las variables dentro de la estación Belisario.

4.2.2.2 Carapungo

En la Figura 8, se observa algunas relaciones siendo la más importante la relación positiva fuerte entre la radiación, la velocidad y la relación con la temperatura al igual que en el caso de la estación Belisario. Aquí se muestra una fuerte relación negativa entre la temperatura y la humedad del aire; además, de las relaciones negativas leves entre la radiación y la humedad como la lluvia con la temperatura. Por parte del IQGCDA muestra leves relaciones, siendo la más importante aquella con la dirección del viento, con la humedad, la radiación y la temperatura.

Figura 8

Diagramas Dispersión, Elipse e Índices de Correlación - Carapungo



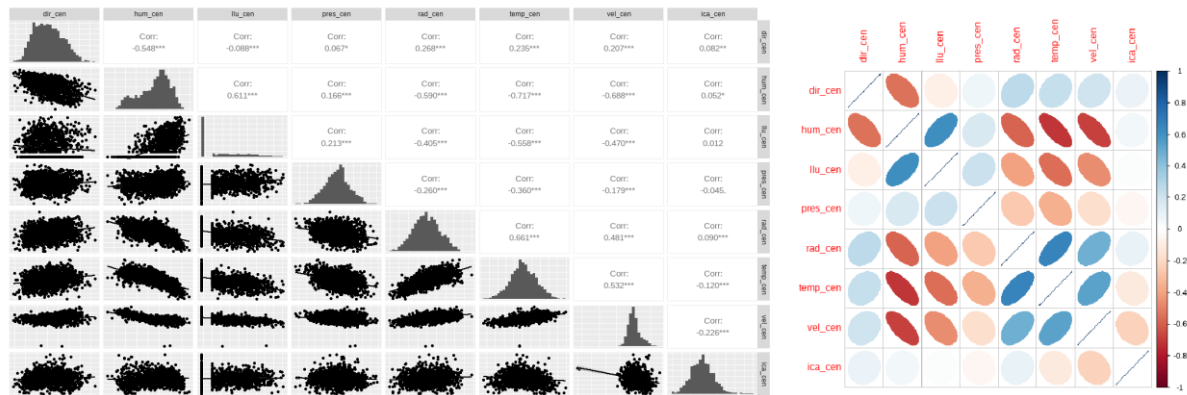
Nota: Visualización gráfica sobre la correlación que tienen las variables dentro de la estación Carapungo.

4.2.2.3 Centro

En la estación Centro se pudo observar algunas relaciones importantes, como la relación positiva fuerte entre la radiación, la velocidad y la radiación con la temperatura como en el caso de las anteriores estaciones; además, se une una relación positiva fuerte entre la lluvia y la humedad del aire. En las relaciones fuertes negativas tenemos aquella entre la humedad con relación a la temperatura, velocidad del aire, humedad y la radiación solar. Por parte del IQGCDA muestra leves relaciones, con las variables del clima, siendo aquella con la velocidad del viento la más predominante (ver Figura 9).

Figura 9

Diagramas Dispersión, Elipse e Índices de Correlación - Centro



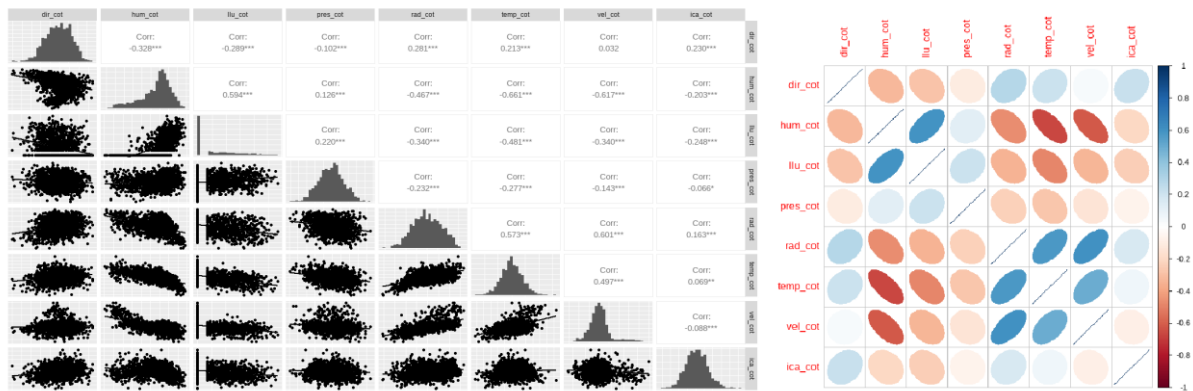
Nota: Visualización gráfica sobre la correlación que tienen las variables dentro de la estación Centro.

4.2.2.4 Cotocollao

En la estación Cotocollao se hallan algunas relaciones a resaltar, por ejemplo, como relaciones positivas fuertes tenemos aquella entre la lluvia y la humedad, como a su vez la relación entre la radiación con la temperatura y la velocidad. En las relaciones negativas fuerte se encuentra las existentes entre la humedad en relación con la radiación, temperatura y velocidad. Hablando sobre el IQGCDA se visualiza que hay ligeras relaciones con la dirección del viento, la humedad y las precipitaciones mostradas en todo el tiempo (ver Figura 10).

Figura 10

Diagramas Dispersión, Elipse e Índices de Correlación - Cotocollao



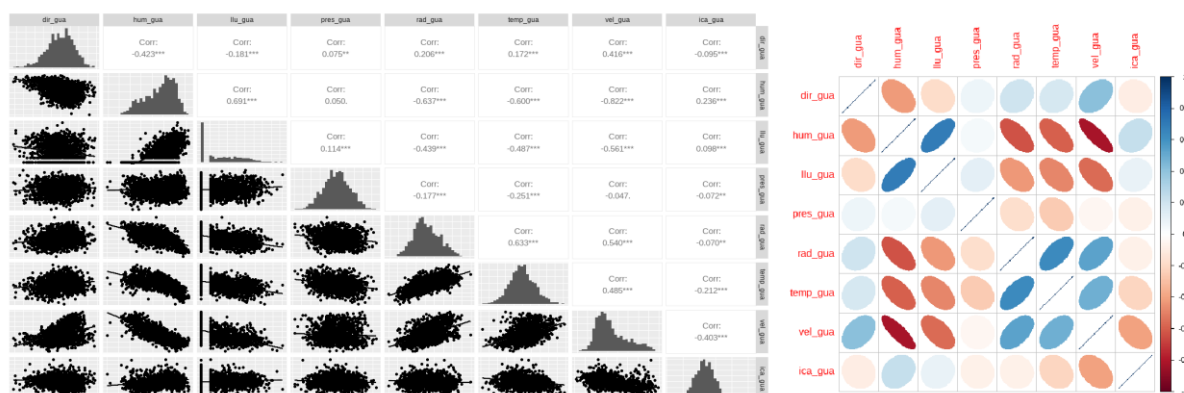
Nota: Visualización gráfica sobre la correlación que tienen las variables dentro de la estación Cotocollao.

4.2.2.5 Guamaní

Dentro de la estación Guamaní hay una fuerte relación negativa entre la velocidad del viento y la humedad ambiental, además de otras relaciones con respecto a la temperatura y la radiación. En el ámbito de relaciones positivas fuertes destaca aquella entre la lluvia y la humedad, como la existente entre la radiación y la temperatura, además de la radiación con la velocidad. Por parte del IQGCDA se puede ver que tiene una relación negativa comparándola con la velocidad del viento, como con la temperatura, y una relación positiva con la humedad del ambiente (Figura 11).

Figura 11

Diagramas Dispersión, Elipse e Índices de Correlación - Guamaní



Nota: Visualización gráfica sobre la correlación que tienen las variables dentro de la estación Guamaní.

4.2.3 Correlación de Pearson y Spearman entre calidad de Aire y Clima

Tras revisar la normalidad en la distribución de las variables, haber constatado la simetría antes y luego de la transformación (ver Anexo 4, Anexo 5) se procede con la elección del índice indicado para medir la correlación. Para aquellas variables que sean ambas normales se procede con la obtención del índice de Correlación de Pearson. Sin embargo, para las restantes así sea que solo una variable no es normal se procede con el Índice de Correlación de Spearman mismo que es más robusto frente a la falta de normalidad (Ramírez, 2017). En este apartado se centró directamente en la relación de cada una de las variables meteorológicas con relación al IQGCDA.

En Tabla 9, se visualiza que en la estación Belisario las correlaciones más importantes entre el IQGCDA al analizar el p-value ($p \leq 0,001$) se dan con la presión ambiente, la temperatura y la velocidad del viento. En la estación de Carapungo hay correlación significativa con la humedad relativa, la precipitación, la radiación y la temperatura. Siendo negativa en las tres primeras variables señaladas y positiva con las dos últimas variables. En el Centro hay correlación en los casos de la dirección del viento, radiación, temperatura y velocidad de viento, donde con relación a la dirección y radiación solar es una correlación positiva. En Cotocollao la correlación positiva se da con la dirección del viento y

radiación solar, mientras que la correlación negativa se da con la humedad relativa y la precipitación. Finalmente, en Guamaní solo no hay correlación significativa con la presión del ambiente, siendo las positivas las de humedad relativa, precipitación, mientras que con las demás hay una correlación negativa. En todos los casos que se observa correlación negativa, se puede entender que, al incrementar cualquier variable meteorológica antes mencionadas, tenderá a bajar el índice de calidad de aire, por lo que la contaminación reducirá su impacto, mientras que sí la correlación es positiva será lo contrario (Picardo, 2022).

Resumiendo, se puede entender que en algunas estaciones la correlación varía, esto puede deberse a la ubicación de la estación a lo largo de Quito, pero en las variables: radiación solar y la velocidad del viento se mantiene constante, e indicando que la correlación con la primera variable mencionada es positiva mientras que en la segunda variable existe una relación negativa. Esto se puede explicar incluso físicamente hablando ya que a mayor velocidad va a haber una mejor dispersión de los contaminantes, haciendo que se mejore la calidad del aire, mientras que en lugares estáticos o que no corre viento con velocidad baja, los contaminantes pueden llegar a estacionarse por lo que la contaminación del aire aumenta. En el caso de la radiación solar se debe a que al existir una incidencia directa de ella se producen reacciones fotoquímicas que dan lugar a la formación de ozono troposférico (O_3) y otros contaminantes secundarios, que contribuyen a una mala calidad del aire y a valores más elevados del índice de calidad del aire. Además, de ello se puede tener inversiones de temperatura que producen un mal flujo de contaminantes por lo que la calidad del aire empeorará.

Tabla 9

Prueba de Correlación de Pearson y Spearman

Variables		Belisario	Carapungo	Centro	Cotocollao	Guamaní
Dirección del viento-IQGDCA	rho	-0,0462	-0,4060	0,0975	0,2301	-0,1174
	P-val	0,0741	0,0000	0,0002	0,0000	0,0000
Humedad Relativa-IQGDCA	rho	0,0683	-0,4501	0,0363	-0,2757	0,2416
	P-val	0,0083	0,0000	0,1616	0,0000	0,0000
Precipitación-IQGDCA	rho	-0,0113	-0,3072	0,0227	-0,2426	0,1212
	P-val	0,6633	0,0000	0,3815	0,0000	0,0000
Presión-IQGDCA	rho	-0,1390	-0,0597	-0,0451	-0,0661	-0,0716
	P-val	0,0000	0,0211	0,0819	0,0106	0,0057
Radiación-IQGDCA	rho	0,0643	0,3100	0,0901	0,1630	-0,0729
	P-val	0,0130	0,0000	0,0005	0,0000	0,0048
Temperatura-IQGDCA	rho	-0,1294	0,2365	-0,1196	0,0692	-0,2121
	P-val	0,0000	0,0000	0,0000	0,0075	0,0000

Velocidad del viento-IQGDCA	rho	-0,1811	0,0354	-0,2292	-0,0649	-0,4129
	P-val	0,0000	0,1723	0,0000	0,0122	0,0000

Nota: rho y p-val en prueba de Pearson (sin resaltar) y prueba de Spearman (resaltado) entre las variables climatológicas y el IQGCDA.

4.2.4 Teoría Cópula

En la Tabla 10 se evidencia que los valores observados muestran una consistencia con las pruebas de Pearson y de Spearman anteriormente revisadas, por lo cual se vuelven más robustas pudiendo determinar que los casos donde varía la correlaciones se pueden deber a su ubicación o algún factor externo que no se tomó en cuenta dentro de este estudio. Por otro lado, en las correlaciones con las variables radiación solar y velocidad del viento se mantienen teniendo concordancia con lo mencionado previamente en el estudio. Para mayor detalle sobre este apartado se encuentran en el Anexo 6.

Tabla 10

Teoría Cópula – Relación entre Clima y IQGCDA

Variables		Belisario	Carapungo	Centro	Cotocollao	Guamaní
Dirección del viento-IQGDCA	Rho	-0,0383	-0,2733	0,0083	0,1549	-0,0677
	P-val	0,0901	0,0000	0,0001	0,0000	0,0000
Humedad Relativa-IQGDCA	rho	0,0844	-0,3148	0,0910	-0,1980	0,1895
	P-val	0,0087	0,0000	0,1949	0,0000	0,0000
Precipitación-IQGDCA	rho	-0,0384	-0,2074	0,0191	-0,0570	0,1257
	P-val	0,6281	0,0000	0,3458	0,0000	0,0000

Presión- IQGDCA	rho	-0,0898	-0,0535	-0,0396	-0,0570	-0,0560
	P-val	0,0000	0,0484	0,0244	0,0146	0,0075
Radiación- IQGDCA	rho	0,0387	0,1945	0,0577	0,0941	-0,0676
	P-val	0,0526	0,0000	0,0008	0,0000	0,0051
Temperatura- IQGDCA	rho	-0,1137	0,1685	-0,1064	0,0725	-0,1263
	P-val	0,0000	0,0000	0,0000	0,0193	0,0000
Velocidad del viento- IQGDCA	rho	-0,1258	0,0322	-0,1568	-0,0558	-0,2916
	P-val	0,0000	0,1743	0,0000	0,0136	0,0000

Nota: Valores para correlación en base a la teoría Cúpula para observar el comportamiento entre las variables climatológicas y el IQGDCA.

4.3 Modelos estadísticos

En la Tabla 8, se evidenció que las estaciones de Carapungo y Cotocollao tienen una tendencia negativa estadísticamente significativa, por lo cual se toma a este como el Primer Momento con tendencia. Para los siguientes dos periodos se muestra que existe una tendencia positiva en ambos casos para la mayoría de las estaciones por lo que se ocupa a estos dos periodos como el Segundo Momento con tendencia. Por lo tanto, los nuevos momentos quedaron establecidos de la siguiente manera:

- El primer Momento con Tendencia (Tendencia Negativa) es entre las fechas 01/01/2019 a 15/03/2020, que fue donde empezaron las paralizaciones por el COVID-19, es decir, durante este tiempo la calidad de aire estaba mejorando en algunas estaciones del estudio. Dentro de este primer momento se hace la separación tanto de la base de datos para calibración (80%) que va desde el 01/01/2019 al 18/12/2019 y como para

la validación (20%) desde 19/12/2019 hasta el fin del momento el 15/03/2020.

- A partir del 16/03/2020 hasta el 31/01/2023 se convierte en el Segundo Momento con Tendencia (Tendencia Positiva) que es donde muestra un aumento en la mayoría de las estaciones, es decir se tiene una peor calidad del aire. Dentro de este segundo momento se hace la separación de la base siendo que la calibración (80%) parte desde el 16/03/2020 al 05/07/2022 y la validación (20%) va desde 06/07/2022 hasta el fin del momento el 31/01/2023.

Una vez definidos los dos momentos con tendencia se proceda unificar las variables de cada estación a fin de encontrar por medio de la media un valor indicativo que sirve como referencia a todo el Distrito Metropolitano de Quito.

A continuación, se indica en la Tabla 11 y Tabla 12 los resultados tanto de la modelación por medio de una Regresión Lineal Múltiple como por medio de análisis de series de tiempo con ARIMAX.

4.3.1. Modelo de Regresión Lineal Múltiple (RLM)

En la Tabla 11, se ilustran los valores de cada uno de los coeficientes de la Regresión Lineal Múltiple relacionados a cada variable climatológica, como a su vez los puntos de intercepción con el eje Y. Tras realizar un planteamiento jerárquico no se llega a tener una significancia en todos los coeficientes por lo cual se realiza una selección de los mejores predictores por medio del modelo Akaike (AIC) que toma en cuenta tanto la bondad del ajuste como la complejidad del modelo, penalizando los modelos con más parámetros (Cavanaugh y Neath, 2019). En el Primer Momento con Tendencia se tiene coeficientes que son significativos ($p < 0,05$) en todas las variables a excepción de la humedad relativa, la precipitación y la presión atmosférica. Sin embargo, al comparar con otros modelos, la precipitación es importante para ayudar a encajar el modelo con los valores observados, por lo cual se toma su coeficiente dentro de la función a utilizar.

De igual manera se nota en la Tabla 11, los coeficientes del segundo momento con tendencia aplicando la Regresión Lineal Múltiple (RLM), en ella se observa

que ganan significancia las anteriores variables mencionadas, perdiendo únicamente significancia en la dirección de viento, pero que ayuda a mejorar el modelo planteado. Cabe indicar que al final de realizar la modelización se revisaron supuestos con los residuos para analizar qué tan confiable es nuestro modelo, mismos que los pueden revisar en el Anexo 7.

Tabla 11

Valores Regresión Lineal Múltiple

Variables	Regresión Lineal 1^{er} Momento Tendencia	Regresión Lineal 2^{do} Momento Tendencia
Intercepto	4,22790 *** (0,00000)	4,03963*** (0,00000)
Dirección Viento	0,11178 ** (0,00143)	0,03421 (0,146043)
Humedad Relativa		-0,07836*** (0,00019)
Precipitación	- 0,04791. (0,09214)	-0,09131 *** (0,00000)
Presión Atmosférica		-0,04503*** (0,00002)
Radiación Solar	0,16172 *** (0,00000)	0,20430*** (0,000000)
Temperatura	- 0,08565 ** (0,00111)	-0,14620*** (0,00000)
Velocidad Viento	-0,16436*** (0,00000)	-0,15931*** (0,00000)
AIC	-1722,77	-2598,49
Residual Estándar	0,216	0,2127
Error		
Adjusted R-Squared	0,216	0,2289

Código Significancia: 0 (***) 0,001 (**) 0,01 (*) 0,05 (.) 0,1 () 1

Nota: Valores obtenidos para la Regresión Lineal Múltiple para los dos momentos con tendencia.

4.3.2. Modelo ARIMAX

El modelamiento mediante series de tiempo se lleva a cabo mediante el modelo ARIMAX, que es autorregresivo como el modelo ARIMA, pero que toma en cuenta las variables extra (el clima) como variables exógenas o explicativas que ayudan a mejorar el sistema. El software R se encarga de realizar los cálculos de los coeficientes del Modelo ARIMAX; sin embargo, para indicar cuales variables fueron seleccionadas se toma en cuenta aquellas que fueron significativas en el modelo de Regresión Lineal Múltiple anteriormente realizada. Los coeficientes mostrados como “ar” son los valores que representan una regresión de la variable contra valores pasados de ella misma, es decir nos ayudan a conocer la parte autorregresiva del modelo. De igual manera los valores “ma” representan la media móvil ponderado de los errores pasados del pronóstico (Kongcharoen y Kruangpradit, 2013). Existe variación como en el modelo de Regresión Lineal Múltiple sobre cuáles fueron las variables que nos ayudan a predecir de una mejor manera y lo puede revisar en la Tabla 12. Para conocer acerca del supuesto de independencia se debe revisar el Anexo 7, siendo este el supuesto más importante dentro de los modelos de series de tiempo y que para este estudio se termina por aceptar los supuestos de independencia, además de visualizar valores de pruebas de normalidad también realizados. El primer momento de tendencia tuvo valores (3, 0, 2) correspondientes a (p = parte autorregresiva, d = grado de diferenciación y q = parte media móvil), mientras que en el segundo momento de tendencia lo que se obtiene de manera automática fue que lo mejor para realizar el modelamiento era que estos valores variarían a (2, 1, 2).

Tabla 12

Valores ARIMAX

Variabes	1^{er} Momento Tendencia - ARIMAX (3, 0, 2)	2^{do} Momento Tendencia - ARIMAX (2, 1, 2)
ar1	2,3021 . (0,0697)	- 0,2505 (0,1436)
ar2	-1,8325 (0,1089)	0,3069 . (0,0866)

ar3	0,5199* (0,0494)	
ma1	-1,848 . (0,0667)	-0,1819 (0,1350)
ma2	0,9005 . (0,0668)	-0,7020 (0,1287)
Intercepto	4,2584** (0,00453)	
Dirección Viento	0,0946* (0,0289)	0,0386* (0,0184)
Humedad Relativa		0,0084* (0,0235)
Precipitación	- 0,1038** (0,0232)	(0,1124* (0,0143)
Presión Atmosférica		-0,0060* (0,0143)
Radiación Solar	0,1291* (0,0239)	0,1243* (0,0150)
Temperatura	- 0,0867* (0,0267)	-0,0502* (0,0208)
Velocidad Viento	- 0,1768* (0,0209)	-0,1329* (0,0141)
AIC	-218.91	-621,66

Código Significancia: 0 (***) 0,001 (**) 0,01 (*) 0,05 (.) 0,1 () 1

Nota: Valores obtenidos para el modelamiento ARIMAX para los dos momentos con tendencia.

4.3.3. Validación

La selección de cuál fue el mejor modelamiento posible con las variables climatológicas se realiza mediante la comparación de valores como RMSE, MSE, MAE y R-cuadrado previamente indicados en la sección 3.4 (Procedimiento), cabe indicar que se obtuvo estos valores tanto para la base de datos planteada para la calibración (80% de datos) como para la base de datos de validación (20% de datos). De los resultados obtenidos en la fase de calibración (Tabla 13), el modelamiento del IQGCDA por Regresión Lineal Múltiple presenta los valores

más bajos de RMSE, menor diferencia cuadrática promedio, menor distancia promedio entre los datos reales y los datos pronosticados y relación entre el sesgo y varianza en la construcción de los dos modelos (Tabla 11 vs Tabla 12). Esto sucede tanto para el Primer y Segundo Momento con Tendencia, se puede tener en cuenta la relación entre el sesgo y varianza en la construcción de los dos modelos, en donde la Regresión Lineal Múltiple tiene el mínimo AIC en ambos momentos de tendencia (Tabla 11 vs Tabla 12).

Tabla 13

Validación de la Regresión Lineal Múltiple y ARIMAX - Primer y Segundo Momento Tendencia, Etapa de calibración.

Calibración del Modelo	1 ^{er} Momento de Tendencia		2 ^{do} Momento de Tendencia	
	Regresión	Arimax	Regresión	Arimax
	Lineal Múltiple		Lineal Múltiple	
RMSE	14,25	14,79	13,76	16,61
MSE	203,17	218,85	189,27	275,91
MAE	10,84	11,66	9,42	11,35
R SQUARED	0,20	0,19	0,19	0,15

Nota: Valores Validación del modelo por medio de RMSE, MSE, MAE y R-cuadrado, etapa Calibración.

El IQGCDA simulado para la etapa de validación revela un mejor ajuste para la etapa de validación tanto para el Primer y Segundo Momento con Tendencia, sin embargo, el R-cuadrado en el Primer Momento con Tendencia es mejor para el primer modelo ARIMAX (Tabla 14), mostrando un mejor ajuste entre los datos pronosticados y los observados. De igual manera si se considera la relación entre el sesgo y varianza en la construcción de los dos modelos, la Regresión Lineal Múltiple tiene el mínimo AIC (Tabla 11 vs Tabla 12).

Tabla 14

Evaluación del modelo de Regresión Lineal Múltiple y ARIMAX - Primer y Segundo Momento Tendencia, Etapa validación.

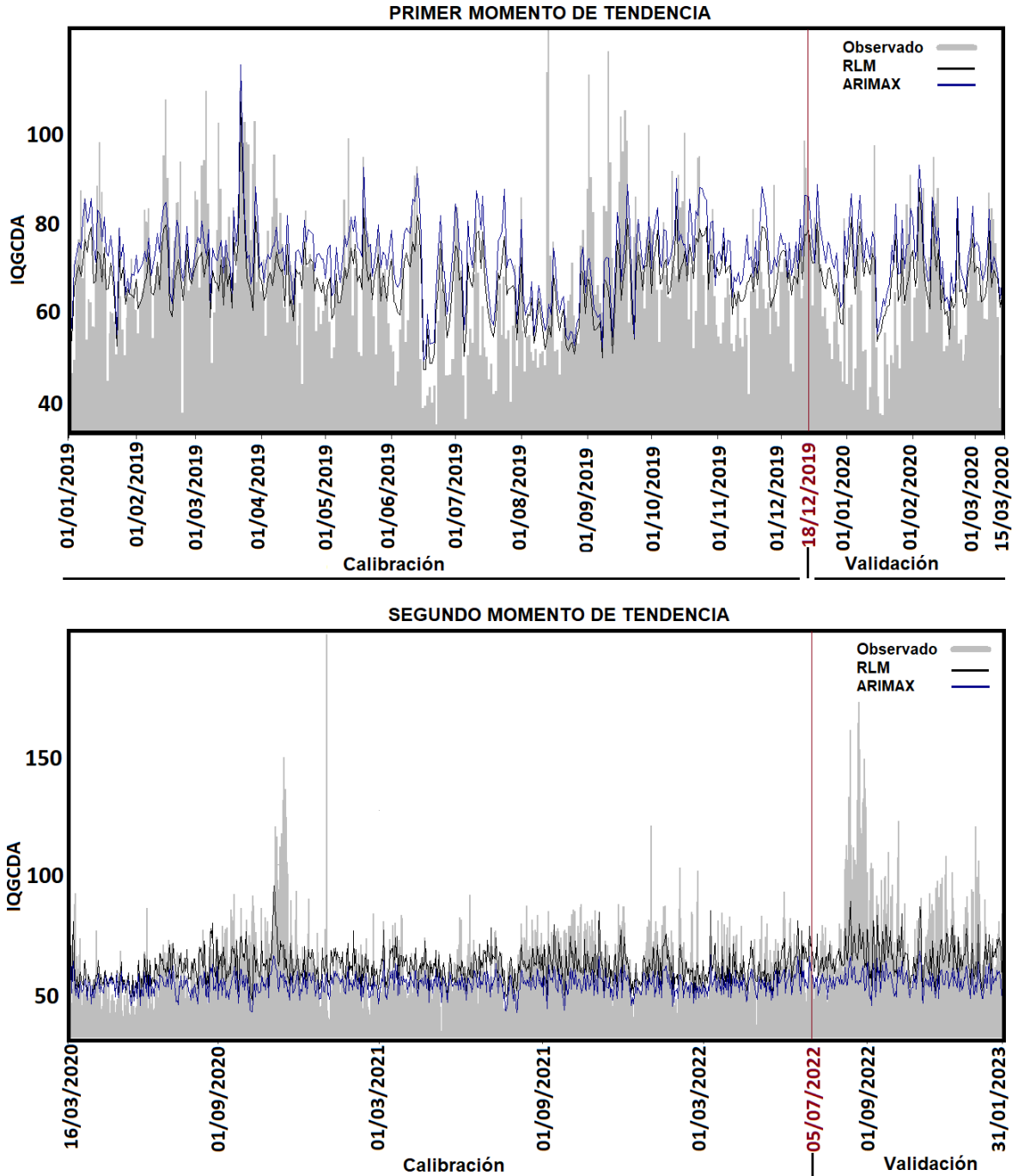
Validación del Modelo	1 ^{er} Momento de Tendencia		2 ^{do} Momento de Tendencia	
	Regresión		Regresión	
	Lineal Múltiple	Arimax	Lineal Múltiple	Arimax
RMSE	12,85	15,02	26,46	34,07
MSE	165,02	225,51	700,25	1160,69
MAE	10,61	12,62	18,55	25,57
R SQUARED	0,33	0,38	0,16	0,08

Nota: Valores Validación del modelo por medio de RMSE, MSE, MAE y R-cuadrado, etapa Validación.

En la Figura 12 se evidencia que el IQGCDA simulado por Regresión Lineal Múltiple para el primer momento de tendencia, fase de calibración y validación revelan un bajo ajuste, porque no logra representar la distribución diaria de la contaminación en el municipio de Quito. Es decir, sobrestima la dinámica de los días con menor concentración, y sobrestima los días de mayor concentración del IQGCDA. El modelo ARIMAX, también presenta un bajo ajuste, con una distribución temporal tanto en calibración como validación, sobrestimando los valores mínimos y subestimando los valores máximos diarios. Al comparar los dos modelos en los dos momentos de tendencia en la Figura 12, se evidencia que los dos modelos no logran representar la dinámica de la contaminación, aunque el modelo ARIMAX presenta un poco más de dificultad para interpretar el IQGCDA en la ciudad de Quito, resultados similares fueron encontrados en un estudio de serie de tiempos de PM10 (Zafra *et al.*, 2019). Queda a juicio del investigador el inclinarse por uno u otro modelo, en próximos estudios se podría probar con Machine Learning ya que ha tenido un gran impacto tanto en mejorar los modelos como en predecir los datos (Lei *et al.*, 2022; Liu *et al.*, 2022). En este estudio únicamente se muestra la modelación con los dos procedimientos previamente indicados; además, de que se recuerda que la variabilidad climática es causada principalmente por la temperatura y la precipitación (Useros, 2012).

Figura 12

Simulación de los modelos de RLM y ARIMAX en comparación con lo observado



Nota: Gráficas de lo simulado versus lo observado en cada uno de los modelos de acuerdo con el Primer y Segundo Momento con Tendencia.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

El estudio realizado en este informe de investigación lleva a obtener algunos resultados importantes de los cuales podemos destacar los siguientes:

1. Los valores de concentración de distintos contaminantes del aire como el Monóxido de Carbono (CO), Partículas ($PM_{2.5}$), Dióxido de Azufre (SO_2), Dióxido de Nitrógeno (NO_2) y Ozono (O_3) se analizaron e imputaron para datos faltante menores al 8%. Con estos valores se obtuvo el Índice Quiteño General obtenido para ir en concordancia con lo mencionado por la Secretaría de Ambiente de Quito. Esto sirve para un análisis antes, durante y post Pandemia COVID-19. La estación del Centro parte siendo una de las menos contaminadas ($62,56 \pm 20,40$), y termina siendo una de las más contaminadas ($65,36 \pm 23,20$) solo superada por la estación Guamaní ($71,21 \pm 23,70$). De igual manera se pudo visualizar que Carapungo ($59,98 \pm 21,90$) y Cotocollao ($59,59 \pm 20,50$) fueron de las menos contaminadas en todo el tiempo. Cabe destacar que hubo días y etapas específicas donde se notaba una mayor contaminación como en las festividades de Fin de Año o por las clases de la región Sierra. En el periodo Durante la Pandemia si bien se ve un descenso de la contaminación al inicio de las paralizaciones, se tuvo un aumento posterior por el auge de empresas de plásticos, transporte, materiales del hogar, entre otros. Para el periodo Después de la Pandemia en la mayoría de las estaciones se nota aumento de valores del Índice Quiteño General de Calidad Diaria de Aire.

2. El análisis de Tendencia del Índice Quiteño General de Calidad Diaria de Aire por la prueba no paramétrica de Mann – Kendall evidencia únicamente dos momentos destacables de tendencia, uno que es el Primer Momento con Tendencia (Negativa) que va desde 01/01/2019 hasta el 15/03/2020 y un Segundo Momento con Tendencia (Positiva) que va desde 16/03/2020 al 31/01/2023. Esto se puede destacar haciendo referencia que antes de la pandemia ya se habían estado tomando medidas preventivas dentro del Distrito Metropolitano; es decir, se buscaba mejorar la calidad del aire (Primer Momento con Tendencia). Pero posteriormente debido a la pandemia se produjo aumento de empresas que empezaron a contaminar el aire de manera descontrolada, siendo que se disminuye el transporte público por las paralizaciones, pero se incrementa el transporte por medio de motos, quienes ahora estaban activos para hacer la repartición de alimentos, productos de hogar, medicina, entre otros. A más de ello, las empresas de plásticos se convierten en beneficiarios económicamente hablando, pero traen a la luz la creciente contaminación por su producción con quema de restos fósiles y su posterior distribución (Segundo Momento con Tendencia).
3. El análisis de Correlación de Pearson y de Spearman, evidencia una relación estadísticamente significativa positiva entre el IQGCDA con la radiación solar y una correlación negativa con la velocidad de viento. Quedando claro que con un control y predicción de estas variables se puede llegar a simular la contaminación del aire. En concordancia se recuerda la relación que mantienen estas con las variables de la precipitación y temperatura que por lo general suelen definir la variabilidad climática. La relación entre estas variables se pudo explicar físicamente puesto que una mayor velocidad del viento produce mejor dispersión de contaminantes por lo que se tiene una mejor calidad del aire. Mientras que una mayor radiación solar produce reacciones fotoquímicas con los contaminantes del aire, además de que produce un aumento de temperatura causando a su vez inversiones térmicas que producen cambios en la calidad del aire en la ciudad. Para las otras variables hay discrepancia en los valores y varían dependiendo de la estación, esto

puede explicarse por la ubicación de cada uno de los sensores o por alguna variable externa que no se toma en cuenta dentro de cada estación. Los resultados fueron comparados y afirmados con un análisis por medio de la Teoría Cópula, que evidencia que existe relación evidente entre las variables meteorológicas y la calidad del aire.

4. Finalmente, para los dos Momentos marcados por parte de la prueba de Tendencias, se realizaron modelaciones del IQGCDA mediante los métodos estadísticos de Regresión Lineal Múltiple y el método de tratamiento de series temporales ARIMAX que toma en cuenta variables exógenas. La validación de los modelos evidencia que la Regresión Lineal Múltiple para los momentos antes (Primer Momento de Tendencia), durante y después de la pandemia (Segundo Momento de Tendencia) presentan los menores valores de RMSE, MSE, MAE, AIC, aunque con un ajuste de R-cuadrado relativamente bajo (menor al 33%) es decir nos ayuda a predecir la calidad del aire, pero al momento de mirar su ajuste aún es pequeño. Únicamente por parte del modelo ARIMAX se tiene un ajuste mayor (38%), sin embargo, los demás indicadores para aceptar el modelo son altos por lo cual no se llega a hacerlo por lo que el modelamiento de Regresión Lineal Múltiple si bien no es específicamente un modelo para predicción posterior, ayuda a tener una idea de cómo se van a distribuir la contaminación del aire con relación a las variables climatológicas. Hay que tener en cuenta que para predicciones o pronósticos de incidencia climática se necesita por lo menos un listado de datos de 10 años o un período de tiempo más extenso, para justificar la variabilidad climática, como nos lo mencionó la Organización Meteorológica Mundial (OMM) en 2018, por lo que en este trabajo se realizó únicamente una comparación de causa y efecto de las variables implícitas.

Recomendaciones

Una de las principales limitaciones de este estudio es la falta de datos en tiempo real sobre la calidad del aire en Quito. Los datos de calidad del aire utilizados en este estudio se obtuvieron de estaciones de control, que no proporcionan datos en tiempo real y que algunos no son tomados directamente del ambiente. Esto puede limitar la capacidad del estudio para captar los cambios a corto plazo en la calidad del aire, por lo que se recomienda consultar sobre posibles mejoras en la obtención de datos como se ha venido desarrollando en otros países.

Es necesario que se busque a su vez bases de datos que de igual manera no tengan valores perdidos para de esta manera reducir el sesgo que pudiera surgir al imputar los datos. Por lo que además se recomienda trabajar con otras fuentes potenciales de contaminación atmosférica, como la quema de biomasa y la contaminación del aire en interiores. Estas fuentes pueden contribuir a la contaminación atmosférica en Quito, pero es posible que su contribución no quede reflejada en los datos utilizados en este estudio.

Además, se recomendaría el estudio en un conjunto más grande de estaciones ya que en este estudio únicamente se han utilizado solo 5 estaciones. El tener mayor cantidad de estaciones puede brindar un soporte mejor al estudio para abarcar mayoritariamente el Distrito Metropolitano de Quito y así al momento de realizar las predicciones se puede buscar nuevos métodos que ayuden a incluir más información como bien puede ser el uso de modelos econométricos espaciales que aparte de estudiar a las variables propiamente dichas, estudia como la distancia que tienen entre estaciones afecta la dispersión que pueden tener los modelos. De ser posible también se recomienda recurrir e investigar métodos de Machine Learning que son más robustos frente a falta de normalidad y que pueden ayudar a tener predicciones más exactas del índice de calidad del aire, siendo que tomarán mayor énfasis en detalles que quizá los métodos estadísticos estén ignorando.

REFERENCIAS

- Abdi, H., y Molin, P. (2007). *Lilliefors/Van Soest's test of normality*. *Encyclopedia Meas.*
https://www.researchgate.net/publication/246690791_LillieforsVan_Soest's_test_of_normality
- Almeida, M., Acosta-Vargas, P., y González, M. (2020). *Quito air quality modeling and prediction using meteorological and pollution data*.
<https://www.researchgate.net/publication/344672315>
- Álvarez Torres, E. (2011). *Contaminacion Ambiental Generada Por La Emision De Gases*.
- Angel, J., y Casas, R. (2023). *La identificación de la tendencia, estocástica o determinista, en series temporales económicas*.
- Ansari, M., y Alam, M. (2023). An Intelligent IoT-Cloud-Based Air Pollution Forecasting Model Using Univariate Time-Series Analysis. *Arabian Journal for Science and Engineering*. <https://doi.org/10.1007/s13369-023-07876-9>
- Anusha, N., Sai Chaithanya, M., y Jithendranath Reddy, G. (2019). Weather Prediction Using Multi Linear Regression Algorithm. *IOP Conference Series: Materials Science and Engineering*, 590(1).
<https://doi.org/10.1088/1757-899X/590/1/012034>
- Bakola, M., Hernandez Carballo, I., Jelastopulu, E., y Stuckler, D. (2022). The impact of COVID-19 lockdown on air pollution in Europe and North America: a systematic review. *European Journal of Public Health*, 32(6), 962–968. <https://doi.org/10.1093/eurpub/ckac118>
- Barzyk, T., Huang, H., Williams, R., Kaufman, A., y Essoka, J. (2018). Advice and frequently asked questions (FAQs) for citizen-science environmental health assessments. *International Journal of Environmental Research and Public Health*, 15(5). <https://doi.org/10.3390/ijerph15050960>
- Borck, R. (2019). Public transport and urban pollution. *Regional Science and Urban Economics*, 77, 356–366.
<https://doi.org/10.1016/j.regsciurbeco.2019.06.005>

- Breiman, L. (2001). *Random Forests* (Vol. 45).
- Cabral Júnior, B., y Luna Lucena, R. (2020). *Analysis Of Precipitation Using Mann-Kendall And Kruskal-Wallis Non-Parametric Tests*. <https://www.redalyc.org/articulo.oa?>
- Cavanaugh, J., y Neath, A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. In *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 11, Issue 3). Wiley-Blackwell. <https://doi.org/10.1002/wics.1460>
- Central Pollution Control Board. (2015). *National Air Quality Index*.
- Chai, T., y Draxler, R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Charalampopoulos, I. (2020). The R language as a tool for biometeorological research. In *Atmosphere* (Vol. 11, Issue 7). MDPI AG. <https://doi.org/10.3390/atmos11070682>
- Chicco, D., Warrens, M., y Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/PEERJ-CS.623>
- Consejo Nacional de Planificación. (2021). *PLAN NACIONAL DE DESARROLLO 2021-2025*.
- Cordova, C., Portocarrero, M., Salas, R., Torres, R., Rodrigues, P., y López-Gonzales, J. (2021). Air quality assessment and pollution forecasting using artificial neural networks in Metropolitan Lima-Peru. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-03650-9>
- Cornejo-Vasconez, D., Rodriguez-Espinosa, F., Guasumba, A., y Toulkeridis, T. (2022). Contrasting Effects Of Air Pollution Assessment In Two Areas Of The Quito Metropolitan District, Ecuador. *Granja*, 36(2). <https://doi.org/10.17163/lgr.n36.2022.08>
- Coronel Vargas, G., Au, W., y Izzotti, A. (2020). Public health issues from crude-oil production in the Ecuadorian Amazon territories. In *Science of the Total Environment* (Vol. 719). Elsevier B.V. <https://doi.org/10.1016/j.scitotenv.2019.134647>

- Crocker, D. (1972). Some Interpretations of the Multiple Correlation Coefficient. In *Source: The American Statistician* (Vol. 26, Issue 2).
- Cruz-Roa, A., y Barrios, M. (2018). Estimación de datos faltantes de lluvia mensual a través de la asimilación de información satelital y pluviométrica en una cuenca andina tropical. *Idesia (Arica), ahead*, 0–0. <https://doi.org/10.4067/s0718-34292018005001601>
- Dalal, M., Li, A., y Taori, R. (2019). *Autoregressive Models: What Are They Good For?* <http://arxiv.org/abs/1910.07737>
- D'Amato, G., Annesi-Maesano, I., Biagioni, B., Lancia, A., Cecchi, L., D'Ovidio, M., y D'Amato, M. (2023). New Developments in Climate Change, Air Pollution, Pollen Allergy, and Interaction with SARS-CoV-2. In *Atmosphere* (Vol. 14, Issue 5). MDPI. <https://doi.org/10.3390/atmos14050848>
- Dastres, R., y Soori, M. (2021). Artificial Neural Network Systems. In *International Journal of Imaging and Robotics (IJIR)* (Vol. 2021, Issue 2). www.ceserp.com/cp-jour
- De Amorim, L., Cavalcanti, G., y Cruz, R. (2022). *The choice of scaling technique matters for classification performance.* <https://doi.org/10.1016/j.asoc.2022.109924>
- De Nijs, R., y Klausen, T. (2013). On the expected difference between mean and median. *Electronic Journal of Applied Statistical Analysis*, 6(1), 110–117. <https://doi.org/10.1285/i20705948v6n1p110>
- Dey, S., y Dhal, G. (2019). Materials progress in the control of CO and CO2 emission at ambient conditions: An overview. In *Materials Science for Energy Technologies* (Vol. 2, Issue 3, pp. 607–623). KeAi Communications Co. <https://doi.org/10.1016/j.mset.2019.06.004>
- Di, Y., y Li, R. (2019). Correlation analysis of AQI characteristics and meteorological conditions in heating season. *IOP Conference Series: Earth and Environmental Science*, 242(2). <https://doi.org/10.1088/1755-1315/242/2/022067>
- Emperatriz, S., y Murillo, P. (2018). *Impact Of Atmospheric Contamination In Two Main Cities Of Ecuador.* <http://rus.ucf.edu.cu/index.php/rus>
- EPA. (2023, March 10). *Air Pollution: Current and Future Challenges.* <https://www.epa.gov/clean-air-act-overview/air-pollution-current-and-future-challenges>

- Evgeniou, T., y Pontil, M. (2001). Support vector machines: Theory and applications. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2049 LNAI, 249–257. https://doi.org/10.1007/3-540-44673-7_12
- Florian, F. (2021). *El uso exagerado del plástico durante la pandemia de COVID-19 afecta a los más vulnerables*.
- Fowler, D., Brimblecombe, P., Burrows, J., Heal, M., Grennfelt, P., Stevenson, D., Jowett, A., Nemitz, E., Coyle, M., Lui, X., Chang, Y., Fuller, G., Sutton, M., Klimont, Z., Unsworth, M., y Vieno, M. (2020). A chronology of global air quality: The development of global air pollution. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (Vol. 378, Issue 2183). Royal Society Publishing. <https://doi.org/10.1098/rsta.2019.0314>
- García Nieto, P., Combarro, E., Del Coz Díaz, J., y Montañés, E. (2013). A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study. *Applied Mathematics and Computation*, 219(17), 8923–8937. <https://doi.org/10.1016/j.amc.2013.03.018>
- García-Villacís, K., Ramos-Guerrero, L., Canga, J., Hidalgo-Lasso, D., y Vargas-Jentzsch, P. (2021). Environmental Impact Assessment of Remediation Strategy in an Oil Spill in the Ecuadorian Amazon Region. *Pollutants*, 1(4), 234–252. <https://doi.org/10.3390/pollutants1040019>
- Gholamy, A., Kreinovich, V., y Kosheleva, O. (2018). *Why 70/30 or 80/20 Relation Between Training and Testing Sets: Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation*. https://scholarworks.utep.edu/cs_techrephttps://scholarworks.utep.edu/cs_techrep/1209
- Gouveia, N., Kephart, J., Dronova, I., McClure, L., Granados, J., Betancourt, R., O’Ryan, A., Texcalac-Sangrador, J., Martinez-Folgar, K., Rodriguez, D., y Diez-Roux, A. (2021). Ambient fine particulate matter in Latin American cities: Levels, population exposure, and associated urban factors. *Science of the Total Environment*, 772. <https://doi.org/10.1016/j.scitotenv.2021.145035>

- Guía, P., Mario, I., González, S., Autor, R., Nicolás, M., y Gachet, A. (2019). *Quito Air Quality Modeling And Prediction Using Meteorological And Pollution Data*.
- Guo, J., y Feng, Z. (2019). Study on Spatial Temporal Distribution Characteristics of Air Quality Index in Beijing and Its Correlation with Local Meteorological Conditions. *Discrete Dynamics in Nature and Society*, 2019. <https://doi.org/10.1155/2019/1462034>
- Handhayani, T. (2023). An integrated analysis of air pollution and meteorological conditions in Jakarta. *Scientific Reports*, 13(1), 5798. <https://doi.org/10.1038/s41598-023-32817-9>
- Hernández Travieso, J. (2018). *55 Modelos De Predicción Meteorológica*. Universidad De Las Palmas De Gran Canaria.
- Huang, Y., Unger, N., Harper, K., y Heyes, C. (2020). Global Climate and Human Health Effects of the Gasoline and Diesel Vehicle Fleets. *GeoHealth*, 4(3). <https://doi.org/10.1029/2019GH000240>
- Husaini, D., Reneau, K., y Balam, D. (2022). Air pollution and public health in Latin America and the Caribbean (LAC): a systematic review with meta-analysis. In *Beni-Suef University Journal of Basic and Applied Sciences* (Vol. 11, Issue 1). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1186/s43088-022-00305-0>
- Iskandaryan, D., Ramos, F., y Trilles, S. (2021). Features exploration from datasets vision in air quality prediction domain. In *Atmosphere* (Vol. 12, Issue 3). MDPI AG. <https://doi.org/10.3390/atmos12030312>
- Ji, M., Jiang, Y., Han, X., Liu, L., Xu, X., Qiao, Z., y Sun, W. (2020). Spatiotemporal Relationships between Air Quality and Multiple Meteorological Parameters in 221 Chinese Cities. *Complexity*, 2020. <https://doi.org/10.1155/2020/6829142>
- Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., y Kolehmainen, M. (2004). Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, 38(18), 2895–2907. <https://doi.org/10.1016/j.atmosenv.2004.026>
- Kalisa, E., Fadlallah, S., Amani, M., Nahayo, L., y Habiyaemye, G. (2018). *Temperature and air pollution relationship during heatwaves in Birmingham, UK*.

- Kanchan, C., Gorai, A., y Goyal, P. (2015). A review on air quality indexing system. In *Asian Journal of Atmospheric Environment* (Vol. 9, Issue 2, pp. 101–113). Korean Society for Atmospheric Environment. <https://doi.org/10.5572/ajae.2015.9.2.101>
- Komorowski, M., Marshall, D., Saliccioli, J., y Crutain, Y. (2016). Exploratory data analysis. In *Secondary Analysis of Electronic Health Records* (pp. 185–203). Springer International Publishing. https://doi.org/10.1007/978-3-319-43742-2_15
- Kongcharoen, C., y Kruangpradit, T. (2013). *Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export*.
- Larasati, A., Hajji, A., y Dwiastuti, A. (2019). The relationship between data skewness and accuracy of Artificial Neural Network predictive model. *IOP Conference Series: Materials Science and Engineering*, 523(1). <https://doi.org/10.1088/1757-899X/523/1/012070>
- Lemeshko, B. (2015). Chi-Square-Type Tests for Verification of Normality. *Measurement Techniques*, 58(6), 581–591. <https://doi.org/10.1007/s11018-015-0759-2>
- Lei, T. M. T., Siu, S., Monjardino, J., Mendes, L., y Ferreira, F. (2022). Using Machine Learning Methods to Forecast Air Quality: A Case Study in Macao. *Atmosphere*, 13(9). <https://doi.org/10.3390/atmos13091412>
- Liu, T., Lau, A. K. H., Sandbrink, K., y Fung, J. C. H. (2018). Time Series Forecasting of Air Quality Based On Regional Numerical Modeling in Hong Kong. *Journal of Geophysical Research: Atmospheres*, 123(8), 4175–4196. <https://doi.org/10.1002/2017JD028052>
- Liu, Y., Wang, P., Li, Y., Wen, L., y Deng, X. (2022). Air quality prediction models based on meteorological factors and real-time data of industrial waste gas. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-13579-2>
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A., y Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. In *Frontiers in Public Health* (Vol. 8). Frontiers Media S.A. <https://doi.org/10.3389/fpubh.2020.00014>
- Marlier, M., Jina., Kinney, P., y DeFries, R. (2016). Extreme Air Pollution in Global Megacities. In *Current Climate Change Reports* (Vol. 2, Issue 1, pp. 15–27). Springer. <https://doi.org/10.1007/s40641-016-0032-z>

- Massey, F. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253), 68–78. <https://doi.org/10.1080/01621459.1951.10500769>
- Mila, F., y Yáñez, K. (2019). “The Environmental Constitutionalism In Ecuador.”
- Ministerio del Ambiente. (2010). *Plan Nacional de la Calidad del Aire*. <https://www.ambiente.gob.ec/wp-content/uploads/downloads/2012/10/libro-calidad-aire-1-final.pdf>
- Mohajan, Haradhan (2020): Quantitative Research: A Successful Investigation in Natural and Social Sciences. Published in: *Journal of Economic Development, Environment and People*, Vol. 9, No. 4 (31 December 2020): pp. 52-79
- Moran-Zuloaga, D., Merchan-Merchan, W., Rodríguez-Caballero, E., Hernick, P., Cáceres, J., y Cornejo, M. (2021). Overview and Seasonality of PM10 and PM2.5 in Guayaquil, Ecuador. *Aerosol Science and Engineering*, 5(4), 499–515. <https://doi.org/10.1007/s41810-021-00117-2>
- Moreira-Romero, Á. (2018). Contaminación del aire en el medio ambiente por las emisiones de gases tóxicos de empresas industriales en Ecuador. *Polo Del Conocimiento*, 3(7), 299. <https://doi.org/10.23857/pc.v3i7.553>
- Municipio del Distrito Metropolitano de Quito. (2020, December 30). *Celebremos el Fin de Año sin contaminar el ambiente de Quito*. <http://www.quitoinforma.gob.ec/2020/12/30/celebremos-el-fin-de-ano-sin-contaminar-el-ambiente-de-quito/>
- Nakamura, A., Nakatani, N., Maruyama, F., Fujiyoshi, S., Márquez-Reyes, R., Fernández, R., y Noda, J. (2022). Characteristics of PM2.5 Pollution in Osorno, Chile: Ion Chromatography and Meteorological Data Analyses. *Atmosphere*, 13(2). <https://doi.org/10.3390/atmos13020168>
- Ocles, M. (2020). *RESOLUCIÓN-COE-16.04.2020*.
- OMM. (2014). *Guía de Instrumentos y Métodos de Observación Meteorológicos*.
- OMS. (2006). *Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre*.
- Orcan, F. (2020). Parametric or Non-parametric: Skewness to Test Normality for Mean Comparison. *International Journal of Assessment Tools in Education*, 236–246. <https://doi.org/10.21449/ijate.656077>

- Palomino Parra, J., Torres Cruz, O., y Angulo Méndez, Y. (2020). Dispositivo basado en modelo arima para predicción de variables ambientales (temperatura, humedad, velocidad del aire) en el área agrícola del departamento del Meta. *Revista GEON (Gestión, Organizaciones y Negocios)*, 7(2), 1–12. <https://doi.org/10.22579/23463910.193>
- Parra, R. (2022). Effect of Global Atmospheric Datasets in Modeling Meteorology and Air Quality in the Andean Region of Ecuador. *Aerosol and Air Quality Research*, 22(11). <https://doi.org/10.4209/aaqr.210292>
- Parra, R., y Espinoza, C. (2020). Insights for air quality management from modeling and record studies in Cuenca, Ecuador. *Atmosphere*, 11(9). <https://doi.org/10.3390/atmos11090998>
- Pélabon, C., Hilde, C., Einum, S., y Gamelon, M. (2020). On the use of the coefficient of variation to quantify and compare trait variation. *Evolution Letters*, 4(3), 180–188. <https://doi.org/10.1002/evl3.171>
- Picardo, E. (2022). *Negative Correlation: How it Works, Examples And FAQ*. <https://www.investopedia.com/terms/n/negative-correlation.asp>
- Quinteros, M., Lu, S., Blazquez, C., Cárdenas-R, J., Ossa, X., Delgado-Saborit, J., Harrison, R., y Ruiz-Rudolph, P. (2019). Use of data imputation tools to reconstruct incomplete air quality datasets: A case-study in Temuco, Chile. *Atmospheric Environment*, 200, 40–49. <https://doi.org/10.1016/j.atmosenv.2018.11.053>
- Ramírez, O. (2017, October 9). *Correlación y Regresión Lineal*. <https://rpubs.com/osoramirez/316691>
- Rani Das, K. (2016). A Brief Review of Tests for Normality. *American Journal of Theoretical and Applied Statistics*, 5(1), 5. <https://doi.org/10.11648/j.ajtas.20160501.12>
- Reshu Yadav, Tripathi, S., Pranuthi, G., y Dubey, S. (2014). Trend analysis by Mann-Kendall test for precipitation and temperature for thirteen districts of Uttarakhand. *Journal of Agrometeorology*, 16(2), 164–171. <https://doi.org/10.54386/jam.v16i2.1507>
- Riojas-Rodríguez, H., Soares Da Silva, A., Luis Texcalac-Sangrador, J., y Litai Moreno-Banda, G. (2016b). Air pollution management and control in Latin America and the Caribbean: implications for climate change Special report. In *Rev Panam Salud Publica* (Vol. 40, Issue 3).

- Rodríguez-Guerra, A., y Cuvi, N. (2019). Air pollution and environmental justice in Quito, Ecuador. *Fronteiras*, 8(3), 13–46. <https://doi.org/10.21664/2238-8869.2019v8i3.p13-46>
- Salazar, R. (2021). *RESOLUCIÓN-COE-21.04.2021*.
- Sampath, V., Aguilera, J., Prunicki, M., y Nadeau, K. (2023). Mechanisms of climate change and related air pollution on the immune system leading to allergic disease and asthma. *Seminars in Immunology*, 67, 101765. <https://doi.org/10.1016/j.smim.2023.101765>
- Secretaría de Ambiente de Quito. (2011). *Indice Quiteño De La Calidad Del Aire IQCA*.
- Secretaria del Ambiente MDMQ. (2023). *Generalidades Red de Monitoreo Atmosférico*.
- Sedgwick, P. (2012). Pearson's correlation coefficient. In *BMJ (Online)* (Vol. 345, Issue 7864). <https://doi.org/10.1136/bmj.e4483>
- Setiawan, I. (2020). Time series air quality forecasting with R Language and R Studio. *Journal of Physics: Conference Series*, 1450(1). <https://doi.org/10.1088/1742-6596/1450/1/012064>
- Souza, M., Rocha, B., Adeyemi, J., Nadal, M., Domingo, J., y Barbosa, F. (2022). Legacy and emerging pollutants in Latin America: A critical review of occurrence and levels in environmental and food samples. In *Science of the Total Environment* (Vol. 848). Elsevier B.V. <https://doi.org/10.1016/j.scitotenv.2022.157774>
- Suman. (2019). Air quality indices: A review of methods to interpret air quality status. *Materials Today: Proceedings*, 34, 863–868. <https://doi.org/10.1016/j.matpr.2020.07.141>
- Thangavel, P., Park, D., y Lee, Y. (2022). Recent Insights into Particulate Matter (PM_{2.5})-Mediated Toxicity in Humans: An Overview. In *International Journal of Environmental Research and Public Health* (Vol. 19, Issue 12). MDPI. <https://doi.org/10.3390/ijerph19127511>
- UNEP. (2020). *Actions on Air Quality in Latin America and the Caribbean Executive Summary*. https://wedocs.unep.org/bitstream/handle/20.500.11822/36699/AAQLAC_ES.pdf
- UNICEF Ecuador. (2020). *El Aire Que Respiramos*. www.unicef.org/ecuador

- Urrutia-Pereira, M., Mello-da-Silva, C. A., y Solé, D. (2020). COVID-19 and air pollution: A dangerous association? *Allergologia et Immunopathologia*, 48(5), 496–499. <https://doi.org/10.1016/j.aller.2020.05.004>
- US EPA. (2022a, June 14). *Ground-level Ozone Basics*. <https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics>
- US EPA. (2022b, August 2). *Basic Information about Carbon Monoxide (CO) Outdoor Air Pollution*. <https://www.epa.gov/co-pollution/basic-information-about-carbon-monoxide-co-outdoor-air-pollution>
- US EPA. (2023a, January 18). *Nitrogen Oxides (NOx) Control Regulations*. <https://www3.epa.gov/region1/airquality/nox.html>
- US EPA. (2023b, February 16). *Sulfur Dioxide Basics*. <https://www.epa.gov/so2-pollution/sulfur-dioxide-basics>
- Useros, J. (2012). El Cambio Climático: Sus Causas Y Efectos Medioambientales
Climate Change: Causes And Environmental Effects.
- Uyanık, G., y Güler, N. (2013). A Study on Multiple Linear Regression Analysis. *Procedia - Social and Behavioral Sciences*, 106, 234–240. <https://doi.org/10.1016/j.sbspro.2013.12.027>
- Veenhoven, R. (2010). Life is getting better: Societal evolution and fit with human nature. *Social Indicators Research*, 97(1), 105–122. <https://doi.org/10.1007/s11205-009-9556-0>
- Velayarce, D., Bustos, Q., García, M., Timaná, C., Carbajal, R., Salvatierra, N., Horna, D., y Murray, V. (2022). Air Quality Analysis in Lima, Peru Using the NO₂ Levels during the COVID-19 Pandemic Lockdown. *Atmosphere*, 13(3). <https://doi.org/10.3390/atmos13030373>
- Vilá, R., Torrado, M., y Reguant, M. (2019). Análisis de regresión lineal múltiple con SPSS: un ejemplo práctico. *REIRE Revista d Innovación i Recerca En Educació*n, 12 (2). <https://doi.org/10.1344/reire2019.12.222704>
- Vinueza, D., Ochoa-Herrera, V., Maurice, L., Tamayo, E., Mejía, L., Tejera, E., y Machado, A. (2021). Determining the microbial and chemical contamination in Ecuador's main rivers. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-96926-z>
- Walker Hanlon, W., Alsan, M., Atkin, D., Boustan, L., Clay, K., Costa, D., Donaldson, D., Foquet, R., Goldin, C., Hoffman, P., Kahn, M., Katz, L.,

- Lleras-Muney, A., Nafziger, S., y Rosenthal, J. (2015). *Pollution and Mortality in the 19th Century* (Vol. 53). <http://www.nber.org/papers/w21647>
- Wang, Q., y Li, X. (2022). Correlation Analysis between Meteorological Factors and Pollutants Based on Copula Theory. *Journal of Physics: Conference Series*, 2168(1). <https://doi.org/10.1088/1742-6596/2168/1/012028>
- Wang, W., y Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, 324(1). <https://doi.org/10.1088/1757-899X/324/1/012049>
- Watthanacheewakul, L. (2021). *Transformations for Left Skewed Data*.
- Wei, W. (2019). *Multivariate time series regression models* (pp. 105–138). <https://doi.org/10.1002/9781119502951.ch3>
- West, R. (2022). Best practice in statistics: The use of log transformation. *Annals of Clinical Biochemistry*, 59(3), 162–165. <https://doi.org/10.1177/00045632211050531>
- Wood, D. (2022). Local integrated air quality predictions from meteorology (2015 to 2020) with machine and deep learning assisted by data mining. *Sustainability Analytics and Modeling*, 2, 100002. <https://doi.org/10.1016/j.samod.2021.100002>
- Xing, Q., y Sun, M. (2022). Characteristics of PM_{2.5} and PM₁₀ Spatio-Temporal Distribution and Influencing Meteorological Conditions in Beijing. *Atmosphere*, 13(7). <https://doi.org/10.3390/atmos13071120>
- Zafra-Mejía, C., Alarcón-Hincapié, J., y Echeverry-Prieto, L. (2019). Análisis de series de tiempo para concentraciones de PM₁₀: Influencia de la cobertura superficial urbana. *Scientia Et Technica*, 24(1), 146–153.
- Zalakeviciute, R., Vasquez, R., Bayas, D., Buenano, A., Mejia, D., Zegarra, R., Diaz, V., y Lamb, B. (2020). Drastic improvements in air quality in ecuador during the COVID-19 outbreak. *Aerosol and Air Quality Research*, 20(8), 1783–1792. <https://doi.org/10.4209/aaqr.2020.05.0254>
- Zhang, Q. (2023). *On relationships between Chatterjee's and Spearman's correlation coefficients*. <http://arxiv.org/abs/2302.10131>
- Zhu, Z., Qiao, Y., Liu, Q., Lin, C., Dang, E., Fu, W., Wang, G., y Dong, J. (2021). The impact of meteorological conditions on Air Quality Index under different urbanization gradients: a case from Taipei. *Environment*,

Development and Sustainability, 23(3), 3994–4010.
<https://doi.org/10.1007/s10668-020-00753-7>

Zinzendoff Okwonu, F., Laro Asaju, B., y Irimisose Arunaye, F. (2020). Breakdown Analysis of Pearson Correlation Coefficient and Robust Correlation Methods. *IOP Conference Series: Materials Science and Engineering*, 917(1). <https://doi.org/10.1088/1757-899X/917/1/012065>

ANEXOS

Anexo A. Certificado del abstract por parte de idiomas

Figura 13

Certificado Abstract

Tulcán, 26 de noviembre de 2023

CERTIFICADO

Yo, María Teresa Rodríguez de Verenzuela, con cédula de ciudadanía N° 1759039371, docente del Centro de Idiomas Extranjeros y Lenguas Nativas CIDEN – UPEC, certifico que revisé y corregí la traducción del Abstract del Trabajo "Incidencia de la meteorología en la calidad del aire antes, durante y después de la pandemia COVID-19 en Quito – Ecuador.", mismo que pertenece al señor Pablo Neptalí Escobar Tibán con cédula de ciudadanía N.º 1600526378; estudiante del Programa de Maestría en Estadística aplicada de la Universidad Politécnica Estatal del Carchi.

Atentamente,



Dra. María Teresa Rodríguez de Verenzuela

C.I: 1759039371

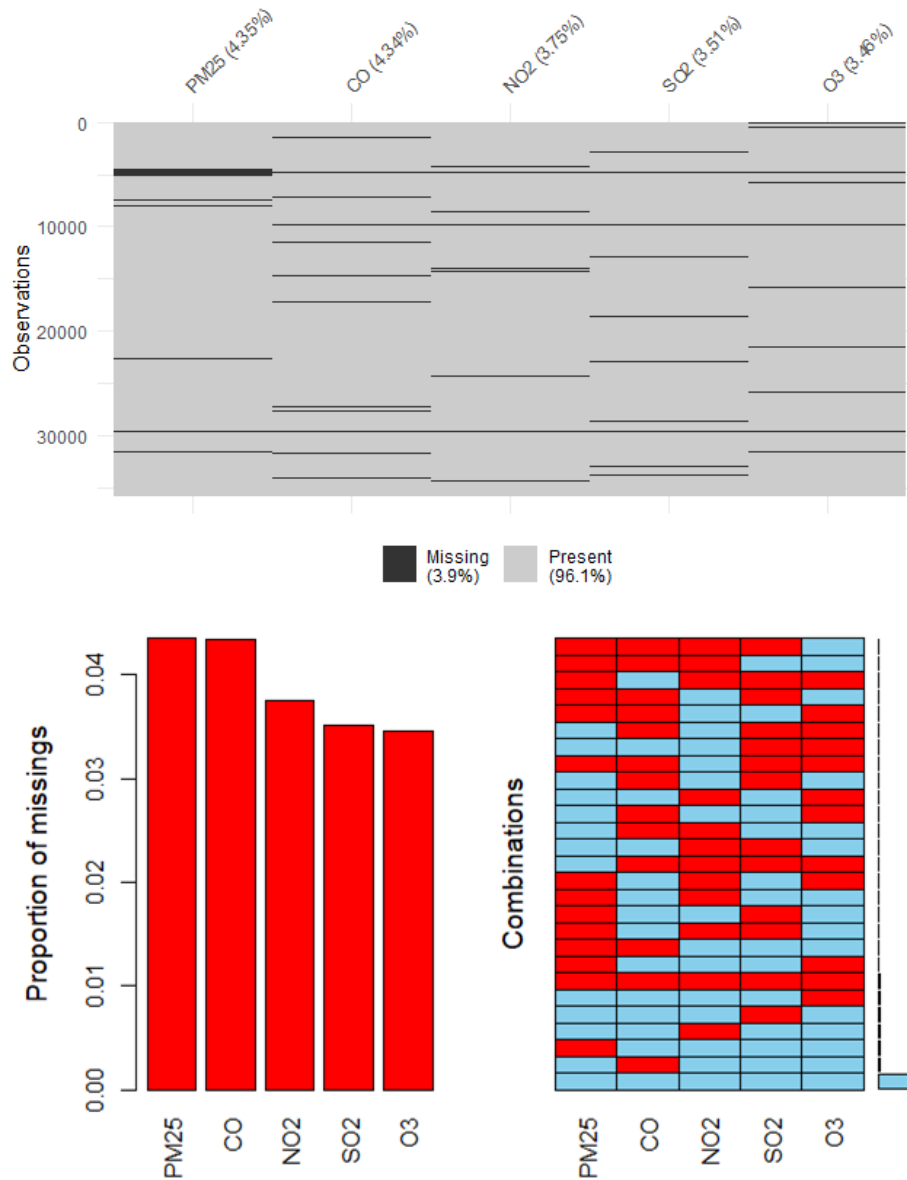
DOCENTE CIDEN – UPEC

Nota: Certificado de Abstract por parte del CIDEN - UPEC.

Anexo B. Resultados Imputación De Datos - Belisario

Figura 14

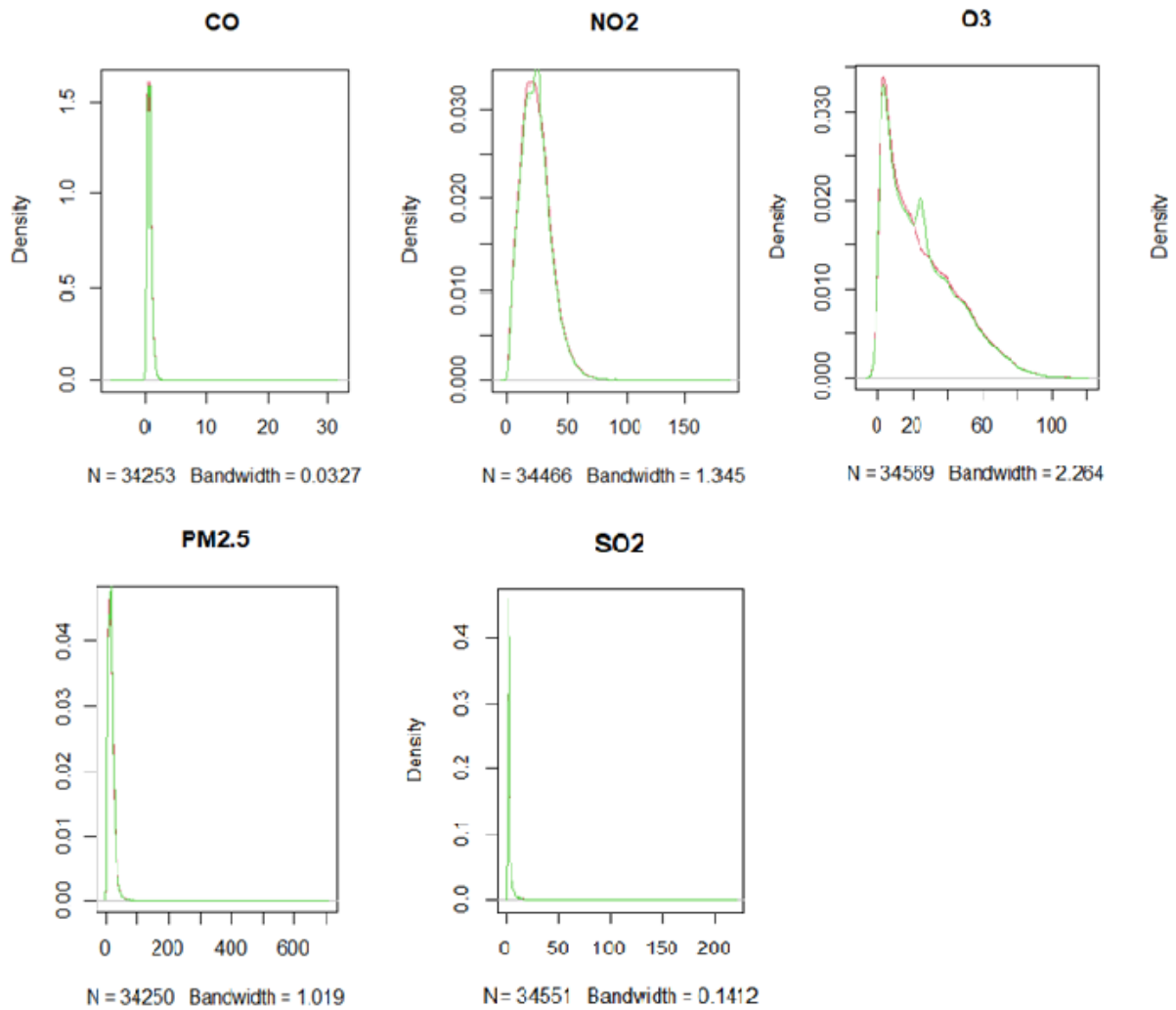
Datos Perdidos dentro de la Base - Belisario



Nota: Valores perdidos encontrados en la base de datos correspondiente a la estación Belisario.

Figura 135

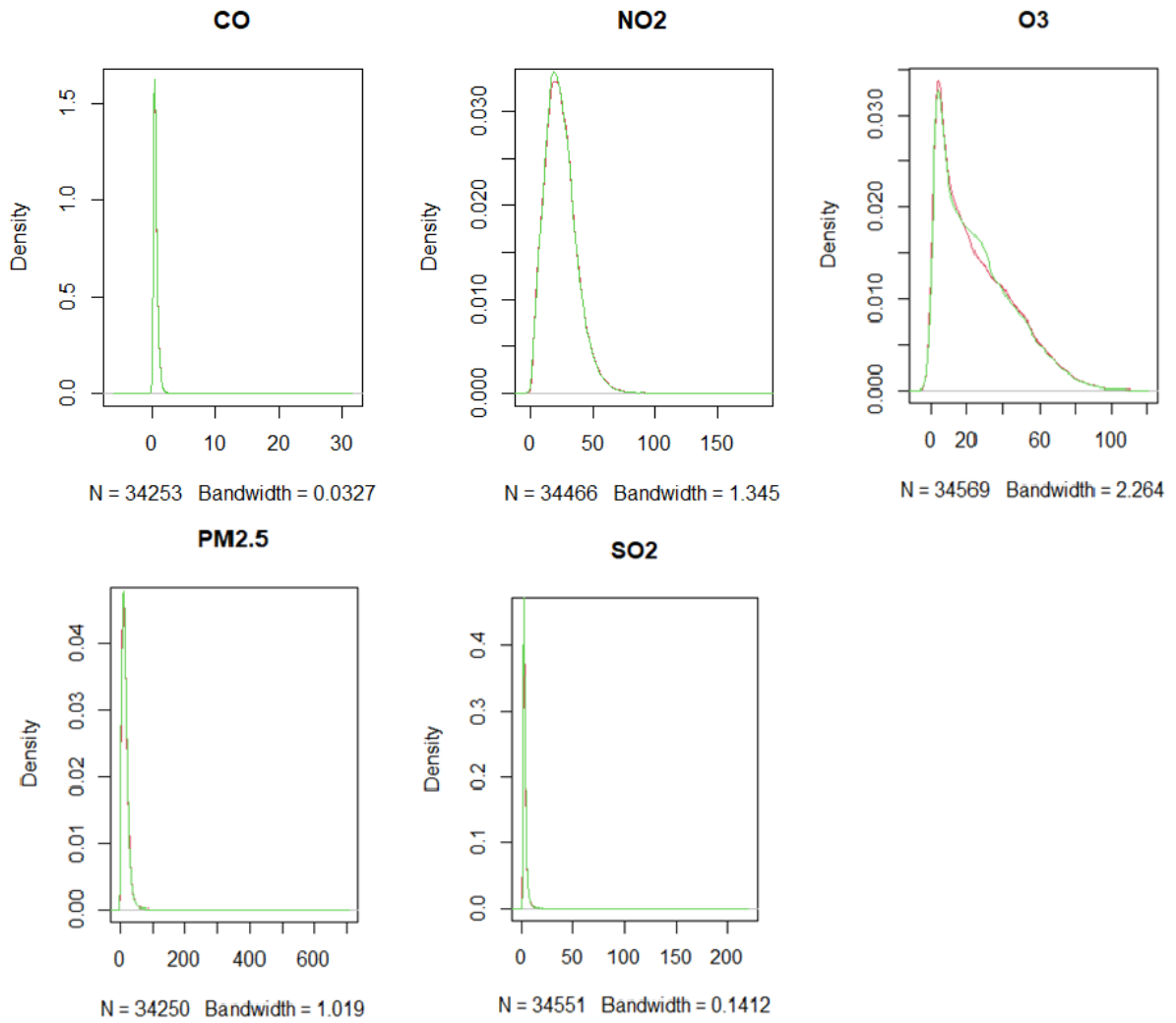
Imputación con la media en Belisario



Nota: Datos antes y después de la imputación con la media en la estación Belisario.

Figura 146

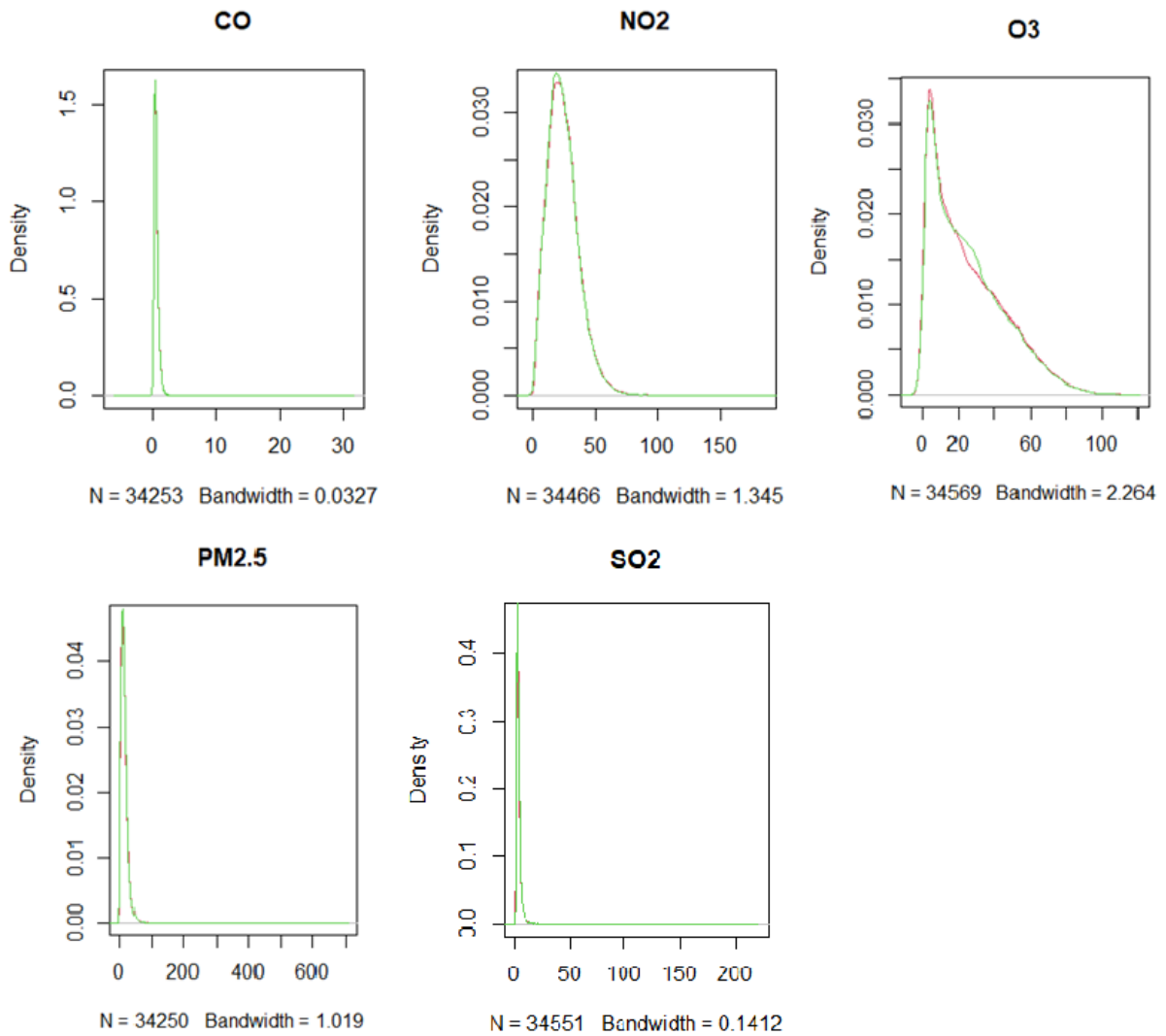
Imputación mediante regresión lineal en Belisario



Nota: Datos antes y después de la imputación mediante regresión lineal en la estación Belisario.

Figura 157

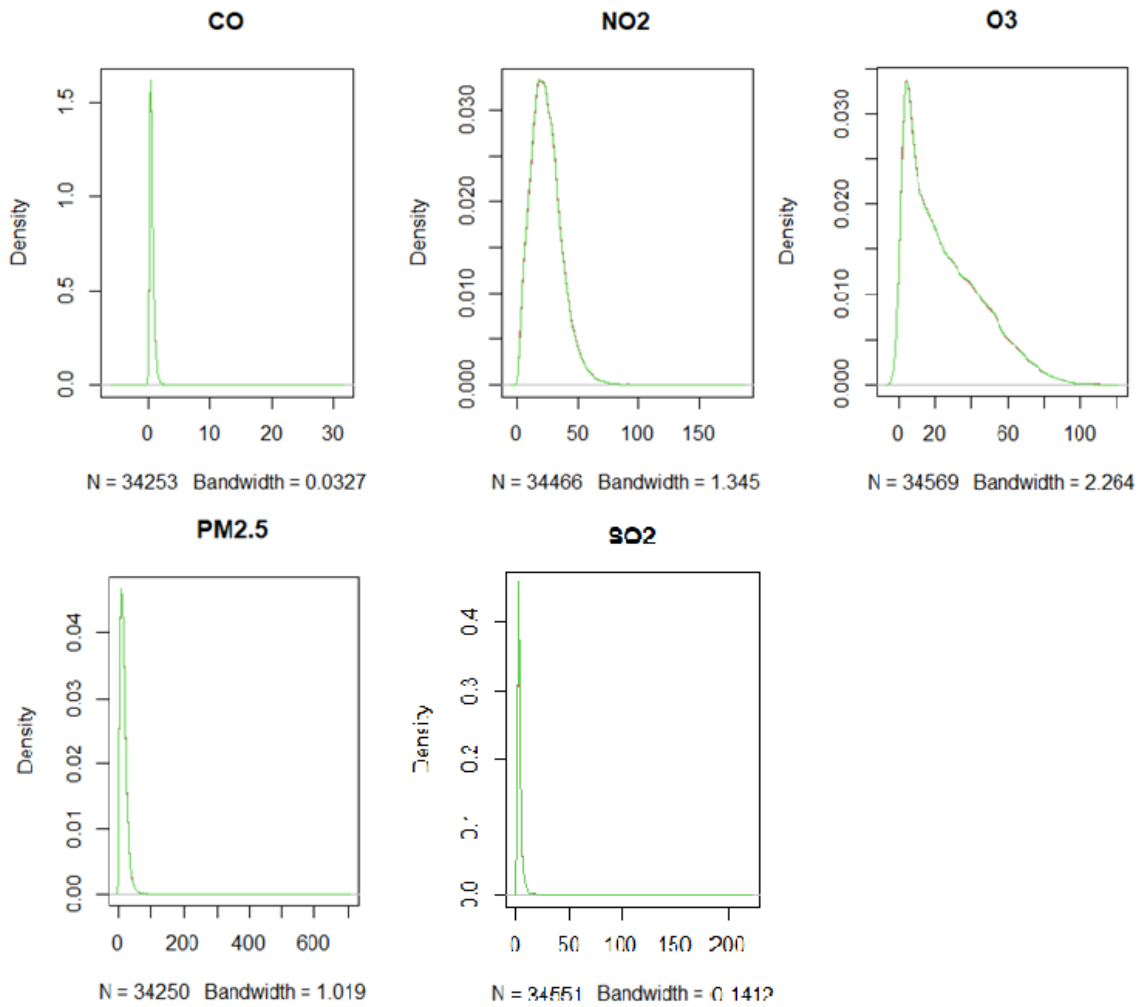
Imputación mediante regresión estocástica en Belisario



Nota: Datos antes y después de la imputación mediante regresión estocástica en la estación Belisario.

Figura 168

Imputación Múltiple en Belisario



Nota: Datos antes y después de la imputación múltiple en la estación Belisario.

Anexo C. Normalidad

Tabla 15

Pruebas de Normalidad

	IQG CDA. stati stic	IQGCDA .p.value	Dir.st atisti c	Dir.p .valu e	Hum. statis tic	Hum. p.val ue	Llu.s tatis tic	Llu. p.val ue	Pres. statis tic	Pres. p.val ue	Rad. stati stic	Rad. p.val ue	Temp .statis tic	Temp .p.val ue	Vel.s tatis tic	Vel. p.val ue
Kol_Be l	0,08 67	0,0000	0,047 1	0,00 27	0,087 6	0,000 0	0,31 39	0,00 00	0,016 6	0,806 6	0,030 4	0,12 70	0,016 1	0,833 9	0,16 24	0,00 00
Lillie_B el	0,08 67	0,0000	0,047 1	0,03 42	0,087 6	0,000 0	0,31 39	0,00 00	0,016 6	0,410 2	0,030 4	0,00 26	0,016 1	0,457 8	0,16 24	0,00 00
Chi-2 _Bel	0,00 02	0,0000	0,000 1	0,00 04	0,024 7	0,000 0	0,28 47	0,00 00	0,430 1	0,045 5	0,408 6	0,07 08	0,284 7	0,493 0	0,07 78	0,00 00
Kol_Ca ra	0,08 48	0,0000	0,014 2	0,92 45	0,124 5	0,000 0	0,34 46	0,00 00	0,015 0	0,888 4	0,029 4	0,15 15	0,023 4	0,387 2	0,05 07	0,00 00
Lillie_C ara	0,08 48	0,0000	0,014 2	0,66 02	0,124 5	0,000 0	0,34 46	0,00 00	0,015 0	0,569 0	0,029 4	0,00 43	0,023 4	0,054 2	0,00 05	0,00 00
Chi- 2_Cara	0,02 25	0,0000	0,004 8	0,01 65	0,005 4	0,000 0	0,08 99	0,00 00	0,032 1	0,314 9	0,037 7	0,12 93	0,037 9	0,125 3	0,62 69	0,00 00

Kol_Ce	0,09	0,0000	0,048	0,00	0,089	0,000	0,30	0,00	0,022	0,456	0,020	0,52	0,019	0,611	0,12	0,00
n	17		4	18	6	0	16	00	1	9	9	93	7	0	72	00
Lillie_C	0,00	0,0000	0,004	0,01	0,009	0,000	0,30	0,00	0,022	0,084	0,020	0,11	0,019	0,174	0,00	0,00
en	92		8	03	6	0	16	00	1	0	9	61	7	6	01	00
Chi-2	0,00	0,0000	0,095	0,00	0,000	0,000	0,00	0,00	0,040	0,074	0,030	0,40	0,023	0,745	0,04	0,00
_Cen	00		0	60	3	0	15	00	6	2	1	86	7	5	38	00
Kol_Co	0,07	0,0000	0,039	0,02	0,118	0,000	0,32	0,00	0,016	0,813	0,026	0,24	0,017	0,726	0,18	0,00
t	78		1	11	5	0	10	00	5	6	5	36	9	4	58	00
Lillie_C	0,00	0,0000	0,039	0,01	0,011	0,000	0,32	0,00	0,016	0,422	0,026	0,01	0,017	0,294	0,18	0,00
ot	78		1	50	8	0	10	00	5	0	5	56	9	3	58	00
Chi-2	0,00	0,0000	0,050	0,00	0,040	0,000	0,00	0,00	0,034	0,231	0,034	0,22	0,028	0,509	9,25	0,00
_Cot	02		3	83	6	0	14	00	2	5	3	99	2	0	78	00
Kol_Gu	0,07	0,0000	0,039	0,01	0,088	0,000	0,27	0,00	0,017	0,741	0,048	0,00	0,030	0,131	0,18	0,00
a	57		4	96	9	0	13	00	6	9	3	19	2	2	30	00
Lillie_G	0,00	0,0000	0,039	0,12	0,008	0,000	0,27	0,00	0,017	0,314	0,048	0,01	0,030	0,029	0,18	0,00
ua	76		4	42	9	0	13	00	6	4	3	20	2	4	30	00
Chi-2	0,00	0,0000	0,005	0,00	0,002	0,000	0,15	0,00	0,029	0,433	0,078	0,21	0,045	0,024	0,00	0,00
_Gua	16		3	45	4	0	03	00	6	3	1	81	8	7	23	00

Nota: Valores del estadístico y p-value para la prueba de Kolmogorov Smirnov, Lilliefors y Chi-cuadrado.

Anexo D. Simetría

Tabla 16

Simetría Sin Transformación

	IQGCDA	DIR	HUM	LLU	PRES	RAD	TEMP	VEL
Belisario	1,8038	0,0632	-0,6109	3,1535	-0,0345	0,1163	-0,0344	1,7976
Carapungo	1,4582	0,0378	-1,1164	4,8916	-0,0434	-0,1701	0,1014	0,7544
Centro	2,0369	0,2884	-0,5612	2,7682	-0,0455	0,0689	0,0344	1,3732
Cotocollao	1,6023	-0,2253	-0,9365	3,2122	-0,0333	-0,0696	0,0155	2,8418
Guamaní	1,7839	-0,3005	-0,5556	2,5999	-0,0363	0,2512	0,1127	1,5631

Nota: Valores de la prueba de Asimetría de las distintas variables de estudio antes de aplicar alguna transformación.

Tabla 17

Simetría después de Transformación

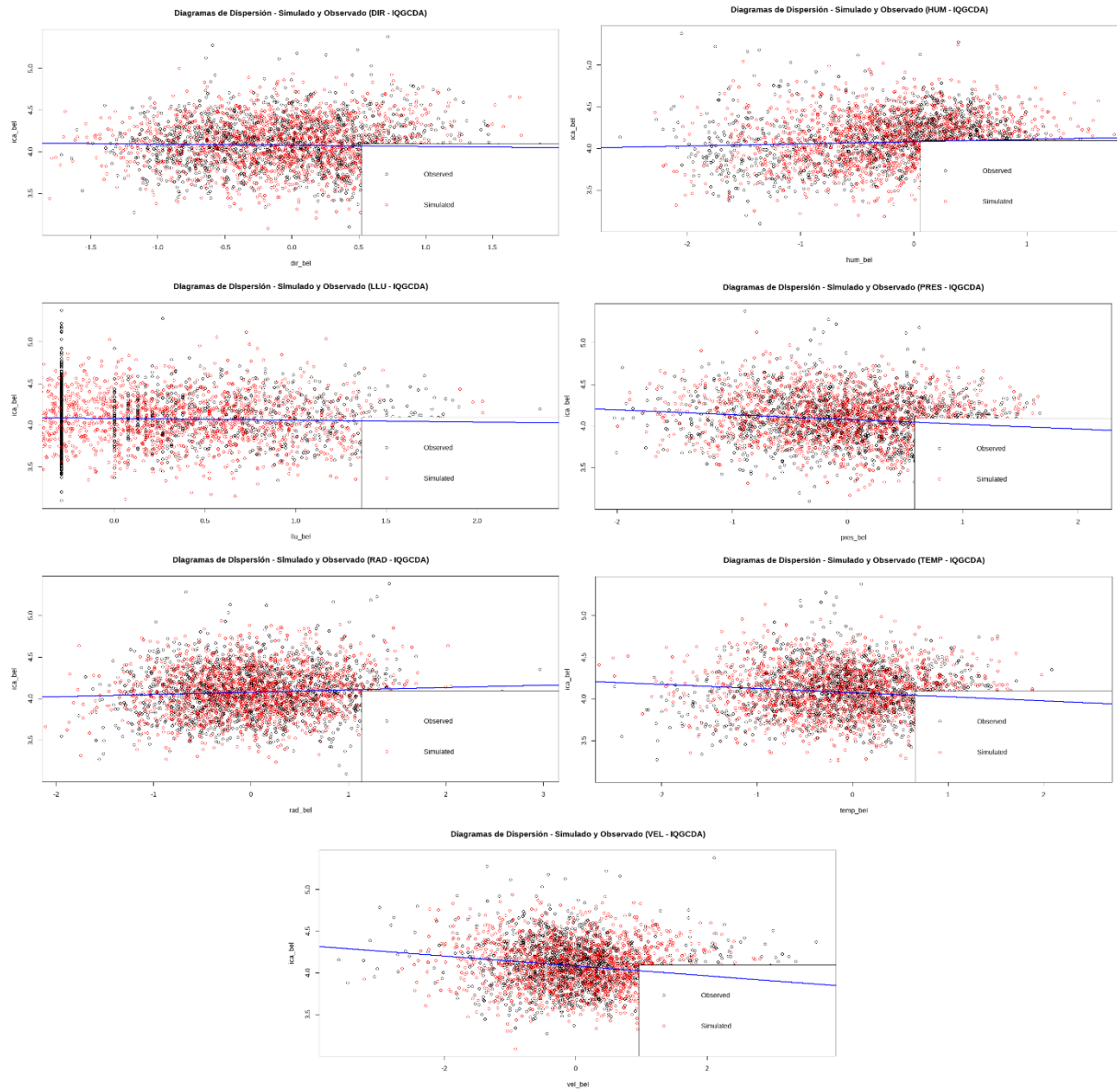
	IQGCDA	DIR	HUM	LLU	PRES	RAD	TEMP	VEL
Belisario	0,2971	0,0632	-0,6109	0,9085	-0,0345	0,1163	-0,0344	0,5122
Carapungo	0,2704	0,0378	-1,1164	1,1818	-0,0434	-0,1701	0,1014	-0,0810
Centro	0,3342	0,2884	-0,5612	0,7971	-0,0455	0,0689	0,0344	-0,2585
Cotocollao	0,2141	-0,2253	-0,9365	0,9900	-0,0333	-0,0696	0,0155	1,2144
Guamaní	0,1883	-0,3005	-0,5556	0,4422	-0,0363	0,2512	0,1127	0,7996

Nota: Valores de la prueba de Asimetría de las distintas variables de estudio después de aplicar alguna transformación.

Anexo E. Teoría Copula

Figura 179

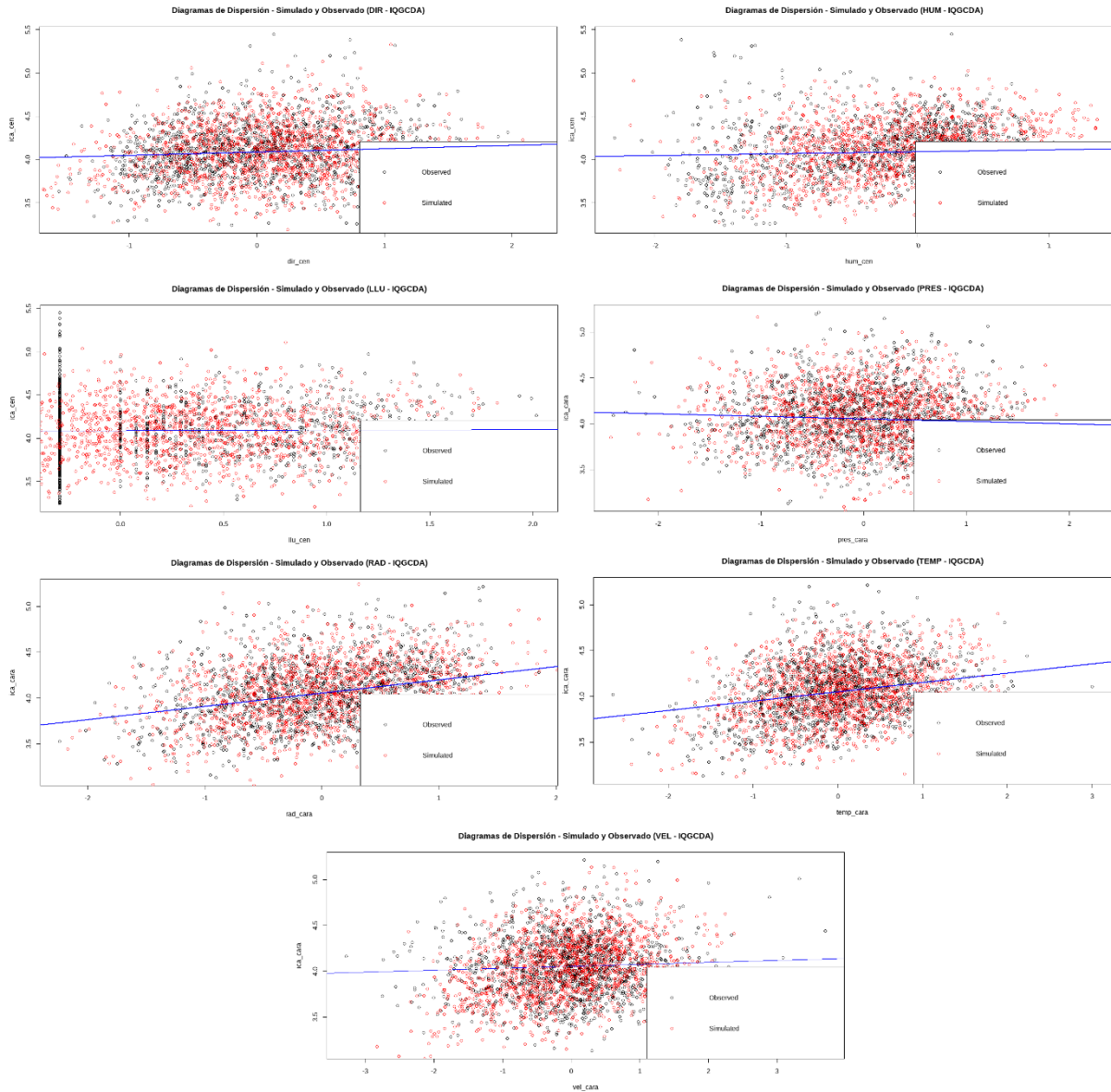
Belisario - Relación Mediante Funciones Copula



Nota: Gráfica de datos simulados y observados por medio del uso de la Teoría Copula de la estación Belisario

Figura 20

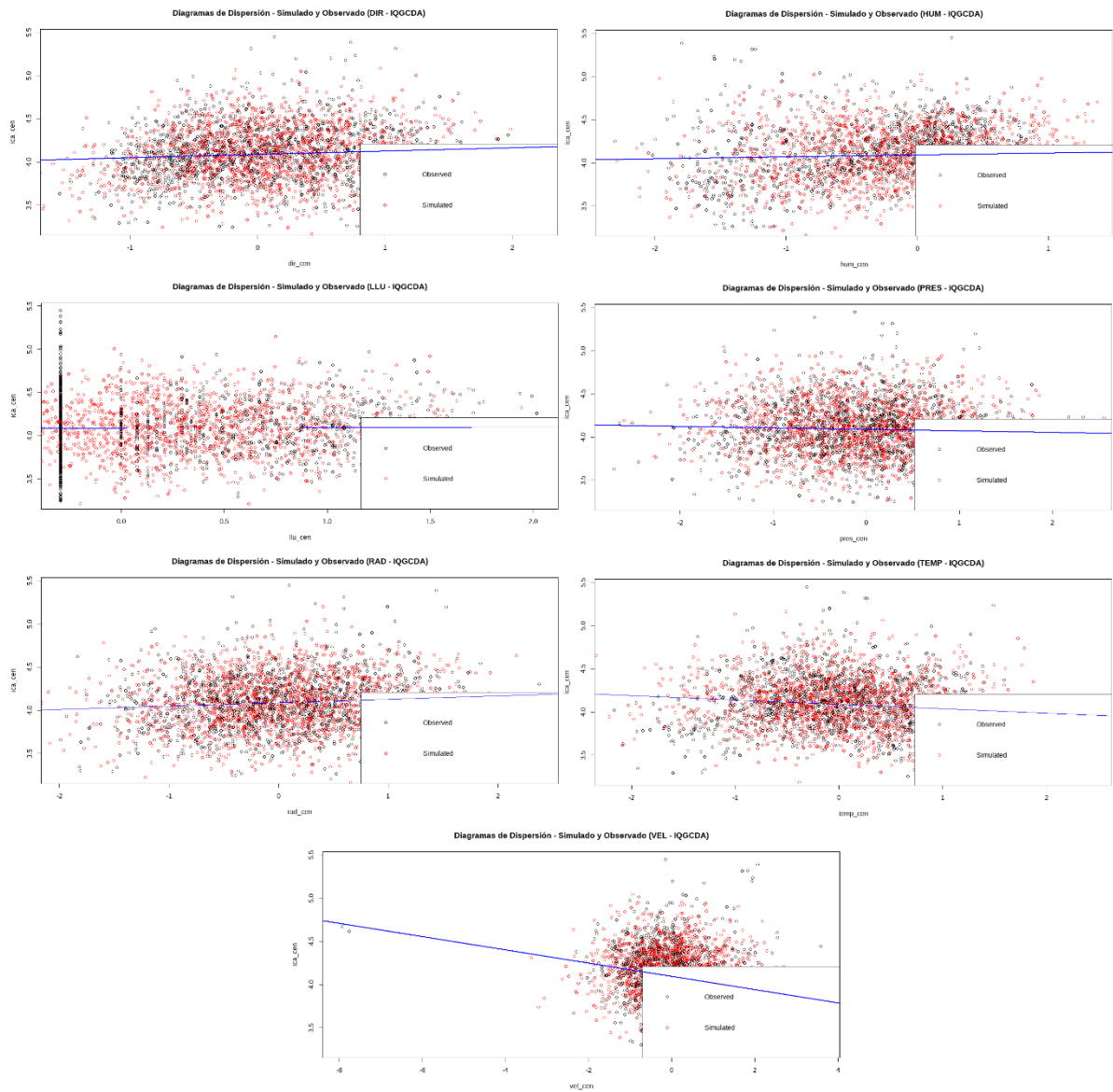
Carapungo - Relación Mediante Funciones Copula



Nota: Gráfica de datos simulados y observados por medio del uso de la Teoría Copula de la estación Carapungo.

Figura 181

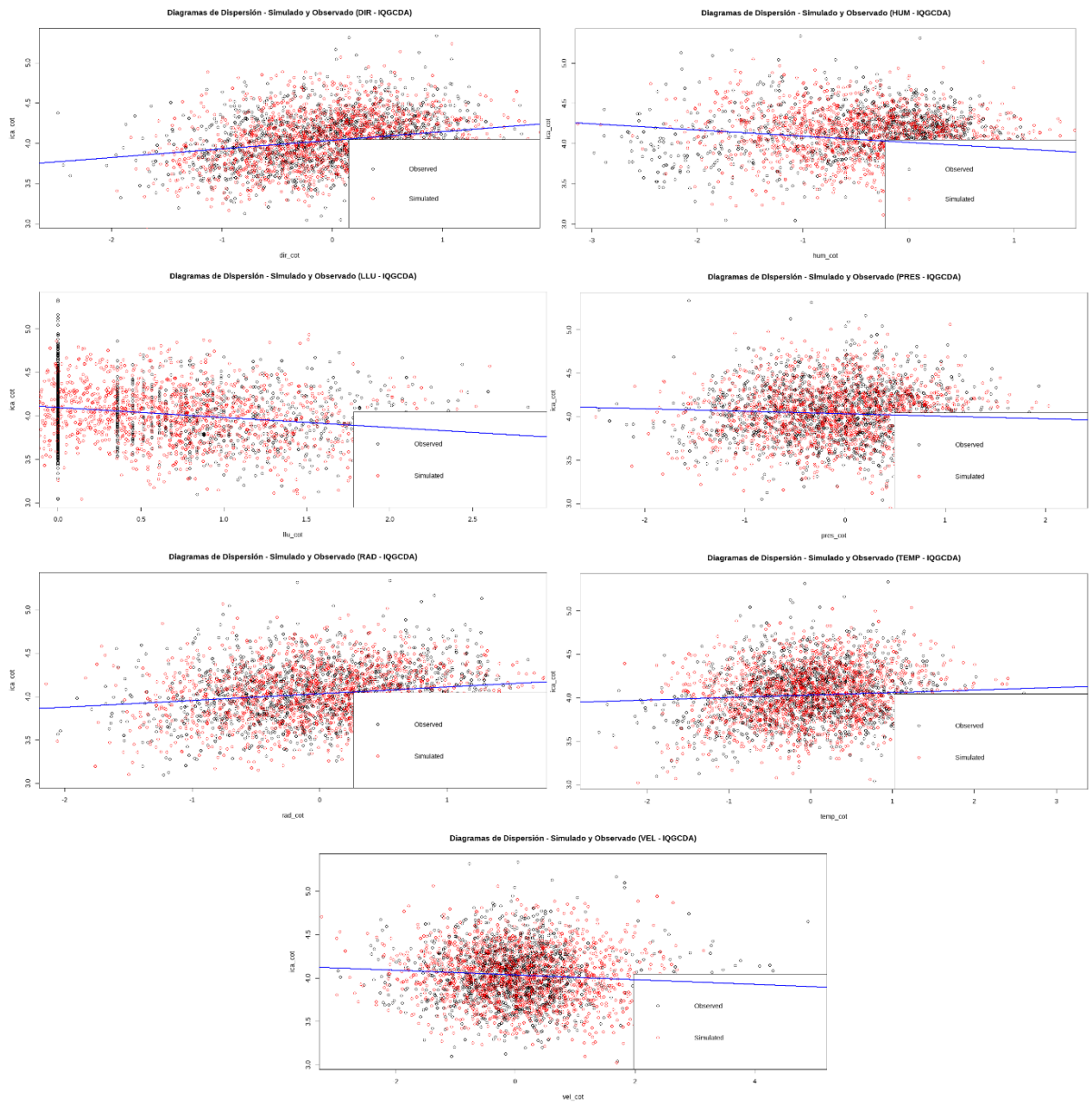
Centro - Relación Mediante Funciones Copula



Nota: Gráfica de datos simulados y observados por medio del uso de la Teoría Copula de la estación Centro.

Figura 192

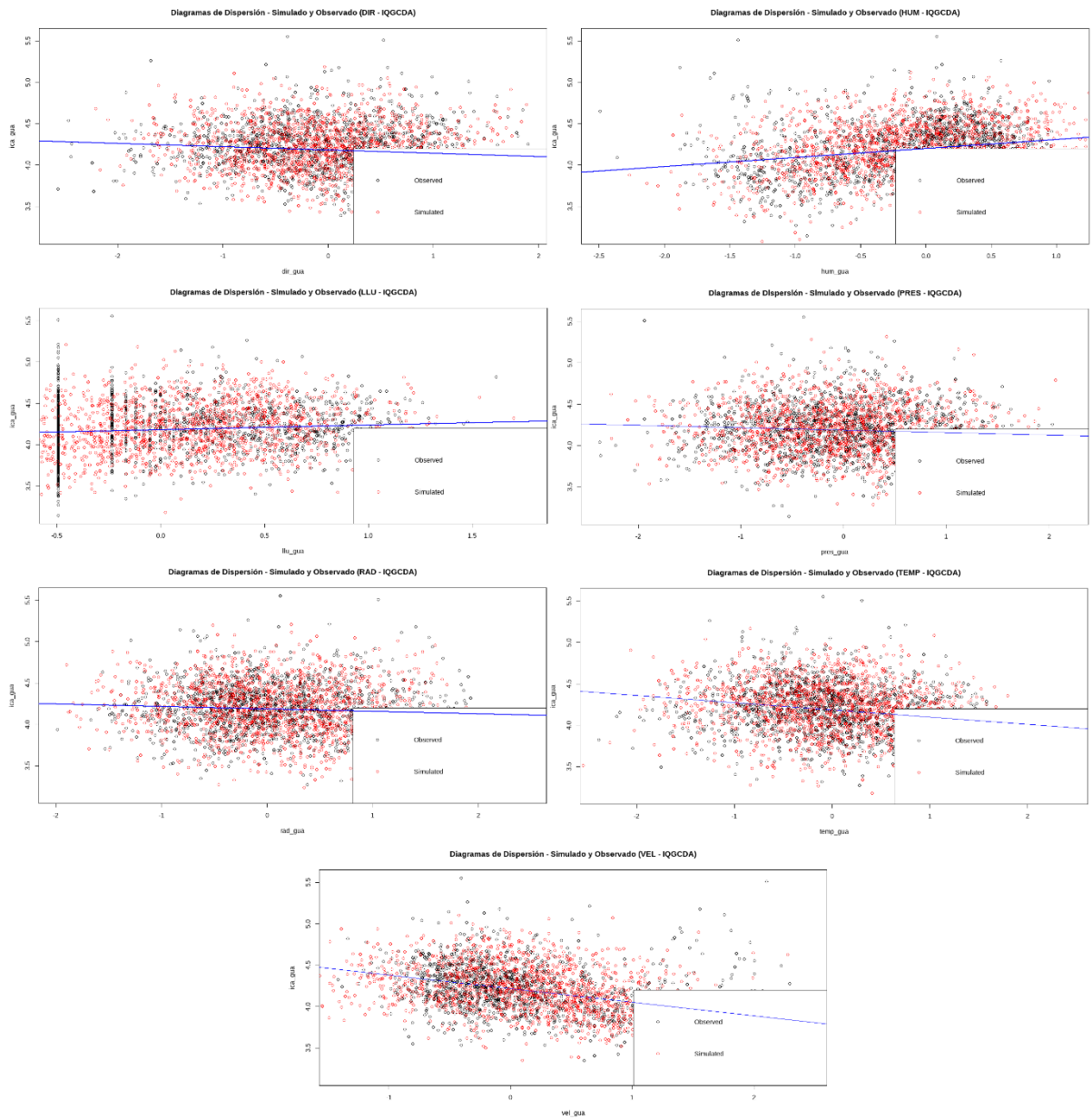
Cotocollao - Relación Mediante Funciones Copula



Nota: Gráfica de datos simulados y observados por medio del uso de la Teoría Copula de la estación Cotocollao.

Figura 203

Guamaní - Relación Mediante Funciones Copula



Nota: Gráfica de datos simulados y observados por medio del uso de la Teoría Copula de la estación Guamaní.

Anexo F. Supuestos Regresiones

Tabla 18

Revisión de Supuestos - Regresión Lineal Múltiple

	1er Momento Tendencia	2do Momento Tendencia
S. Linealidad.statistic	0,2304	5,0355
S. Linealidad.p.value	0,7943	0,0067
S. Normalidad.statistic	0,0304	0,0421
S. Normalidad.p.value	0,9019	0,1009
S. Homoceda.statistic	0,4414	0,5976
S. Homoceda.p.value	0,0640	1,0000
S. Independencia.statistic	0,9198	0,8318
S. Independencia.p.value	0,0000	0,0000

Nota: Valores para comprobar supuestos en Regresión Lineal Múltiple en cada uno de los momentos con Tendencia.

Tabla 19

Revisión de Supuestos - ARIMAX

	1er Momento Tendencia	2do Momento Tendencia
S. Independencia.statistic	0,05623695	0,000395199
S. Independencia.p.value	0,81254561	0,984139404
S. Normalidad.statistic	0,04612374	0,025996971
S. Normalidad.p.value	0,44229309	0,619831513

Nota: Valores para comprobar supuestos en series de tiempo ARIMAX13 en cada uno de los momentos con Tendencia.