



**UNIVERSIDAD CÉSAR VALLEJO**

FACULTAD DE INGENIERÍA Y ARQUITECTURA  
ESCUELA PROFESIONAL INGENIERÍA AMBIENTAL

**Revisión sistemática: estudio de métodos para la sustitución de  
datos meteorológico no registrados en calidad de aire.**

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniera Ambiental

**AUTORA:**

Carrasco Villena, Gloria Alejandra (ORCID: 0000-0003-2657-6035)

**ASESOR:**

Mg. Ugarte Alvan Carlos Alfredo (ORCID: 0000-0001-6017-1192)

**LÍNEA DE INVESTIGACIÓN:**

Sistema de Gestión Ambiental

LIMA – PERÚ

2020

## **Dedicatoria**

A mis padres por haberme forjado como la persona que soy en la actualidad y estar siempre apoyándome, muchos de mis logros se lo debo principalmente a mi madre Rosa Villena Ramírez, a mi amada abuelita Alejandrina Ramírez, a mis sobrinos Diego y Andrea, a mis hermanos Omar, Arturo y Fanny, a toda mi familia, a mis amigos que siempre estuvieron a mi lado, hasta llegar a lo que se concreta hoy, el primer peldaño del éxito.

## **Agradecimiento**

Principalmente a Dios y la Virgen, por bendecir cada día de mi vida. A mis padre Alfredo Carrasco Rodríguez y mi madre Rosa Villena Ramírez, gracias por creer en mí siempre, por su confianza, enseñándome a valorar todo lo que tengo, por hacer de mi la persona y profesional que soy en estos momentos por darme ese gran ejemplo de lucha, compromiso y perseverancia, que hubiera sido sin ustedes muchas gracias por todo, a mi abuelita Alejandrina Ramírez, gracias por siempre estar a mi lado en los buenos y malos momentos gracias por tus enseñanzas, a mi asesor Mg. Ugarte Alvan Carlos Alfredo y el Ingeniero Fernando Antonio Sernaque Auccahuasi, por el apoyo, consejos y asesoramiento para esta investigación, también a mis queridos hermanos Omar, Arturo y mi cuñada Fanny, gracias por sus consejos y compañía, a Diego y Andrea por hacer mis días tan felices, a toda mi familia muchas gracias y en especial a la familia Ramírez Torres, gracias tíos queridos por sus consejos y por su apoyo de siempre, gracias también a mis buenos amigos Karen M, y Bryan A., que siempre estuvieron a mi lado en las buenas y malas, por la compañía en estos últimos años, los cuales recordare siempre.

## ÍNDICE DE CONTENIDOS

Carátula .....	i
Dedicatoria .....	ii
Agradecimiento .....	iii
Índice de contenidos.....	iv
Índice de tablas.....	v
Índice de gráficos .....	vi
RESUMEN.....	vii
ABSTRACT.....	viii
I. INTRODUCCIÓN.....	1
II. MARCO TEÓRICO .....	9
2.1. Calidad de aire.....	9
2.2. Contaminantes.....	9
2.3. Datos faltantes.....	10
2.4. Sustitución de datos.....	11
2.5. Tipos de métodos.....	12
2.6. Indicadores de adecuación.....	16
III. METODOLOGÍA .....	17
3.1 Tipo y diseño de investigación .....	17
3.2 Categorías, subcategorías y matriz de categorización apriorística .....	19
3.3 Escenario de estudio .....	22
3.4 Participantes .....	22
3.5 Técnicas e instrumentos de recolección de datos .....	22
3.6 Procedimientos .....	24
3.7 Rigor científico.....	25
3.8 Método de análisis de datos.....	27
3.9 Aspectos éticos .....	28
IV. RESULTADOS Y DISCUSIÓN .....	28
V. CONCLUSIONES .....	48
VI. RECOMENDACIONES .....	49
REFERENCIAS .....	50
ANEXOS.....	58

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Antecedentes.....	4
<b>Tabla 2.</b> Matriz de categorización .....	19
<b>Tabla 3.</b> Clasificación de métodos .....	29
<b>Tabla 4.</b> Métodos analizados en los artículos encontrados .....	31
<b>Tabla 5.</b> Métodos más utilizados y presentaron mejores resultados en las investigaciones.....	37
<b>Tabla 6.</b> Métodos que se ajustan y aplican mejor al contaminante PM10, para relleno de data faltante.....	41
<b>Tabla 7.</b> Métodos que se ajustan y aplican mejor al contaminante PM2,5, para relleno de data faltante.....	46
<b>Tabla 8.</b> Métodos que se ajustan y aplican mejor a los contaminantes NOX,O3,SO2,C,CO,NO2, para relleno de data faltante.....	48

## ÍNDICE DE GRÁFICOS

<i>Gráfica 1. Cantidad de artículos por País .....</i>	<i>30</i>
<i>Gráfica 2. Cantidad de veces que los métodos fueron aplicados en las investigaciones ...</i>	<i>35</i>
<i>Gráfica 3. Los métodos más utilizados para la sustitución de datos meteorológicos no registrados en calidad de aire .....</i>	<i>50</i>

## RESUMEN

Debido a los problemas que pueden generarse en las estaciones de vigilancia hay datos que no se registraron o se perdieron por algún problema técnico o error humano, generando así datos incompletos, la evaluación de la calidad del aire mediante la medición de la concentración de contaminantes debe ser precisa para evitar daños a la salud y al medio ambiente. Por lo tanto, es necesario completar la información que falta para realizar cualquier análisis, a fin de homogeneizarla y validarla mediante diferentes métodos. Los objetivos del presente estudio fueron identificar, clasificar y detallar los métodos más utilizados para la sustitución de los datos meteorológicos no registrados en la calidad del aire y a partir de ellos poder encontrar en los diferentes artículos los métodos más utilizados que generaron mejores resultados para ser aplicados en otros estudios en el futuro. La conclusión final es que estos métodos se dividen en dos tipos: paramétricos y no paramétricos, los métodos que dieron mejores resultados son la interpolación lineal (LI) siendo este de tipo paramétrico y el método del promedio por encima y por debajo de este último es de tipo no paramétrico.

**Palabras claves:** *Métodos de sustitución, data de calidad de aire, contaminantes del aire.*

## **ABSTRACT**

Due to the problems that can be generated in monitoring stations there are data that were not recorded or were lost due to some technical problem or human error, thus generating incomplete data, the assessment of air quality by measuring the concentration of pollutants must be accurate to avoid damage to health and the environment. Therefore, it is necessary to complete the missing information to perform any analysis, in order to homogenize and validate it through different methods. The objectives of the present study were to identify, classify and detail the most used methods for the substitution of the meteorological data not registered in the air quality and from them to be able to find in the different articles the most used methods that generated better results to be applied in other studies in the future. The final conclusion is that these methods are divided into two types: parametric and non-parametric, the methods that gave better results are the linear interpolation (LI) being this one of parametric type and the method of the average above and below this last one is of non-parametric type.

**Keywords:** *Replacement methods, air quality data, air pollutants .*

## I. INTRODUCCIÓN

Actualmente en el área de la meteorología y climatología, según Zakaria Y Noor. (2018) se presentan problemas en el análisis de la base de datos, debido a que, usualmente los registros diarios, mensuales o anuales, de calidad de aire dependen del uso de instrumentos que pueden presentar fallas, ya sean por errores técnico o humano proporcionando así los datos incompletos (p.6)

Es por ello, que la ORGANIZACIÓN METEOROLÓGICA MUNDIAL se encarga de recomendar guías para el buen manejo de la información, con el propósito de estimar datos faltantes, puesto que existen procesos estadísticos que permiten considerar propiedades físicas de la estación de vigilancia, con el fin de permitir la obtención de valores o datos coherentes y realistas, para evitar daños a la salud (OMM,2011, p.47)

En este contexto, a nivel nacional se cuenta con el Servicio Nacional de Meteorología e Hidrología (SENAMHI), la cual se encargada de brindar los datos referentes a la calidad del aire, sin embargo, estos pueden presentar fallas en calidad de sus datos por lo que es necesario completar la información faltante para realizar cualquier análisis, con el fin de homogenizarlo y validarlo por medio de diferentes métodos. Consecuentemente, según Carslaw y Ropkins. (2012) con esta información recopilada se puede confirmar el acatamiento de los estándares de calidad del aire permitido, y así establecer las respectivas pautas y restricciones en este ámbito, por lo que se puede obtener demasiada información del análisis de datos innovando con nuevas técnicas. (p.1)

Uno de los principales problemas actualmente, según, Emanuel, Alves y Gomes. (2019) para analizar series largas de datos (series temporales) es el déficit de una base de datos diarios, mensuales y anuales, consistentes de estaciones meteorológicas, dado que esto resulta un problema a la hora de analizar, pues al tener valores incompletos la base de datos resulta poco fiable (p.2)

Las mediciones proporcionadas de las estaciones de vigilancia de la calidad de aire, para Sozzi et al. (2017) son la principal fuente de información, pero la disponibilidad de estos datos es a menudo limitada por varios problemas técnicos y operacionales. (pp.5), para el IGP, uno de los principales dilemas, es la carencia

de reconocimientos de las herramientas de los modelos numéricos climáticos, debido a que sin el conocimiento de estos no se permitirá poder generar futuros escenarios de cambio climáticos. (IGP, 2020, pp.1)

Por lo tanto, en este trabajo se presentaron varios métodos que se utilizan la sustitución de datos meteorológicos no registrados en calidad de aire. Asimismo, el presente estudio pretende dar respuesta al siguiente problema general ¿Cuáles son los métodos más utilizados para la sustitución de datos meteorológicos no registrados en calidad de aire? además de saber los problemas específicos:

- ¿Cómo se clasifican los métodos para la sustitución de datos meteorológicos no registrados en calidad de aire?
- ¿Cuáles son los métodos que se utilizan para la sustitución de datos meteorológicos no registrados en calidad de aire?

Según la Dra. Silva, Yamina. (2020) investigadora peruana del Instituto Geofísico del Perú (IGP) las principales herramientas para poder generar futuros escenarios de cambio climáticos, son los modelos numéricos climáticos, cabe resaltar que al tener datos no registrados, no se podría confirmar el acatamiento de los estándares de calidad del aire permitido, se debe tener conocimientos de estos métodos, ya que así se podrían dar una predicción más real y apropiada, además de comprender mejor los fenómenos naturales y poder contribuir a una buena gestión del riesgo de desastres, así mejorar los pronósticos meteorológicos en cualquier lugar del mundo (pp.3) el presente trabajo permitirá mostrar que métodos se generaron mejores resultados y se podrían utilizar para rellenar los datos faltantes de registros tomados para contaminantes como el PM10, PM2.5, CO, SO2, C,O3, NOX y NO2,

De acuerdo a la problemática expuesta anteriormente, este estudio presenta como objetivo general identificar los métodos más utilizados para la sustitución de datos meteorológicos no registrados en calidad de aire. Para ello se presentan los siguientes objetivos específicos:

- Clasificar los métodos utilizados para la sustitución de datos meteorológicos no registrados en calidad de aire.
- Detallar los métodos utilizados para la sustitución de datos meteorológicos no registrados en calidad de aire.

Por lo tanto, se realizó una exhaustiva revisión teórica la cual se presenta en el siguiente cuadro.

**Tabla 1. Trabajos Previos: Antecedentes Internacionales.**

N°	PAÍS/AÑO	MÉTODOS USADOS	CONTAMINANTE	PORCENTAJE DE DATA FALTANTE	OBSERVACIONES	AUTOR
1	Colombia/2020	1. Análisis de Componente principales funcional ACPF	(concentraciones de PM10)	Autor no específico	En el estudio se utilizó el método ACPF, para imputar los datos faltantes de las concentraciones de PM10, obtenido como resultado que este método se puede aplicar y ajustar muy bien a este tipo de estudio.	(Meléndez, Bolívar y Rojano 2020)
2	China/2020	1. La matriz de bajo rango 2. El vecino más cercano 3. La sustitución de la media 4. El método basado en la regresión Lineal 5. La interpolación spline	(NOX, O3, SO2, PM2.5, PM10)	Data faltante 30%	El estudio comparo al método de la matriz de bajo rango con otros 6 métodos, con el objetivo de demostrar el rendimiento de este, obteniendo como resultado que el método de la matriz de bajo rango supera a los métodos comparadas, generando buenos resultados al remplazar los datos faltantes.	(Liu et al. 2020)

		6. El método espectral				
		7. El algoritmo de maximización de las expectativas				
3	China/2020	1. Función ortogonal empírica restringida de ciclo diurno DCCEOF	(contracciones de PM2.5)	Data faltante 42.6 %	En el estudio se utilizó el método DCCEOF, con el objetivo de llenar los datos faltantes de PM 2.5 por hora, presentando una buena precisión de los picos diarios o mínimos diarios que no pueden ser restaurados por enfoques de interpolación convencionales	(Bai et al. 2020)
4	Malaysia/2020	1. Maximización de las expectativas (EM) 2. Método del Vecino Más Cercano (NNM)	(contracciones de PM10)	Data faltante en 5%, 10%, 15%, 25% y 40%.	En este estudio se utilizaron 2 métodos, con el objetivo de comprar estos métodos, usando datos faltantes de la calidad de aire de partículas de PM10 de 5 estaciones, llegando a la conclusión de que en método más preciso es el método del vecino más cercano.	(Rumaling et al. 2020)
5	China/2018	1. Minimización de la norma nuclear global (GNNM) 2. Minimización de la norma	(contracciones de PM2.5)	Autor no específico	En este estudio se utilizaron 2 métodos, basados en la norma nuclear de una matriz, con el objetivo de determinar la eficacia de ambos, en relación de rellenar los datos faltantes de calidad de aire	(Chen y Xiao 2018)

		nuclear local (LNNM)		(PM 2.5), llegaron a la conclusión ambos métodos son eficaces, sin embargo, el de mejor rendimiento resulto ser el de minimización de la norma nuclear local (LNNM)		
6	Malaysia /2018	<ol style="list-style-type: none"> <li>1. La media por encima y por debajo</li> <li>2. El vecino más cercano</li> <li>3. Imputación múltiple</li> <li>4. Regresión lineal</li> </ol>	(PM10, CO, SO2, O3, NOX, NO2)	Datos faltantes 5%, 10%, 15% y 20%.	<p>En el presente estudio se utilizaron 4 métodos para rellenar los registros de datos perdidos de calidad de aire, con el objetivo de determinar cuál es el método más apropiado, el resultado del análisis demostró que el más eficiente y con menos margen de error, es el método la media por encima y por debajo</p> <p>En el estudio se aplicó el método de estimación lineal imparcial (THE BLUE ESTIMATOR), aplicado en datos faltantes de mediciones de calidad del aire en tierra de PM10, con el objetivo de probar su efectividad, el cual fue aprobado.</p>	(Zakaria y Noor 2018)
7	Italia/2017	<ol style="list-style-type: none"> <li>1. Estimación lineal imparcial (THE BLUE ESTIMATOR)</li> </ol>	(contracciones de PM10)	Autor no específico	<p>En este estudio se utilizaron 9 métodos, para rellenar los datos faltantes de la calidad de aire para concentraciones de PM10, con el objetivo de determinar el método más sencillo y preciso, el resultado</p>	(Sozzi et al. 2017)
8	Malaysia/2015	<ol style="list-style-type: none"> <li>1. Interpolación lineal (LI)</li> <li>2. La media por encima y por debajo (MAB)</li> </ol>	(contracciones de PM10)	Datos faltantes de 5 % y 15 %.		(Noor, Yahaya, et al. 2015)

				<ul style="list-style-type: none"> <li>3. El vecino más cercano</li> <li>4. Media de la fila (RM)</li> <li>5. Año anterior (PY)</li> <li>6. Media diaria (DM)</li> <li>7. 12-h media(12M)</li> <li>8. 6-h media (6M)</li> <li>9. Imputación múltiple (MI)</li> </ul>	<p>del análisis determinó que el método de interpolación lineal (LI) y la media de 6 horas (6M) es adecuada para llenar los huecos que faltan</p>	
9	Chile/2015	1. Método kriging ordinario	(contracciones de PM10)	Autor no específico	<p>En este estudio se utilizó el Método kriging ordinario para rellenar data faltante de PM10, con el objetivo de poder encontrar los valores estimados de exposición, dando como resultado que este método ayuda mucho a este tipo de predicción.</p>	(Fuenzalida, D, 2015)
10	Malaysia/ 2015	<ul style="list-style-type: none"> <li>1. Interpolación Lineal</li> <li>2. El método de la media</li> </ul>	PM10, SO2, O3 y C.	Data faltante en 5%, 10%, 15%, 25% y 40%.	<p>En este estudio se realizaron 2 métodos para rellenar datos faltantes, con el objetivo de encontrar el mejor de ellos, el estudio concluyó que método de interpolación lineal dio mejores</p>	(Noor, Al Bakri Abdullah, et al. 2015)

11	España/2015	1. Método kriging universal	(contracciones de PM10)	Autor no específico	<p>resultados para todos los porcentajes de los valores perdidos. En este estudio se realizó el método Kriging universal, teniendo como objetivo demostrar la efectividad de este método para rellenar los datos faltantes de PM10, la cual tuvo como resultado afirmativo.</p>	(Cañada, M, Moreno,A. Y Gonzales, H. 2015)
12	Malaysia/2014	<ol style="list-style-type: none"> <li>1. La media por encima y por debajo (MAB)</li> <li>2. Media por encima (MA)</li> <li>3. Método de la media</li> </ol>	(contracciones de PM10)	Data faltante en 5%, 10%, 15%, 25% y 40%.	<p>En este estudio se realizaron 3 tipos de métodos para rellenar datos faltantes de concentraciones de PM10, teniendo como objetivo encontrar la adecuada forma de estimar los datos faltantes, el estudio concluyo que el método de la media por encima y por abajo (MAB), es el mejor para predecir estos datos faltantes.</p>	(Noor et al. 2014b)
13	Malaysia/ 2014	<ol style="list-style-type: none"> <li>1. Interpolaciones lineales</li> <li>2. Interpolación cuadráticos</li> <li>3. Interpolación cúbicos</li> </ol>	(concentraciones de PM10)	Data faltante 2,1% de 8577 Concentraciones horarias.	<p>En este estudio se utilizó 3 métodos, con el objetivo de llenar datos faltantes de las concentraciones de PM10, el cual obtuvo como resultado que los tres métodos son apropiados generando un valor de R2 muy alto.</p>	(Noor et al. 2014a)

Fuente: Elaboración propia

## II. MARCO TEÓRICO

Según la OMS, si se habla de la calidad de aire, se basa en el acatamiento de los estándares de calidad ambiental de aire, los que determinan los niveles para la concurrencia de los contaminantes en el aire, puesto que si estos se mantienen a bajos niveles no simbolizan un riesgo para el ambiente ni a la salud (OMS,2005,pp.9) algunos de estos contaminantes los cuales son monitoreados en las estaciones de vigilancia mediante instrumentos, como el PM10 Y PM2.5, según Joanne Y Sánchez.(2013) es aquella mezcla de partículas minúsculas ya sean sólidas o líquidas, las cuales se localizan en el aire que respiramos, en cuanto al Ozono (O<sub>3</sub>), es un gas puede acusar problemas en el sistema respiratorio aun estando a bajos niveles, este es constituido en la atmósfera mediante reacciones fotoquímicas, y en presencia de la luz solar, como también compuestos orgánicos volátiles y por el óxido de nitrógeno (NO<sub>x</sub>) (pp.9)

Otro contaminante para Joanne Y Sánchez.(2013) es el dióxido de nitrógeno (NO<sub>2</sub>) que se forma por la quema de combustibles fósiles a altas temperaturas, este gas presenta a una de sus principales fuentes de emisión al sector automotriz, otro gas que se forma gracias al sector automotriz pero en menor grado, es el SO<sub>2</sub>, también presenta otra fuente de emisión que es combustión de combustible fósiles pero esta vez en las plantas que generan energías, de mismo modo como en las instalaciones industriales (pp.9) estos contaminantes son monitoreado en las estaciones de vigilancias, uno de los problemas que presentan estos registros son los datos faltantes, para Norazian et al. (2008) normalmente se presentan debido a muchos factores como fallas de los instrumentos, cambios en la ubicación los monitores, el mantenimiento de rutina y el error humano. El conjunto de datos incompletos suele causar sesgos debido a diferencias entre lo observado y lo no observado datos. Por lo tanto, es importante asegurar que los datos analizados sean de alta calidad (pp.1), para ello, se aplican procesos que parten de la sustitución de datos, para Medina Galván, es un proceso que posibilita generar estimadores inclinados que alteran las relaciones de origen entre las variables.

Durante los últimos años se han propuesto distintas metodologías con la finalidad de sustituir datos faltantes; no obstante, es frecuente que estos procedimientos se

utilicen sin tener en cuenta sus limitaciones prácticas y fundamentos teóricos, (Medina, F. y Galván, M. 2007, pp.9)

Este tipo de proceso da lugar a muchos métodos, los cuales se usan con el propósito de poder sustituir, aquellos valores faltantes de una base de datos.

Según Viada, Bouza y Fors, estos métodos no acostumbra tener una base teórica sólida, pese a que, son fáciles de llevar a cabo, así como presentan en general importantes inconvenientes y carencias, especialmente en contextos multivariantes (Viada, C., Bouza, C. y Fors, M. ,2016, pp.114)

Para, EK et al. (1997) se conoce como métodos paramétricos y no paramétricos, aquellos que son utilizados como predicción puesto que la imputación permite reemplazar los valores faltantes o datos no registrados de cualquiera unidad de estudio con mediciones otra unidad con similares características (pp.129)

Para Salas et al. (2010) entre los métodos paramétricos tenemos a la regresión lineal, regresión múltiple, los mínimos cuadrados, estos métodos se ajusta a algunas condiciones estrictas, como el requisito de que los datos de su muestra vengan de una base de valores normalmente distribuidos. Entre los métodos más usados tenemos a la regresión lineal, análisis de covarianza, de varianza factorial y unidireccional, la regresión de mínimos cuadrados parciales son métodos, la regresión lineal múltiple y de efectos mixtos son algunos de los métodos paramétricos (pp.11) en cuanto a los no paramétricos al vecino consiste en generar la tendencia de un punto llamado X considerando así un promedio ponderado de respuestas llamado Y, de esta manera da menor peso a las respuestas lejanas y mayor peso a las cercanas, este método es aplicable a variables de tipo original y nominal, estas no necesariamente tiene que estar medidas en intervalos o razón y si requieres un análisis estas variables deben ser categóricas, los métodos más utilizados con el chi cuadrado, coeficiente de correlación para tabulaciones cruzadas, los métodos necesitan el uso de computadoras para poder así realizar sus cálculos, estos permiten analizar el efecto de 2 a más variables independientes sobre las variables dependientes (pp.14)

En cuanto a los métodos de interpolación, para Chapra Y Canale. (1989) el método de Interpolación lineal, consiste en unir 2 puntos de datos con una línea recta,

además de ser la forma más simple de interpolación, de esta manera el valor que falta se puede calcular usando la ecuación lineal, donde  $x$  simboliza la variable dependiente mientras  $y$  simboliza la variable independiente (pp.20) mientras que el método de Interpolación cuadráticos, utiliza un polinomio de segundo orden, que dispone de 3 puntos de datos, donde  $x$  es la variable independiente,  $x_0$  y  $x_1$  son valores conocidos de la variable independiente,  $b_0$  y  $b_1$  son los coeficientes desconocidos y  $f_2(x)$  es el polinomio interpolador de segundo orden se hallar con la ecuación lineal (pp.23), el método de Interpolación cúbicos, es la estimación al igual que las anteriores dispone de 4 puntos de datos, en un polinomio llamado cubio o de tercer orden, donde la ecuación para hallar los coeficientes  $b_0$ ,  $b_1$  y  $b_2$  es el mismo que de los anteriores (pp.25). Para Ramsay y Silverman, la interpolación spline, son los polinomios de orden en un punto de vista alternativo, es dividir el intervalo de aproximación en un conjunto de subintervalos y construir un polinomio de acercamiento distinto en cada subintervalo. Esto se llama acercamiento polinomial por tramos. (Ramsay, J. y Silverman, B. 2005, pp.48)

Según la OMM, las ventajas de una media consisten en un patrón de cifra práctico por lo que se refiere a las fluctuaciones de las observaciones (dado que la adición de las desviaciones con respecto de una media es 0), que se calcula desde luego, también de que pueden combinarse las medias relativas a diferentes conjuntos no superpuestos del letanía terminado de las observaciones, y que el descuido en una apreciación de una media a partir de una muestra es menor que otras medidas de la tendencia cardinal (OMM,1990.pp.23)

Para Noor et al. (2015) los métodos de la media se presentan los siguientes, si hablamos del método de media por encima y por debajo, este permite calcula los datos perdidos con la media del punto por encima de la zona faltante y el punto debajo de la zona faltante (pp.926), según, Noor et al. (2014) en otro artículo publicado en el 2014, habla acerca del método de Media por encima, donde dice que este sustituye todos los valores faltantes por la media de todos los datos disponibles por encima de los valores faltantes (pp.905) con relación al método de la media, sustituye todos los valores que faltan por la media de todos los datos disponibles (pp.904) así mismo, para Plaia Y Boni, la Media de la fila (RM), utiliza

información horaria conocida sobre los niveles de concentración de PM10 en la otra estación, este método de imputación rellena un valor perdido calculando la media de los registros conocidos sobre la misma matriz de datos, quise decir que computa la media de los valores registrados al mismo tiempo por otro sitio de monitoreo (Plaia, A. y Bondi, A. 2006, pp.5)

La sustitución de la media, según, Cohen, M. (1996) este método permite sustituir el valor faltante mediante la media de los valores existentes y válidos, utilizando la media muestral de los datos disponible como donador en cada uno de los datos perdidos, este método es simple pero también atractivo ya que facilita estimaciones inclinadas para la media poblacional. Su modificación presenta aumentar más variabilidad a los datos imputados mediante el método de la media usando la variabilidad de los datos muestrales, consiste en imputar la mitad de los datos faltantes con los valores. (pp.293), para Weiss y Hays. (2005) los métodos de media en horas se presentan la Media diaria (dm), este método es la media de los registros calculados cada 24 horas, las observaciones faltantes son reemplazadas por el valor medio de las 24 horas de esa hora en particular del dato que falta (pp.57) ,respecto a la media 12-h (12M), este método es la media calculada cada doce horas, los valores perdidos se imputan con el valor de la media de 12 horas para esa hora específica perdida de la brecha. (pp.63) en cuanto a la media 6-h (6M), este método es la media calculada cada 6 horas, los valores perdidos se imputan con el valor de la media de 6 horas para esa hora específica pérdida de la brecha. (pp.65)

Según, Olivier y Hanginang, El vecino más cercano (NN), reemplaza los datos que faltan por el valor del punto de partida de cada zona, este permite ajustar los datos con valores conocidos, para así identificar y eliminar los valores que se consideran irrelevantes (Olivier, R. Y Hanqiang, C. 2012, pp.1), referente al método del Año anterior (PY), este utiliza la información conocida de las concentraciones del año anterior para así rellenar el vacío que faltante. Según Roderick y Donal, estos datos se imputan con el valor conocido de las observaciones según las estaciones para la hora específica de la desaparición. (pp.219) sobre la Imputación múltiple (MI), este sustituye cada valor que falta por un grupo de valores posibles que introdujeron

un error aleatorio apropiado en el proceso de imputación de hace que sea posible obtener cálculos aproximado e imparciales de todos los parámetros (Roderick, L. y Donal, R. 2019, pp.232)

En cuanto a métodos lineales con relación a la estimación lineal Imparcial (The Blue Estimator), Para, Albert, el método lineal con los valores determinados, permite extraer una muestra aleatoria de un conjunto y usarla para estimar los valores faltantes, las estimaciones del coeficiente de mínimos cuadrados siguen la distribución de muestreo más ajustada posible de estimaciones imparciales (Albert, A. 1973, pp. 182), además de la regresión lineal, que permite la relación de dos variables ajustadas en una ecuación lineal a los datos observados, los valores que faltan de los datos se remplazasen por una regresión de variables no observadas frente a las observadas para ese conjunto de datos. Según Tusell, F. (2003) esto se debe a que los modelos que dependen linealmente de sus parámetros desconocidos son más fáciles de encajar que los modelos que no están relacionados (pp.120)

Según Chen y Xiao, de métodos de la norma nuclear se presentan los siguientes Minimización de la norma nuclear global (GNNM), este método es simple, pero puede producir una estimación subóptima de valores faltantes, especialmente cuando los datos no son muy fuertes dependientes de una manera global, permitiendo minimizar la norma nuclear directamente sobre los datos de calidad del aire con el fin de imputar los valores que falten. Entonces este método trata de imputar los valores perdidos minimizando directamente la norma nuclear de toda la matriz de la muestra, al mismo tiempo maximizando la dependencia lineal de las muestras, en consideración a la Minimización de la norma nuclear local (LNNM), este se concentra más en cada muestra y sus muestras más similares que se estiman a partir de los resultados de imputación del método anterior. De esta manera, la minimización de la norma nuclear puede realizarse en esas muestras altamente correlacionadas en lugar de toda la matriz de la muestra como en el GNNM, reduciendo así el impacto adverso de las muestras irrelevantes (Chen, X. y Xiao, Y. 2018, pp.3)

El método de maximización de las expectativas (EM), emplea un conjunto de ecuaciones iterativas para así estimar la matriz de covarianza de la multivariante distribución de la familia exponencial y el vector medio, este método maximiza la probabilidad de encontrar parámetros cuando hay valores perdidos. (Schneider, T 2001, pp.34)

Si hablamos del método de función ortogonal empírica restringida de ciclo diurno DCCEOF, según Bai et al, este se desarrolló para manejar mejor las lagunas de datos presentes en las series temporales con marcada variabilidad en el espacio y el tiempo, tomando el patrón de variación diurna como una restricción crítica en la predicción de valores perdidos. A nuestro conocimiento, ninguno de los métodos de relleno de lagunas existentes ha dado cuenta del patrón de variación diurna de los datos dados en sus planes de restauración de valor perdido, y por lo tanto los valores previstos de esos métodos estarían sujetos a un gran sesgo (Bai et al. 2020, pp.3)

Para Candès y Recht, la matriz de bajo rango, reconstruye una matriz partir del subconjunto observado de su entrada basada en la propiedad de bajo rango de la matriz original. Llena a los datos no observados que hace que la matriz tenga un rango inferior con los datos observados sin cambio, dada una matriz espacial incompleta de datos con algunos valores ausentes. (Candès, E. y Recht, B. 2009, pp.278)

Según, Ramsay Y Silverman. (2005) el método basado en la regresión, se basa en la regresión predicen los valores utilizando otras variables como predictores. La limitación común del modelo de regresión es que los parámetros en el modelo son sensibles al muestreo y a los valores irregulares (valores atípicos o anomalías presentadas en los conjuntos de datos de los contaminantes atmosféricos) que son difíciles de eliminar por medio del propio modelo (pp.32) Así mismo para Chávez, Sánchez y De La Cruz, el método de análisis de componente principales funcional ACPF, se pueden ver sencillamente como una sustitución de los vectores por funciones, las matrices de covarianza por operadores de covarianza, las matrices por operadores lineales y los productos escalares en espacios vectoriales por productos escalares en espacios funcionales. Este análisis permite demostrar la forma en que dicho conjunto de datos funcionales puede variar y este sentido de

variación se cuantifica la discrepancia con respecto a la media de cada dato función (Chávez, C., Sánchez, J. y De La Cruz, J. 2005, pp.15) para Ramsay Y Silverman. (2005) este método no paramétrico, donde se utilizan la estimación de Kernel. (pp.15)

Según, Ramón, et. al, el método Kriging universal es un estimador permite reducir la varianza del error de predicción, ya que este utiliza al estimar la correlación espacial y variabilidad del fenómeno estudiado, la cual presenta una metodología de interpolación que estaba basada en un grupo de algoritmos de regresión coincidente por mínimos cuadrados este presenta una función aleatoria y esta se descompone en una componente de tendencia y otra residual. (Ramón, et. al, 2010, pp. 68)

Kriging ordinario, según Cañada, et al, este método se utiliza para aquellos datos que no presentan tendencia y presentan medidas desconocidas, este uso covarianza o también llamada semivariogramas, los cuales permiten transformar esos datos, eliminar y proporcionar medidas de error, cuando la forma de la curva se presenta cerca al origen y este difiere significativamente, quiere decir que cuando más empinada sea esta curva y este cerca al origen, más predominio a los vecinos más cercanos tendrá la predicción (Cañada, et al., 2006,pp. 823)

Según, Sheldon, M. (2010) para poder evaluar la bondad de estos métodos utilizados en las investigaciones, se utilizaron indicadores de adecuación tal como, el coeficiente de determinación ( $r^2$ ) este valor se utiliza como indicador para ver el modelo de regresión utilizado se acopla bien a los datos, ya que si presenta un valor cercano a 1 nos indica que es un buen ajuste y mientras se acerca a 0 indica que es un mal ajuste. (pp.578)

Con respecto al segundo indicador, error cuadrático de la raíz (RMSE), según, Simón, et al, permite comparar errores de predicción de diferentes modelos de una variable en particular y no entre variables, porque depende de la escala, por eso es una buena medida de precisión. (Simón, et al, 2018, pp.43).

Error absoluto medio (MAE), según, Wang y Lu, es la diferencia de dos variables continuas que, este utiliza la misma escala que los datos que se miden, y no se

puede utilizar para comparar mediciones de diferentes escalas, su valor disminuye cada que redondeamos el índice previsto al vecino más cercano. (Wang, W. y Lu, Y. 2018, pp.3)

### **III. METODOLOGÍA**

#### **3.1. Tipo y diseño de investigación**

Este estudio presento una investigación de tipo aplicada, para Salinas, P, la investigación aplicada, es llamada así debido a que sus resultados se pueden utilizar para una solución inmediata y directa de los problemas que les afecta, de tal forma que este tipo de investigación es empleada en las ciencias como la ingeniería, medicina, la agronomía, la arquitectura, la mineralogía y muchas más. (Salinas, P. 2010, pp.17)

Según Vargas, Z. (2009) la investigación de tipo aplicada, puede ser llamada también, investigación empírica dado que esta busca una aplicación de los conocimientos adquiridos, y asegurar el saber para resolver una situación, o llamada investigación práctica, se identifica porque busca la aplicación de los conocimientos ya adquiridos (pp.23), no obstante, según Eisenhower, W. et al, está enfocada básicamente a un objetivo practico especifico. (Eisenhower, W. et al, 2018, pp.20), así mismo, para Rodríguez, D. Se centra particularmente de qué manera se pueden llevar a la práctica las teorías generales, su enfoque va directamente a la resolución de problemas que se trazan en un momento dado, de manera que estos problemas son conocidos y establecidos por el investigador, de modo que este utiliza la investigación aplicada para dar respuesta a aquellas preguntas específicas (Rodríguez, D.2019, pp.1)

Asimismo, Rodríguez, D. esta guarda una corta relación con la investigación de tipo básica, en vista que depende mucho de los descubrimientos de la investigación básica y potencia de dichos hallazgos y así poder conocer las realidades con una prueba científica esto quiere decir que esta se basa en sus resultados obtenidos. (Rodríguez, D.2019, pp.3)

Según, Rodríguez, D. (2019) los objetivos principales de este tipo de investigación es pronosticar un comportamiento determinado en una situación precisa. Del mismo modo una de sus características resaltantes es que requiere necesariamente de un marco teórico, ya que de él se basara para generar de esta manera una solución a dicho problema específico que se desee resolver. Este tipo de investigación se centra específicamente en un análisis y una solución a los problemas de varias condiciones de la vida real, nutriéndose así de los avances científico (pp.5)

Por lo tanto esta investigación es de tipo aplicada ya que, se tomó las investigaciones de varios autores sobre el estudio de métodos para la sustitución de datos meteorológicos no registrados en calidad de aire, los cuales se identificaron, clasificaron, y detallaron, y así dieron respuesta a aquellas preguntas específicas propuestas anteriormente, finalmente se consideraron los métodos más adecuados, con la intención de que sean utilizados para solucionar el problema que causan los datos faltantes en los registros de monitoreo.

En cuanto al diseño de investigación de narrativos de tópicos, según, Moncada, L. Nos dice que este diseño es una subcategoría del diseño narrativo, este recaba datos de experiencias, para analizadas y descritas, ayudan a aclarar preguntas que no están claras, esta información se obtiene de entrevistas, documentos, biografías, artículos en la prensa, e investiga los sucesos, hechos y los resultados del proceso de investigación, donde el investigador revisa documentos como libros, registros de archivos (Moncada, L. 2018, pp.1)

Para Hernández, R. (2016) el contexto se localiza de acuerdo con el planteamiento del problema, para revisar estos sucesos es muy importante tener varias fuentes de datos, este diseño demanda que se investiguen muchas fuentes útiles para resolver el problema, y así poder llegar a tener su propio instrumento de investigación, ya que el investigador de por si debe estar analizando constantemente los datos que este genere de la investigación y lo describa durante su recolección (pp.468)

De modo que, esta investigación presenta un diseño narrativo de tópicos antes mencionado ya que se recaudó información de investigaciones antes realizadas las cuales sirvieron como escenario de estudio, permitiéndonos generar una solución

a la problemática mencionada al inicio para ello, las fuentes utilizadas son artículos de importancia publicados en revistas científicas reconocidas mundialmente, lo que permite generar resultados veraces.

### 3.2. Categoría, subcategoría y matriz de categorización apriorística

**Título:** Revisión bibliográfica: Estudio de métodos para la sustitución de datos meteorológicos no registrados en calidad de aire.

**Objetivo General:** Identificar los métodos más utilizados para la sustitución de datos meteorológicos no registrados en calidad de aire.

TABLA 2: MATRIZ DE CATEGORIZACIÓN						
OBJETIVOS ESPECÍFICOS	PROBLEMAS ESPECÍFICOS	CATEGORÍA	SUBCATEGORÍA	CRITERIO 1	CRITERIO 2	CRITERIO 3
Clasificar los métodos utilizados para la sustitución de datos meteorológicos no registrados en calidad de aire	¿Cómo se clasifican los métodos para la sustitución de datos meteorológicos no registrados en calidad de aire?	Clasificación de los métodos	Métodos Paramétricos (SALAS ET AL. 2010, PP,11)	De acuerdo al nivel de complejidad	De acuerdo al modelo de ecuación	De acuerdo a aplicación
			Métodos no paramétricos (SALAS ET AL. 2010, PP,14)	(GONZALEZ,2009, PP.3)	(REYES,2010, PP.12)	(ZAKARIA Y NOOR,2018, PP.10)
Detallar los métodos utilizados para la	¿Cuáles son los métodos que se utilizan	Métodos	-Análisis de Componente principales funcional ACPF	De acuerdo al nivel de complejidad	De acuerdo al modelo de ecuación	De acuerdo a su aplicación
			-La matriz de bajo rango -El vecino más cercano -La sustitución de la media. -El método basado en la regresión lineal. -La interpolación spline,	(GONZALEZ,2009, PP.3)	(REYES,2010, PP.12)	(ZAKARIA Y NOOR,2018, PP.10)

---

<p>sustitución de datos meteorológicos no registrados en calidad de aire.</p>	<p>para la sustitución de datos meteorológicos no registrados en calidad de aire?</p>	<ul style="list-style-type: none"> <li>-El método espectral</li> <li>-El método kriging ordinario y universal</li> <li>-El algoritmo de maximización de las expectativas.</li> <li>-Función ortogonal empírica restringida de ciclo diurno DCCEOF</li> <li>-Maximización de las expectativas (EM)</li> <li>-Minimización de la norma nuclear global (GNNM)</li> <li>-Minimización de la norma nuclear local (LNNM)</li> <li>-La media por encima y por debajo</li> <li>-Regresión lineal</li> <li>-Estimación lineal imparcial (THE BLUE ESTIMATOR)</li> <li>-Interpolación lineal (LI)</li> <li>-Media de la fila (RM)</li> <li>-Año anterior (PY)</li> <li>-Media diaria (DM)</li> <li>-12-h media(12M)</li> <li>-6-h media (6M)</li> <li>-Imputación múltiple</li> <li>-El método de la media</li> <li>-Interpolaciones lineales</li> <li>-Interpolación cuadráticos</li> <li>-Interpolación cúbicos</li> </ul>
---	---	--

---

*Fuente: Elaboración propia*

### **3.3. Escenario de estudio.**

Para la investigación se realizó una meticulosa revisión sobre los métodos utilizados para la sustitución de datos meteorológicos no registrados en calidad de aire, con el cual se recurrió a diferentes fuentes de información sobre el tema de estudio, puesto que, ya que se empleó el diseño narrativo de tópicos. Según las investigaciones se desarrollaron de las estaciones de vigilancia de calidad de aire, las cuales presentan valores de datos faltantes en sus registros de monitoreo, los parámetros que se miden en estas estaciones con los de PM10, PM2.5, NOX, O3, SO2 y Co2.

Estas estaciones estudiadas están en distintas ciudades, así como la de Ciudad de Beijing en China, Seberang Perai, Penang, Península de Malasia, La Guajira, Colombia, Malacca, Malasia, Sabah, Malasia, Región del Lazio en Italia y Changzhou, China.

### **3.4. Participantes.**

Para efectos del análisis de la información en esta investigación, se recurrió al uso de artículos científicos, estas fuentes se obtuvieron de Revistas Científicas como Scopus, ScienceDirect, Scielo y Hinari, estas plataformas digitales, las cuales permitieron poder consultar todas las publicaciones más importantes en el mundo, además de libros sobre estadística como Functional Data Analysis y Statistical Analysis with missing data. Por último, se utilizó información de organizaciones Internacionales y nacionales como la Organización Mundial Meteorológica, Organización Mundial de la Salud y el Instituto Geofísico del Perú.

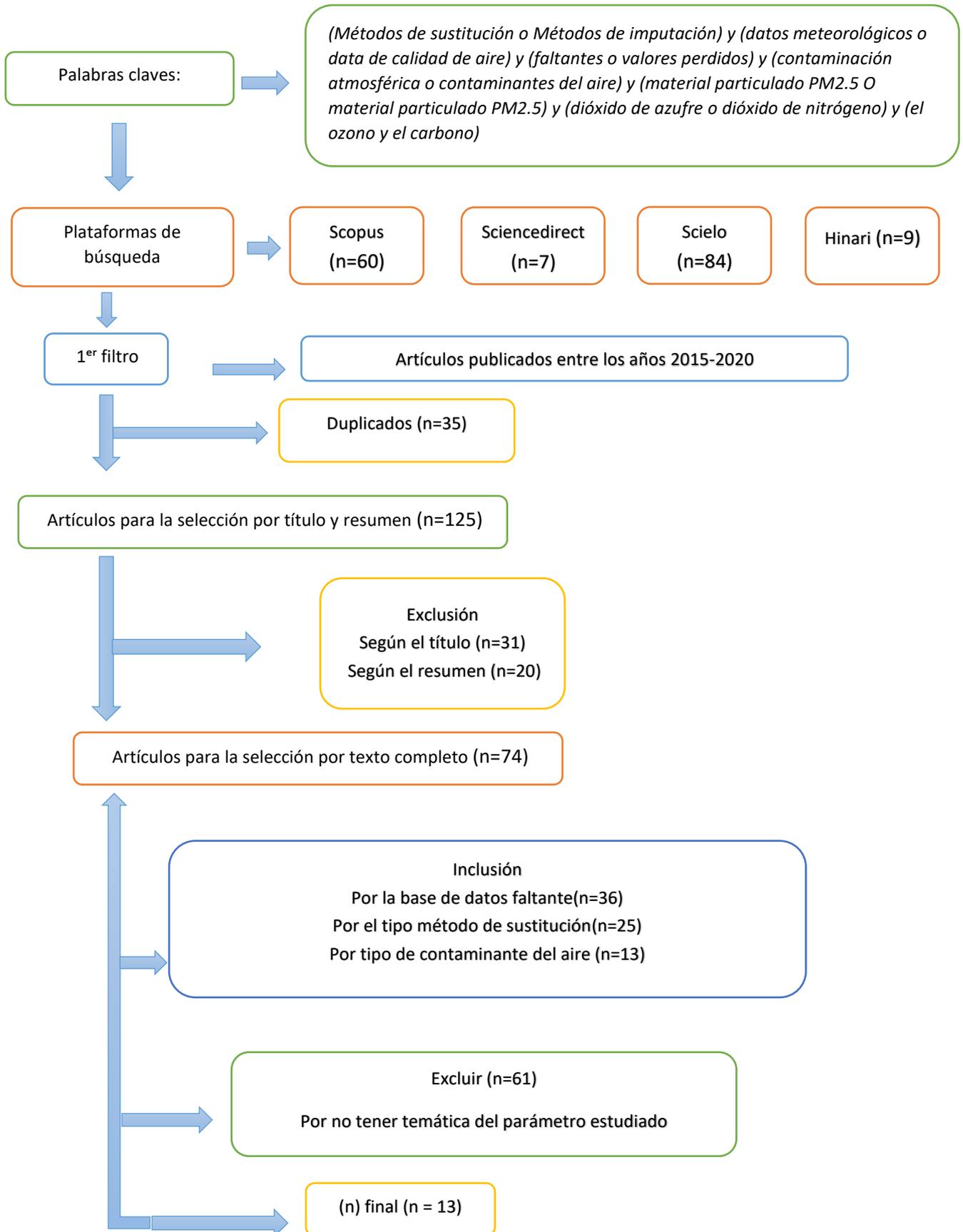
### **3.5. Técnicas e instrumentos de recolección de datos**

La técnica que se empleó en esta investigación es la del análisis de documentos, esta es una parte fundamental de un proceso sistemático de una investigación científica, según, Rodríguez, M, está técnica se realiza mediante la consulta de distintos tipos de documentos como revistas, libros, registros además de revisiones de tipo de diseños, metaanálisis, bibliográfica, revisión de evidencias y narrativas, (Rodríguez, M. 2013, pp.2) siendo esta última la que se utilizó en esta investigación.

El método de recojo de datos según, Orozco y Diaz, nos dicen que se debe seguir un proceso sistemático específico y definido, donde se deben citar las fuentes consultadas y como se obtuvo el acceso a ellas, donde se detalle cómo se realizó la elección de cada uno de los documentos (Orozco, A. y Diaz, pp. 2018, pp.71)

En cuanto a la ficha de análisis de contenido (Ver anexo I), se usó para la recolección de datos, la cual se aplicaron a cada artículo seleccionado para esta investigación, en esta ficha se aprecia datos importantes y útiles como el autor, lugar y año de publicación, así también como la información del contenido de estos, en este caso fue el tipo de método; parámetros de monitoreo, data faltante, resultados y conclusiones para cada uno de ellos, se contó con una información precisa y que apoyo a la investigación.

**3.6. Procedimiento:** Diagrama del proceso de búsqueda de los artículos de investigación.



En la presente investigación, se indagó 160 artículos de forma general, debidamente antes de ello se utilizaron ciertos criterios de inclusión y exclusión, de los cuales 13 artículos quedaron. La selección de estos documentos aporta la información de acuerdo con los objetivos dados, iniciando desde las palabras claves, las cuales fueron utilizadas como guía para la búsqueda de artículos referentes al tema de investigación, que se extrae desde las mencionadas bases de datos, así mismo se excluyeron, los artículos que no estaban dentro de los años 2015 al 2020, como también artículos duplicados, se buscó artículos en los cuales se enfocaron en revisar título y resumen y descartando los menos relevantes, por último los artículos por selección de texto completo y en inglés, de igual manera se tomaron en cuenta artículos que contengan información sobre datos faltantes, métodos de sustitución y tipo de contaminantes, excluyendo los que no presentan una temática del parámetro estudiado, al cumplir con los criterios de inclusión y exclusión se llegó a extraer una información coherente, detallada y garantizada para el estudio.

### **3.7. Rigor científico**

Durante el proceso de indagación cualitativa se ejecutó un trabajo de índole que cumpliera con el rigor de la metodología de la investigación, por lo que se utilizaron los criterios de rigor científico para así lograr la calidad de la investigación.

La credibilidad, Según, Cádiz, J, es la conformación de diferentes fuentes de datos, y la confrontación en el procedimiento y tratamiento de los datos, en la investigación cualitativa, se da un contraste sistemático de la información con audiencias colaboradoras, la cual se somete a juicio crítico de otros investigadores de las interpretaciones y observaciones realizadas (Cádiz, J. 2006, pp. 14)

Por lo tanto, esta investigación, presento el criterio de credibilidad, ya que los artículos consultados son artículos que se realizaron de diferentes fuentes de datos sistemáticos los cuales se han sometido a juicio crítico con otros investigadores, al ser consultados una vez se obtuvieron sus conclusiones, además de que estos fueron extraídos de una base de datos de las revistas científicas de renombre.

La conformabilidad, Según Castillo Y Vásquez. (2003) permite a los investigadores puedan seguir pautas para analizar el proceso de investigaciones y ver como otro

autor puedo llegar al resultado, lográndolo esto mediante la revisión de investigaciones pertinentes y con óptimos estándares de orden y claridad y así varios investigadores puedan llegar a conclusiones similares y explicar la posición de los investigadores (pp.166)

Para Morse, et al. este criterio permite asegurar la calidad del proceso de investigación, como los datos hallados, interpretaciones y recomendaciones, siempre y cuando estos sean coherentes. (Morse, et al. 1994, pp.154)

Por consiguiente, el criterio de conformabilidad, también está presente en la investigación, ya que se siguió putas para analizar cada uno de los procesos de investigación, se reanudo información de revisiones sistemáticas y llegando así a una conclusión similar y explicar la posición de cada uno de los investigadores.

El criterio de transferibilidad, según Guba y Lincoln, este criterio nos indica la capacidad de poder extender los resultados obtenidos de los estudios a diferentes poblaciones (Guba, E. y Lincoln, Y. 1982, pp. 234)

Según, Morse, et al. permite demostrar hasta qué punto pueden llegar adecuarse los resultados con otros ámbitos y también las condiciones con las que se pueden considerar válidos los datos. (Morse, et al. 1994, pp.153)

Por esta razón también se aplicó el criterio de transferibilidad, puesto que se extrajo los resultados obtenidos de otros estudios, y a la vez se demuestra cómo estos resultados se adecuan con otros del mismo ámbito.

Dependencia según, Castillo y Vásquez, es el criterio determina la consistencia lógica de los resultados coherentes obtenidos, siempre y cuando los investigadores no tengan relación personal, se debe evitar, las opiniones y creencias que afecten la sistematización de las interpretaciones así mismo establecer conclusiones, antes de recolectar los datos. (Castillo y Vásquez, 2003, pp.164)

Para, Arias y Giraldo, la dependencia presenta, los procedimientos que aseguran la estabilidad, permitirán así conseguir una menor inestabilidad de los diferentes datos y llegar al objetivo de contar con los mismos resultados a partir de una información similar, mediante los diferentes métodos de recolecciones de datos. (Arias, M Y Giraldo, C 2011, pp. 514)

De manera que el criterio de dependencia permitió, contar con los mismos resultados a partir de una información similar, ya que los investigadores no tienen relación personal, ya que así se tiene una mejor estabilidad en los diferentes datos.

### **3.8. Método de análisis de datos**

El análisis de información de esta investigación, se rigió en 03 criterios, para las categorías, 1) Clasificación de métodos y 2) Métodos, y así mismo sus categorías, como se puede apreciar en la (TABLA II) matriz apriorística.

En cuanto a la primera categoría, clasificación de los métodos, se tomaron de los artículos de investigación encontrados, donde la subcategoría muestra que se clasifican en métodos paramétricos y no paramétricos. Se analizó a partir de criterios los cuales ayudaran a agrupar la información es así que se presentan los siguientes criterios, de acuerdo al nivel de complejidad, al modelo de ecuación y por último de acuerdo a aplicación.

En cuanto a segunda categoría, métodos, de igual manera se tomaron de los artículos de investigación encontrados, donde se identificaron, El vecino más cercano, Interpolaciones lineales, la sustitución de la media, e método basado en la regresión, la interpolación spline, el método espectral, el algoritmo de maximización de las expectativas, función ortogonal empírica restringida de ciclo diurno DCCEOF, maximización de las expectativas (EM), minimización de la norma nuclear global (GNNM), minimización de la norma nuclear local (LNNM), la media por encima y por debajo, regresión lineal, estimación lineal imparcial (THE BLUE ESTIMATOR), interpolación lineal (LI), media de la fila (RM), año anterior (PY), media diaria (DM), 12-h media (12M), 6-h media (6M), imputación múltiple, interpolación cuadráticos e interpolación cúbicos, Análisis de Componente principales funcional ACPF, La matriz de bajo rango y El método de la media. De igual manera la primera categoría estos se analizaron a partir de criterios los cuales ayudaran a agrupar la información es así que se presentan los siguientes criterios, de acuerdo al nivel de complejidad, al modelo de ecuación y por último de acuerdo a aplicación.

### **3.9. Aspectos éticos.**

Esta investigación cumplió a totalidad con los estándares máximos los cuales están establecidos en el código de ética, con resolución N ° 0126-2017/UCV, utilizándola con honestidad y cumpliendo con cada estándar, además de que se respetó toda información extraída de artículos de investigación, citando a los autores de acuerdo a la norma internación ISO 690, la cual nos guía al momento de citar y realizar las referencias bibliográfica, respetando así la autonomía de cada artículo consultado, además de cumplir con valores éticos como responsabilidad, compromiso y disciplina.

## **IV. RESULTADOS Y DISCUSIÓN**

Se obtuvieron 13 artículos los cuales contenían métodos que su utilizaron para el relleno de data faltante presente en registros de monitoreo de calidad del aire. En estos, se encontraron 19 métodos como se puede apreciar en la (TABLA 4) que se compararon entre sí con el objetivo de demostrar cual es el que más apropiado para cada caso. La mayoría de estos artículos utilizaron de 1 a 4 métodos, dos de ellos utilizaron 7 y 9 métodos, estos fueron: (Liu et al. 2020), utilizó 7 métodos, y (Noor, Al Bakri Abdullah, et al. 2015) quien comparo entre 9 métodos (TABLA 4)

Según las investigaciones revisadas, estos métodos se clasifican paramétricos y no paramétricos, aquellos que son utilizados, según EK, et al. como predicción puesto que la imputación permite reemplazar los valores faltantes o datos no registrados de cualquiera unidad de estudio con mediciones otra unidad con similares características (Ek et al.1997, pp.129)

Para, EK, et al. (1997) entre los métodos paramétricos tenemos a la regresión lineal, regresión múltiple, los mínimos cuadrados, estos métodos se ajusta a algunas condiciones estrictas, como el requisito de que los datos de su muestra vengan de una base de valores normalmente distribuidos (pp.132)

Según, Salas et al. entre los métodos más usados tenemos a la regresión lineal, análisis de covarianza, de varianza factorial y unidireccional, la regresión de mínimos cuadrados parciales son métodos, la regresión lineal múltiple y de efectos mixtos son algunos de los métodos paramétricos (Salas et al. 2010, pp.11)

Así mismo los no paramétricos al vecino consiste en generar la tendencia de un punto llamado X considerando así un promedio ponderado de respuestas llamado Y, de esta manera da menor peso a las respuestas lejanas y mayor peso a las cercanas, este método es aplicable a variables de tipo original y nominal, estas no necesariamente tiene que estar medidas en intervalos o razón y si requieres un análisis estas variables deben ser categóricas, los métodos más utilizados con el chi cuadrado, coeficiente de correlación para tabulaciones cruzadas, los métodos necesitan el uso de computadoras para poder así realizar sus cálculos, estos permiten analizar el efecto de 2 a más variables independientes sobre las variables dependientes (pp.14)

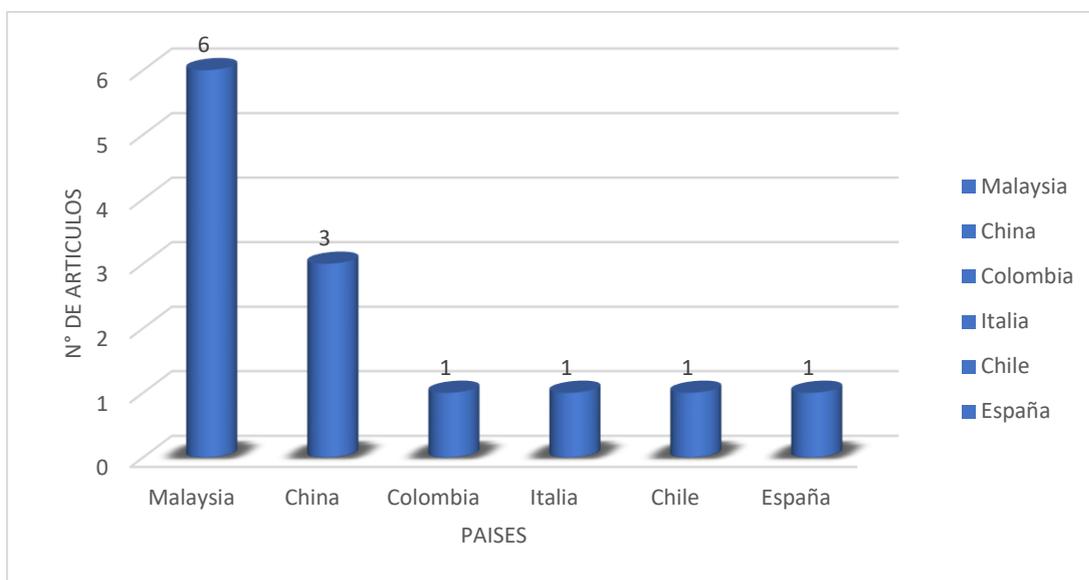
<b>TABLA 3: CLASIFICACIÓN DE LOS MÉTODOS</b>	
<i>PARAMÉTRICOS</i>	<i>NO PARAMÉTRICOS</i>
Regresión Lineal	Vecino más cercano
Interpolación Lineal	Media por encima y por debajo
El Método Espectral	Imputación múltiple
Interpolación Cuadráticos	Matriz de bajo rango
Interpolación Cúbicos	Análisis de Componente
	Principales funcional ACPF
	Maximización de las expectativas (EM)
	Estimación lineal imparcial (THE BLUE ESTIMATOR)
	Media de la fila
	Método kriging ordinario
	Método kriging universal
	La interpolación spline
	12-h media
	6-h media
	Año anterior
	La sustitución de la media
	Minimización de la norma
	Nuclear global (GNNM)

*Fuente: Elaboración propia*

De la **TABLA 3**, se observa que la mayoría de métodos utilizados en las investigaciones son de clasificación no paramétricos.

De las investigaciones revisadas se realizaron en diferentes países, los cuales utilizaron los datos de monitoreo de calidad de aire realizados en sus distintas ciudades, así como la de Ciudad de Beijing en China, Seberang Perai, Penang, Península de Malasia, La Guajira, Colombia, Malacca, Malasia, Sabah, Malasia, Región del Lazio en Italia y Changzhou, China. Siendo Malasia el país que presento el mayor número de investigaciones en cuanto al tema mencionado, líneas abajo se mostrara una gráfica para mejor interpretación.

**Gráfica 1:** Cantidad de artículos por País



*Fuente: Elaboración propia*

En la **GRÁFICA 1** se observa que, en el país de Malaysia se realizaron más artículos de este tipo de investigación, seguido de la China, en cuanto a Colombia Italia, Chile y Colombia, se utilizaron la misma cantidad de artículos.

**TABLA 4: MÉTODOS ENCONTRADOS EN LOS ARTÍCULOS ANALIZADOS.**

MÉTODO USADO	ESTUDIO	CONTAMINANTE	PORCENTAJE DE DATA FALTANTE	REFERENCIA
<b>Análisis de Componente principales funcional ACPF</b>	En el estudio se utilizó el método ACPF, para imputar los datos faltantes de las concentraciones de PM10, obtenido como resultado que este método se puede aplicar y ajustar muy bien a este tipo de estudio	Calidad del aire (concentraciones de PM10)	Autor no específica.	(Meléndez, Bolívar y Rojano 2020)
<b>La matriz de bajo rango</b>	El estudio comparo al método de la matriz de bajo rango con otros 6 métodos, con el objetivo para demostrar el rendimiento de este, obteniendo como resultado que el método de a matriz de bajo rango supera a los métodos comparadas, generando buenos resultados al remplazar los datos faltantes.	Calidad del aire (NOX, O3, SO2, PM2.5, PM10)	Data faltante 30%	(Liu et al. 2020)
<b>El vecino más cercano</b>				
<b>La sustitución de la media</b>				
<b>El método basado en la regresión</b>				
<b>La interpolación spline</b>				
<b>El método espectral</b>				
<b>El algoritmo de maximización de las expectativas</b>				
<b>Función ortogonal empírica restringida de ciclo diurno DCCEOF</b>	En el estudio se utilizó el método DCCEOF, con el objetivo de llenar los datos faltantes de PM 2.5 por hora, presentando una buena precisión de los picos diarios o mínimos diarios que no pueden ser restaurados por enfoques de interpolación convencionales	Calidad del aire (contracciones de PM2.5)	Data faltante 42.6 %	(Bai et al. 2020)

<b>Maximización de las expectativas (EM)</b>	En este estudio se utilizaron 2 métodos, con el objetivo de comprar estos métodos, usando datos faltantes de la calidad de aire de partículas de PM10 de 5 estaciones, llegando a la conclusión de que en método más preciso es el método del vecino más cercano.	Calidad del aire (contracciones de PM10)	Data faltante en 5%, 10%, 15%, 25% y 40%.	(Rumaling et al. 2020)
<b>Método del Vecino Más Cercano (NNM)</b>				
<b>Minimización de la norma nuclear global (GNNM)</b>	En este estudio se utilizaron 2 métodos, basados en la norma nuclear de una matriz, con el objetivo de determinar la eficacia de ambos, en relación de rellenar los datos faltantes de calidad de aire (PM 2.5), llegaron a la conclusión ambos métodos son eficaces, sin embargo, el de mejor rendimiento resulto ser el de minimización de la norma nuclear local (LNNM)	Calidad del aire (contracciones de PM2.5)	Autor no especifica.	(Chen y Xiao 2018)
<b>Minimización de la norma nuclear local (LNNM)</b>				
<b>La media por encima y por debajo</b>	En el presente estudio se utilizaron 4 métodos para rellenar los registros de datos perdidos de calidad de aire, con el objetivo de determinar cuál es el método más apropiado, el resultado del análisis demostró que el más eficiente y con menos margen es el método la media por encima y por debajo	Calidad del aire (PM10, CO, SO2, O3, NOX, NO2)	Dats faltantes 5%, 10%, 15% y 20%.	(Zakaria y Noor 2018)
<b>El vecino más cercano</b>				
<b>Imputación múltiple</b>				
<b>Regresión lineal</b>				
<b>Estimación lineal imparcial (THE BLUE ESTIMATOR)</b>	En el estudio se utilizó el método de estimación lineal imparcial (THE BLUE ESTIMATOR), aplicado en datos faltantes de mediciones de calidad del aire en tierra de PM10, con el objetivo de probar su efectividad, el cual fue aprobado.	Calidad del aire (contracciones de PM10)	Autor no especifica	(Sozzi et al. 2017)
<b>Interpolación lineal (LI)</b>	En este estudio se utilizaron 9 métodos, para rellenar los datos faltantes de la calidad de aire para concentraciones de PM10, con el objetivo de determinar el método más sencillo y preciso, el resultado del análisis determino que	Calidad del aire	Dats faltantes	(Noor, Yahaya,
<b>La media por encima y por debajo (MAB)</b>				
<b>El vecino más cercano</b>				

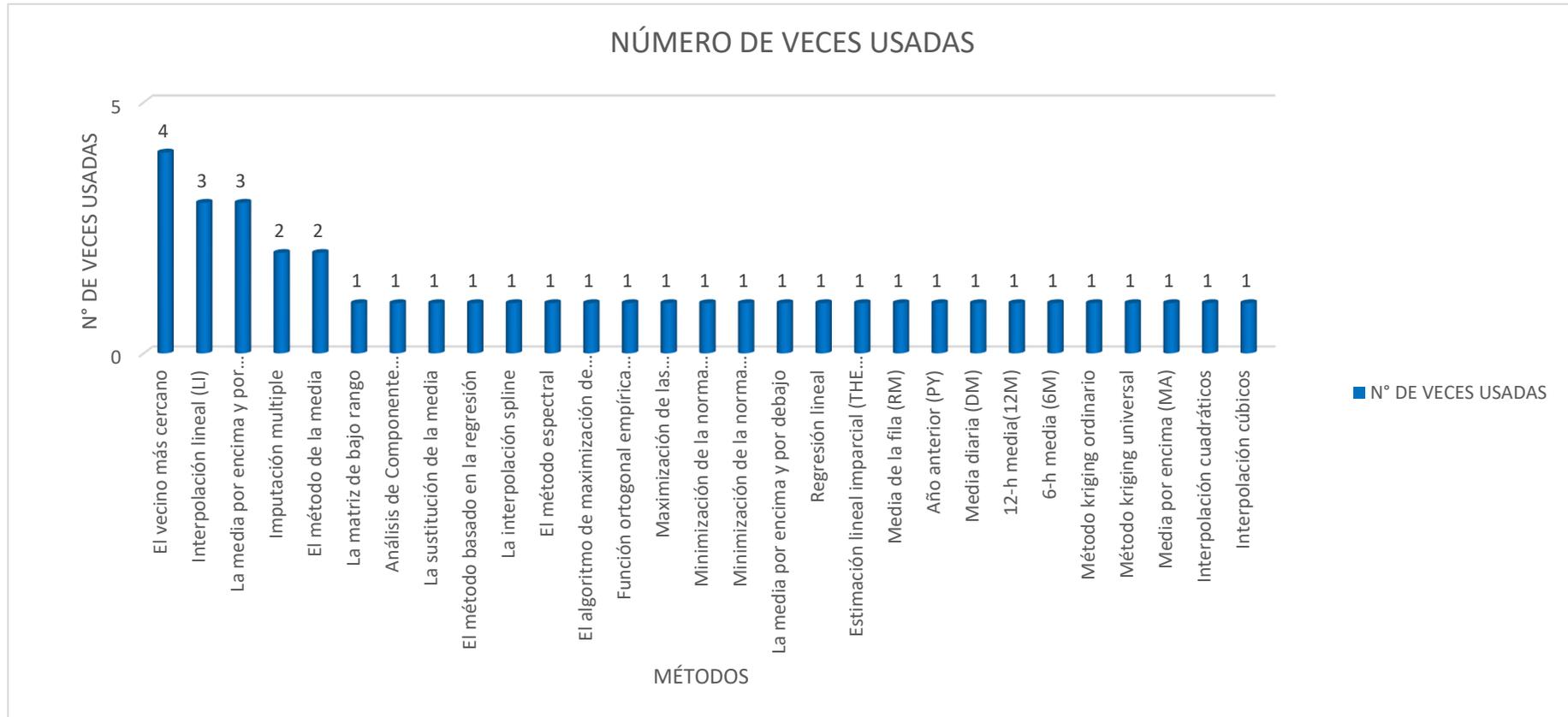
<b>Media de la fila (RM)</b>	el método de interpolación lineal (LI) y la media de 6 horas (6M) es adecuada para llenar los huecos que faltan	(contracciones de PM10)	de 5 % y 15 %.	et al. 2015)
<b>Año anterior (PY)</b>				
<b>Media diaria (DM)</b>				
<b>12-h media(12M)</b>				
<b>6-h media (6M)</b>				
<b>Imputación múltiple (MI)</b>				
<b>Método kriging ordinario</b>	En este estudio se utilizó el Método kriging ordinario para rellenar data faltante de PM10, con el objetivo de poder encontrar los valores estimados de exposición, dando como resultado que este método ayuda mucho a este tipo de predicción.	Calidad del aire (contracciones de PM10)	Autor no específica	(Fuenzalida, D, 2015)
<b>Interpolación Lineal</b>	En este estudio se realizaron 2 métodos para rellenar datos faltantes, con el objetivo de encontrar el mejor de ellos, el estudio concluyó que método de interpolación lineal dio mejores resultados para todos los porcentajes de los valores perdidos.	PM10, el dióxido de azufre, el ozono y el carbono	Data faltante en 5%, 10%, 15%, 25% y 40%.	(Noor, Al Bakri Abdullah, et al. 2015)
<b>El método de la media</b>				
<b>Método kriging universal</b>	En este estudio se realizó el método Kriging universal, teniendo como objetivo demostrar la efectividad de este método para rellenar los datos faltantes de PM10, la cual tuvo como resultado afirmativo.	Calidad del aire (contracciones de PM10)	Autor no específica	(Cañada, M, Moreno, A. Y Gonzales, H. 2015)
<b>La media por encima y por debajo (MAB)</b>	En este estudio se realizaron 3 tipos de métodos para rellenar datos faltantes de concentraciones de PM10, teniendo como objetivo encontrar la adecuada forma de estimar los datos faltantes, el estudio concluyo que el método de la media por encima y por abajo (MAB), es el mejor para predecir estos datos faltantes.	Calidad del aire (contracciones de PM10)	Data faltante en 5%, 10%, 15%, 25% y 40%.	(Noor et al. 2014b)
<b>Media por encima (MA)</b>				
<b>Método de la media</b>				
<b>Interpolaciones lineales</b>	En este estudio se utilizó 3 métodos, con el objetivo de llenar datos faltantes de las concentraciones de PM10, el		Data faltante 2,1% de 8577	(Noor et al.
<b>Interpolación cuadráticos</b>				

<b>Interpolación cúbicos</b>	cual obtuvo como resultado que los tres métodos son apropiados generando un valor de R2 muy alto.	Calidad del aire (concentraciones de PM10)	Concentraciones horarias. 2014a)
------------------------------	---	--	----------------------------------

*Fuente: Elaboración propia*

Con relación a la **TABLA 4** mostradas líneas arriba, se pueden ver que los resultados obtenidos por los investigadores en los estudios realizados, fueron aplicados 29 métodos al realizar el análisis de cada artículo, los que cumplieron con los objetivos de estos y mejor resultado dieron en cuanto a la estimación de los valores faltantes que según cada investigador garantizaron la calidad de los datos analizados, se puede observar que 13 de ellos dieron mejores resultados; en la siguiente tabla se un análisis de estos.

**Gráfica 2:** Cantidad de veces que los métodos fueron aplicados en las investigaciones



Fuente: Elaboración propia

De la **GRÁFICA 2**, se observa que los estudios revisados se encontraron que, de estos 29 métodos encontrados, 5 de estos fueron aplicados en la mayoría de las investigaciones analizadas, como se puede observar a continuación. A su vez el método que más se usó en las investigaciones fue el del vecino más cercano, seguido de la media por encima y por debajo.

De los métodos utilizados en los estudios revisados, se presenta la siguiente Tabla, en la cual, se muestra que métodos se utilizaron más y a la vez presentaron mejores resultados en las investigaciones.

**TABLA 5. MÉTODOS MAS UTILIZADOS Y PRESENTARON MEJORES RESULTADOS EN LAS INVESTIGACIONES**

<i>N°</i>	<i>Métodos con mejores resultados</i>	<i>referencia 1</i>	<i>Referencia 2</i>	<i>referencia 3</i>
<b>1</b>	Análisis de Componente principales funcional ACPF	(Meléndez, Bolívar y Rojano 2020)		
<b>2</b>	La matriz de bajo rango	(Liu et al. 2020)		
<b>3</b>	El vecino más cercano	(Rumaling et al. 2020)		
<b>4</b>	Minimización de la norma nuclear local (LNNM)	(Chen y Xiao 2018)		
<b>5</b>	La media por encima y por debajo (MAB)	(Noor et al. 2014b)	(Zakaria y Noor 2018)	
<b>6</b>	Estimación lineal imparcial (THE BLUE ESTIMATOR)	(Sozzi et al. 2017)		
<b>7</b>	Interpolación lineal (LI)	(Noor, Yahaya, et al. 2015)	Noor et al. 2014a)	(Noor, Al Bakri Abdullah, et al. 2015)
<b>8</b>	6-h media (6M)	(Noor, Yahaya, et al. 2015)		
<b>9</b>	Método kriging ordinario	(Fuenzalida, D, 2015)		
<b>10</b>	Método kriging universal	(Cañada, M, Moreno, A. Y Gonzales, H. 2015)		
<b>11</b>	Interpolación cuadráticos	(Noor et al. 2014a)		
<b>12</b>	Interpolación cúbicos	(Noor et al. 2014a)		
<b>13</b>	Función ortogonal empírica restringida de ciclo diurno DCCEOF	(Bai et al. 2020)		

*Fuente: Elaboración propia*

Se observa en la (TABLA 5), que el método de interpolación lineal fue el más utilizado y generó buenos resultados en los estudios que lo aplicaron. Tres artículos utilizaron este método siendo estos, el de Noor, Yahaya, et al (2015), que comparó 9 métodos, como se puede apreciar en la (TABLA 4), de los cuales concluyo, que indica que la interpolación lineal le generó un buen resultado siempre y cuanto más

corta sea la hora promedio mejor resultado presentaran los valores faltantes, este estudio utilizo como base el 5% y 15% de data faltantes del monitoreo del contaminante PM10, este método se comparó con otros 8 métodos siendo estos, el método la media por encima y por debajo y el vecino más cercano, media de la fila, año anterior, media diaria, 12h media, 6h media e imputación múltiple, concluyendo que el método del año anterior género el peor resultado en este estudio, demostrando que este se puede utilizar en porcentajes más grandes de data faltante. Comparando con los demás artículos analizados se observó no se mencionó ni se utilizaron, en ninguno de ellos, presentando este mayor porcentaje de data faltante como en los artículos de (Noor, Al Bakri Abdullah, et al. 2015) que presento un porcentaje de data faltante de hasta el 40%, de igual manera que el de (Noor et al. 2014b) y el de (Rumaling et al. 2020), además el de (Liu et al. 2020) con un porcentaje de data faltante del 30% como se puede apreciar en la **(TABLA 4)**.

Otro artículo el de Noor, et al. 2014a, quien utilizó también el método de interpolación lineal, el cual fue comparado con otros dos métodos como el método de interpolación cuadráticos y cúbicos como se puede apreciar en la **(TABLA 4)**, el cual obtuvo como resultado que los tres métodos presentados un mejor ajuste en sus datos de valores faltantes de PM10, con un porcentaje 2,1%, siendo esta el menor porcentaje de data comprando con los otros 12 estudios, como se puede apreciar en la **(TABLA 4)**.

Por último, para Noor, Al Bakri Abdullah, et al. (2015), en este estudio comparo dos métodos siendo el segundo el método de la media, llegando a la conclusión de que el método de interpolación lineal dio mejor resultados para los porcentajes de valores faltantes con una base del 5%,15%,25% y 40% de data faltantes del monitoreo de los contaminantes PM10, SO2, C y el O3, a comparación de los otros dos artículos mencionados anterior mente como el de Noor, Yahaya et al, (2015) Y Noor, et al. (2014<sup>a</sup>), que utilizaron menos valores del porcentaje de data faltante y solo aplicaron valores registrados del contaminante PM10, como se puede apreciar en la **(TABLA 4)**, demostrando así que este método de interpolación lineal, también se puede utilizar para registro de monitoreos de contaminantes como dióxido de azufre, el ozono y el carbono, generando buenos resultados.

En cuanto al otro método que también dio buenos resultados en dos investigaciones, es la media por encima y por debajo, según, Noor et al. (2014b), este método fue el que mejor predijo los valores de data faltante del 5%,10%,15%,25% y 40% del registro de monitoreo del contaminante de PM10 por hora, como se puede apreciar en la (TABLA 4), este presentó como resultado que el RMSE ( error cuadrático medio) arrojó un valor más pequeño que en los otros métodos, según Barnston, A. (1992), este es indicador de desempeño RMSE, se utiliza comúnmente en pronósticos análisis de regresión y climatología, ya que este nos indica que tan cerca están los datos de la línea de regresión, esto nos muestra la medida de dispersión de los residuos (pp. 4) y en cuanto al R2 presento el valor más alto, según, Simón et al. (2018) este indicador de desempeño, predice los resultados de los valores faltantes, entre más alto sea el R2 mayor será el ajuste que se desea encontrar en este caso el valor faltante (pp.43)

Comparando con el artículo de Noor et al. (2014b), Zakaria y Noor. (2018), tuvo como resultado que este método de la media por encima y por debajo, predijo el mejor ajuste para la data faltante, utilizó 3 métodos, el vecino más cercano, imputación múltiple y la regresión lineal, con una data faltante del 5%,10%,15% , y 20% de los contaminantes PM10, SO2, CO, O3, NOX y NO2, mientras que el Noor et al. (2014b), que solo aplicaron valores registrados del contaminante PM10, como se puede apreciar en la (TABLA 4), para Zakaria Y Noor. (2018), el método de la media por encima y por debajo, le genero RMSE más pequeño y el valor R2 más alto para una base de datos al igual que el estudio anterior.

Si hablamos de los contaminantes como el PM10 Y PM2.5, según Joanne y Sánchez. (2013), es aquella mezcla de partículas minúsculas ya sean sólidas o líquidas, las cuales se localizan en el aire que respiramos, para Carslaw y Ropkins (2012) es necesario saber la calidad del aire en cuanto a estos contaminantes y si éstos cumplen con los estándares de calidad del aire permitido, cabe resaltar que al existir datos no registrados para estos parámetros se tienen que utilizar métodos que permitan rellenar la data faltante y predecir un mejor resultado (pp.1)

Según los resultados de los artículos revisados, los métodos que se adecuaron apropiadamente a rellenar la data faltante y presentar una predicción

más real para los registros de monitoreo para los contaminantes como el PM10 y el PM2.5, fueron los siguientes:

**TABLA 6. MÉTODOS QUE SE AJUSTAN Y APLICAN MEJOR AL CONTAMINANTE PM10, PARA RELLENO DE DATA FALTANTE.**

<b>Métodos</b>	<b>Contaminante</b>	<b>Porcentaje de Data faltante</b>	<b>Referencia</b>
<b>1</b> Análisis de Componente principales funcional ACPF	Calidad del aire (concentraciones de PM10)	Autor no específico.	(Meléndez, Bolívar y Rojano 2020)
<b>2</b> Método del Vecino Más Cercano (NNM)	Calidad del aire (contracciones de PM10)	Data faltante en 5%, 10%, 15%, 25% y 40%.	(Rumaling et al. 2020)
<b>3</b> Estimación lineal imparcial (THE BLUE ESTIMATOR)	Calidad del aire (contracciones de PM10)	Autor no específico	(Sozzi et al. 2017)
<b>4</b> Interpolación lineal (LI) 6-h media (6M)	Calidad del aire (contracciones de PM10)	Datas faltantes de 5 % y 15%.	(Noor, Yahaya, et al. 2015)
<b>5</b> Método kriging ordinario	Calidad del aire (contracciones de PM10)	Autor no específico	(Fuenzalida, D, 2015)

<b>6</b>	Método kriging universal	Calidad del aire (contracciones de PM10)	Autor no específico	(Cañada, M, Moreno, A. Y Gonzales, H. 2015)
<b>7</b>	La media por encima y por debajo (MAB)	Calidad del aire (contracciones de PM10)	Data faltante en 5%, 10%, 15%, 25% y 40%.	(Noor et al. 2014b)
<b>8</b>	Interpolaciones lineales Interpolación cuadráticos Interpolación cúbicos	Calidad del aire (concentraciones de PM10)	Data faltante 2,1% de 8577 Concentraciones horarias.	(Noor et al. 2014a)

*Fuente: Elaboración propia*

Como se puede observar en la (TABLA 6), los métodos que se utilizaron y dieron como resultado favorable su aplicación para poder estimar los datos faltantes monitoreo en cuanto al contaminante PM10, para porcentajes de data faltante según los autores del 2.1%, 5%, 10%, 15%, 25% y 40%, se observa también una carencia de información ya que en cuatro de estos artículos mencionados, no se especifica el porcentaje de data faltante que presentaron sus bases de datos; estos son los siguientes:

Para, Meléndez, Bolívar Y Rojano. (2020) los resultados indicaron que el método del análisis de componentes principales funcionales (ACPF) utilizado para la imputación datos semanales faltantes, (no se especificó cuanta data faltante tenía la base analizada), lo que si se menciona es que fueron obtenidos de estaciones automatizadas semanales de partículas PM10, como se puede apreciar en la **(TABLA 4)**, en el artículo el método ACPF busca imputar estos datos, utilizando los valores propios de una matriz de covarianza de aquellos datos para así encontrar las direcciones en el espacio de estudio de las cuales, estos datos presentan mayor variabilidad. El resultado que se obtuvo en este estudio demostró que el método ACPF, se aplica y ajusta bien a este tipo de investigación (pp.36)

Según Fuenzalida, M. (2015) al igual que el estudio de Meléndez, Bolívar Y Rojano. (2020), no especifica cuanta data faltante tenía la base analizada, el método kriging ordinario modelo que aplicó Fuenzalida, M.(2015) , para la estimación de valores faltantes de concentraciones de PM10, de ciento ocho estaciones de monitoreo, como se puede apreciar en la **(TABLA 4)**, para así obtener predicciones que permitan obtener mapas de predicción, resultado este muy útil para presentar así los valores de exposición de este contaminante (pp.11)

Para Cañada, M, Moreno, A. y Gonzales, H. al igual que los estudios de Meléndez, Bolívar Y Rojano. (2020), tampoco se especifica cuanta data faltante tenía la base analizada, aquí se menciona el método Kriging universal, este método que aplicó Cañada, M, Moreno, A, en este estudio con el objetivo de predecir la historia del contaminante PM10, en ubicaciones no observadas, como se puede apreciar en la **(TABLA 4)**, con el cual se puede obtener así los niveles de concentraciones de este contaminante (pp.23)

El último estudio al igual que los anteriores mencionados no se especifica el porcentaje de data faltante analizada es el de Sozzi et al. 2017, menciona que el método estimación lineal imparcial (THE BLUE ESTIMATOR), al ser aplicado a una base de datos faltante de las concentración diarias de PM10, en Lazio , Italia, como se puede apreciar en la **(TABLA 4)**, este estudio es utilizado en concentraciones diarias a diferencias de los artículos de Cañada, M, Moreno, A. y Gonzales, H, quien lo utilizó en concentraciones mensuales, mientras que Meléndez, Bolívar Y Rojano, en concentraciones semanales y Fuenzalida, M. en concentraciones

anuales. El método de utilizo Sozzi et al. 2017, permitió aumentar el rendimiento de la red produciendo así una serie temporal prácticamente completa ya que ayudo a reconstruir una serie temporal con valores faltantes, Sozzi et al, nos indica en el futuro este método puede también ser aplicado a diferentes contaminantes como PM2.5, NO2, O3 y policíclico hidrocarburo aromáticos (pp.17)

A comparación con los estudios anteriores donde no se especifica la cantidad de data faltante, Rumaling et al. 2020, en su estudio si menciona la cantidad de data faltante presente en la base de datos que fueron tomados de cinco estaciones meteorológicas de calidad de aire para el parámetro PM10 con un porcentaje de data faltante del 5%,10%,15%,25% y 40%, como se puede apreciar en la **(TABLA 4)**, en este artículo a diferencia de los anteriores se compararon entre dos métodos, el vecino más cercano y el método el algoritmo de maximización de expectativas (EM), se realizó la comparación entre ambos dando como resultado que el vecino más cercano (NNM), resulto ser el que el mejor método de imputación de datos, para las cinco estaciones, indicando que es importante para investigaciones futuras la imputación de datos utilizando este método ya que permite un mejor análisis sobre los datos de calidad del aire y sean más fiables (Rumaling et al. 2020, pp.12)

Según, Noor, Yahaya, et al. (2015), que a diferencia de Rumaling et al. Este comparo nueve métodos, para estimación de datos faltantes con un porcentaje de 5% y 15% para el contaminante PM10, de ocho estaciones de monitoreo en la península de Malasia, como se puede apreciar en la **(TABLA 4)**, Dando como resultado que dos de los nueve métodos, presentaron mejores resultados como son el método de interpolación lineal y el 6-h media (6M), cabe resaltar que entre más pequeña sea la hora promedio siendo así que estas pérdidas de datos se limita a < 12 horas, este método también se utiliza para concentraciones diarias al igual que en el estudio de Sozzi et al. 2017 (pp.13)

Para Noor et al. (2014b), encontró que el método de la media por encima y por debajo (MAB), es el que mejor resultado género, usando en este estudio dos métodos como la media por encima y el método de la media, en este caso se menciona un porcentaje de data faltante del 5%,10%,15%,25% y 40% al igual que el artículo de Rumaling et al. (2020), con la diferencia que solo se obtuvo esta data

de una estación de monitoreo con registros anuales de monitoreo por hora del contaminante PM10, como se puede apreciar en la (TABLA 4), el resultado de la investigación de Noor et al. (2014b), demostró que el error cuadrático medio RMSE, arrojó un valor más pequeño y en cuanto a R2 presento valores más alto aplicando el método de la media por encima y por debajo, así garantizar la alta calidad de estos datos analizados (pp.25)

El último artículo que utiliza un base de datos con valores del contaminante PM10, es de Noor et al. (2014a), en este estudio se compararon tres métodos, el de Interpolaciones cuadráticos, cúbicos y lineales, según los resultados que obtuvo Noor et al. (2014a), demostraron que los tres métodos de imputación de datos aplicados se ajustaron muy bien, generando una buena predicción de los valores faltante de un total de 8577 observaciones de concentraciones de PM10, como se puede apreciar en la **(TABLA 4)**, durante un año con un porcentaje 2,1%, siendo esta el menor porcentaje de data comprando con los otros artículos anteriormente mencionados en la **(TABLA 6)** (pp.12)

**TABLA 7. MÉTODOS QUE SE AJUSTAN Y APLICAN MEJOR AL CONTAMINANTE PM2.5, PARA RELLENO DE DATA FALTANTE.**

Método Usado	Contaminante	Porcentaje de Data faltante	Referencia
1 La matriz de bajo rango	Calidad del aire (contracciones de PM2.5)	Data faltante 30%	(Liu et al. 2020)
2 Función ortogonal empírica restringida de ciclo diurno DCCEOF	Calidad del aire (contracciones de PM2.5)	Data faltante 42.6 %	(Bai et al. 2020)
3 Minimización de la norma nuclear local (LNNM)	Calidad del aire (contracciones de PM2.5)	Autor no específico.	(Chen y Xiao 2018)

*Fuente: Elaboración propia*

En la **TABLA 7**, nos menciona a los tres métodos que son adecuados según los resultados de las investigaciones realizadas por los autores, que la matriz de bajo rango, la función ortogonal empírica restringida de ciclo diurno DCCEOF y la minimización de la norma nuclear local (LNNM), son métodos que se pueden aplicar para rellenar la data faltante de una base de datos de concentraciones de PM2.5, para un porcentaje de data faltantes de un rango del 30% y un 42.6%,

En cuanto a los métodos aplicados en los estudios, para las concentraciones de PM2.5, fueron tres métodos que resultaron presentar mejores resultados.

El estudio de Liu et al. (2020), donde se comparan 7 métodos, de los cuales el método de la matriz de bajo rango presento mejores resultados, en este estudio se analizó una base de datos con valores faltantes del 30%, como se puede apreciar en la **(TABLA 4)**, con el objetivo de demostrar el rendimiento de este, obteniendo como resultado que el método de a matriz de bajo rango supera a los métodos comparados, generando buenos resultados al remplazar los datos faltantes (pp.16)

Chen y Xiao. (2018), a comparación del estudio de Liu et al. (2020), este comparo solo dos métodos, de Minimización de norma nuclear global (GNNM) y el método de Minimización de la norma nuclear local (LNNM), para rellenar data faltante de un registro de concentración de 2.5PM por hora, de veintidós estaciones de monitoreo, como se puede apreciar en la **(TABLA 4)**, en este estudio el autor no menciona el porcentaje de data faltante, a comparación del estudio de Liu et al. (2020), que si lo menciona, los resultados se compararon con los valores reales del terreno y así evaluar el desempeño de la imputación de este método, como resultado la efectividad de este método dando buenos resultados y poder obtener futuros datos que permitir una buena calidad de aire (Chen y Xiao 2018,pp.27)

Por último, el estudio de Bai et al. 2020, a comparación de los estudios anteriores este utilizó solo el método de la función ortogonal empírica restringida de ciclo diurno DCCEO, para rellenar los valores faltantes de un registro de monitoreo de concentraciones de PM2.5, como se puede apreciar en la **(TABLA 4)**, el cual redujo del 42,6% de la data faltante a un 5,7%, demostrando así que este método presento una buena precisión de predicción para mínimos diarios los cuales no pueden ser restaurados con métodos de interpolación convencionales (Bai et al. 2020, pp.11)

**TABLA 8. MÉTODOS QUE SE AJUSTAN Y APLICAN MEJOR A LOS CONTAMINANTES NOX, O3, SO2, C, CO, NO2 PARA RELLENO DE DATA FALTANTE.**

Métodos	Contaminante	Porcentaje de Data faltante	Referencia
1 La matriz de bajo rango	Calidad del aire (NOX, O3, SO2)	Data faltante 30%	(Liu et al. 2020)
2 La media por encima y por debajo	Calidad del aire (CO, SO2, O3, NOX, NO2)	Datas faltantes 5%, 10%, 15% y 20%.	(Zakaria y Noor 2018)
3 Interpolación Lineal	Calidad del aire (SO2, O3 y C)	Data faltante en 5%, 10%, 15%, 25% y 40%.	(Noor, Al Bakri Abdullah, et al. 2015)

*Fuente: Elaboración propia*

En cuanto a la **TABLA 8**, esta nos presenta también tres métodos que se utilizaron y dieron como resultado, que se adecuan muy bien, en cuanto a rellenar la data faltante para monitoreos de NOX, O3, SO2, C, con data faltante del 5%,10%,15%,25%, 30% y el 40%.

Como se presentó en la tabla (**TABLA 8**) de los estudios revisados, se encontró que tres métodos, utilizados por los investigadores, presentaron mejores resultados en cuanto fueron aplicados para estimar los valores faltantes de registros de datos de contaminantes como (NOX, O3, SO2, C, CO, NO2)

El artículo de Liu et al. 2020, donde se comparan 7 métodos, de los cuales el método de la matriz de bajo rango presento mejores resultados, en este estudio se analizó una base de datos con valores faltantes del 30%, del registro de monitoreos de los contaminantes NOX, O3, SO2, como se puede apreciar en la (**TABLA 4**), con el objetivo de demostrar el rendimiento de este, obteniendo como resultado que el método de a matriz de bajo rango supera a los métodos comparados, generando buenos resultados al remplazar los datos faltantes, no solo ser mejor método de predicción concentraciones de PM10 y PM2.5, sino también para las contaminantes como NOX, O3, SO2 (Liu et al. 2020,PP.17)

Comparando con Liu et al. 2020, el estudio de Zakaria y Noor. (2018), este comparo 4 métodos, siendo el método de la media por encima y por debajo, predijo el mejor ajuste para la data faltante, para una data faltante del 5%,10%,15%, y 20% del monitoreo de los contaminantes PM10, SO2, CO, O3, NOX y NO2, como se puede apreciar en la (**TABLA 4**), para Zakaria y Noor. (2018), el método de la media por encima y por debajo, le genero RMSE más pequeño y el valor R2 más alto para una base de datos al igual que el estudio anterior, a su vez demostró no solo ser mejor método de predicción concentraciones de PM10, sino también para los contaminantes antes mencionados.

El último artículo el de Noor, Al Bakri Abdullah, et al. (2015), este comparo solo 2 métodos, la media y el método de Interpolación Lineal, siendo este el último el que demostró ser mejor para este tipo de estudio, para los porcentajes de valores faltantes de 5%,10%,15%,25% y 40% para los registros de datos de los contaminantes SO2, O3 y C, como se puede apreciar en la (**TABLA 4**), ya que genero menos error en la correlación de la matriz (pp.23)

al igual que los otros dos estudios el de Liu et al. (2020) y, Zakaria y Noor (2018) también demostró no solo ser mejor método de predicción para concentraciones de PM10, sino también para contaminantes como el SO<sub>2</sub>, O<sub>3</sub> y C.

## V. CONCLUSIONES

1. El método de interpolación lineal y el método de la media por encima y por debajo fueron los métodos que más se utilizaron, así mismo estos al ser evaluados mediante los indicadores de adecuación (MAE, RMSE Y R<sup>2</sup>), presentaron un mejor ajuste, además permiten comparar mediciones de diferentes escalas, evitando así que se cometan errores de predicción en los valores faltantes.
2. En cuanto a la clasificación de los métodos estos se clasifican en paramétricos los cuales presentan valores con una distribución normal y los no paramétricos que presentan valores con distribución no normal. En el caso de la interpolación lineal este método es de tipo paramétrico y el método de la media por encima y por debajo es de tipo no paramétrico.
3. De los métodos antes presentados, nueve de ellos se utilizaron específicamente para rellenar los datos faltantes de concentraciones de PM10 como; Análisis de Componente principal funcional ACPF, el método del vecino más cercano, interpolación lineal (LI), la media de 6 horas (6M), estimación lineal imparcial (THE BLUE ESTIMATOR), el método de la media por encima y por abajo (MAB), interpolaciones lineales, interpolación cuadráticos e interpolación cúbicos, otros tres de ellos fueron utilizados para rellenar datos faltantes de concentraciones de PM2.5 como función ortogonal empírica restringida de ciclo diario DCCEOF, minimización de la norma nuclear global (GNNM) y minimización de la norma nuclear local (LNNM), en el caso de los métodos usados contaminantes NO<sub>x</sub>, O<sub>3</sub>, SO<sub>2</sub>, NO<sub>2</sub>, fueron el método de la matriz de bajo rango, la media por encima y por debajo y método de interpolación lineal estas a su vez se utilizaron también para los contaminantes PM10 y PM2.5.

## **VI. RECOMENDACIONES**

Se recomienda aplicar estos métodos, no solamente en los registros de la calidad del aire en base a contaminantes, sino también los registros de precipitación, humedad y otros que sean necesarios.

Se debe tener en cuenta la cantidad de data faltante presente en ellos, ya que así se podrá elegir mejor el método el cual genere un mejor resultado.

Otro punto importante es establecer qué tipo de valores presentan las bases de dato, ya sean de distribución normal o no normal, mediante las pruebas de normalidad, esto generará que se elija un método adecuado para el caso de estudio ya que así se podría obtener un mejor rendimiento en cuanto a los resultados.

## REFERENCIAS

ALBERT, ARTHUR. (.1973). The Gauss-Markov Theorem for Regression Models with Possibly Singular Covariances. *SIAM Journal on Applied Mathematics*, 24(2), 182-187. Retrieved May 12, 2020, from [www.jstor.org/stable/2099674](http://www.jstor.org/stable/2099674)

ARIAS VALENCIA, M; GIRALDO MORA, C. El rigor científico en la investigación cualitativa. *Investigación y educación en enfermería*, 2011; 29(3): 500-514. (<https://aprendeonline.udea.edu.co/revistas/index.php/iee/article/view/5248/9922>)

BAI, K., LI, K., GUO, J., YANG, Y. y CHANG, N. Bin, 2020. Filling the gaps of in situ hourly PM<sub>2.5</sub> concentration data with the aid of empirical orthogonal function analysis constrained by diurnal cycles. *Atmospheric Measurement Techniques*, vol. 13, no. 3, pp. 1213-1226. ISSN 18678548. DOI 10.5194/amt-13-1213-2020.

BARNSTON, AG, 1992: Correspondencia entre las medidas de correlación, RMSE y Heidke Forecast Verification; Refinamiento de la puntuación de Heidke. *Wea. Forecasting*, 7, 699–709, [https://doi.org/10.1175/1520-0434\(1992\)007<0699:CATCRA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007<0699:CATCRA>2.0.CO;2).

CÁDIZ, J. (2006). Rigurosidad científica y principios orientadores para el evaluador/ investigador, en: *Ars Médica*, 14. Disponible en: [Http://escuela.med.puc.cl/publ/arsmedica/arsmedica9/ars3.html/](http://escuela.med.puc.cl/publ/arsmedica/arsmedica9/ars3.html/)

CAÑADA TORRECILLA, M.R., MORENO JIMÉNEZ, A. y GONZÁLEZ LORENZO, H., 2014. Modelado de la calidad del aire urbano. Un experimento meteorológico con técnicas de interpolación espacial. *Boletín de la Asociación de Geógrafos Españoles*. 2014. pp. 317-342. ISSN 0212-9426.

CANDÈS, E.J. y RECHT, B., 2009. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717-772. ISSN 16153375. DOI 10.1007/s10208-009-9045-5.

CARSLAW, D.C y ROPKINS, K.,2012. Environmental Modelling & Software openair d An R packege for air quality data analysis. *Environmental Modelling and Software* [en línea], vol 27-28, pp 52-62. ISSN 1364-8152.DOI

10.1016/J.envsoft.2011.09.008.

CASTILLO, E. y VÁSQUEZ, M. (2003). El rigor metodológico en la investigación cualitativa, en: Colombia Médica, 34:164-167.

CHAPRA, Steven y CANALE, Raymond. Methods for Engineers. Singapore:McGrawHill. [en línea].1 a ed. Estados Unidos: National Library Board Singapur, 1989. [fecha de consulta: 29 de abril de 2020].Disponible en: [https://eservice.nlb.gov.sg/data2/BookSG/publish/2/2404b03b-b7a0-43f0-b166-93919a0b51d0/web/html5/index.html?opf=tablet/BOOKSG.xml&launchlogo=tablet/BOOKSG\\_BrandingLogo\\_.png](https://eservice.nlb.gov.sg/data2/BookSG/publish/2/2404b03b-b7a0-43f0-b166-93919a0b51d0/web/html5/index.html?opf=tablet/BOOKSG.xml&launchlogo=tablet/BOOKSG_BrandingLogo_.png) ISBN: 0-07-079984-9

CHAVEZ, C, SANCHEZ,J.y DE LA CRUZ,J., 2015. Analisis de componentes principales funcionales en series de tiempo economicas., vol 3 N° 2. Gecontec: Revista Internacional de Gestión del Conocimiento y la Tecnología, vol. 3 (2) 2015 ISSN 2255-5684. Disponible en SSRN: <https://ssrn.com/abstract=2737561>

CHEN, X. y XIAO, Y., 2018. A Novel Method for Air Quality Data Imputation by Nuclear Norm Minimization. Journal of Sensors, vol. 2018, no. 2. ISSN 16877268. <https://doi.org/10.1016/j.envint.2020.105713>.<https://www.hindawi.com/journals/js/2018/7465026/>

COHEN, M.P. (1996). A new approach to imputation. American Statistical Association Proceeding of the Section on Survey Research Methods,ELSOVIER. THE NETHERLANDS. pp. 293–298.ISBN 9780444529589

EISENHOWER RENGEL, Wilmer., et al, 2018. Publicar investigación científica Metodología y desarrollo [en línea]. Ecuador, 1era edición [fecha de consulta 28 mayo 2020] ISBN: 978-9942-775-16-0 Disponible en: [https://issuu.com/marabiertouleam/docs/publicar\\_investigacion\\_cientifica](https://issuu.com/marabiertouleam/docs/publicar_investigacion_cientifica)

EK,ALAN R., ANDREW P. ROBINSON, PHILIP J. RADTKE, DAVID K. WALTERS, Development and testing of regeneration imputation models for forests in Minnesota[en línea]. Estado Unidos: Forest Ecology and Management 1997 [fecha de consulta: 18 de octubre de 2005]. Disponible en:

<http://www.sciencedirect.com/science/article/pii/S0378112796039709> ISSN 0378-1127

EMANUEL, L., ALVES, R. y GOMES, H.B., 2019. Anuário do Instituto de Geociências - UFRJ Validação da Imputação Múltipla via Predictive Mean Matching para Preenchimento de Falhas nos Dados Pluviométricos da Bacia do Médio São Francisco Validation of Multiple Imputation by Predictive Mean Matching for. , vol. 43, pp. 199-206

FUENZALIDA-DÍAZ, M., 2015. Evaluación de modelos geoestadísticos aplicados a la exposición al contaminante atmosférico PM10 en Chile. Ciencias Espaciales, vol. 8, no. 1, pp. 441-457. ISSN 2225-5249. DOI 10.5377/ce.v8i1.2060.

GONZÁLEZ, J., 2009. LA TEORÍA DE LA COMPLEJIDAD. , vol. 76, pp. 243-245. Disponible en: <https://www.redalyc.org/pdf/496/49611942024.pdf>

GUBA, E. Y LINCOLN, Y. Epistemological and methodological bases of naturalistic inquiry. Educ Tech Res Dev. 1982;30(4):233-52.

HERNANDEZ, Roberto., 2016. Diseños del proceso de investigación cualitativa. ,.Revista: wordpress, Disponible: <https://administracionpublicaub.files.wordpress.com/2016/03/hernc3a1ndez-samipieri-cap-15-disec3b1os-del-proceso-de-investigac3b3n-cualitativa.pdf>, pp. 468-506

JOANNE, G. y SÁNCHEZ, S., 2013. La Calidad del Aire en América Latina: Una Visión Panorámica. Clean Air Institute [en línea], Disponible en: [https://www.minambiente.gov.co/images/AsuntosambientalesySectorialyUrban/a/pdf/contaminacion\\_atmosferica/La\\_Calidad\\_del\\_Aire\\_en\\_América\\_Latina.pdf](https://www.minambiente.gov.co/images/AsuntosambientalesySectorialyUrban/a/pdf/contaminacion_atmosferica/La_Calidad_del_Aire_en_América_Latina.pdf) .

LIU, X., WANG, X., ZOU, L., XIA, J. y PANG, W., 2020. Spatial imputation for air pollutants data sets via low rank matrix completion algorithm. Environment International [en línea], vol. 139, no. March. ISSN 0160-4120. DOI 10.1016/j.envint.2020.105713. Disponible en: <https://doi.org/10.1016/j.envint.2020.105713>.

MEDINA, F. y GALVÁN, M., 2007. Imputación de datos: teoría y práctica [en línea]. S.l.: s.n. ISBN 9213217919. Disponible en: <http://www.cepal.org/publicaciones/xml/8/14038/lc2024e.pdf>.

MELÉNDEZ, R., BOLÍVAR, S. y ROJANO, R., 2020. Imputación de valores perdidos y detección de valores atípicos en datos funcionales: Una aplicación con datos de PM10. REVISTA UIS INGENIERÍAS, vol. 19, no. 2. DOI 10.18273/revuin.v19n2-2020001. ISSN en línea: 2145 – 8456.

MONCADA, Laura, 2018. Diseño Narrativo. *Laurapatriciamh.blogspot* [En línea]. Disponible en: <http://laurapatriciamh.blogspot.com/disenos-narrativos.html> [Consulta: 13/06/2020]

MORSE J. Qualitative research methods. Beverly Hills: Sage Publications; 1994

NOOR, M.N., YAHAYA, A.S., RAMLI, N.A. y BAKRI, A.M.M. Al, 2014a. Filling Missing Data Using Interpolation Methods: Study On The Effect Of Fitting Distribution. , vol. 595. DOI 10.4028/www.scientific.net/KEM.594-595.889.

NOOR, M.N., YAHAYA, A.S., RAMLI, N.A. y BAKRI, A.M.M. Al, 2014b. Mean Imputation Techniques for Filling the Missing Observations in Air Pollution Dataset. , vol. 595. DOI 10.4028/www.scientific.net/KEM.594-595.902.

NOOR, N.M., AL BAKRI ABDULLAH, M.M., YAHAYA, A.S. y RAMLI, N.A., 2015. Comparison of linear interpolation method and mean method to replace the missing values in environmental data set. Materials Science Forum, vol. 803, pp. 278-281. ISSN 02555476. DOI 10.4028/www.scientific.net/MSF.803.278.

NOOR, N.M., YAHAYA, A.S., RAMLI, N.A., MUSTAFA, M. y BAKRI, A., 2015. Filling the Missing Data of Air Pollutant Concentration using Single Imputation Methods. , vol. 755, pp. 923-932. DOI 10.4028/www.scientific.net/AMM.754-755.923.

NORAZIAN, M.N., SHUKRI, Y.A., AZAM, R.N., MUSTAFA, A.M. y BAKRI, A., 2008. ESTIMATION OF MISSING DATA USING INTERPOLATION TECHNIQUE : FITTING ON WEIBULL DISTRIBUTION problem in many scientific field including. , pp. 1-4.

OLIVIER, R. y HANQIANG, C., 2012. Nearest Neighbor Value Interpolation. International Journal of Advanced Computer Science and Applications, vol. 3, no. 4. ISSN 2158107X. DOI 10.14569/ijacsa.2012.030405.

ORGANIZACIÓN MUNDIAL DE LA SALUD, 2005. Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre [en línea]. Disponible en: [https://apps.who.int/iris/bitstream/handle/10665/69478/WHO\\_SDE\\_PHE\\_OEH\\_06.02\\_spa.pdf?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/69478/WHO_SDE_PHE_OEH_06.02_spa.pdf?sequence=1) [consulta:23 abril 2020].

ORGANIZACIÓN MUNDIAL METEOROLOGICA, 1990. On the Statistical Analysis of Series of Observations, WMO-No. 415, [en línea]. Ginebra: Secretaría de la Organización Meteorológica Mundial, [consulta:24 abril 2020]. Disponible: [http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo\\_100\\_es.pdf](http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo_100_es.pdf). ISBN 978-92-63-10415-1

ORGANIZACIÓN MUNDIAL METEOROLOGICA, 2011. Guías de prácticas climatológicas, [en línea]. Disponible en [http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo\\_100\\_es.pdf](http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo_100_es.pdf) [consulta:23 abril 2020].

OROZCO ALVARADO, J.C. y DÍAZ PÉREZ, A.A., 2018. ¿Cómo redactar los antecedentes de una investigación cualitativa? Revista Electrónica de Conocimientos, Saberes y Prácticas, vol. 1, no. 2, pp. 66-82. ISSN 2616-8294. DOI 10.30698/recsp.v1i2.13.

PLAIA, A. y BONDÌ, A.L., 2006. Single imputation method of missing values in environmental pollution data sets. Atmospheric Environment, vol. 40, no. 38. ISSN 13522310. DOI 10.1016/j.atmosenv.2006.06.040.

RAMSAY, J.O. & SILVERMAN, B.W. Functional Data Analysis (2nd edition), Springer. [en línea]. Nueva York, 2005. [fecha de consulta: 08 de mayo de 2020]. Disponible en: [https://books.google.com.pe/books?id=mU3dop5wY\\_4C&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.pe/books?id=mU3dop5wY_4C&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false) ISSN 0172-739

REYES, M., 2010. Estimación Paramétrica y No Paramétrica de la Tendencia en Datos con Dependencia Espacial. Un Estudio de Simulación. [en línea], pp. 58. Disponible en:

[http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto\\_401.pdf](http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_401.pdf).

RODERICK, Little y DONALD, Rubin. Statistical Analysis with missing data. [en línea]. 3 a ed. Estados Unidos: Edition History, 2019. [fecha de consulta: 29 de abril de 2020]. Disponible en: [https://books.google.com.pe/books?id=BemMDwAAQBAJ&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.pe/books?id=BemMDwAAQBAJ&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false). ISBN: 9781118596012

RODRIGUEZ, Daniela, 2019. Investigación Aplicada. En Liferder [en línea]. Disponible en: <https://www.liferder.com/investigacion-aplicada/> [consulta: 28 de mayo 2019]

RODRIGUEZ, M, Acerca de la investigación bibliográfica y documental, 2013. Plataforma de Metodología de la Investigación y Guía de Tesis de Grado [En línea]. Disponible en: <https://guiadetesis.wordpress.com/2013/08/19/acerca-de-la-investigacion-bibliografica-y-documental/> [fecha de consulta: 13 de junio del 2020]

RUMALING, M.I., CHEE, F.P., DAYOU, J., HIAN, J. y CHANG, W., 2020. Missing Value Imputation for PM 10 Concentration in Sabah using Nearest Neighbour Method ( NNM ) and Expectation-Maximization ( EM ) Algorithm. , vol. 14, no. 1, pp. 62-72.

SALAS, C., ENE, L., OJEDA, N., SOTO, H. y DE CORRESPONDENCIA:, A., 2010. Métodos estadísticos paramétricos y no paramétricos para predecir variables de rodal basados en Landsat ETM+: una comparación en un bosque de Araucaria araucana en Chile Parametric and non-parametric statistical methods for predicting plotwise variables b. *Bosque* [en línea], vol. 31, no. 3, pp. 179-194. Disponible en:

<https://scielo.conicyt.cl/pdf/bosque/v31n3/art02.pdf>.

SALINAS, P. J. 2010. Metodología de la investigación científica. Un enfoque integrador. Venezuela, pp.450.

SCHNEIDER, T., 2001: Análisis de datos climáticos incompletos: estimación de valores medios y matrices de covarianza e imputación de valores perdidos. *J. Climate* , 14 ,853871, [https://doi.org/10.1175/15200442\(2001\)014<0853:AOICDE>2.0.CO;2](https://doi.org/10.1175/15200442(2001)014<0853:AOICDE>2.0.CO;2) .

SHELDON M. ROSS. *Introductory Statistics, (Third Edition)* [en línea]. Estado Unidos: Academic Press,2010[fecha de consulta: 18 de octubre de 2005]. Capítulo 12. Linear regression. Disponible en: <http://www.sciencedirect.com/science/article/pii/B9780123743886000120>, ISBN 9780123743886.

SILVA VIDAL, YAMINA, INSTITUTO GEOFÍSICO DEL PERÚ, 2020. IGP lanza su primer curso virtual sobre modelado de la atmosfera. Plata forma digital única del estado. [en línea]13 de mayo. Disponible en: <https://www.gob.pe/institucion/igp/noticias/154669-igp-lanza-su-primer-curso-virtual-sobre-modelado-de-la-atmosfera?fbclid=IwAR2xcDQncaoSzHgSjRcRjfHGBKcJBThjxpw2cFZ74g39BVoCfbT4VktZ-zc>

SILVENTE, VANESA BERLANGA. “Clasificación De Pruebas No Paramétricas. Cómo Aplicarlas En SPSS.” *REIRE. Revista d’Innovació i Recerca en Educació* (2012): n. pag. Web. [https://www.academia.edu/42342311/Clasificaci%C3%B3n\\_de\\_pruebas\\_no\\_param%C3%A9tricas\\_C%C3%B3mo\\_aplicarlas\\_en\\_SPSS](https://www.academia.edu/42342311/Clasificaci%C3%B3n_de_pruebas_no_param%C3%A9tricas_C%C3%B3mo_aplicarlas_en_SPSS)

SIMON P. NEILL, M. REZA HASHEMI,2005. *Fundamentals of Ocean Renewable Energy*[en línea]. Estado Unidos: Academic Press,2018[fecha de consulta: 18 de octubre de 2005]. Chaper 8. Ocean Modelling for Resource Characterization. Disponible en: <http://www.sciencedirect.com/science/article/pii/B9780128104484000082> ISBN 9780128104484

SOZZI, R., BOLIGNANO, A., CERADINI, S. y MORELLI, M., PETENKO, I., & ARGENTINI, S. 2017. Quality control and gap-filling of PM 10 daily mean concentrations with the best linear unbiased estimator. *Environmental monitoring and assessment*, 189(11), 562. <https://doi.org/10.1007/s10661-017->

6273-z

TUSELL, FERNANDO., 2011. Analisis de Regresi Introduccion Teorica y Practica basada en R Indice general. , pp. 264.

VARGAS CORDERO, Z.R., 2009.La Investigación aplicada: Una forma de conocer las realidades con evidencia científica. Revista Educación, vol.33, no.1, pp. 155. ISSN 0379-7082.DOI 10.15517/revedu. v33i1.538.

VIADA, C., BOUZA, C.N. y FORS, M., 2016. REVISIÓN SISTEMÁTICA DE LOS MÉTODOS DE IMPUTACIÓN DE DATOS FALTANTES. , RESEARCHGATE. Edition: Tomo, Chapter: Capítulo 12. Disponible en [https://www.researchgate.net/publication/289248594\\_REVISION\\_SISTEMATICA\\_DE\\_LOS\\_METODOS\\_DE\\_IMPUTACION\\_DE\\_DATOS\\_FALTANTES](https://www.researchgate.net/publication/289248594_REVISION_SISTEMATICA_DE_LOS_METODOS_DE_IMPUTACION_DE_DATOS_FALTANTES)

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 Conf. Series: Materials Science and Engineering 324 (2018) 012049 doi:10.1088/1757-899X/324/1/012049j

WEISS, Albert and Hays, Cinthia. Calculating daily mean air temperatures by different methods: implications from a non-linear algorithm. Agricultural and Forest Meteorology. [en línea]. 20 de enero 2005, [Fecha de consulta: 27 de abril de 2020]. Disponible en <http://www.sciencedirect.com/science/article/pii/S0168192304002199>. ISSN: 0168-1923,

ZAKARIA, N.A. & Noor, N.M.. (2018). Imputation methods for filling missing data in urban air pollution data for Malaysia. Journal Urbanism Architecture Constructions. Vol. 9,Nr.2,pp.159-166. ISSN 2069-6469

## ANEXOS

### ANEXO I

#### 3.1. Ficha de análisis de contenido

	<b>FICHA DE ANÁLISIS DE CONTENIDO</b>
---	---------------------------------------

#### TÍTULO:

<b>REVISTA CIENTÍFICA:</b> <input type="text"/>	<b>AÑO DE PUBLICACIÓN</b> <input type="text"/>	<b>LUGAR DE PUBLICACIÓN</b> <input type="text"/>
--	---	---

<b>TIPO DE INVESTIGACIÓN:</b>	<b>AUTOR (ES):</b>
-------------------------------	--------------------

<b>CÓDIGO:</b>	
<b>PALABRAS CLAVES:</b>	
<b>TIPO DE MÉTODO:</b>	
<b>PARÁMETRO DE MONITOREO:</b>	
<b>DATA FALTANTE:</b>	
<b>RESULTADOS:</b>	
<b>CONCLUSIONES:</b>	



**UNIVERSIDAD CÉSAR VALLEJO**

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

**Declaratoria de Originalidad del Autor**

Yo, CARRASCO VILLENA GLORIA ALEJANDRA estudiante de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA AMBIENTAL de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, declaro bajo juramento que todos los datos e información que acompañan la Tesis titulada: "REVISIÓN SISTEMÁTICA: ESTUDIO DE MÉTODOS PARA LA SUSTITUCIÓN DE DATOS METEOROLÓGICO NO REGISTRADOS EN CALIDAD DE AIRE.", es de mi autoría, por lo tanto, declaro que la Tesis:

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

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

<b>Nombres y Apellidos</b>	<b>Firma</b>
GLORIA ALEJANDRA CARRASCO VILLENA <b>DNI:</b> 70110759 <b>ORCID</b> 0000-0003-2657-6035	Firmado digitalmente por: GCARRASCOV1 el 21-12- 2020 21:53:07

Código documento Trilce: TRI - 0090775