



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

**COMPARACIÓN DE TÉCNICAS DE RELLENO DE DATOS
FALTANTES DE VARIABLES METEOROLÓGICAS EN LA
PROVINCIA DE CHIMBORAZO.**

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:
INGENIERO ESTADÍSTICO

AUTORES:

CRISTOFER JAVIER ALTAMIRANO ESPINEL
PABLO FERNANDO CARRILLO FREIRE

DIRECTORA: ING. CRISTINA ESTEFANÍA RAMOS ARAUJO, MSG.

Riobamba – Ecuador

2023

©2023, Altamirano Espinel Cristofer Javier y Carrillo Freire Pablo Fernando

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.

Nosotros, CRISTOFER JAVIER ALTAMIRANO ESPINEL Y PABLO FERNANDO CARRILLO FREIRE, declaramos que el presente trabajo de integración curricular es de nuestra autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

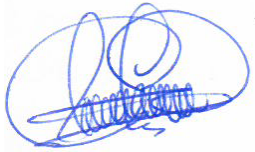
Como autores asumimos la responsabilidad legal y académica de los contenidos de este trabajo de integración curricular; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 21 de enero de 2023



Altamirano Espinel Cristófer Javier

172609705-6



Carrillo Freire Pablo Fernando

180531097-4

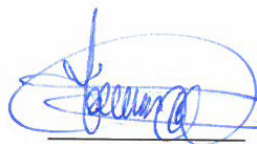
ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Integración Curricular certifica que: el Trabajo de Integración Curricular; Tipo: Proyecto de Investigación, **COMPARACIÓN DE TÉCNICAS DE RELLENO DE DATOS FALTANTES DE VARIABLES METEOROLÓGICAS EN LA PROVINCIA DE CHIMBORAZO**, realizado por los señores: **CRISTOFER JAVIER ALTAMIRANO ESPINEL Y PABLO FERNANDO CARRILLO FREIRE**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Integración Curricular, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

FIRMA

FECHA

Ing. Johanna Enith Aguilar Reyes, Mgs.
PRESIDENTE DEL TRIBUNAL



13/07/2023

Ing. Cristina Estefanía Ramos Araujo, Mgs.
DIRECTORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR



13/07/2023

Ing. Natalia Alexandra Pérez Londo, Mgs
ASESORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR



13/07/2023

DEDICATORIA

Este trabajo va dedicado a Dios por brindarme su protección y compañía en esta meta que me he propuesto y lo he logrado, también se la dedico a mi hija Citlaly Altamirano que con su llegada se convirtió en el motor y motivo de mis estudios, además agradecer a mis padres Amparo Espinel y Javier Altamirano por haber confiado en mí, y haberme ayudado de mil maneras a cumplir mi sueño en especial a mi madre que a pesar de tantas lagrimas hoy estamos viendo las recompensas, a mi novia Karla que ha sido una pieza fundamental en mi vida gracias por todo su apoyo, comprensión y paciencia que me brindo te amo, además dedicar ese triunfo a toda mi familia que estuvo a mi lado aportando con un granito de arena en especial a mi segundo padre Luis Altamirano que ha sido un ejemplo a seguir y me apoyado siempre, por ultimo agradecer a todos mis compañeros en especial a mi hermano de otra madre Fer gracias bro por todas esas anécdotas y travesías que hemos pasado y de lo cual hemos aprendido mucho.

Cristofer

Este trabajo va dedicado a Dios por haberme propinado la fuerza, el valor y las ganas de seguir luchando por alcanzar el sueño de ser ingeniero estadístico, agradezco a mis padres quienes con su esfuerzo, su cariño y amor incondicional me han brindado las energías para poder llegar hasta donde estoy en este momento y siempre tener el objetivo de ser mejor persona, a mi tío que está en el cielo y que desde ahí el está velando por toda la familia y especialmente por verme cumplir la promesa que le hice en sus últimos días de vida, a todos mis familiares que estuvieron pendientes de mi desarrollo como estudiante, a mi novia Samy Granizo que desde que llegó a mi vida siempre estuvo pendiente de que no me rinda en el camino quien siempre supo extender su mano cuando la necesitaba y por brindarme todo su amor desde aquel primer día de relación, además agradezco encarecidamente a todos mis compañeros y compañeras con quienes desde el primer semestre hemos compartido momentos llenos de alegrías, y finalmente mi gran amigo Cristo con el cual hemos pasado por buenos y malos momentos a lo largo de la carrera en los cuales hemos crecido como personas y nos hemos forjado como profesionales.

Fernando

AGRADECIMIENTO

En primer lugar, queremos agradecer a la Ing. Cristina Ramos, Ing. Natalia Pérez e Ing. Johanna Aguilar que nos han acompañado, guiado y aconsejado a lo largo del desarrollo de nuestro proyecto de investigación, además un sincero agradecimiento a todos los docentes que han sabido impartirnos sus conocimientos a lo largo de nuestro período como estudiantes de esta prestigiosa carrera, un especial agradecimiento además a todos nuestros compañeros, compañeras con los y las que hemos compartido infinidad de vivencias dentro y fuera de lo académico.

Cristofer y Fernando

TABLA DE CONTENIDOS

RESUMEN	viii
ABSTRACT	ix
INTRODUCCIÓN	1

CAPÍTULO I

1. PROBLEMA DE INVESTIGACIÓN	4
1.1. Antecedentes	4
1.2. Planteamiento del Problema	4
1.2.1. <i>Problema general de la investigación</i>	5
1.2.2. <i>Problemas específicos de la investigación</i>	5
1.3. Objetivos	6
1.3.1. <i>Objetivo general</i>	6
1.3.2. <i>Objetivos específicos</i>	6
1.4. Justificación	6
1.4.1. <i>Justificación Teórica</i>	6
1.4.2. <i>Justificación Metodológica</i>	6
1.4.3. <i>Justificación Práctica</i>	7
1.5. Hipótesis De La Investigación	7

CAPÍTULO II

2. MARCO TEÓRICO	8
2.1. Bases teóricas	8
2.1.1. <i>Variables Meteorológicas de Estudio</i>	8
2.1.2. <i>Factores del Clima</i>	8
2.1.3. <i>Estaciones meteorológicas</i>	9
2.1.4. <i>Clima en la provincia de Chimborazo</i>	10
2.1.5. <i>Precipitación</i>	10
2.1.6. <i>Datos faltantes</i>	10
2.1.7. <i>Índices de cambio climático recomendados por la ECTCCDI</i>	11

2.2.	Métodos, técnicas e instrumentos de investigación	11
2.2.1.	<i>Método de imputación con MissForest</i>	11
2.2.2.	<i>Método de imputación por medias</i>	11
2.2.3.	<i>Método de la Razón q</i>	11
2.2.4.	<i>Regresión Lineal Entre Variables</i>	12
2.2.5.	<i>Método de imputación por imágenes satelitales</i>	13

CAPÍTULO III

3.	MARCO METODOLÓGICO	14
3.1.	Tipo de investigación	14
3.2.	Diseño de la investigación	15

CAPÍTULO IV

4.	MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS	17
4.1.	Procesamientos	17
4.1.1.	<i>Obtención y manipulación de las bases de datos</i>	17
4.1.2.	<i>Estadísticos Descriptivos</i>	20
4.2.	Análisis e interpretación de resultados	22
4.3.	Estación Cumandá Imputación De Datos	22
4.3.1.	<i>Estación Cumandá Datos Faltantes</i>	22
4.3.2.	<i>Relleno de datos faltantes usando MissForest</i>	24
4.3.3.	<i>Relleno de datos mediante Medias</i>	26
4.3.4.	<i>Relleno De Datos Usando La Razón Q</i>	27
4.3.5.	<i>Relleno De Datos Usando Regresión Lineal Simple</i>	29
4.3.6.	<i>Relleno De Datos Mediante Imágenes Satelitales</i>	33
4.4.	Estación San Juan Imputación De Datos	35
4.4.1.	<i>Estación San Juan Datos Faltantes</i>	35
4.4.2.	<i>Relleno de datos faltantes mediante MissForest</i>	37
4.4.3.	<i>Relleno De Datos Mediante Las Medias</i>	38
4.4.4.	<i>Relleno De Datos Usando La Razón Q</i>	40
4.4.5.	<i>Relleno De Datos Usando Regresión Lineal Simple</i>	41
4.4.6.	<i>Relleno De Datos Mediante Imágenes Satelitales</i>	45

4.5.	Selección de la mejor técnica de imputación de datos faltantes	48
4.5.1.	<i>Matrices de Correlación Estación Cumandá</i>	48
4.5.2.	<i>Matrices de Correlación Estación San Juan</i>	55
4.5.3.	<i>Correlación Promedio Estación Cumandá</i>	62
4.5.4.	<i>Correlación Promedio Estación San Juan</i>	63
4.5.5.	<i>Errores Cuadráticos Medios Estación Cumandá</i>	65
4.5.6.	<i>Errores Cuadráticos Medios Estación San Juan</i>	67
4.6.	Representación gráfica de los resultados	69
4.6.1.	<i>Estación Cumandá</i>	69
4.6.2.	<i>Estación San Juan</i>	71

CAPÍTULO V

5.	CONCLUSIONES Y RECOMENDACIONES	75
5.1.	Conclusiones	75
5.2.	Recomendaciones	76

BIBLIOGRAFÍA

ANEXOS

ÍNDICE DE ANEXOS

Anexo A: Base de datos Originales

Anexo B: Obtención de las gráficas de los Datos Imputados en R

Anexo C: Matrices de Correlación y Errores Cuadráticos Medios

RESUMEN

Las variables meteorológicas han remarcado su importancia. Su estudio ha ayudado a comprender el comportamiento climático. Se instaló una red de 11 estaciones meteorológicas en la provincia de Chimborazo. La frecuencia de mantenimiento y calibración de los equipos ha dependido de recursos económicos que cada vez son más escasos, generando como consecuencia pérdida de información. Sin embargo, se evidenció una gran cantidad de datos faltantes para cada estación, se presentan datos registrados cada hora desde 2014 hasta el 2021. Mediante un análisis estadístico descriptivo se observó que la estación San Juan presenta aproximadamente el 0.32 % de datos faltantes en todo el periodo de tiempo. Por otro lado, la estación de Cumandá tiene aproximadamente un 32,42 % de datos faltantes, definiéndola como el peor escenario. El objetivo de esta investigación fue determinar la mejor técnica o método para imputación de datos faltantes en las variables: Temperatura Ambiente, Humedad Relativa y Precipitación. En el desarrollo de los objetivos, se propone varias técnicas de imputación de datos encontradas en bibliografía para poder rellenar los datos ausentes que presentan las diferentes estaciones meteorológicas. Se han comparado las siguientes técnicas: relleno de datos por medias, relleno de datos usando la razón q, relleno de datos aplicando regresión lineal simple, relleno de datos usando MISSFOREST en R y relleno de datos usando imágenes satelitales. Una vez aplicadas estas técnicas y métodos sobre las dos estaciones (Cumandá y San Juan) se ha determinado cuál de estas resulta ser la mejor al momento de rellenar los datos faltantes, a través de métricas estadísticas tales como la media, la varianza, los errores cuadráticos y además mediante series temporales para analizar el comportamiento de los datos antes y después del relleno de datos. Con esto se llega a la conclusión que el método MISSFOREST es apropiado para el relleno de variable Temperatura Ambiente y Humedad Relativa y la técnica de imágenes satelitales satisface la variable Precipitación.

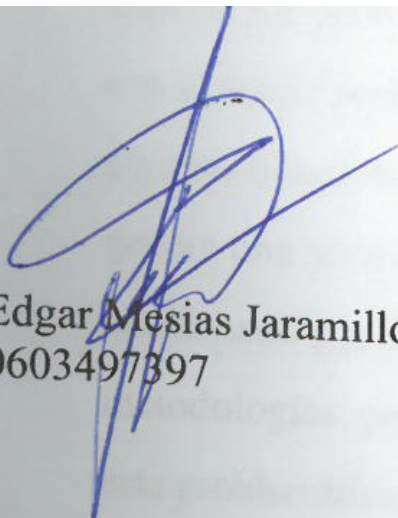
Palabras clave: <METEOROLOGÍA>, <MÉTODO DE IMPUTACIÓN>, <MÉTODO DE MISSFOREST>, <IMÁGENES SATELITALES>, <TÉCNICAS DE RELLENO DE DATOS FALTANTES>



ABSTRACT

Meteorological variables have highlighted their importance. Their study has helped to understand climate behavior. A network of 11 meteorological stations was installed in the province of Chirnborazo. The frequency of maintenance and calibration of the equipment has depended on economic resources that are increasingly scarce, generating as a consequence loss of information. However, a large amount of missing data was evidenced for each station, data recorded every hour from 2014 to 2021 is presented. Through a descriptive statistical analysis, it was observed that the San Juan station presents approximately 0.32 *hand, the Cumand station has approximately 32.42* development of the objectives, several data imputation techniques found in the bibliography are proposed to be able to fill in the missing data presented by the different meteorological stations. The following techniques have been compared: data filling by means, data filling using the q ratio, data filling using simple linear regression, data filling using MISSFOREST in R, and data filling using satellite imagery. Once these techniques and methods have been applied to the two stations (Cumandá and San Juan), it has been determined which of these turns out to be the best when filling in the missing data, through statistical metrics such as the mean, the variance, the errors quadratics and also through time series to analyze the behavior of the data before and after data filling. With this, it is concluded that the MISSFOREST method is appropriate for filling in the Ambient Temperature and Relative Humidity variable and the satellite image technique satisfies the Precipitation variable.

Palabras clave: <METEOROLOGY>, <IMPUTATION METHOD>, <MISSFOREST METHOD>, <SATELLITE IMAGES>, <DATA FILLING TECHNIQUES>



Edgar Mesias Jaramillo Moyano
0603497397

INTRODUCCIÓN

El presente proyecto tiene un enfoque a la meteorología por lo cual podemos decir que, los orígenes de la Organización Meteorológica Internacional (OMI) remontan a 1873, año en que se celebró en Viena el primer Congreso Meteorológico Internacional. El Congreso encomendó a un Comité Meteorológico Permanente la tarea de redactar los reglamentos y estatutos de una organización meteorológica internacional para facilitar el intercambio transfronterizo de información meteorológica. Esa tarea se finalizó en 1878, en Utrecht, y el año siguiente se fundó la OMI en el Congreso Meteorológico Internacional de Roma. Siguió funcionando hasta 1950, año en que la OMI se convirtió oficialmente en la Organización Meteorológica Mundial (OMM) (Organización 2016). Tras varios años en el Congreso de Roma se aprobó una resolución por la que se creaba un Comité Meteorológico Internacional, la cual sería sustituida por un mandato similar al que ya existía. Heinrich Wild, de origen suizo director de Servicios Meteorológicos Internacional, y Robert Scott director del Observatorio Geofísico Central de San Petersburgo. El mandato consistía en promover la cooperación internacional en meteorología, fomentar las investigaciones meteorológicas y dar uniformidad a las prácticas operativas, en particular en lo que se refería a las observaciones e informes meteorológicos, así como al intercambio y la difusión de datos. El compromiso de Wild y de Scott con la meteorología internacional queda demostrado por todos los años dedicados al desempeño de sus funciones: Wild hasta 1896 y Scott hasta 1900. Llenos de entusiasmo y con gran competencia, utilizaron como base la labor realizada por Buys Ballot y otros, y trabajaron incansablemente para formular y promover un mecanismo de cooperación internacional en meteorología. De hecho, el compromiso y el entusiasmo demostrados han sido un sello distintivo de quienes han ayudado a configurar la historia de la OMI y de la institución que la sustituyó: la Organización Meteorológica Mundial (Organización 2016). Es un hecho conocido por los profesionales que lidian con datos en servicios meteorológicos/hidrológicos que los sistemas de medición en ocasiones presentan problemas técnicos, los cuales pueden comprometer la integridad de los datos generados. Los datos faltantes son problemáticos, debido a que la mayoría de los métodos estadísticos no pueden ser aplicados de forma directa a datos incompletos (Josse y col. 2012). Uno de los problemas que se presenta en climatología al momento de procesar bases de datos con el fin de realizar los análisis estadísticos, es la ausencia de registros diarios o mensuales en algunas estaciones climatológicas que se mantienen en funcionamiento de forma regular, lo que genera una serie de datos faltantes en dichas bases de datos. Como solución a este problema que es ocasionado por errores humanos o informáticos, se han creado una serie de técnicas, métodos, metodologías, procedimientos, modelos estadísticos y matemáticos que permiten dar solución a esta problemática, este trabajo muestra una metodología que se utilizan para la imputación de datos

en series climáticas (Urrutia y col. 2010). Los datos faltantes son un problema familiar en todas las áreas de investigación. Si los datos son recolectados y almacenados por alguno sistema automático, como es el caso de la recogida de datos en las actuales estaciones meteorológicas, pueden ocurrir fallos en el automatismo, ocasionados por la falta de energía, o por algún tipo de mala configuración del mecanismo. Estos tipos de falta de datos que suelen escapar al control del investigador. A fin de definir los años-tipo, se necesitan de muestras completas de datos. Así, hay que de algún modo, intentar completar las bases de datos, si no con los datos reales, pero al menos con datos generados por alguna metodología estadística que garantice la aleatoriedad de los datos imputados, y que se ajuste a la estructura del proceso estocástico subyacente (Mateus Ferreira 2003). La serie de tiempo anual, simulado series incompletas con la eliminación aleatoria de 10 % y 20 % de los datos de velocidad del viento, en rachas de 4, 8, 12, 16, 20 y 24 horas, y han estimado los datos faltantes por diversas metodologías (matrices de transición de Markov, análisis de la tendencia aumento o disminución de la velocidad de una hora a la siguiente, y modelo autorregresivo), y han concluido que las técnicas ensayadas han conducido a resultados muy parecidos a la serie original. Los errores en las estimaciones de la velocidad media han sido inferiores a unos 5 %, y los errores en las estimaciones teóricas de la energía producida por un aerogenerador han sido aún menores, de la orden de unos 2 % o menos (Kline y col. 1998).

La climatología es aquella que busca analizar el comportamiento y a su vez la interacción entre variables las cuales estén relacionadas con la perturbación climática, teniendo en cuenta la variabilidad con la que se presenten los datos obtenidos mediante un estudio de zona como puede ser su ubicación y las escalas que se implementen en su medición. Los eventos climáticos afectan la producción y abastecimiento de recursos por lo que cada país está en la obligación de monitorear el clima y sus variaciones, esto se consigue mediante el uso de estaciones meteorológicas que permite recopilar información que se utiliza para el estudio del clima, su comportamiento y predicción a corto, mediano y largo plazo (Toro Trujillo y col. 2018).

Por esta razón se sugiere realizar mediciones en diferentes escalas temporales mediante una estación meteorológica la cual permita la recolección de dicha información ya sea de forma manual o automática la cual nos permita estar al tanto de los eventos climáticos pertinentes en las diferentes zonas. En Ecuador, el INAMHI (Instituto Nacional de Meteorología e Hidrología) es la organización que tiene a su cargo el monitoreo y recolección de información de las estaciones meteorológicas, que, a pesar de su importancia, varias de estas estaciones se encuentran en abandono o no están completamente operativas lo que genera datos erróneos e incompletos. Estos datos deberán ser verificados y rellenados para fines académicos con el mínimo error posible (Carrera-Villacrés y col. 2016).

Las series que se busca completar son las de temperatura, precipitación a diferentes escalas, pudiendo ser diaria o mensual, etc., esto dependerá del tipo de análisis, los resultados y objetivos que se persigan durante la investigación. Para ello, existen varios métodos que permiten el rellenado de datos meteorológicos entre ellos se puede mencionar el método de regresión-correlación, que depende del supuesto de linealidad entre los datos de las estaciones que se encuentren dentro de un área de influencia cercana y establece el nivel de relación mediante un análisis de dispersión utilizando un coeficiente de correlación, posterior a esto se obtiene una ecuación que se usa para calcular el valor faltante en función de los datos de las estaciones de referencia (Matovelle y col. [2022](#)). Otro método es el denominado promedio aritmético simple (SAA), se utiliza cuando el valor anual en cada uno de los datos a promediar difiere por lo menos de 10% y llega a presentar estimaciones confiables si el dato analizado no tiene variabilidad espacial y si existe correlación entre las estaciones utilizadas para el análisis. El método de ponderación de distancia inversa IDW se utiliza cuando se dispone de registros en la misma escala temporal de tres estaciones cercanas a la estación en estudio. El IDW se utiliza en cualquier delta de tiempo y para su utilización se requiere una cantidad suficiente de datos para producir un semi vario grama fiable (Toro Trujillo y col. [2018](#)).

CAPÍTULO I

1. PROBLEMA DE INVESTIGACIÓN

1.1. Antecedentes

En los repositorios de la biblioteca de la Escuela Superior Politécnica de Chimborazo en la Facultad de Ciencias, Carrera de Estadística se encontró un proyecto de investigación que tiene relación con nuestro trabajo de titulación, el estudio se realizó empleando la librería Factoshiny de R y los datos corresponden a los obtenidos de las estaciones meteorológicas Alao, Atillo, Cumandá, Espoch, Matus, Multitud, Quimiag, Tunshi y Urbina; durante el año 2015 en los meses de enero a noviembre. El análisis de componentes principales permitió identificar dos tipos de comportamientos en la velocidad de viento: el primero lo determina la región norte de la provincia, con zonas ubicadas a una altura superior que los 1000 m.s.n.m; y el segundo corresponde a la región del subtrópico, misma que posee mayor diversidad, con fauna y flora totalmente diferente a la primera (Haro Rivera y col. [2020](#)).

El Grupo de Energías Alternativas y Ambiente de la Facultad de Ciencias de la ESPOCH tiene como principal objetivo contribuir con conocimientos al desarrollo de la investigación, ciencia y tecnología, en el ámbito de las energías alternativas y ambiente; pero a lo largo del tiempo se ha notado que las estaciones meteorológicas distribuidas en la provincia de Chimborazo han recolectado gran cantidad de datos faltantes o de tipo inválidos esto debido a diversos factores tales como fallas en los sensores, apagones en las estaciones, entre otros.

1.2. Planteamiento del Problema

Hoy en día una de las grandes preocupaciones para el ámbito meteorológico se presenta al momento de analizar las matrices de datos ya que en la gran mayoría de estas se presentan datos faltantes, o a su vez éstas matrices cuentan con datos atípicos o de tipo inválidos, que resultan ser datos que no sirven para estudios posteriores tales como predicción climática a corto y largo plazo, análisis de los estadísticos de las variables existentes, entre otros, ya que para poder efectuar el procesamiento de datos de tipo meteorológico se necesita tener la mayor cantidad de datos válidos para que los estudios sean acertados y de gran ayuda para el Grupo de Energías Alternativas y Ambiente (GEAA), La Escuela Superior Politécnica De Chimborazo (ESPOCH), para estudios generados en la provincia de Chimborazo, y a gran escala para brindar apoyo a diversas estaciones meteorológicas distribuidas a nivel nacional.

Dentro del Grupo de Energías Alternativas y Ambiente (GEAA) el principal problema que se ha identificado es la existencia de datos faltantes o inválidos recolectados por las estaciones meteorológicas ubicadas en toda la provincia de Chimborazo, debido a problemas en las estaciones meteorológicas tales como: fallas al momento de calibrar los sensores de las estaciones, daños irreparables en los sensores, apagones en las estaciones, falla en la recolección de datos, fallas en la manipulación de los datos, entre otros, y es por eso que se ha visto la necesidad de identificar una técnica que permita efectuar el relleno de datos faltantes o inválidos en las variables Temperatura Ambiente, Humedad Relativa y Precipitación que son las variables de interés del presente proyecto de investigación.

1.2.1. *Problema general de la investigación*

¿Qué tan efectivas serán las técnicas establecidas para la imputación de datos faltantes en las variables meteorológicas Temperatura Ambiente, Humedad Relativa y Precipitación?

1.2.2. *Problemas específicos de la investigación*

¿Cómo ejecutar la comparativa entre las técnicas consideradas para el relleno de datos faltantes en las variables de interés?

¿Cuál o cuales son los estimadores que se van a considerar para determinar la mejor técnica de relleno de datos, una vez que se han comparado todas las técnicas?

1.3. Objetivos

1.3.1. *Objetivo general*

Comparar las técnicas de relleno de datos faltantes de variables meteorológicas sobre las bases de datos proporcionadas por el Grupo de Energías Alternativas y Ambiente (GEAA) para poder determinar la técnica que mejores resultados genera.

1.3.2. *Objetivos específicos*

- Realizar una depuración de datos inválidos, y elaborar nuevas bases de datos donde solamente se usen las variables consideradas para el estudio a partir de las bases de datos originales.
- Realizar un análisis estadístico mediante el cálculo de los totales de datos faltantes para cada una de las estaciones con la finalidad de poder determinar cual es la estación que presenta mayor y menor cantidad de datos faltantes.
- Comparar las técnicas usadas para relleno de datos faltantes a través de la experimentación y el análisis de cada una de ellas para saber si es viable y factible emplear dicha técnica sobre las variables de estudio.
- Obtener estimadores estadísticos que permitan determinar cual es la mejor técnica para la imputación de datos faltantes en las variables consideradas para el presente proyecto de investigación.

1.4. Justificación

1.4.1. *Justificación Teórica*

El presente proyecto de investigación busca efectuar la comparativa entre diferentes técnicas, las cuales son relleno de datos mediante MissForest, Medias, Razón Q, Regresión Lineal Simple e Imágenes Satelitales, las cuales serán comparadas sobre las estaciones meteorológicas que presenten mayor y menor cantidad de datos faltantes en las variables Temperatura Ambiente, Humedad Relativa y Precipitación.

1.4.2. *Justificación Metodológica*

La presente investigación servirá para lograr mejoras tanto para una variable como para varias variables de tipo meteorológico al momento de encontrar valores que reemplacen a los datos

faltantes que estas presenten, una vez determinada cual es la mejor técnica de relleno de datos a través de la comparación entre estas, la que presente mejores indicadores estadísticos será seleccionada como la mejor técnica para imputación de datos sobre las variables Temperatura Ambiente, Humedad Relativa y Precipitación.

1.4.3. *Justificación Práctica*

A partir de los objetivos planteados se va a determinar la mejor técnica para el relleno de datos faltantes en variables meteorológicas, lo cual servirá principalmente para el Grupo de Energías Alternativas y Ambiente (GEAA), para la Carrera de Estadística, para la Escuela Superior Politécnica de Chimborazo, entre otras instituciones y organizaciones ligadas al ámbito meteorológico, ya que mediante esto se obtendrá un resultado válido y de fácil acceso para el relleno de datos faltantes en variables de tipo meteorológico, con lo cual los estudios estadísticos posteriores se presentarán de mejor manera al tener bases de datos completas.

1.5. Hipótesis De La Investigación

Se buscará comprobar que el comportamiento de los datos rellenados no varía con el comportamiento que presentan los datos observados en las variables meteorológicas originales.

CAPÍTULO II

2. MARCO TEÓRICO

2.1. Bases teóricas

2.1.1. *Variables Meteorológicas de Estudio*

Temperatura

Es la cantidad de energía calorífica que posee el aire en un momento determinado. Se mide mediante termómetros, habitualmente en grados Celsius y determina las sensaciones de calor y frío (Navarra 2023).

Precipitación

Es la caída al suelo del agua contenida en la atmósfera. Puede ser en forma de agua, de nieve, de brumas o de rocío y se produce cuando la atmósfera no puede contener más agua y esta se condensa y precipita. Se mide en litros por metro cuadrado de superficie (l/m^2), o su medida equivalente milímetros de altura del agua caída (mm). El instrumento de medición es el pluviómetro (Navarra 2023).

Humedad

La humedad absoluta es la cantidad de vapor de agua presente en el aire y se mide en g/m^3 . La humedad relativa es la relación entre la cantidad de vapor de agua que contiene el aire y la máxima cantidad de vapor de agua que puede contener a una determinada temperatura. Cuanto mayor es la temperatura del aire, más cantidad de vapor de agua disuelto admite. La humedad relativa se mide en porcentaje: un valor de 100% indica que el aire está saturado de vapor de agua y ya no puede retener más, lo que da lugar a la formación de nubes, nieblas, rocío o si la temperatura es lo suficientemente baja, escarcha. (Navarra 2023).

2.1.2. *Factores del Clima*

Latitud

Es la distancia angular que hay desde un punto en la superficie de la tierra hasta el ecuador. Se mide en grados, minutos y segundos. Cuanto más cerca se encuentre un lugar del ecuador menos variación habrá en la duración de los días y la energía solar incidirá más vertical y con más intensidad sobre él, por lo que más cálidas serán las temperaturas. Cuanto más lejos se encuentre ese punto del ecuador la energía solar le alcanzará con un ángulo menor, además durante el invierno los días

serán cortos acentuando el frío (Navarra 2023).

Si el punto se encuentra cerca de los polos los rayos solares le llegarán con un ángulo muy pequeño y solo durante algunos meses ya que en invierno allí no sale el sol. Esto hará que las temperaturas sean muy bajas todo el año (Navarra 2023).

La latitud determina las grandes franjas climáticas de la Tierra:

- Una zona cálida o tropical a ambos lados del ecuador entre los trópicos de Cáncer y Capricornio.
- Dos zonas templadas norte y sur, entre estos trópicos y los círculos polares ártico y antártico respectivamente.
- Dos zonas frías o polares en torno a los polos norte y sur.

En general se puede decir que la zona tropical tiene temperaturas elevadas (excepto en las regiones de mucha altitud) y que varían poco a lo largo del año, que las zonas templadas tienen cuatro estaciones y que las zonas polares tienen siempre temperaturas muy bajas.

Cambios Climáticos

El cambio climático se conoce como grandes variaciones en los promedios del clima. Según el IPCC Grupo Intergubernamental sobre el Cambio Climático, (2013) es la variación estadísticamente significativa en las condiciones climáticas medias o en su variabilidad durante un período prolongado, típicamente décadas o más. Aunque estas variaciones se producen a escala mundial, sus efectos a menudo varían de una región a otra. Las condiciones climáticas en el Ecuador, por ejemplo, son influenciadas por dos variables o factores principales la temperatura y la precipitación, que dan lugar a marcados cambios temporales y espaciales en las diferentes regiones del país. A diferencia de los países con latitudes altas, en el Ecuador se observan dos épocas bien diferenciadas por la distribución temporal de las precipitaciones, una época lluviosa y otra seca, excepto en la Amazonía, ya que las lluvias son consideradas durante todo el año, al igual que en el resto del planeta, las observaciones de temperatura muestran un leve cambio que se ha ido registrando de manera paulatina a través del tiempo (Pérez y col. 2020).

2.1.3. Estaciones meteorológicas

La red meteorológica del IDEAM está compuesta por un conjunto de estaciones de observación, medición y registro de diferentes fenómenos atmosféricos, con diferentes finalidades, distribuidas en el territorio (IDEAM, 2016). Esta estación meteorológica cuenta con dos piranómetros SR11-10-Hukseflux, uno de ellos provisto de un anillo de sombra para el registro de radiación difusa, pluviómetro, anemómetro ultrasónico 8500 de marca Vaisalia, barómetro QML 201C de marca

Vaisalia, termómetro HMP155 de marca Vaisalia, además de sensores de perfil de temperatura de suelo (Pérez y col. 2020).

2.1.4. *Clima en la provincia de Chimborazo*

La provincia de Chimborazo está ubicada en la región centro de Ecuador, por estar situada en la cordillera de los Andes presenta grandes elevaciones como el Chimborazo (6.310 m), tiene fértiles valles como los de Guano, Riobamba y Penipe. La provincia tiene importantes afluentes; entre los principales están: el Chambo, Chanchán y Chimbo. Debido a la presencia de masas de aire 27 amazónicas y oceánicas; y sobre todo la altura, se presentan cuatro tipos de climas; a los 4.600 m.s.n.m. es de tipo glacial, entre los 3.000 y 4.000 m.s.n.m. es páramo, bajando; a los 2.000 m.s.n.m. el clima es mesotérmico seco; y en zonas cercanas a la costa el clima es mesotérmico húmedo y semihúmedo. La provincia se caracteriza por ser ganadera y agrícola, por tal razón es importante analizar el comportamiento meteorológico de la velocidad de viento (Toainga Oñate 2015).

2.1.5. *Precipitación*

Se denomina precipitación al agua que proviene de la humedad atmosférica y cae a la superficie terrestre, principalmente en estado líquido, (lluvia) o solido (nieve o granizo) (Russi y Larrahondo, 2017). Los dos principales dispositivos de medición de la precipitación son el pluviómetro y el pluviógrafo. El primero mide el volumen total de lluvia caída durante el día meteorológico (7 A.M. a 7 A.M. del día siguiente). El segundo es un aparato de registro continuo que permite determinar la intensidad de la precipitación en un intervalo de tiempo dado (Russi y col. 2017).

2.1.6. *Datos faltantes*

En un sentido amplio, el termino datos faltantes significa que falta algún tipo de información sobre los fenómenos en los que se está interesado. Existen situaciones inherentes que pueden afectar la calidad de los datos o registros de las variables de estudio, tal es el caso de la información de tipo climático, específicamente, las series históricas de precipitación, construidas a partir de las mediciones realizadas en distintas estaciones meteorológicas, las cuales pueden verse afectadas por el cambio de lugar o movimiento de instrumentos, el cambio del espacio físico del entorno donde se encuentra la estación, o no registrar el dato en el momento adecuado, lo que conlleva a la inconsistencia o ausencia de datos (Russi y col. 2017).

2.1.7. Índices de cambio climático recomendados por la ECTCCDI

El ECTCCDI aprobó una lista de 40 índices, de los cuales 27 fueron elegidos como los básicos para un análisis de eventos climáticos extremos, tomando en cuenta valores diarios de temperatura, cantidad de precipitación diaria, y umbrales fijos de relevancia para aplicaciones particulares. (Serrano Vincenti y col. 2012).

2.2. Métodos, técnicas e instrumentos de investigación

2.2.1. Método de imputación con MissForest

MissForest es un algoritmo de imputación de bosque aleatorio para datos faltantes, implementado en R en el paquete `missForest()`. Inicialmente imputa todos los datos faltantes usando la media/moda, luego, para cada variable con valores faltantes, MissForest ajusta un bosque aleatorio en la parte observada y luego predice la parte faltante. Este proceso de entrenamiento y predicción se repite en un proceso iterativo hasta que se cumple un criterio de parada establecido por la programación de la paquetería misma o se alcanza un número máximo de iteraciones especificadas por el usuario. El motivo de las iteraciones múltiples es que, a partir de la iteración 2 en adelante, los bosques aleatorios que realizan la imputación se entrenarán con datos cada vez de mejor calidad que se han imputado de manera predictiva (Morgan 2020).

2.2.2. Método de imputación por medias

Consiste en imputar los datos faltantes con el valor promedio de la variable que se desea imputar, pero para poder usar esta técnica se debe tener en consideración que el porcentaje de datos faltantes presentado en la variable de interés deberá ser menor o igual al 10%.

$$y = \frac{\sum_{i=1}^n (a_i)}{n} \quad (1)$$

n : Número total de datos presentados en la variable.

a_i : Dato i -ésimo de la variable.

y : Valor de la media, que servirá para imputación de datos faltantes.

2.2.3. Método de la Razón q

Consiste en obtener un estimador llamado razón, denotado por: q_i , el cual resulta de el cociente entre la sumatoria de los datos existentes en la variable A (Variable con datos faltantes) y a sumatoria de

los datos de la variable B (Variable sin datos faltantes), este estimador está definido por:

$$q = \frac{\sum_{i=1}^n (a_i)}{\sum_{i=1}^n (b_i)} \quad (2)$$

Donde:

N : Número total de datos de la serie

a_i : Dato i -ésimo de la estación A

b_i : Dato i -ésimo de la estación B

Una vez que se ha determinado el estimador q_i para determinar los valores para imputación se debe multiplicar el valor de la razón q por cada uno de los i -ésimos valores de la variable que no presenta valores faltantes.

$$y = q * a_i \quad (3)$$

Donde:

y : Nuevos datos que servirán para imputación.

q : Razón Q para la variable a imputar.

a_i : Dato i -ésimo de la estación A (Sin datos faltantes).

2.2.4. Regresión Lineal Entre Variables

La guía de prácticas climatológicas de la Organización Meteorológica Mundial (OMM, 2011) propone diversas metodologías utilizadas para el relleno de datos faltantes de bases de datos meteorológicas. Una de ellas es la metodología empleada por Carrera et al. (2016), la cual supone una relación lineal entre variables, para esto se establece una regresión y correlación lineal entre una estación auxiliar y la estación en estudio con carencia de información, mediante una ecuación de dos variables:

$$y = a + bx \quad (4)$$

Donde:

y = Valor estimado de la variable para la variable carente.

x = Valor de la variable registrada por la variable auxiliar.

a, b = Parámetros de regresión.

Este método se recomienda para la estimación de datos diarios, mensuales y anuales. El principal indicador de este método es el coeficiente de determinación el cual puede adquirir resultados que oscilan entre 0 y 1. Así, cuando adquiere resultados más cercanos a 1, mejor resultará el ajuste del

modelo a la variable, pero si los valores de este estimador se aproximan a cero el ajuste del modelo a la variable será menos adecuado (Ayala y col. 2018).

2.2.5. Método de imputación por imágenes satelitales

Los satélites artificiales que orbitan en el espacio tienen varias funciones, como emitir señales de telecomunicaciones, recoger información con fines estratégicos o para predecir el comportamiento del clima. En cualquier caso, los satélites recogen imágenes que permiten analizar todo tipo de realidades. Se trata de imágenes que se pueden obtener de forma muy rápida e inmediata. De hecho, actualmente puede hacerse de forma instantánea a través de internet. Permiten conocer áreas de gran extensión y a un bajo costo. Es posible acceder a lugares muy remotos, puesto que desde el espacio desaparecen las fronteras y esto permite que sea posible captar cualquier lugar del planeta por remoto que sea (Navarro 2017).

Este método consiste en obtener las bases de datos generadas por los satélites artificiales, y depurar los datos de forma que se obtengan solamente las variables de interés en el estudio, para el siguiente reemplazar los datos faltantes por los valores presentados en las bases de datos satelitales, logrando así la imputación.

CAPÍTULO III

3. MARCO METODOLÓGICO

3.1. Tipo de investigación

Dentro del tipo de investigación el presente proyecto será de tipo cuantitativo ya que es conocido que el enfoque cuantitativo es secuencial y probatorio, cada etapa procede a la siguiente y no podemos brincar o eludir pasos. En la figura ... se presenta el enfoque que se ha usado en la investigación.

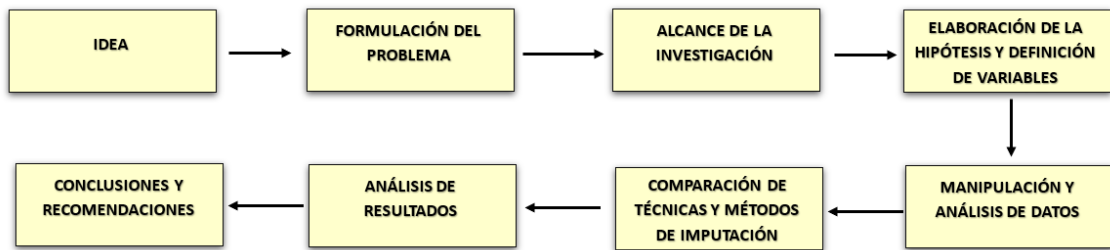


Figura 3-1: Enfoque de la investigación.

Fuente: (Metodología de la Investigación, 2014)

La investigación parte de la idea de buscar una solución para el problema de datos faltantes presentados en las variables Temperatura Ambiente, Humedad Relativa y Precipitación en las estaciones meteorológicas ubicadas en la provincia de Chimborazo siendo este un enorme problema ya que al presentarse datos incompletos o inválidos, los estudios posteriores efectuados por el Grupo de Energías Alternativas y Ambiente (GEAA) no serán efectivos y los resultados obtenidos se verán alterados, y es así que surge la necesidad de encontrar una técnica o un método de imputación de datos para poder solventar este problema y facilitar el trabajo del grupo de investigación, el presente proyecto tiene un alcance a nivel institucional bastante bueno ya que se va a determinar una técnica o método de imputación de datos faltantes con lo cual se van a tener bases de datos donde las variables Temperatura Ambiente, Humedad Relativa y Precipitación se muestren completas en su totalidad con datos válidos, y es por esto que surge la hipótesis de saber si las técnicas o métodos de imputación de datos efectivamente generarán resultados acertados, para esto se han considerado como variables de estudio la correlación y los errores cuadráticos medios ya que mediante estos se va a poder dar respuesta o solución a la hipótesis planteada para la presente investigación. Una vez que se ha mencionado la estructura del proyecto de investigación se deberá realizar una manipulación de las bases de datos facilitadas por el Grupo de Energías Alternativas y Ambiente

(GEAA) de forma que para el proceso de experimentación se cuente solamente con datos válidos de las variables de estudio en el periodo de tiempo comprendido por los años 2014 - 2021 para las variables Temperatura Ambiente y Humedad Relativa y en el periodo de tiempo comprendido entre 2019 - 2021 para la variable Precipitación, determinándose además los totales de datos estimados y datos faltantes para cada variable y cada año de estudio con lo cual se puede determinar cual es la estación meteorológica que presenta mayor y menor cantidad de datos faltantes, ya que de esta forma es mucho más eficiente efectuar el relleno de datos faltantes mediante los métodos y técnicas seleccionadas en esta investigación ya que los datos faltantes serán imputados en este par de estaciones optimizando así el trabajo, una vez que se termine el proceso de imputación mediante métodos y técnicas de relleno de datos faltantes es necesario analizar el comportamiento de los nuevos datos respecto a los datos originales, por lo cual se analiza de forma gráfica mediante series temporales los resultados de la imputación, pero además se considera el cálculo de matrices de correlación entre pares de datos, el cálculo del coeficiente de correlación promedio y los errores cuadráticos medios generados por las nuevas bases de datos respecto a las bases de datos originales, y finalmente se emiten conclusiones en base a los resultados obtenidos en la investigación y además recomendaciones que servirán para la mejora y el desarrollo de investigaciones posteriores.

3.2. Diseño de la investigación

Se utilizará el diseño no experimental, debido a que a lo largo del proyecto de investigación se irá experimentando de forma controlada con diversas técnicas y métodos determinados para el relleno de datos, entre los cuales están: Imputación mediante MissForest, Relleno usando las Medias, Imputación mediante la Razón Q, Relleno mediante Regresión Lineal Simple y finalmente Imputación usando Imágenes Satelitales.

Dentro del proceso de experimentación se va a tomar como referencia dos estaciones modelo, la primera es aquella que presente la mayor cantidad de datos faltantes, y la segunda estación es aquella que presente la menor cantidad de datos faltantes en las tres variables de interés, una vez identificadas estas estaciones se puede aplicar las técnicas y métodos de imputación de datos faltantes.

Paso siguiente una vez que se han identificado cuales son las estaciones con mayor y menor cantidad de datos faltantes y se han imputado sus datos se procede a efectuar un análisis para así determinar cual es la mejor técnica o método para el relleno de datos faltantes, para este proceso se efectuarán gráficas de los datos imputados donde se van a analizar los comportamientos de las series y validar si los trazos son continuos, con esto se podrá ver si las técnicas y métodos efectivamente han sido capaces de imputar los datos faltantes, después una vez se tengan bases de datos completamente

imputadas se procede a efectuar un análisis más amplio para cada una de las variables objeto de estudio, un primer acercamiento a la detección de la mejor técnica o método de imputación de datos será la obtención de las matrices de correlación generadas por cada técnica o método de imputación para saber el grado de relación presentados entre los datos originales y los datos imputados, pero para poder tener una mejor visión de los resultados obtenidos se determinan los coeficientes de correlación promedio generados por pares de datos (observados vs imputados), teniendo como referencia que para que los resultados sean viables se deberá tener coeficientes de correlación iguales o mayores al 70 %, caso contrario los resultados generados por la técnica o método no serán viables y por tanto deberán ser descartados en el estudio, finalmente se calculan los errores cuadráticos medios producidos por cada técnica o método en todos los años con los que se cuenta información en cada variable de estudio, siendo este indicador el último paso para la detección de la mejor técnica o método de imputación de datos, ya que se define como mejor a aquella técnica o aquel método que presente el menor coeficiente de error cuadrático medio.

Finalmente para tener un acercamiento a los resultados obtenidos por la mejor técnica o el mejor método se ejecuta una gráfica con tres secciones, donde se presentan resultados específicos antes y después de la imputación y una gráfica general donde se visualice el comportamiento general de los datos antes y después del relleno.

CAPÍTULO IV

4. MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

4.1. Procesamientos

4.1.1. *Obtención y manipulación de las bases de datos*

El primer paso para la consecución de los objetivos del proyecto de investigación es la obtención de los datos, los cuales fueron facilitados por el Grupo de Energías Alternativas y Ambiente (GEAA), como se muestra en la figura 4-1, las estaciones sobre las que se ha efectuado la comparativa de técnicas de relleno de datos faltantes son: Alao, Atillo, Cumandá, Espoch, Matus, Multitud, Quimiag, San Juan, Tixan, Tunshi y Urbina.

<input checked="" type="checkbox"/>	ALAO	19/11/2022 21:29	Carpeta de archivos
<input checked="" type="checkbox"/>	ATILLO	19/11/2022 21:29	Carpeta de archivos
<input type="checkbox"/>	CUMANDA	19/11/2022 21:29	Carpeta de archivos
<input checked="" type="checkbox"/>	ESPOCH	19/11/2022 21:29	Carpeta de archivos
<input checked="" type="checkbox"/>	MATUS	19/11/2022 21:30	Carpeta de archivos
<input checked="" type="checkbox"/>	MULTITUD	19/11/2022 21:30	Carpeta de archivos
<input checked="" type="checkbox"/>	QUIMIAG	19/11/2022 21:30	Carpeta de archivos
<input checked="" type="checkbox"/>	SAN JUAN	19/11/2022 21:30	Carpeta de archivos
<input checked="" type="checkbox"/>	TIXAN	19/11/2022 21:31	Carpeta de archivos
<input checked="" type="checkbox"/>	TUNSHI	19/11/2022 21:31	Carpeta de archivos
<input checked="" type="checkbox"/>	URBINA	19/11/2022 21:31	Carpeta de archivos

Figura 4-1: Estaciones Meteorológicas.

Realizado por: Altamirano C. y Carrillo P., 2023

Dentro de las carpetas presentadas en la figura 4-1 correspondientes a las 11 estaciones meteorológicas sobre las cuales se van a comparar las técnicas y métodos de imputación de datos faltantes se encuentra la información de los años en los cuales se obtuvo información en las estaciones meteorológicas, en la figura 4-2 se puede ver que las estaciones presentan información para el período de tiempo comprendido entre el año 2014 y el año 2022, notándose que para el año 2022 no se presenta información para todo el año, esto se debe a que en el momento que se han entregado los datos el año no se ha terminado aún y es por ello que los datos para los meses finales del año no han sido leídos aún por las estaciones meteorológicas.

- ESPOCH-2014-L2-ANUAL 17/08/2022 09:58 Carpeta de archivos
- ESPOCH-2015-L2-ANUAL 17/08/2022 09:58 Carpeta de archivos
- ESPOCH-2016-L2-ANUAL 17/08/2022 09:58 Carpeta de archivos
- ESPOCH-2017-L2-ANUAL 17/08/2022 09:58 Carpeta de archivos
- ESPOCH-2018-L2-ANUAL 17/08/2022 09:58 Carpeta de archivos
- ESPOCH-2019-L2-ANUAL 17/08/2022 09:58 Carpeta de archivos
- ESPOCH-2020-L2-ANUAL 17/08/2022 09:58 Carpeta de archivos
- ESPOCH-2021-L2-ANUAL 17/08/2022 09:58 Carpeta de archivos
- ESPOCH-2022-L2-ANUAL 04/10/2022 17:22 Carpeta de archivos

Figura 4-2: Años en los que se presentan datos en las estaciones meteorológicas.

Realizado por: Altamirano C. y Carrillo P., 2023

Cada carpeta almacena información de todo un año, además se ha notado que para cada estación se han medido 51 variables, como se muestra en la figura 4-3.

date	time	status	Stat_TA_1h	Stat_TA_1h	Stat_TA_1h	Stat_RH_1h	Stat_RH_1h	Stat_RH_1h	Stat_RH_1h	Stat_PA_1h	Stat_PA_1h	Stat_PA_1h	Stat_PA_1h	Stat_PA_1h	Stat_SR_Dif_1h	Stat_SR_Dif_1h	Stat_SR_Dif_1h									
			Max	status	Min	status	Max	status	Min	status	Max	status	Min	status	Max	status	Min									
01/01/2014	12:00:00 a. m.	VALID	14.89	VALID	15.083	VALID	14.664	VALID	99.543	VALID	99.623	VALID	99.731	VALID	794.746	VALID	795.038	VALID	794.515	VALID	0.775	VALID	4.498	VALID	0	VALID
01/01/2014	5:00:00 p. m.	VALID	14.766	VALID	15.463	VALID	14.22	VALID	99.796	VALID	99.815	VALID	99.896	VALID	796.787	VALID	797.051	VALID	796.499	VALID	281.98	VALID	482.589	VALID	178.129	VALID
01/01/2014	6:00:00 p. m.	VALID	14.899	VALID	15.936	VALID	14.359	VALID	99.794	VALID	99.821	VALID	99.846	VALID	796.143	VALID	796.505	VALID	795.846	VALID	253.119	VALID	477.04	VALID	83.776	VALID
01/01/2014	7:00:00 p. m.	VALID	14.776	VALID	15.121	VALID	14.41	VALID	99.8	VALID	99.813	VALID	99.784	VALID	795.502	VALID	795.859	VALID	795.138	VALID	139.151	VALID	293.489	VALID	89.892	VALID
01/01/2014	8:00:00 p. m.	VALID	14.768	VALID	15.185	VALID	14.483	VALID	99.814	VALID	99.814	VALID	99.786	VALID	794.794	VALID	795.147	VALID	794.447	VALID	94.796	VALID	137.616	VALID	67.763	VALID
01/01/2014	9:00:00 p. m.	VALID	14.453	VALID	14.833	VALID	14.093	VALID	99.796	VALID	99.827	VALID	99.718	VALID	794.317	VALID	794.441	VALID	794.227	VALID	59.218	VALID	101.947	VALID	32.294	VALID
01/08/2014	3:00:00 p. m.	VALID	15.526	VALID	15.878	VALID	15.08	VALID	99.771	VALID	99.826	VALID	99.572	VALID	796.549	VALID	796.709	VALID	796.392	VALID	408.457	VALID	500.347	VALID	304.477	VALID
01/08/2014	6:00:00 p. m.	VALID	15.626	VALID	16.385	VALID	15.096	VALID	99.81	VALID	99.818	VALID	99.138	VALID	796.044	VALID	796.42	VALID	795.761	VALID	388.43	VALID	609.696	VALID	34.46	VALID
01/08/2014	7:00:00 p. m.	VALID	15.336	VALID	15.754	VALID	14.88	VALID	99.757	VALID	99.828	VALID	99.284	VALID	795.473	VALID	795.783	VALID	795.154	VALID	141.088	VALID	281.028	VALID	53.717	VALID
01/08/2014	8:00:00 p. m.	VALID	15.382	VALID	16.212	VALID	13.137	VALID	99.8	VALID	99.81	VALID	99.018	VALID	794.847	VALID	795.171	VALID	794.549	VALID	217.661	VALID	480.867	VALID	60.094	VALID
01/08/2014	9:00:00 p. m.	VALID	15.015	VALID	15.558	VALID	14.456	VALID	99.785	VALID	99.822	VALID	99.609	VALID	794.509	VALID	794.589	VALID	794.396	VALID	85.743	VALID	139.131	VALID	60.153	VALID
01/08/2014	10:00:00 p. m.	VALID	14.908	VALID	15.224	VALID	14.507	VALID	99.779	VALID	99.82	VALID	99.448	VALID	794.44	VALID	794.589	VALID	794.31	VALID	78.793	VALID	139.159	VALID	26.988	VALID
01/09/2014	3:00:00 p. m.	VALID	14.909	VALID	15.64	VALID	14.335	VALID	99.808	VALID	99.828	VALID	97.751	VALID	797.321	VALID	797.238	VALID	797.278	VALID	223.66	VALID	265.661	VALID	373.2	VALID
01/09/2014	4:00:00 p. m.	VALID	15.783	VALID	16.71	VALID	15.088	VALID	98.027	VALID	99.403	VALID	95.957	VALID	797.104	VALID	797.343	VALID	796.812	VALID	312.944	VALID	507.609	VALID	160.415	VALID
01/09/2014	5:00:00 p. m.	VALID	16.397	VALID	16.905	VALID	15.956	VALID	97.211	VALID	99.245	VALID	94.411	VALID	796.458	VALID	796.785	VALID	796.212	VALID	317.761	VALID	449.504	VALID	203.262	VALID
01/09/2014	6:00:00 p. m.	VALID	16.46	VALID	17.009	VALID	16.032	VALID	98.21	VALID	99.41	VALID	95.604	VALID	795.671	VALID	796.08	VALID	795.566	VALID	201.228	VALID	378.696	VALID	78.91	VALID
01/09/2014	7:00:00 p. m.	VALID	16.148	VALID	16.809	VALID	15.507	VALID	98.544	VALID	99.529	VALID	96.298	VALID	794.955	VALID	795.349	VALID	794.957	VALID	206.292	VALID	297.771	VALID	126.63	VALID
01/09/2014	8:00:00 p. m.	VALID	15.277	VALID	15.906	VALID	14.863	VALID	99.38	VALID	99.818	VALID	98.032	VALID	794.284	VALID	794.539	VALID	794.075	VALID	155.517	VALID	278.635	VALID	30.495	VALID
01/09/2014	9:00:00 p. m.	VALID	15.629	VALID	16.155	VALID	15.151	VALID	99.594	VALID	99.594	VALID	99.076	VALID	793.74	VALID	794.063	VALID	793.714	VALID	138.08	VALID	256.204	VALID	69.294	VALID
01/09/2014	10:00:00 p. m.	VALID	15.264	VALID	15.955	VALID	15.051	VALID	99.672	VALID	99.818	VALID	99.485	VALID	793.861	VALID	793.965	VALID	793.804	VALID	87.461	VALID	168.592	VALID	34.967	VALID
01/09/2014	11:00:00 p. m.	VALID	15.344	VALID	15.955	VALID	14.968	VALID	99.768	VALID	99.816	VALID	95.515	VALID	794.088	VALID	794.268	VALID	793.965	VALID	75.341	VALID	63.807	VALID	13.836	VALID
01/10/2014	2:00:00 p. m.	VALID	15.385	VALID	15.671	VALID	14.873	VALID	99.808	VALID	99.824	VALID	99.694	VALID	795.821	VALID	795.821	VALID	795.795	VALID	197.053	VALID	229.182	VALID	166.011	VALID
01/10/2014	3:00:00 p. m.	VALID	15.379	VALID	15.939	VALID	14.823	VALID	99.787	VALID	99.821	VALID	99.496	VALID	795.967	VALID	796.055	VALID	795.811	VALID	177.433	VALID	223.536	VALID	123.429	VALID
01/10/2014	4:00:00 p. m.	VALID	15.769	VALID	16.184	VALID	15.205	VALID	99.792	VALID	99.835	VALID	99.465	VALID	795.861	VALID	796.016	VALID	795.723	VALID	199.601	VALID	263.207	VALID	171.216	VALID
01/10/2014	5:00:00 p. m.	VALID	15.834	VALID	16.427	VALID	15.321	VALID	99.791	VALID	99.831	VALID	99.494	VALID	795.399	VALID	795.723	VALID	795.038	VALID	199.261	VALID	259.555	VALID	147.048	VALID
01/10/2014	7:00:00 p. m.	VALID	15.671	VALID	15.997	VALID	15.366	VALID	99.8	VALID	99.821	VALID	99.514	VALID	794.571	VALID	795.96	VALID	795.126	VALID	175.126	VALID	235.105	VALID	151.788	VALID
01/10/2014	8:00:00 p. m.	VALID	15.8	VALID	16.128	VALID	15.432	VALID	99.812	VALID	99.834	VALID	99.591	VALID	795.722	VALID	795.964	VALID	795.493	VALID	175.488	VALID	281.101	VALID	114.899	VALID
01/10/2014	9:00:00 p. m.	VALID	15.623	VALID	15.983	VALID	15.262	VALID	99.815	VALID	99.851	VALID	99.686	VALID	795.544	VALID	795.499	VALID	795.213	VALID	154.549	VALID	279.852	VALID	116.979	VALID
01/10/2014	10:00:00 p. m.	VALID	15.332	VALID	15.933	VALID	15.023	VALID	99.811	VALID	99.828	VALID	99.721	VALID	793.308	VALID	793.536	VALID	793.185	VALID	80.325	VALID	127.463	VALID	60.779	VALID
01/10/2014	11:00:00 p. m.	VALID	15.059	VALID	15.326	VALID	14.862	VALID	99.808	VALID	99.822	VALID	99.793	VALID	793.691	VALID	793.828	VALID	793.527	VALID	34.943	VALID	69.734	VALID	16.229	VALID
01/11/2014	2:00:00 p. m.	VALID	15.268	VALID	16.122	VALID	14.577	VALID	98.747	VALID	99.851	VALID	93.526	VALID	797.113	VALID	797.185	VALID	797.618	VALID	207.502	VALID	305.417	VALID	145.734	VALID
01/11/2014	3:00:00 p. m.	VALID	15.832	VALID	17.095	VALID	15.113	VALID	99.257	VALID	99.837	VALID	98.532	VALID	797.113	VALID	797.204	VALID	797.051	VALID	168.662	VALID	472.815	VALID	62.13	VALID
01/11/2014	4:00:00 p. m.	VALID	16.251	VALID	16.866	VALID	15.861	VALID	99.567	VALID	99.844	VALID	98.935	VALID	796.949	VALID	797.045	VALID	796.483	VALID	233.85	VALID	478.488	VALID	87.863	VALID
01/11/2014	5:00:00 p. m.	VALID	16.644	VALID	17.157	VALID	16.021	VALID	99.195	VALID	99.849	VALID	97.341	VALID	796.997	VALID	796.854	VALID	796.875	VALID	325.914	VALID	520.002	VALID	176.611	VALID
01/11/2014	6:00:00 p. m.	VALID	16.291	VALID	16.844	VALID	15.973	VALID	99.5	VALID	99.809	VALID	99.033	VALID	796.036	VALID	796.118	VALID	795.84	VALID	120.002	VALID	187.189	VALID	84.496	VALID
01/11/2014	7:00:00 p. m.	VALID	16.154	VALID	16.603	VALID	15.6	VALID	99.841	VALID	99.841	VALID	99.232	VALID	795.704	VALID	795.916	VALID	795.12	VALID	143.552	VALID	342.897	VALID	164.89	VALID
01/11/2014	8:00:00 p. m.	VALID	15.978	VALID	16.455	VALID	15.588	VALID	99.656	VALID	99.839	VALID	99.182	VALID	795.041	VALID	795.536	VALID	794.808	VALID	154.478	VALID	237.673	VALID	70.844	VALID

Figura 4-3: Variables presentadas en cada estación meteorológica.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-3 se puede ver que los datos han sido leídos con un tiempo de separación de una hora entre cada lectura de datos, tomándose datos para las 24 horas del día, pero la información presentada de esta forma no es de gran ayuda para el objetivo del presente proyecto de investigación, por lo tanto se ha visto la necesidad de manipular las bases de datos que se agrupan en una sola pestaña de Excel todos los años para los que se cuentan con datos (2014 - 2022) y además filtrar solamente las variables que son de interés para el estudio, las cuales son: Temperatura Ambiente, Humedad Relativa y Precipitación, obteniéndose los resultados que se presentan en la figura 4-4, donde se presenta el status (Estado) del dato leído por la estación meteorológica los cuales se presentan como valid (Válido) que son datos que no presentan problemas por parte de la estación y además los datos con status invalid (Inválido) los cuales son datos que no sirven para el estudio y

han sido eliminados de las bases de datos, fuera del status de los datos se nota que se presentan espacios en blanco lo cual representa que en esos intervalos de tiempo las estaciones no recolectaron datos generándose así los datos faltantes que se buscan rellenar de la forma más acertada posible.

		2014		2015		2016		2017		2018		2019		2020		2021		2022	
date	time	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA
01/01/-	00:00:00	VALID	14.859	VALID	15.883	VALID	16.26	VALID	14.484	VALID	14.186	VALID	14.8	VALID	14.205	VALID	13.417	VALID	14.381
01/01/-	01:00:00			VALID	15.062	VALID	15.032	VALID	14.885	VALID	14.006	VALID	14.997	VALID	14.92	VALID	13.41	VALID	14.323
01/01/-	02:00:00			VALID	15.417	VALID	15.953	VALID	14.886	VALID	13.653	VALID	13.671	VALID	14.075	VALID	12.929	VALID	14.389
01/01/-	03:00:00			VALID	15.409	VALID	16.002	VALID	14.439	VALID	13.83	VALID	13.455	VALID	14.325	VALID	12.704	VALID	14.353
01/01/-	04:00:00			VALID	15.102	VALID	15.851	VALID	14.545	VALID	13.78	VALID	13.817	VALID	14.222	VALID	12.843	VALID	14.395
01/01/-	05:00:00			VALID	14.763	VALID	14.996	VALID	14.524	VALID	13.674	VALID	14.026	VALID	14.163	VALID	12.965	VALID	14.113
01/01/-	06:00:00			VALID	14.619	VALID	15.248	VALID	14.479	VALID	13.597	VALID	13.624	VALID	14.075	VALID	12.601	VALID	13.884
01/01/-	07:00:00			VALID	14.663	VALID	15.001	VALID	14.315	VALID	13.505	VALID	13.49	VALID	14.006	VALID	12.645	VALID	13.812
01/01/-	08:00:00			VALID	14.743	VALID	15.192	VALID	14.098	VALID	13.333	VALID	13.456	VALID	14.101	VALID	12.404	VALID	13.564
01/01/-	09:00:00			VALID	14.722	VALID	15.143	VALID	13.597	VALID	13.332	VALID	13.195	VALID	13.984	VALID	12.544	VALID	13.521
01/01/-	10:00:00			VALID	14.587	VALID	14.948	VALID	13.314	VALID	13.285	VALID	12.998	VALID	13.901	VALID	12.996	VALID	12.993
01/01/-	11:00:00			VALID	14.545	VALID	15.198	VALID	13	VALID	13.155	VALID	12.837	VALID	13.854	VALID	13.216	VALID	13.269
01/01/-	12:00:00			VALID	14.739	VALID	14.875	VALID	13.354	VALID	13.341	VALID	13.533	VALID	14.086	VALID	13.847	VALID	13.548
01/01/-	13:00:00			VALID	14.502	VALID	14.738	VALID	14.119	VALID	13.587	VALID	13.098	VALID	14.899	VALID	14.691	VALID	13.827
01/01/-	14:00:00			VALID	15.135	VALID	15.481	VALID	14.391	VALID	14.834	VALID	14.051	VALID	15.789	VALID	15.579	VALID	15.428
01/01/-	15:00:00			VALID	15.303	VALID	20.172	VALID	14.831	VALID	15.847	VALID	15.835	VALID	16.074	VALID	16.388	VALID	17.236
01/01/-	16:00:00			VALID	15.806	VALID	18.145	VALID	15.214	VALID	15.38	VALID	15.59	VALID	16.953	VALID	16.652	VALID	15.721
01/01/-	17:00:00			VALID	15.869	VALID	18.611	VALID	15.834	VALID	16.64	VALID	15.75	VALID	17.24	VALID	15.737	VALID	15.657
01/01/-	18:00:00			VALID	15.98	VALID	17.967	VALID	16.668	VALID	16.735	VALID	15.771	VALID	17.979	VALID	15.228	VALID	16.827
01/01/-	19:00:00			VALID	15.93	VALID	17.983	VALID	16.611	VALID	16.27	VALID	16.061	VALID	17.41	VALID	14.639	VALID	15.692
01/01/-	20:00:00			VALID	15.999	VALID	17.322	VALID	16.115	VALID	16.106	VALID	15.86	VALID	15.968	VALID	14.564	VALID	14.892
01/01/-	21:00:00			VALID	15.713	VALID	16.629	VALID	16.064	VALID	15.702	VALID	15.754	VALID	15.533	VALID	14.228	VALID	14.813
01/01/-	22:00:00			VALID	15.477	VALID	16.508	VALID	15.484	VALID	15.76	VALID	15.743	VALID	15.464	VALID	14.175	VALID	14.528
01/01/-	23:00:00			VALID	15.42	VALID	16.31	VALID	15.327	VALID	16.008	VALID	15.57	VALID	15.055	VALID	14.019	VALID	14.541
01/02/-	00:00:00			VALID	15.387	VALID	16.201	VALID	14.832	VALID	14.919	VALID	15.276	VALID	14.64	VALID	13.846	VALID	14.361
01/02/-	01:00:00			VALID	15.162	VALID	16.262	VALID	14.521	VALID	14.512	VALID	15.026	VALID	14.338	VALID	13.65	VALID	14.426
01/02/-	02:00:00			VALID	15.087	VALID	16.21	VALID	14.912	VALID	14.532	VALID	14.861	VALID	14.153	VALID	13.549	VALID	14.41
01/02/-	03:00:00			VALID	15.086	VALID	16.183	VALID	14.793	VALID	14.229	VALID	14.707	VALID	14.131	VALID	13.489	VALID	14.363
01/02/-	04:00:00			VALID	15.039	VALID	16.144	VALID	14.223	VALID	14.056	VALID	14.333	VALID	14.222	VALID	13.531	VALID	14.371
01/02/-	05:00:00			VALID	14.745	VALID	16.134	VALID	14.061	VALID	13.991	VALID	14.086	VALID	14.037	VALID	13.698	VALID	14.206
01/02/-	06:00:00			VALID	14.63	VALID	16.087	VALID	14.297	VALID	13.871	VALID	13.84	VALID	14.082	VALID	13.858	VALID	14.225
01/02/-	07:00:00			VALID	14.496	VALID	16.021	VALID	13.675	VALID	13.707	VALID	13.952	VALID	14.045	VALID	13.119	VALID	14.023
01/02/-	08:00:00			VALID	14.287	VALID	15.732	VALID	13.822	VALID	13.068	VALID	13.898	VALID	13.908	VALID	13.288	VALID	13.795
01/02/-	09:00:00			VALID	14.087	VALID	15.311	VALID	13.691	VALID	13.065	VALID	13.938	VALID	13.912	VALID	12.926	VALID	13.566
01/02/-	10:00:00			VALID	13.924	VALID	15.053	VALID	13.619	VALID	12.841	VALID	13.939	VALID	13.818	VALID	13.023	VALID	13.117
01/02/-	11:00:00			VALID	13.801	VALID	15.108	VALID	13.825	VALID	12.643	VALID	13.839	VALID	13.814	VALID	12.787	VALID	13.546
01/02/-	12:00:00			VALID	13.83	VALID	15.247	VALID	13.945	VALID	12.762	VALID	13.898	VALID	13.904	VALID	13.187	VALID	14.033
01/02/-	13:00:00			VALID	14.354	VALID	16.201	VALID	15.737	VALID	12.531	VALID	14.267	VALID	14.552	VALID	13.508	VALID	15.456
01/02/-	14:00:00			VALID	14.826	VALID	16.453	VALID	16.688	VALID	13.456	VALID	14.438	VALID	14.918	VALID	15.859	VALID	17.238
01/02/-	15:00:00			VALID	14.47	VALID	16.745	VALID	17	VALID	14.413	VALID	15.118	VALID	15.44	VALID	16.001	VALID	16.001
		Temperatura Ambiente		Humedad Relativa		Precipitación		Temperatura Ambiente		Humedad Relativa		Precipitación							

Figura 4-4: Estación meteorológica con las variables de interés.

Realizado por: Altamirano C. y Carrillo P., 2023

Previamente se ha detallado que el año 2022 no cuenta con datos para todos los meses, entonces al aplicar las técnicas o métodos de imputación de datos lo que se va a generar son predicciones para los meses que no existen registros presentados por las estaciones, por tanto, se ha optado por eliminar dicho año de las matrices de datos, con lo cual se ha logrado reducir la dimensionalidad de las bases de datos pasando de tener 9 años a tener 8 años de información en las variables Temperatura Ambiente y Humedad Relativa, mientras que para la variable Precipitación se ha pasado de 4 años a 3 años de información ya que la variable Precipitación se ha empezado a medir desde el año 2019 en todas las estaciones meteorológicas, generándose así nuevas bases de datos cuyos resultados se presentan en la figura 4-5.

1			2014	2015	2016	2017	2018	2019	2020	2021
2	date	time	T_A_MEDIA	T_A_MEDIA	T_A_MEDIA	T_A_MEDIA	T_A_MEDIA	T_A_MEDIA	T_A_MEDIA	T_A_MEDIA
3	01/01/-	0:00:00		14,973	16,854	16,869		14,107	13,498	14,358
4	01/01/-	1:00:00		14,25	15,982	14,898		12,918	13,219	13,586
5	01/01/-	2:00:00		13,317	14,269	13,559		11,815	13,684	12,953
6	01/01/-	3:00:00		13,175	13,275	12,979		10,8	13,286	12,862
7	01/01/-	4:00:00		12,597	12,436	12,772		10,365	13,284	12,237
8	01/01/-	5:00:00		12,177	12,346	12,091		8,381	13,138	11,845
9	01/01/-	6:00:00		11,722	11,498	11,973		7,295	13,264	11,555
10	01/01/-	7:00:00		11,509	11,348	11,919		7,202	12,48	11,096
11	01/01/-	8:00:00		11,052	11,04	11,63		5,948	12,092	11,313
12	01/01/-	9:00:00		10,835	11,4	11,135		5,546	11,981	10,935
13	01/01/-	10:00:00		10,652	11,616	11,365		5,195	11,933	10,538
14	01/01/-	11:00:00		10,895	11,364	11,131		4,639	11,996	9,577

Figura 4-5: Bases de datos modificadas en los períodos de tiempo establecidos.

Realizado por: Altamirano C. y Carrillo P., 2023

4.1.2. Estadísticos Descriptivos

Dentro del análisis de las estaciones sobre las cuales se va a efectuar el relleno de datos es importante saber a que nos estamos enfrentando, es por eso que se han determinado los resultados presentados en la figura 4-6 donde se muestran las estaciones meteorológicas, cada una de las variables de estudio (Temperatura Ambiente, Humedad Relativa y Precipitación) además de la cantidad de datos estimados, datos faltantes y los datos existentes.

ESTACIONES METEOROLÓGICAS	TEMPERATURA AMBIENTE			HUMEDAD RELATIVA			PRECIPITACIÓN		
	DATOS ESTIMADOS	DATOS FALTANTES	DATOS EXISTENTES	DATOS ESTIMADOS	DATOS FALTANTES	DATOS EXISTENTES	DATOS ESTIMADOS	DATOS FALTANTES	DATOS EXISTENTES
ALAO	70080	507	69573	70080	1021	68545	26280	3671	22609
ATILLO	70080	8929	61151	70080	15123	54957	26280	2680	23600
CUMANDA	70080	19736	50344	70080	22733	47347	26280	11484	14796
ESPOCH	70080	19942	49995	70080	20730	48419	26280	213	26067
MATUS	70080	19555	50525	70080	19448	50632	26280	2398	23882
MULTITUD	70080	8056	62024	70080	12084	54724	26280	7514	18766
QUIMIAG	70080	7635	62253	70080	7635	62253	26280	504	25776
SAN JUAN	70080	229	69851	70080	229	69851	26280	72	26208
TIXAN	70080	1904	67613	70080	1904	67613	26280	2741	23539
TUNSHI	70080	8563	61517	70080	8563	61517	26280	3478	22802
URBINA	70080	7660	62420	70080	7660	62444	26280	3460	22820

Figura 4-6: Datos generales de todas las estaciones meteorológicas

Realizado por: Altamirano C. y Carrillo P., 2023

Una vez determinados los datos que se estima existan en cada variable, cuantos datos faltantes existen por variable y con cuantos datos se cuenta para poder trabajar para la comparativa de las técnicas y métodos de imputación, se puede determinar el total de datos estimados, el total de datos faltantes y el porcentaje de datos faltantes para cada estación considerando las tres variables en estudio, estos resultados se presentan en la figura 4-7.

ESTACIONES METEOROLÓGICAS	TOTAL DATOS ESTIMADOS EN TODAS LAS VARIABLES	TOTAL DATOS FALTANTES EN TODAS LAS VARIABLES	% DE DATOS FALTANTES EN TODA LA ESTACIÓN
ALAO	166440	5199	3.12
ATILLO	166440	26732	16.06
CUMANDA	166440	53953	32.42
ESPOCH	166440	40885	24.56
MATUS	166440	41401	24.87
MULTITUD	166440	27654	16.61
QUIMIAG	166440	15774	9.48
SAN JUAN	166440	530	0.32
TIXAN	166440	6549	3.93
TUNSHI	166440	20604	12.38
URBINA	166440	18780	11.28

Figura 4-7: Datos estimados y faltantes de las estaciones meteorológicas.

Realizado por: Altamirano C. y Carrillo P., 2023

De los resultados obtenidos en la figura 4-7 se puede evidenciar que la estación que presenta mayor cantidad de datos faltantes es la estación de CUMANDÁ ya que en total para las 3 variables en estudio (Temperatura Ambiente, Humedad Relativa y Precipitación) se estima la existencia de 166440 datos y en dicha estación se presentan 56953 datos faltantes lo que representa un total del 32.42 % de datos faltantes, por otra parte se puede ver que la estación que presenta menor cantidad de datos faltantes es la estación de SAN JUAN ya que en total para las 3 variables en estudio (Temperatura Ambiente, Humedad Relativa y Precipitación) se estima la existencia de 166440 datos y en dicha estación se presentan 530 datos faltantes lo que representa un total del 0.32 % de datos faltantes.

Estos resultados son importantes para comprender el panorama al que nos enfrentamos en la presente investigación, pero para poder comprender de mejor manera estos resultados se los va a desglosar de forma que se explique de forma anual la cantidad de datos estimados, la cantidad de datos faltantes y además el porcentaje de datos faltantes existente, recordando que para las variables Temperatura Ambiente y Humedad Relativa se cuenta con datos para el período de tiempo comprendido entre los años 2014 y 2021, a diferencia de la variable Precipitación la cual presenta datos para el período de tiempo comprendido entre el año 2019 y el año 2021.

Desglose anual de datos para la estación Cumandá.

ESTACIÓN CUMANDÁ									
AÑOS	TEMPERATURA AMBIENTE			HUMEDAD RELATIVA			PRECIPITACIÓN		
	DATOS ESTIMADOS	DATOS FALTANTES	% DATOS FALTANTES	DATOS ESTIMADOS	DATOS FALTANTES	% DATOS FALTANTES	DATOS ESTIMADOS	DATOS FALTANTES	% DATOS FALTANTES
2014	8760	152	1.74	8760	152	1.74			
2015	8760	69	0.79	8760	69	0.79			
2016	8760	3505	40.01	8760	3505	40.01			
2017	8760	2605	29.74	8760	2605	29.74			
2018	8760	1678	19.16	8760	1678	19.16			
2019	8760	2019	23.05	8760	3314	37.83	8760	2043	23.32
2020	8760	8469	96.68	8760	8581	97.96	8760	8469	96.68
2021	8760	948	10.82	8760	2650	30.25	8760	972	11.10

Figura 4-8: Datos desglosados para la estación Cumandá

Realizado por: Altamirano C. y Carrillo P., 2023

De la figura 4-8 se nota que para Temperatura Ambiente el año que presenta mayor cantidad de datos faltantes es el año 2020 ya que se presenta un total de 8469 datos faltantes lo cual representa el 96.68 % de los datos estimados, por otra parte, para Humedad Relativa el año que presenta mayor cantidad de datos faltantes sigue siendo el año 2020 ya que se presenta un total de 8581 datos faltantes lo cual representa el 97.96 % de los datos estimados, finalmente, para Precipitación el año que presenta mayor cantidad de datos faltantes sigue siendo el año 2020 ya que se presenta un total de 8469 datos faltantes lo cual representa el 96.68 % de los datos estimados.

El desglose anual de datos para la estación San Juan.

ESTACIÓN SAN JUAN									
AÑOS	TEMPERATURA AMBIENTE			HUMEDAD RELATIVA			PRECIPITACIÓN		
	DATOS ESTIMADOS	DATOS FALTANTES	% DATOS FALTANTES	DATOS ESTIMADOS	DATOS FALTANTES	% DATOS FALTANTES	DATOS ESTIMADOS	DATOS FALTANTES	% DATOS FALTANTES
2014	8760	24	0.27	8760	24	0.27			
2015	8760	0	0.00	8760	0	0.00			
2016	8760	24	0.27	8760	24	0.27			
2017	8760	0	0.00	8760	0	0.00			
2018	8760	157	1.79	8760	157	1.79			
2019	8760	0	0.00	8760	0	0.00	8760	24	0.27
2020	8760	24	0.27	8760	24	0.27	8760	24	0.27
2021	8760	0	0.00	8760	0	0.00	8760	24	0.27

Figura 4-9: Datos desglosados para la estación San Juan

Realizado por: Altamirano C. y Carrillo P., 2023

De la figura 4-9 se nota que para Temperatura Ambiente el año que presenta mayor cantidad de datos faltantes es el año 2018 ya que se presenta un total de 157 datos faltantes lo cual representa el 1.79 % de los datos estimados para este año, para Humedad Relativa el año que presenta mayor cantidad de datos faltantes sigue siendo el año 2018 ya que se presenta un total de 157 datos faltantes lo cual representa el 1.79 % de los datos estimados para este año, finalmente, para Precipitación el año que presenta mayor cantidad de datos faltantes los años 2019, 2020, 2021 presentan un total de 24 datos faltantes lo cual representa el 0.27 % de los datos estimados para dichos años.

4.2. Análisis e interpretación de resultados

Para poder saber si las técnicas y métodos de imputación de datos han funcionado de forma correcta se opta por efectuar gráficas temporales donde se analice los trazos en cada una de estas representaciones ya que se define que si se tienen trazos continuos los datos han sido imputados en su totalidad mientras que si existen segmentos de la gráfica sin trazo representa que las técnicas o métodos no han imputado los datos faltantes en su totalidad.

4.3. Estación Cumandá Imputación De Datos

4.3.1. Estación Cumandá Datos Faltantes

Estación Cumandá Variable Temperatura Ambiente Con Datos Faltantes

En la figura 4-10 se presenta una serie temporal, de los datos de la variable Temperatura Ambiente en los años 2014 hasta 2021 los cuales se presentan de forma horaria, donde se puede notar la discontinuidad en las series temporales obtenidas reflejándose la existencia de datos faltantes.

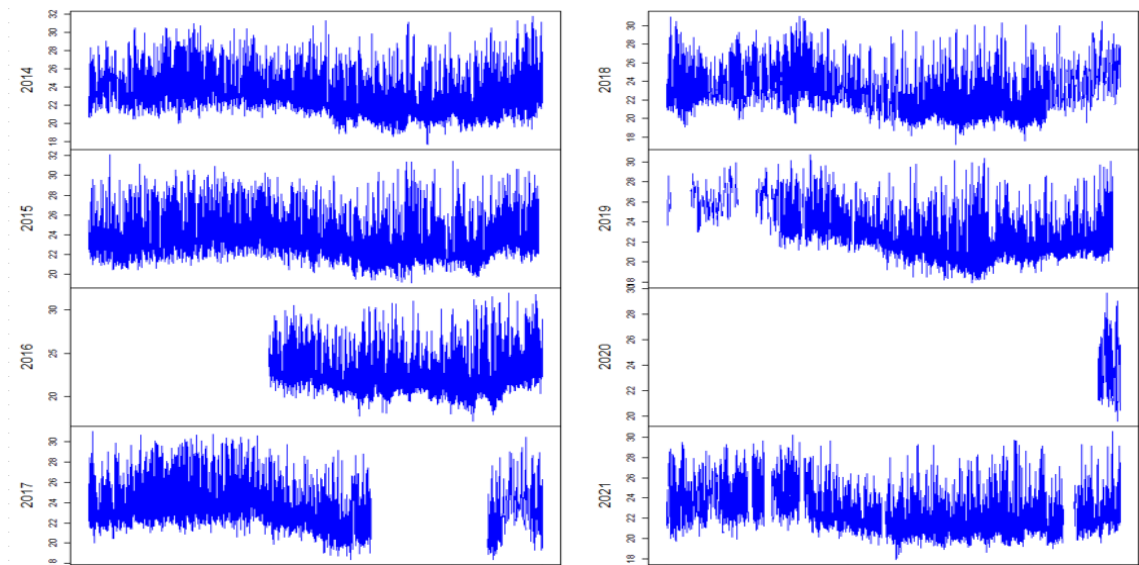


Figura 4-10: Estación Cumandá variable Temperatura Ambiente con datos faltantes.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación Cumandá Variable Humedad Relativa Con Datos Faltantes

En la figura 4-11 se presenta una serie temporal, de los datos de la variable Humedad Relativa en los años 2014 - 2021 los cuales se presentan de forma horaria, donde se puede notar la discontinuidad en los trazos de las series temporales obtenidas reflejándose así la existencia de datos faltantes.

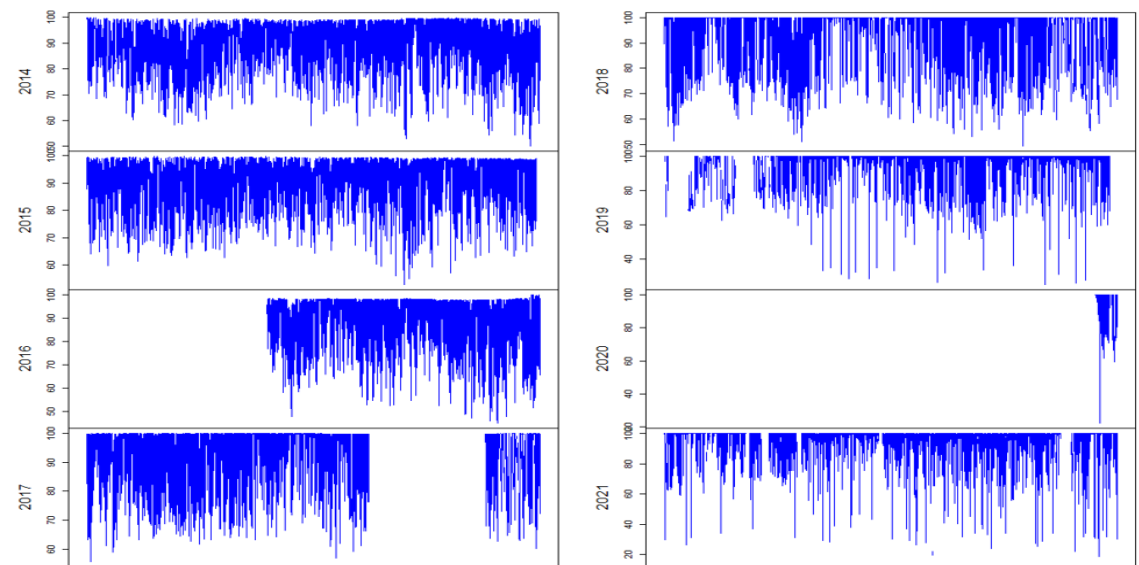


Figura 4-11: Estación Cumandá variable Humedad Relativa con datos faltantes.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación Cumandá Variable Precipitación Con Datos Faltantes

En la figura 4-12 se presenta una serie temporal, de los datos de la variable Humedad Relativa en los años 2019 - 2021 los cuales se presentan de forma horaria, donde se puede notar la discontinuidad en los trazos de las series temporales obtenidas reflejándose así la existencia de



Figura 4-12: Estación Cumandá variable Precipitación con datos faltantes.

Realizado por: Altamirano C. y Carrillo P. ,2023

4.3.2. Relleno de datos faltantes usando MissForest

Estación Cumandá Variable Temperatura Ambiente

En la figura 4-13 se presenta una gráfica con formato de serie temporal de los datos imputados por el método MissForest, se puede ver como las series temporales se han graficado de forma continua, es decir, no se presentan segmentos de gráfica sin trazar, por lo tanto, los datos han sido imputados en su totalidad.

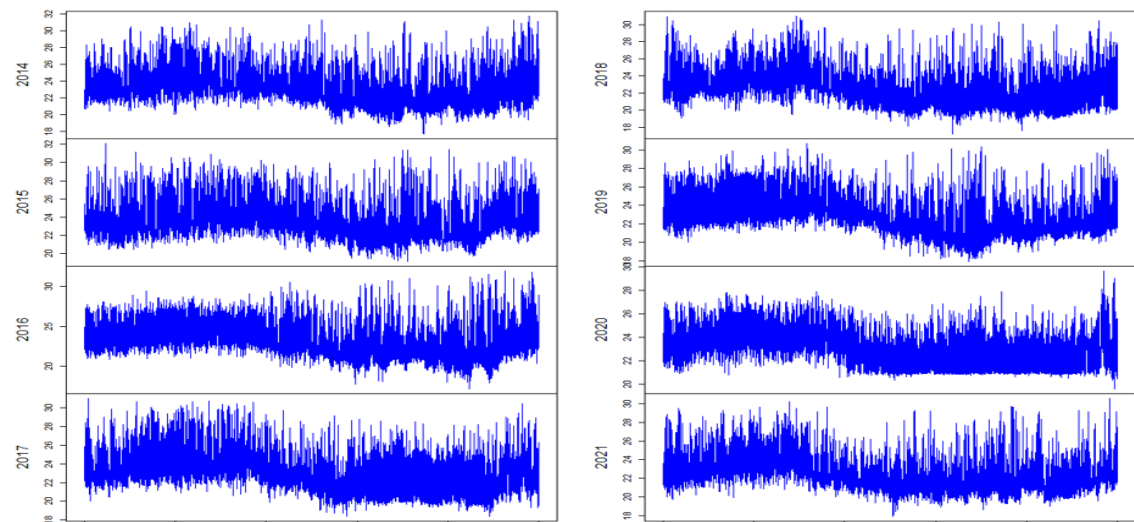


Figura 4-13: Estación Cumandá variable Temperatura Ambiente con datos imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación Cumandá Variable Humedad Relativa

En la figura 4-14 se presenta una gráfica con formato de serie temporal de los datos imputados por el método MissForest, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable a sido imputada en su totalidad.

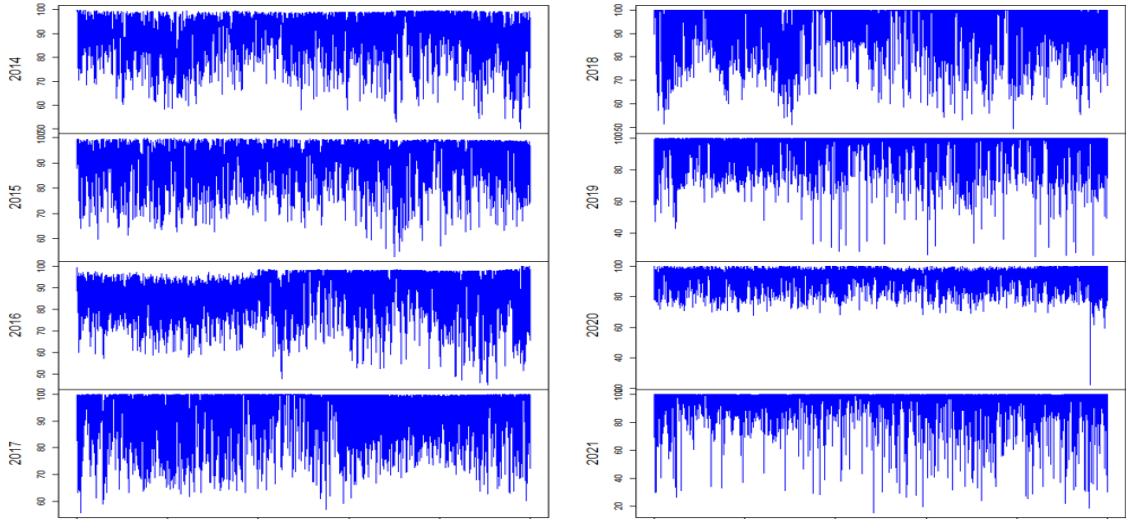


Figura 4-14: Estación Cumandá variable Humedad Relativa con datos imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación Cumandá Variable Precipitación

En la figura 4-15 se presenta una gráfica con formato de serie temporal de los datos imputados por el el método MissForest para la variable precipitación, donde se puede ver que no existe discontinuidad en los trazos de la serie temporal lo que representa que los datos han sido imputados en su totalidad.



Figura 4-15: Estación Cumandá variable Precipitación con datos imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

4.3.3. Relleno de datos mediante Medias

Estación Cumandá Variable Temperatura Ambiente

En la figura 4-16 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de las medias, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable a sido imputada en su totalidad.

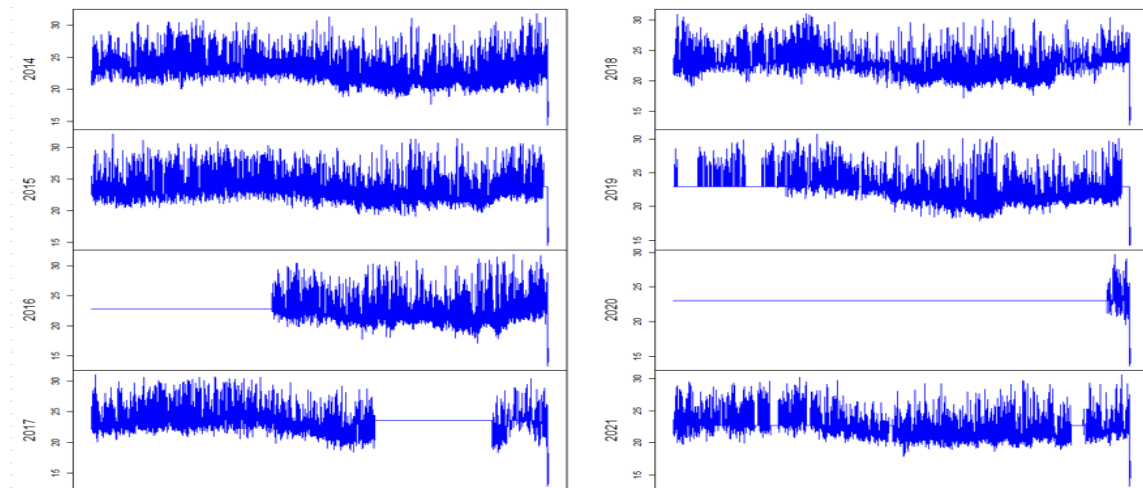


Figura 4-16: Estación Cumandá variable Temperatura Ambiente con datos imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación Cumandá Variable Humedad Relativa

En la figura 4-17 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de las medias, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable a sido imputada en su totalidad.

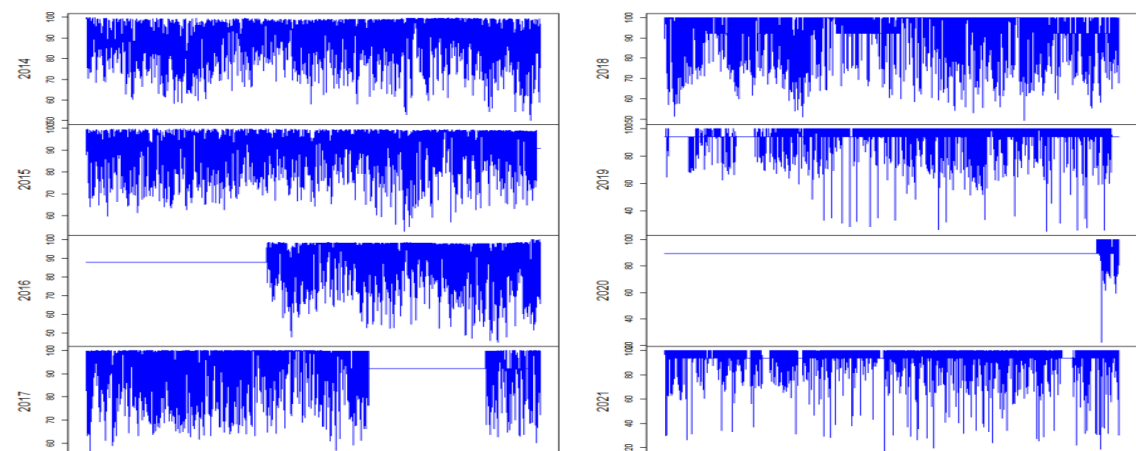


Figura 4-17: Estación Cumandá Variable Humedad Relativa Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación Cumandá Variable Precipitación

En la figura 4-18 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de las medias para la variable precipitación, donde se puede ver que no existen segmentos de gráfica en blanco dentro de la serie temporal lo que representa que los datos han sido imputados en su totalidad.

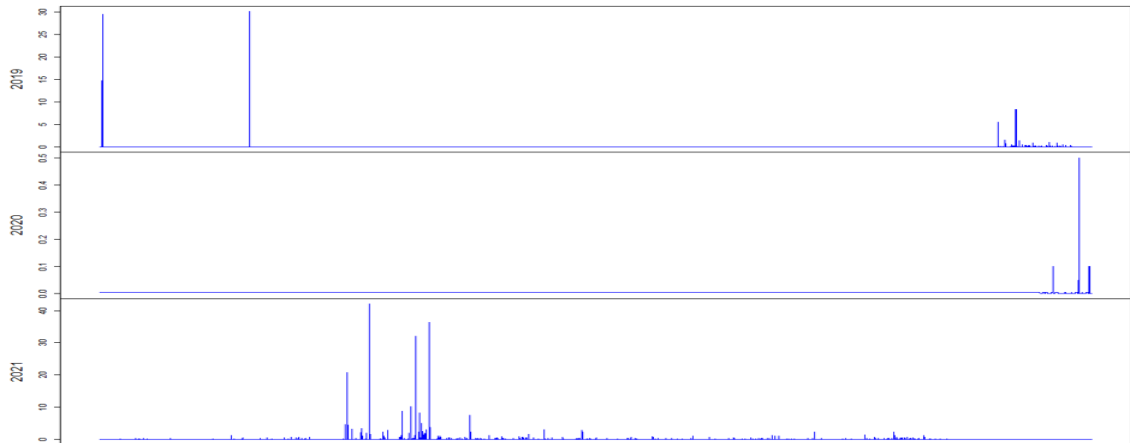


Figura 4-18: Estación Cumandá Variable Precipitación Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

4.3.4. Relleno De Datos Usando La Razón Q

Estación Cumandá Variable Temperatura Ambiente

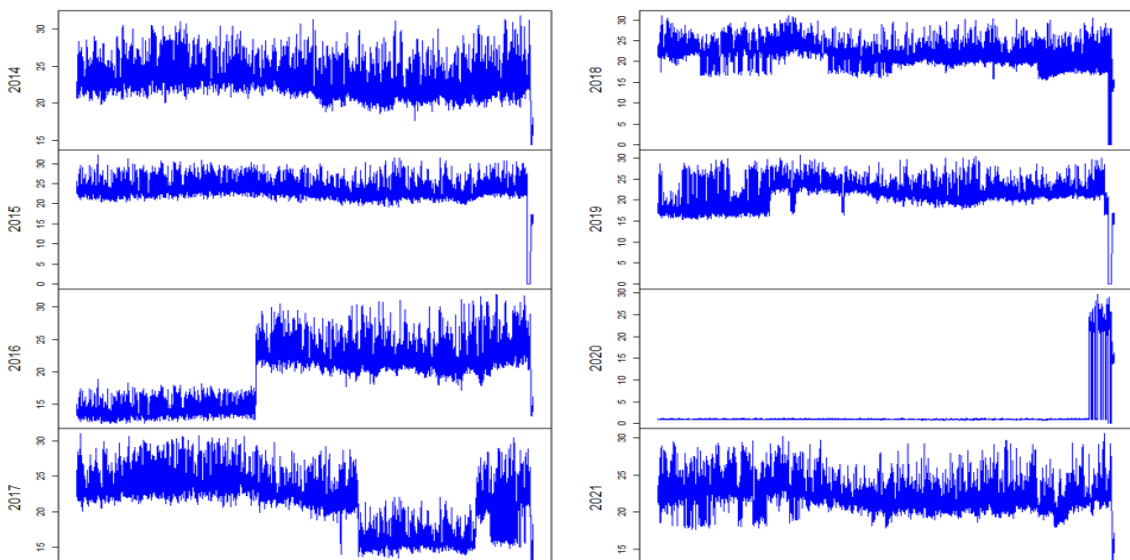


Figura 4-19: Estación Cumandá Variable Temperatura Ambiente Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-19 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de la razón Q, donde se puede ver que no existe discontinuidad en los trazos de la serie temporal lo que representa que los datos han sido imputados en su totalidad.

Estación Cumandá Variable Humedad Relativa

En la figura 4-20 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de la razón Q, se puede ver como las series temporales se han graficado de forma continua, es decir, no se presentan segmentos de gráfica sin trazar, por lo tanto, los datos han sido imputados.

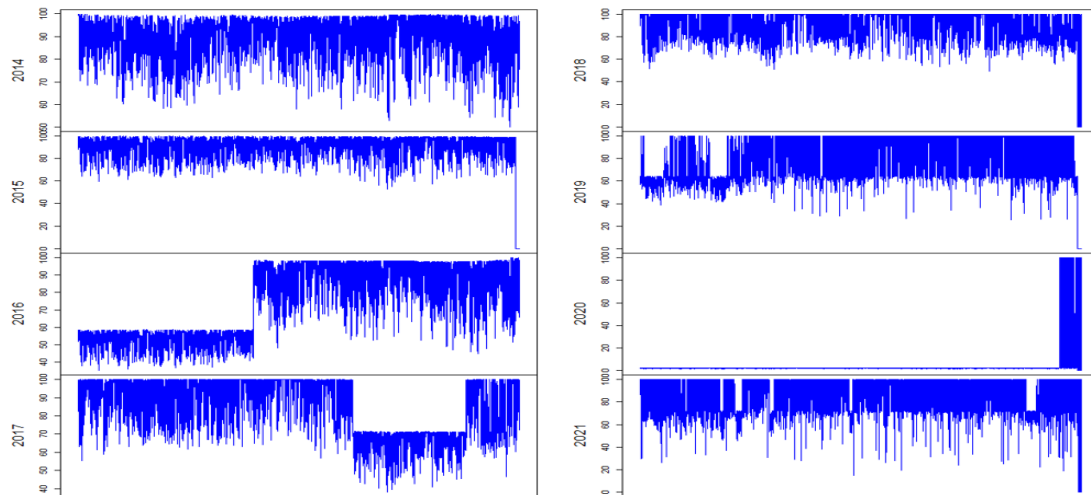


Figura 4-20: Estación Cumandá Variable Humedad Relativa Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

Estación Cumandá Variable Precipitación

En la figura 4-21 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de la razón Q, donde se puede notar que la gráfica es continua, por lo tanto se define que la variable a sido imputada en su totalidad.

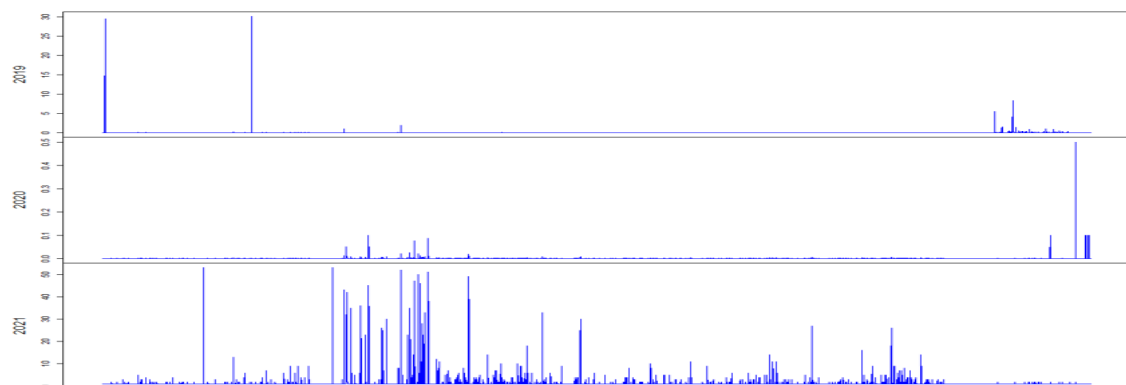


Figura 4-21: Estación Cumandá Variable Precipitación Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

4.3.5. Relleno De Datos Usando Regresión Lineal Simple

Estación Cumandá Variable Temperatura Ambiente

Año 2014		Año 2015		Año 2016		Año 2017	
a	b	a	b	a	b	a	b
0,6926	7,6536	-0,0001	24,323	0,8327	3,2823	0,7692	5,0738
Año 2018		Año 2019		Año 2020		Año 2021	
a	b	a	b	a	b	a	b
0,7314	5,4484	0,7529	4,9616	0,9357	0,2222	0,6507	7,1487

Figura 4-22: Coeficientes de a y b para cada uno de los años de la variable Temperatura Ambiente.

Realizado por: Altamirano C. y Carrillo P., 2023

Una vez que se han determinado los coeficientes que modelan la ecuación pertinente para cada uno de los años, figura 4-22 considerando el año 2015 como año base por ser el que mayor cantidad de datos presenta dentro de la variable Temperatura Ambiente para la estación en estudio, se han obtenido los siguientes modelos de regresión simple:

- **AÑO 2014:**

$$y = 0.6926x + 7.6536$$

- **AÑO 2016:**

$$y = 0.8327x + 3.2823$$

- **AÑO 2017:**

$$y = 0.7692x + 5.0738$$

- **AÑO 2018:**

$$y = 0.7314x + 5.4484$$

- **AÑO 2019:**

$$y = 0.7529x + 4.9616$$

- **AÑO 2020:**

$$y = 0.9357x + 0.2222$$

- **AÑO 2021:**

$$y = 0.6507x + 7.1487$$

Partiendo de los modelos de regresión lineal que definen cada uno de los años de estudio para la variable Temperatura Ambiente se ha obtenido una nueva base de datos imputada en su totalidad, y para verificar este hecho se ha optado por generar gráficas en formato de series temporales para analizar el comportamiento de los nuevos datos.

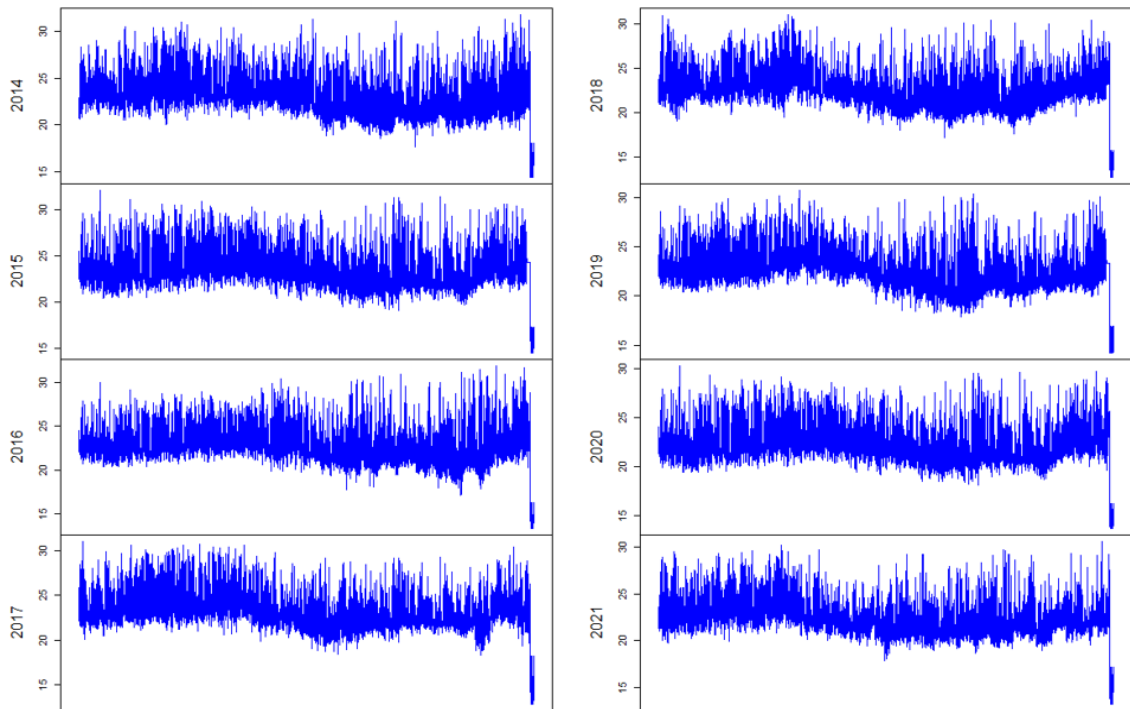


Figura 4-23: Estación Cumandá Variable Temperatura Ambiente Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-23 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de Regresión Lineal Simple, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable a sido imputada en su totalidad.

Estación Cumandá Variable Humedad Relativa

Año 2014		Año 2015		Año 2016		Año 2017	
a	b	a	b	a	b	a	b
0,5215	43,815	0,0002	89,685	0,7018	23,882	0,5924	38,595
Año 2018		Año 2019		Año 2020		Año 2021	
a	b	a	b	a	b	a	b
0,579	40,216	0,3826	59,584	0,4264	53,668	0,3356	63,919

Figura 4-24: Coeficientes de a y b para cada uno de los años de la variable Humedad Relativa.

Realizado por: Altamirano C. y Carrillo P. ,2023

Una vez que se han determinado los coeficientes que modelan la ecuación pertinente para cada uno de los años, figura 4-24 considerando el año 2015 como año base por ser el que mayor cantidad de datos presenta dentro de la variable Humedad Relativa para la estación en estudio, se han obtenido los siguientes modelos de regresión:

- AÑO 2014:

$$y = 0.5215x + 43.815$$

- AÑO 2016:

$$y = 0.7018x + 23.882$$

- AÑO 2017:

$$y = 0.5924x + 38.595$$

- AÑO 2018:

$$y = 0.5790x + 40.216$$

- AÑO 2019:

$$y = 0.3826x + 59.584$$

- AÑO 2020:

$$y = 0.4264x + 53.668$$

- AÑO 2021:

$$y = 0.3356x + 63.919$$

Partiendo de los modelos de regresión lineal que definen cada uno de los años de estudio para la variable Humedad Relativa se ha obtenido una nueva base de datos imputada en su totalidad, y para verificar este hecho se ha optado por generar gráficas en formato de series temporales para analizar el comportamiento de los nuevos datos.

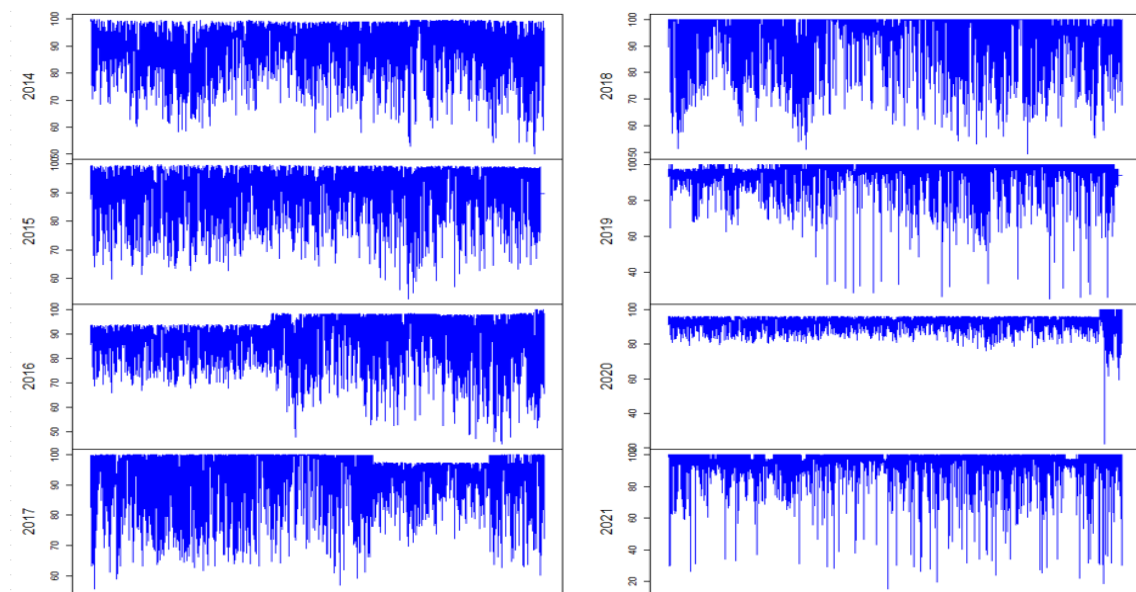


Figura 4-25: Estación Cumandá Variable Humedad Relativa Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-25 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de Regresión Lineal Simple, se puede ver como las series temporales se han graficado de forma continua, es decir, no se presentan segmentos de gráfica sin trazar, por lo tanto, los datos han sido imputados en su totalidad.

Estación Cumandá Variable Precipitación

Año 2019		Año 2020		Año 2021	
a	b	a	b	a	b
-0,0048	0,0814	-0,0117	0,0026	-0,00001	0,1306

Figura 4-26: Coeficientes de a y b para cada uno de los años de la variable Precipitación.

Realizado por: Altamirano C. y Carrillo P. ,2023

Una vez que se han determinado los coeficientes que modelan la ecuación pertinente para cada uno de los años mostrados en la figura 4-26 considerando el año 2021 como año base por ser el que mayor cantidad de datos presenta dentro de la variable Precipitación para la estación en estudio, se han obtenido los siguientes modelos de regresión:

- AÑO 2019:

$$y = -0.0048 x + 0.0814$$

- AÑO 2020:

$$y = -0.0117 x + 0.0026$$

Partiendo de los modelos de regresión lineal que definen los años de estudio para la variable Precipitación se ha obtenido una nueva base de datos imputada en su totalidad..

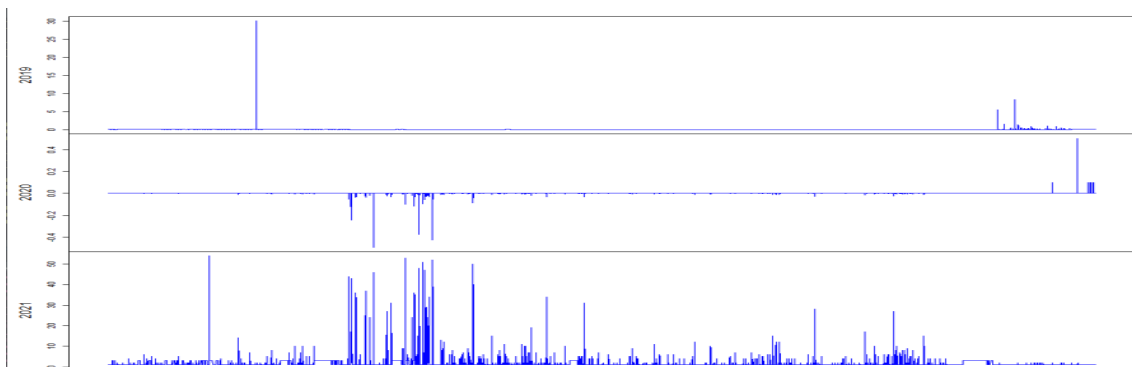


Figura 4-27: Estación Cumandá Variable Precipitación Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-27 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de Regresión Lineal Simple, donde se puede ver que no existe discontinuidad en los trazos de la serie temporal lo que representa que los datos han sido imputados en su totalidad.

4.3.6. *Relleno De Datos Mediante Imágenes Satelitales*

Para poder efectuar el relleno de datos faltantes empleando la técnica de imágenes satelitales ha sido de gran ayuda que el Grupo de Energías Alternativas y Ambiente (GEAA) facilite las bases de datos para las variables Temperatura Ambiente, Humedad Relativa y Precipitación de la estación Cumandá, con lo cual se ha realizado la comparativa entre la base de datos original y la base de datos satelital, para así poder imputar los datos faltantes en la base original para cada una de las variables en estudio.

Estación Cumandá Variable Temperatura Ambiente

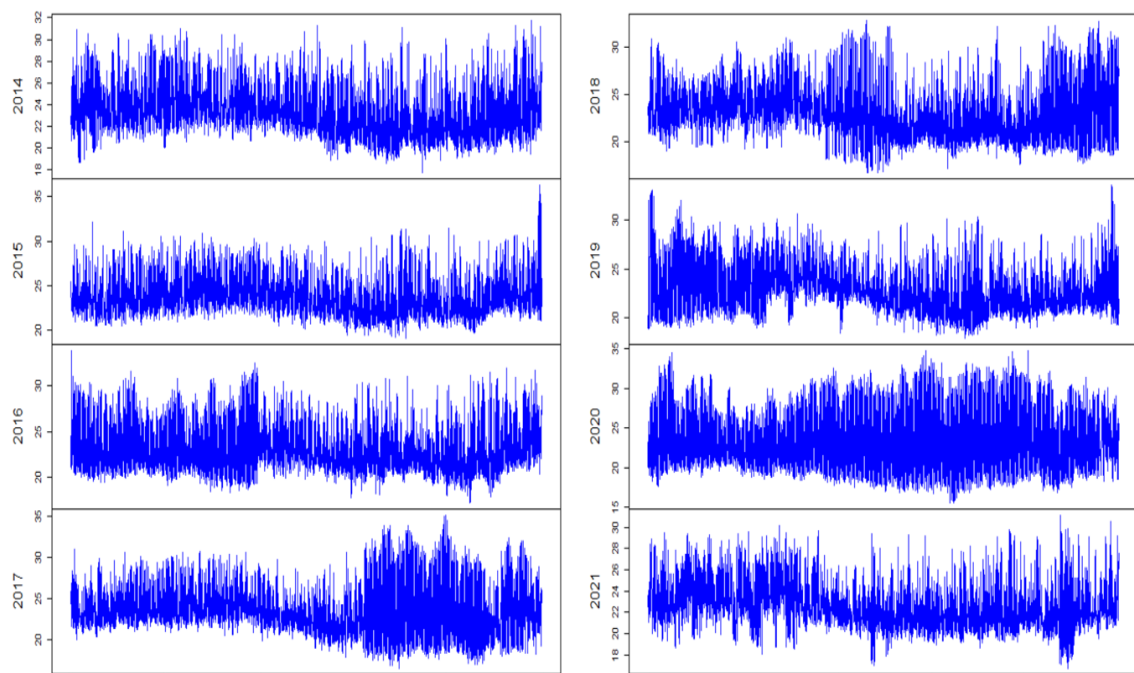


Figura 4-28: Estación Cumandá Variable Precipitación Con Datos Imputados

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-28 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de Imágenes Satelitales, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable ha sido imputada en su totalidad.

Estación Cumandá Variable Humedad Relativa

En la figura 4-29 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de las Imágenes Satelitales, donde se puede ver que no existe discontinuidad en los trazos de la serie temporal lo que indica que los datos han sido imputados en su totalidad.

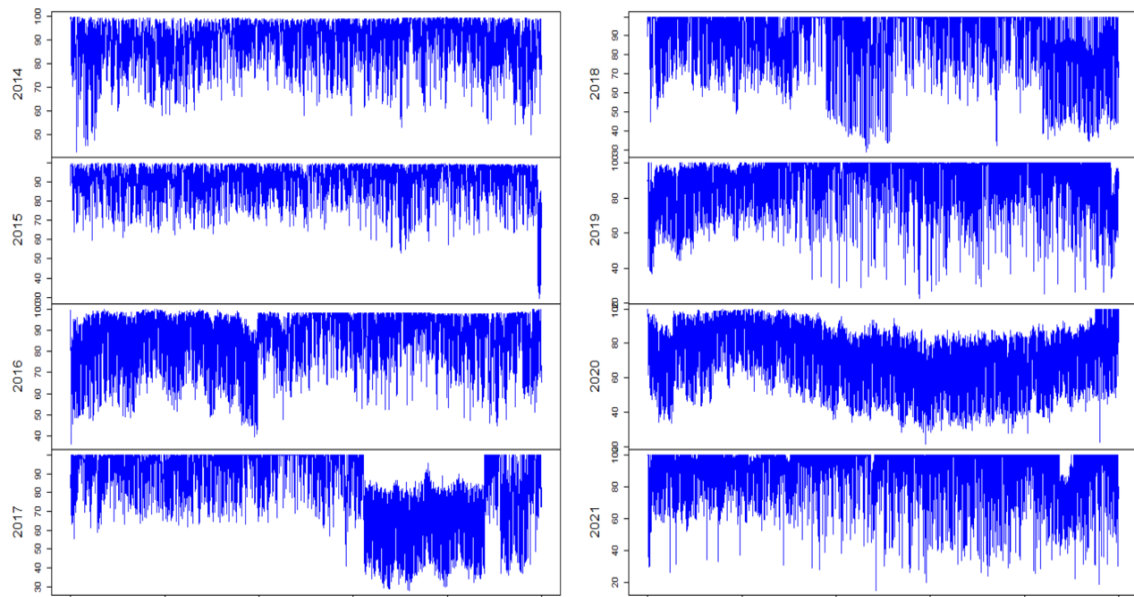


Figura 4-29: Estación Cumandá Variable Humedad Relativa Con Datos Imputados

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación Cumandá Variable Precipitación

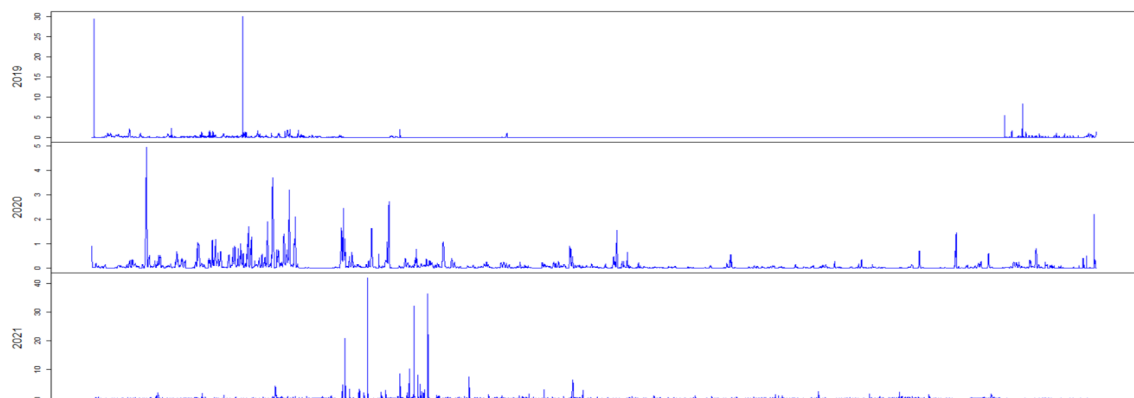


Figura 4-30: Estación Cumandá Variable Precipitación Con Datos Imputados

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-30 se puede notar como los datos efectivamente han sido imputados en su totalidad ya que en las series temporales obtenidas para la variable Precipitación se presentan trazos continuos, pero se puede ver que el comportamiento de las gráficas no lleva un patrón establecido, esto se debe a que los datos imputados no se asemejan a los datos imputados, es por eso que se ha efectuado un nuevo análisis con la base de datos proporcionada por el (GEAA) correspondiente a las imágenes satelitales y así ver como estos datos se comportan en una serie temporal la cual se presenta en la figura 4-31.

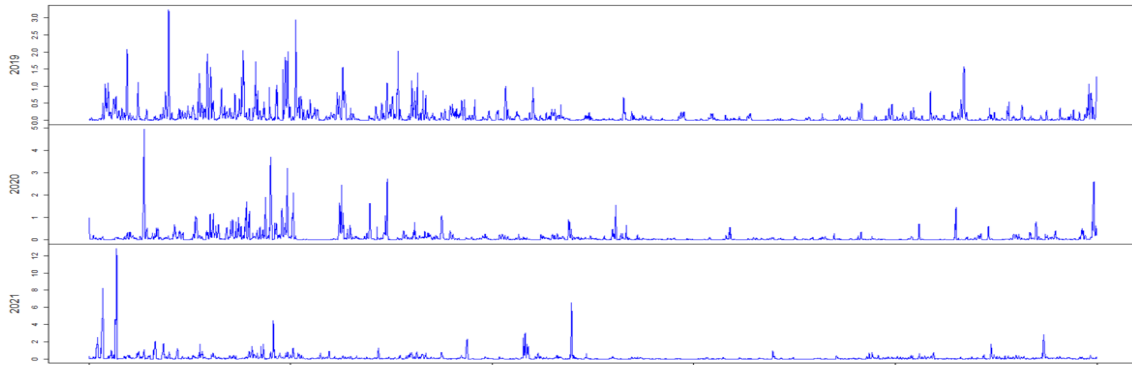


Figura 4-31: Estación Cumandá Variable Precipitación Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-31, se puede ver como el comportamiento de las series temporales para cada uno de los años en los que se ha estudiado la variable Precipitación en la estación Cumandá es mucho más acertada ya que se presentan patrones que no están entorno a cero.

4.4. Estación San Juan Imputación De Datos

4.4.1. Estación San Juan Datos Faltantes

Estación San Juan Variable Temperatura Ambiente Con Datos Faltantes

En la figura 4-32 se presenta una serie temporal de los datos de la variable Temperatura Ambiente en los años 2014 hasta el 2021 que se presentan de forma horaria, en la cual se nota la discontinuidad en los trazos lo cual refleja la existencia de datos faltantes.

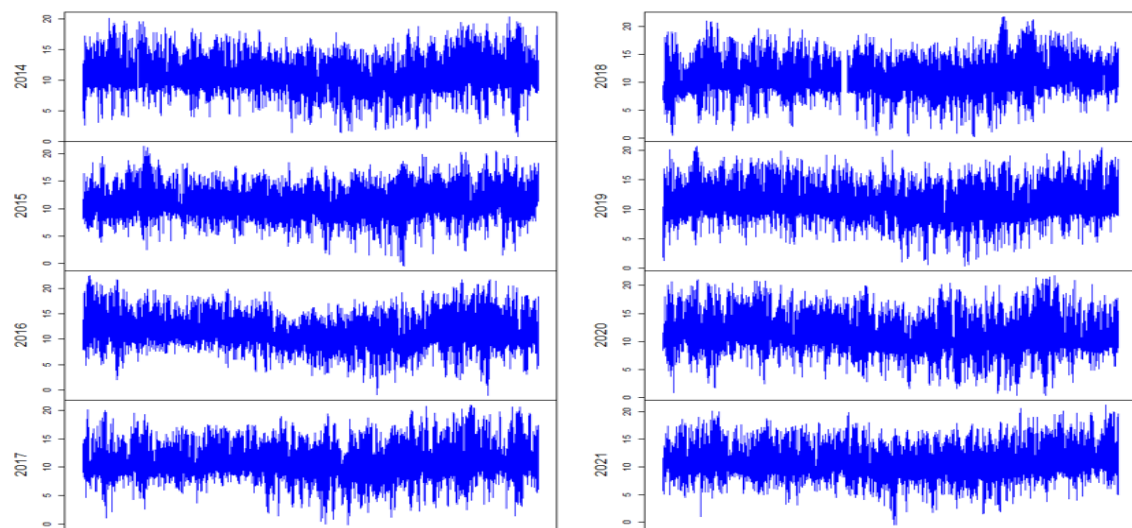


Figura 4-32: Estación San Juan variable Temperatura Ambiente con datos faltantes.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación San Juan Variable Humedad Relativa Con Datos Faltantes

En la figura 4-33 se presenta una serie temporal de los datos de la variable Humedad Relativa en los años que se presentan datos de forma horaria y se nota la discontinuidad en los trazos lo cual refleja la existencia de datos faltantes.

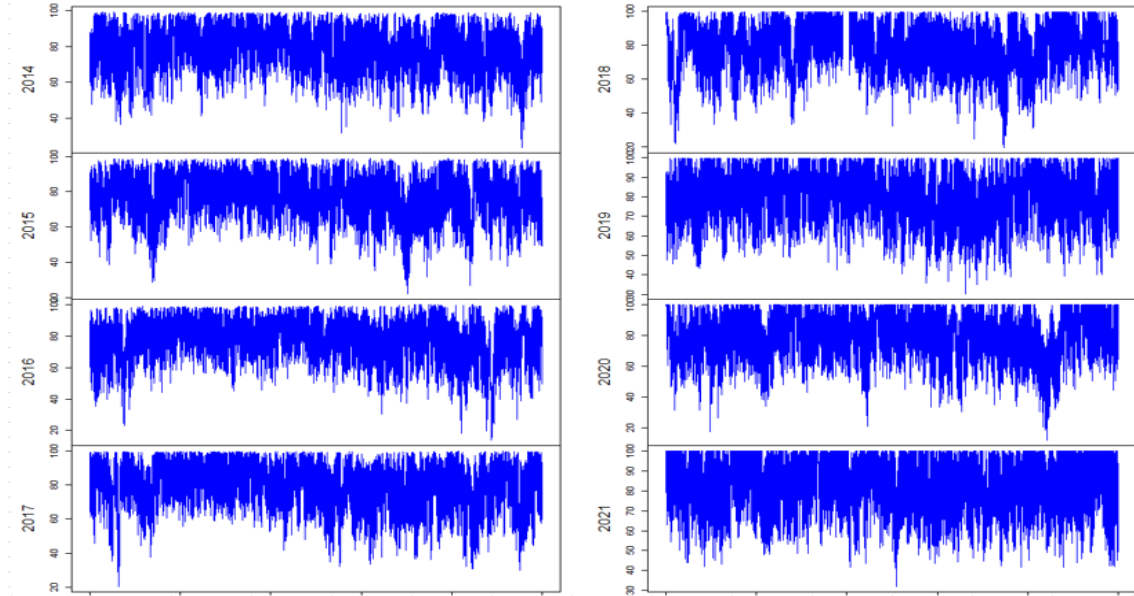


Figura 4-33: Estación San Juan variable Humedad Relativa con datos faltantes.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación San Juan Variable Precipitación Con Datos Faltantes

En la figura 4-34 se presenta una serie temporal de los datos de la variable Humedad Relativa en los años que se presentan datos de forma horaria y se nota la discontinuidad en la gráfica lo cual refleja la existencia de datos faltantes

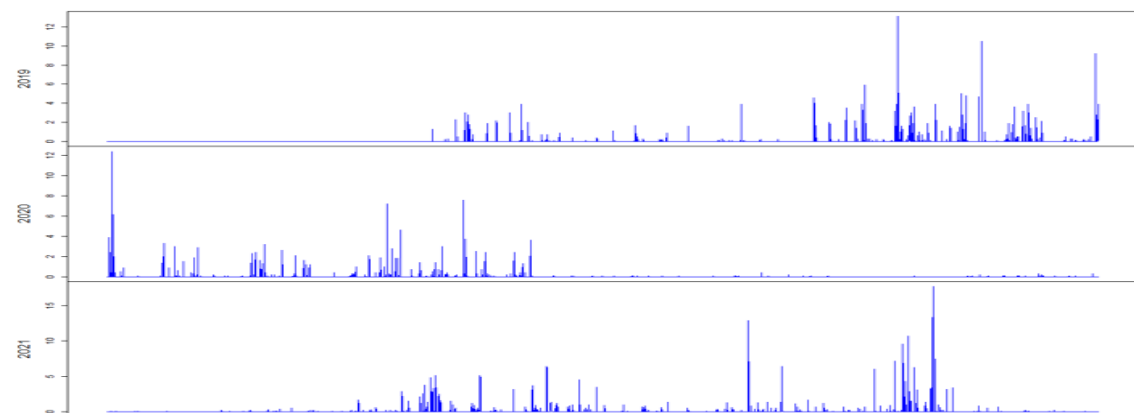


Figura 4-34: Estación San Juan variable Precipitación con datos faltantes.

Realizado por: Altamirano C. y Carrillo P. ,2023

4.4.2. Relleno de datos faltantes mediante MissForest

Estación San Juan Variable Temperatura Ambiente

En la figura 4-35 se presenta una gráfica con formato de serie temporal de los datos imputados por el método MissForest, donde se puede ver que no existe discontinuidad en los trazos de la serie.

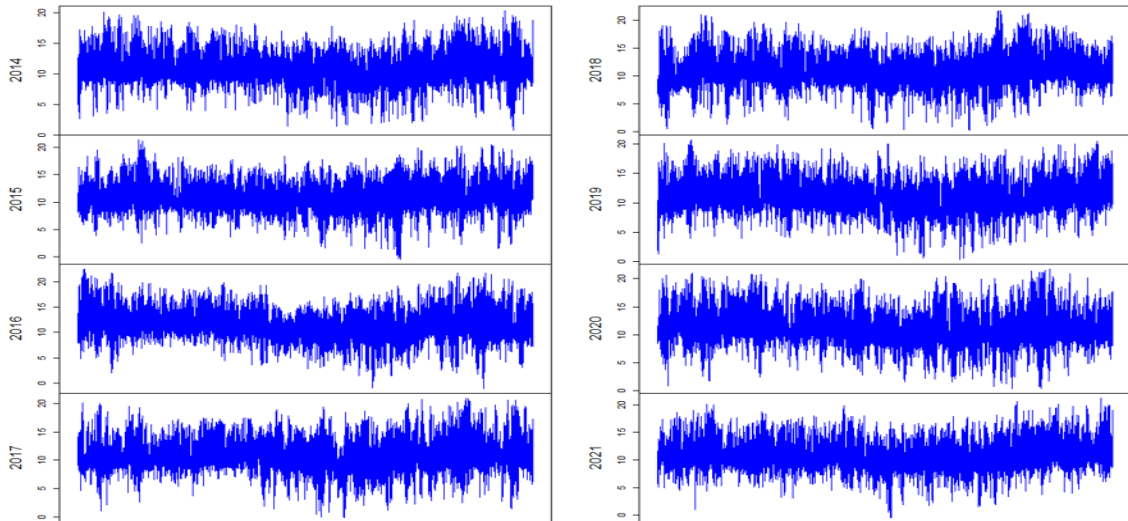


Figura 4-35: Estación San Juan variable Temperatura Ambiente con datos imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

Estación San Juan Variable Humedad Relativa

En la figura 4-36 se presenta una gráfica de los datos imputados por el método MissForest, se puede ver como las series se han graficado de forma continua lo que indica una imputación total.

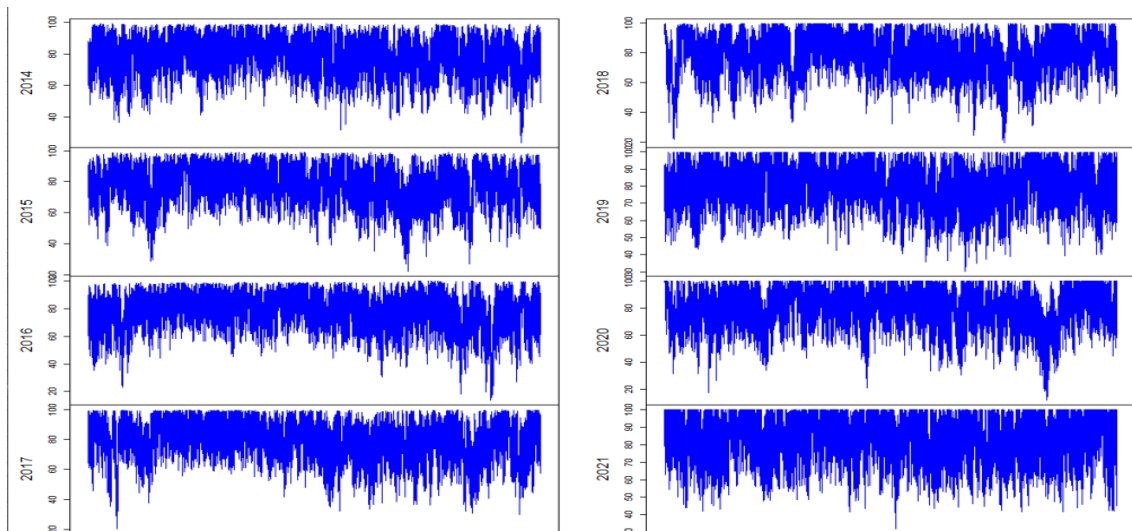


Figura 4-36: Estación San Juan variable Humedad Relativa con datos imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

Estación San Juan Variable Precipitación

En la figura 4-37 se presenta una gráfica con formato de serie temporal de los datos imputados por el método MissForest, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable ha sido imputada en su totalidad.

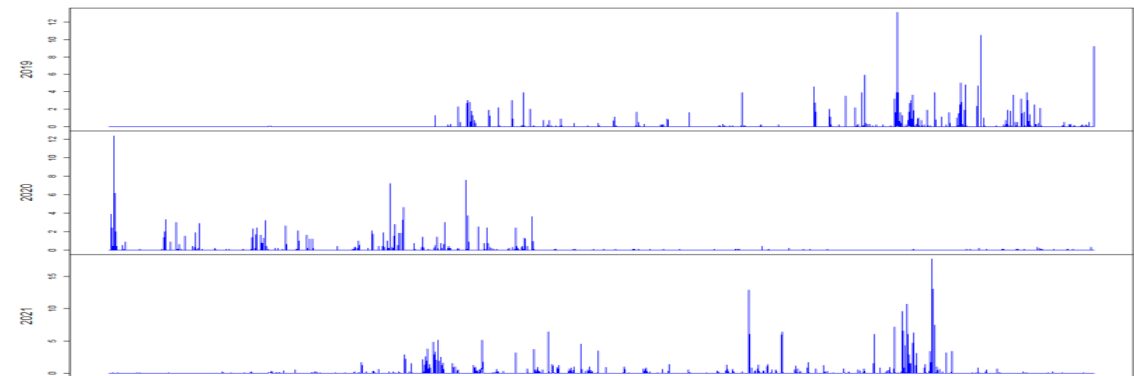


Figura 4-37: Estación San Juan variable Precipitación con datos imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

4.4.3. Relleno De Datos Mediante Las Medias

Estación San Juan Variable Temperatura Ambiente

En la figura 4-38 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de las Medias, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable ha sido imputada en su totalidad.

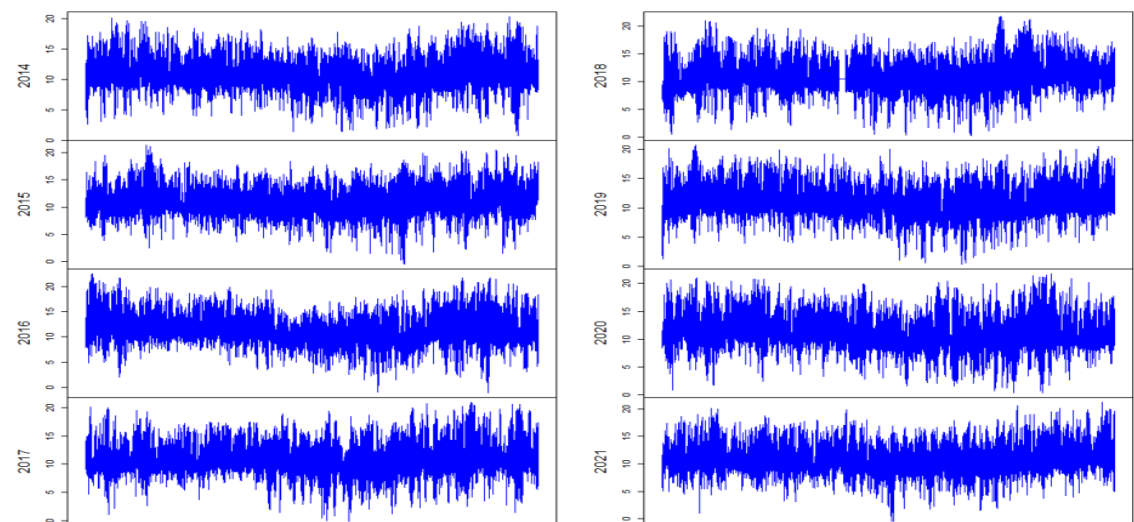


Figura 4-38: Estación San Juan variable Temperatura Ambiente con datos imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

Estación San Juan Variable Humedad Relativa

En la figura 4-39 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de las Medias, se puede ver como las series temporales se han graficado de forma continua, es decir, no se presentan segmentos de gráfica sin trazar, por lo tanto, los datos han sido imputados en su totalidad.

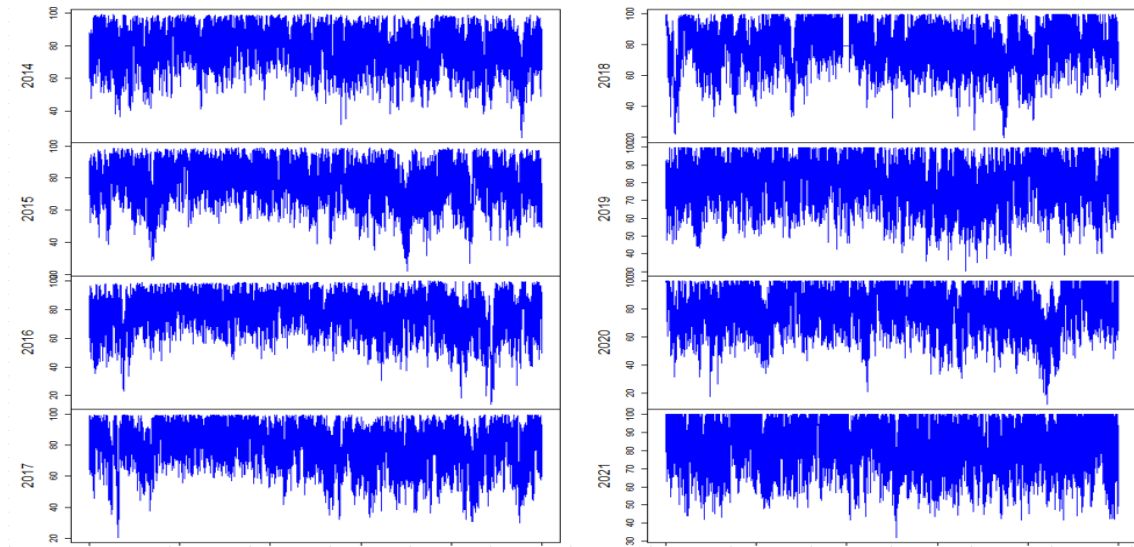


Figura 4-39: Estación San Juan Variable Humedad Relativa Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

Estación San Juan Variable Precipitación

En la figura 4-40 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de las Medias, donde se puede ver que no existe discontinuidad en los trazos de la serie temporal lo que representa que los datos han sido imputados en su totalidad.

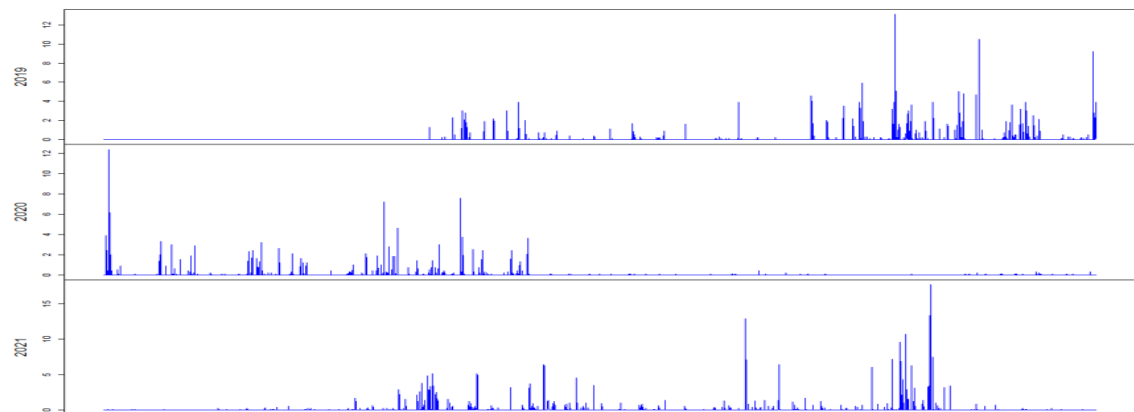


Figura 4-40: Estación San Juan Variable Precipitación Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

4.4.4. Relleno De Datos Usando La Razón Q

Estación San Juan Variable Temperatura Ambiente

En la figura 4-41 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de la razón Q, donde se puede ver que no existe discontinuidad en los trazos de la serie temporal lo que representa que los datos han sido imputados en su totalidad.

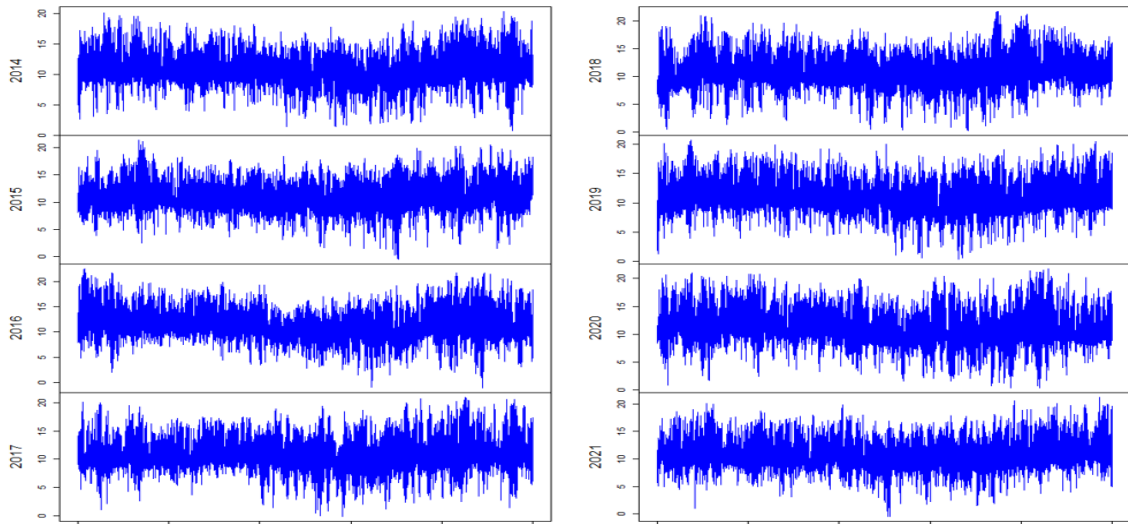


Figura 4-41: Estación San Juan Variable Temperatura Ambiente Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

Estación San Juan Variable Humedad Relativa

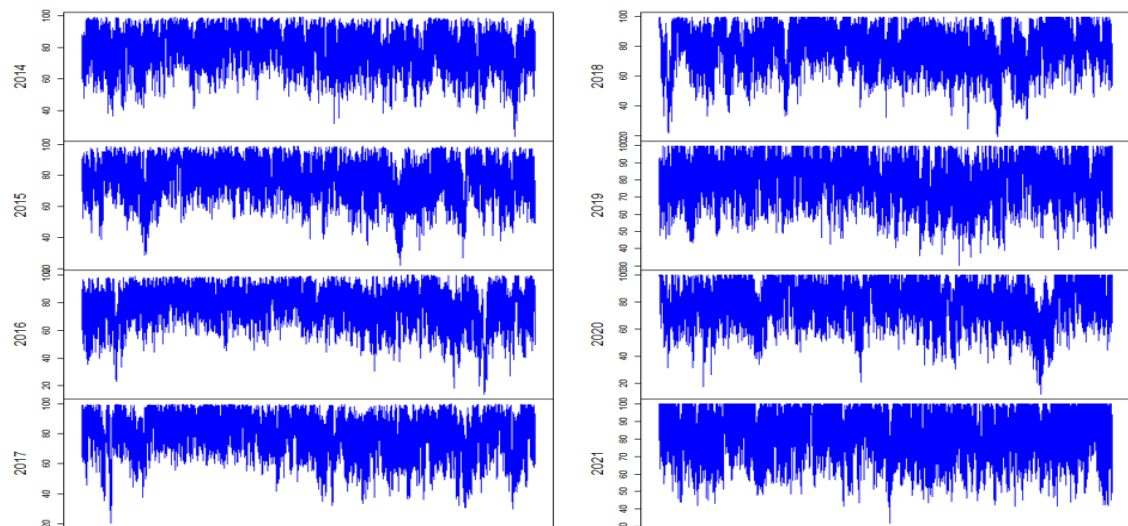


Figura 4-42: Estación San Juan Variable Humedad Relativa Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-42 se puede ver de forma gráfica el comportamiento de las series de tiempo obtenidas

tras la imputación de datos para la variable Humedad Relativa y se nota que los datos han sido imputados en su totalidad ya que las series temporales son continuas para todos los años (no existen segmentos de recta en blanco).

Estación San Juan Variable Precipitación

En la figura 4-43 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de la razón Q, se puede ver como las series temporales se han graficado de forma continua, es decir, no se presentan segmentos de gráfica sin trazar, por lo tanto, los datos han sido imputados en su totalidad.

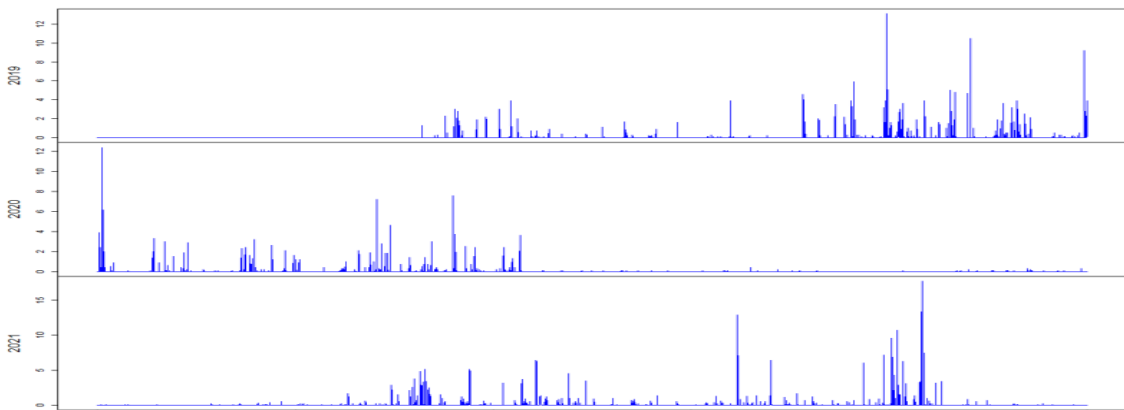


Figura 4-43: Estación San Juan Variable Precipitación Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

4.4.5. Relleno De Datos Usando Regresión Lineal Simple

Estación San Juan Variable Temperatura Ambiente

Año 2014		Año 2015		Año 2016		Año 2017	
a	b	a	b	a	b	a	b
0,769	2,5889	-0,00003	10,465	0,8738	1,7021	0,8101	1,8352
Año 2018		Año 2019		Año 2020		Año 2021	
a	b	a	b	a	b	a	b
0,8263	1,7239	0,8521	1,72	0,8521	1,72	0,7541	2,4528

Figura 4-44: Coeficientes de a y b para cada uno de los años de la variable Temperatura Ambiente.

Realizado por: Altamirano C. y Carrillo P. ,2023

Una vez que se han determinado los coeficientes que modelan la ecuación pertinente para cada uno de los años, mostrados en la figura 4-44 considerando el año 2015 como año base por ser el que mayor cantidad de datos presenta dentro de la variable Temperatura Ambiente para la estación en estudio, se han obtenido los siguientes modelos de regresión simple:

- AÑO 2014:

$$y = 0.7690x + 2.5889$$

- AÑO 2016:

$$y = 0.8738x + 1.7021$$

- AÑO 2017:

$$y = 0.8101x + 1.8352$$

- AÑO 2018:

$$y = 0.8263x + 1.7239$$

- AÑO 2019:

$$y = 0.8521x + 1.7200$$

- AÑO 2020:

$$y = 0.8521x + 1.7200$$

- AÑO 2021:

$$y = 0.7541x + 2.4528$$

Partiendo de los modelos de regresión lineal que definen cada uno de los años de estudio para la variable Temperatura Ambiente se ha obtenido una nueva base de datos imputada en su totalidad, y para verificar este hecho se ha optado por generar gráficas en formato de series temporales para analizar el comportamiento de los nuevos datos.

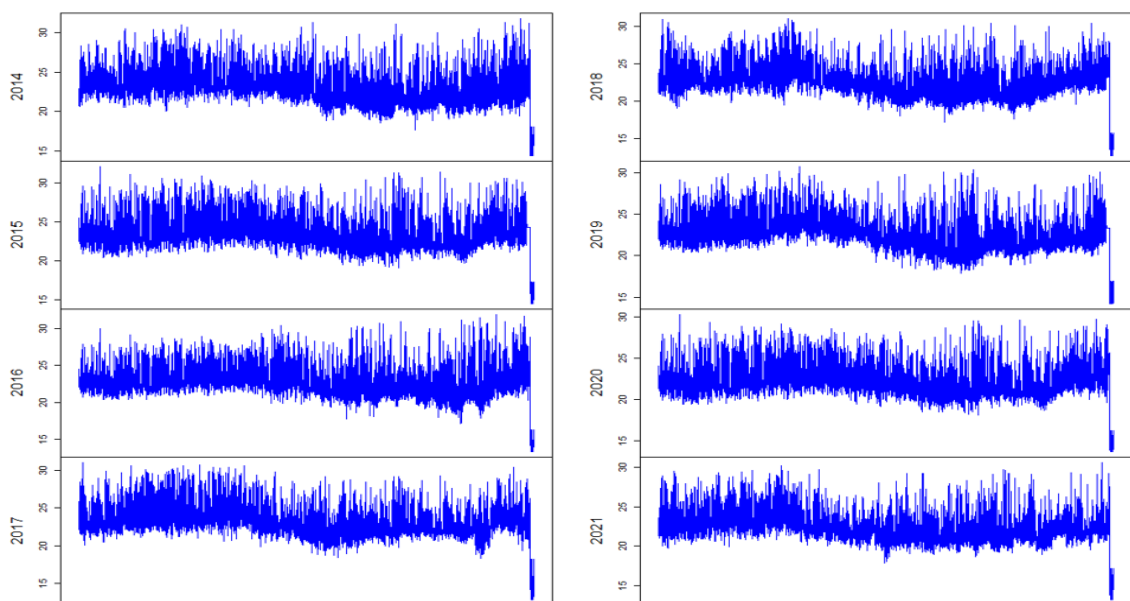


Figura 4-45: Estación San Juan Variable Temperatura Ambiente Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-45 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de Regresión Lineal Simple, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable ha sido imputada en su totalidad.

Estación San Juan Variable Humedad Relativa

Año 2014		Año 2015		Año 2016		Año 2017	
a	b	a	b	a	b	a	b
0,6505	26,876	-0,0007	81,402	0,7137	22,429	0,7125	24,693
Año 2018		Año 2019		Año 2020		Año 2021	
a	b	a	b	a	b	a	b
0,708	23,865	0,7467	22,027	0,7633	20,063	0,7126	27,679

Figura 4-46: Coeficientes de a y b para cada uno de los años de la variable Humedad Relativa.

Realizado por: Altamirano C. y Carrillo P., 2023

Una vez que se han determinado los coeficientes que modelan la ecuación pertinente para cada uno de los años, mostrados en la figura 4-46 considerando el año 2015 como año base por ser el que mayor cantidad de datos presenta dentro de la variable Humedad Relativa para la estación en estudio, se han obtenido los siguientes modelos de regresión:

- AÑO 2014:

$$y = 0.6505 x + 26.8760$$

- AÑO 2016:

$$y = 0.7137 x + 22.4290$$

- AÑO 2017:

$$y = 0.7125 x + 24.6930$$

- AÑO 2018:

$$y = 0.7080 x + 23.8650$$

- AÑO 2019:

$$y = 0.7467 x + 22.0270$$

- AÑO 2020:

$$y = 0.7633 x + 20.0630$$

- AÑO 2021:

$$y = 0.7126 x + 27.6790$$

Partiendo de los modelos de regresión lineal que definen cada uno de los años de estudio para la variable Humedad Relativa se ha obtenido una nueva base de datos imputada en su totalidad, y para verificar este hecho se ha optado por generar gráficas en formato de series temporales para analizar el comportamiento de los nuevos datos.

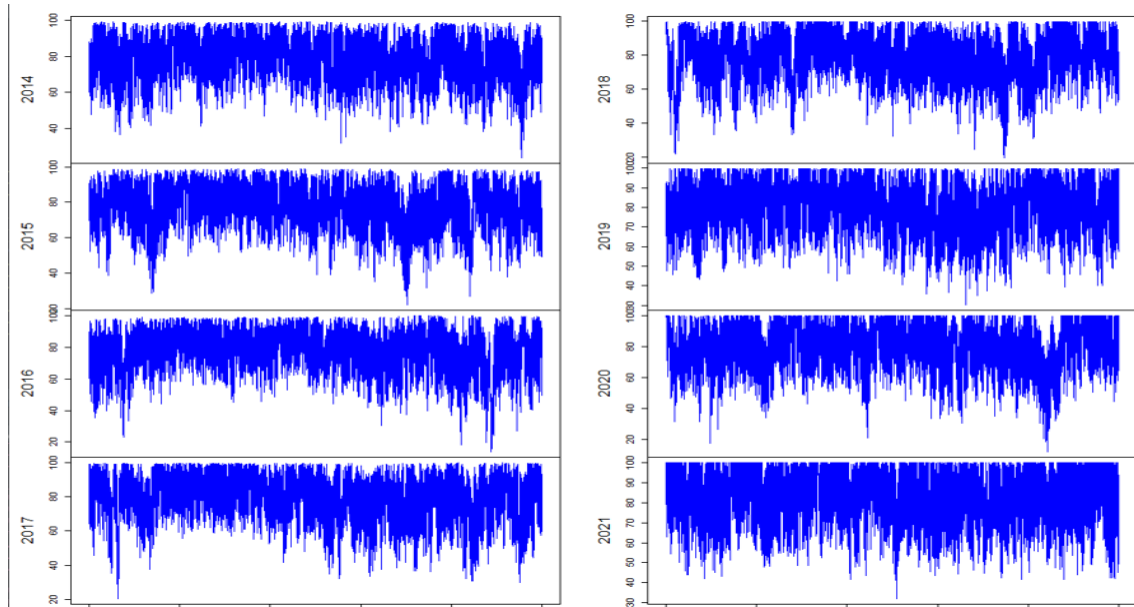


Figura 4-47: Estación San Juan Variable Humedad Relativa Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-47 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de Regresión Lineal Simple, donde se puede ver que no existe discontinuidad en los trazos de la serie temporal lo que representa que los datos han sido imputados en su totalidad.

Estación San Juan Variable Precipitación

Año 2019		Año 2020		Año 2021	
a	b	a	b	a	b
0,0166	0,0587	-0,0188	0,0597	0,000009	0,02

Figura 4-48: Coeficientes de a y b para cada uno de los años de la variable Precipitación.

Realizado por: Altamirano C. y Carrillo P. ,2023

Una vez que se han determinado los coeficientes que modelan la ecuación pertinente para cada uno de los años mostrados en la figura 4-48 considerando el año 2021 como año base por ser el que mayor cantidad de datos presenta dentro de la variable Precipitación para la estación en estudio, se han obtenido los siguientes modelos de regresión:

- AÑO 2019:

$$y = 0.0166x + 0.0587$$

- AÑO 2020:

$$y = -0.0188x + 0.0597$$

Partiendo de los modelos de regresión lineal que definen cada uno de los años de estudio para la variable Precipitación se ha obtenido una nueva base de datos imputada en su totalidad, y para verificar este hecho se ha optado por generar gráficas en formato de series temporales para analizar el comportamiento de los nuevos datos.

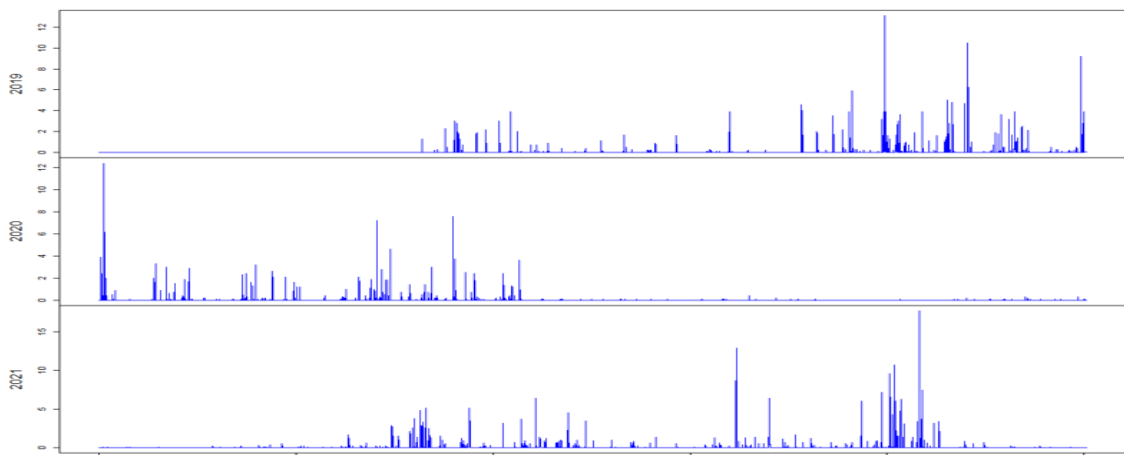


Figura 4-49: Estación San Juan Variable Precipitación Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-49 se presenta una gráfica con formato de serie temporal de los datos imputados por el método de Regresión Lineal Simple, se puede ver como las series temporales se han graficado de forma continua, es decir, no se presentan segmentos de gráfica sin trazar, por lo tanto, los datos han sido imputados en su totalidad.

4.4.6. Relleno De Datos Mediante Imágenes Satelitales

Para poder efectuar el relleno de datos faltantes empleando la técnica de imágenes satelitales ha sido de gran ayuda que el Grupo de Energías Alternativas y Ambiente (GEAA) facilite las bases de datos para las variables Temperatura Ambiente, Humedad Relativa y Precipitación de la estación Cumandá, con lo cual se ha realizado la comparativa entre la base de datos original y la base de datos satelital, para así poder imputar los datos faltantes en la base original para cada una de las variables en estudio.

Estación San Juan Variable Temperatura Ambiente

En la figura 4-50 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de Imágenes Satelitales, donde se puede ver que no existe discontinuidad en los trazos de la serie temporal lo que representa que los datos han sido imputados en su totalidad.

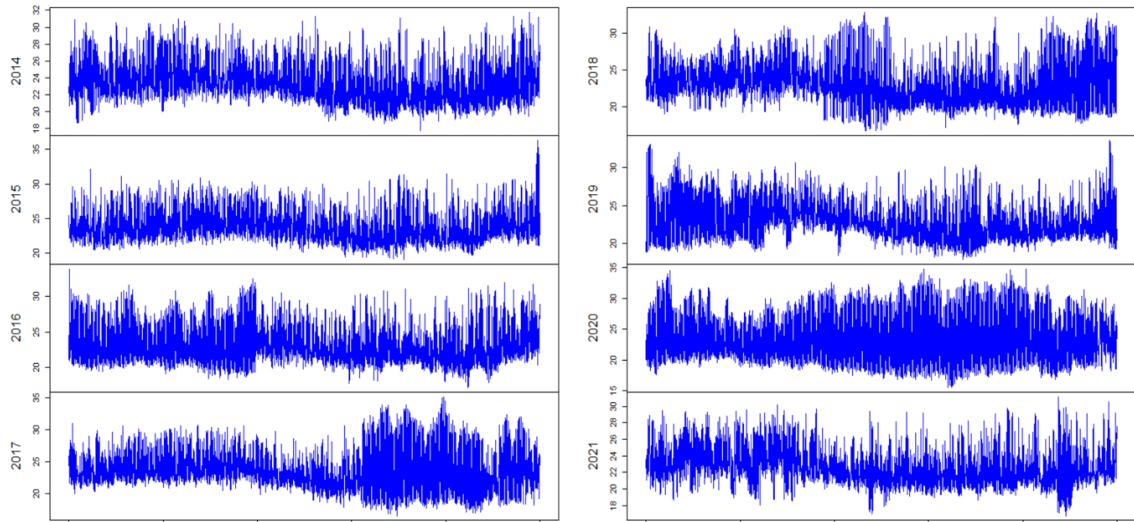


Figura 4-50: Estación San Juan Variable Precipitación Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

Estación San Juan Variable Humedad Relativa

En la figura 4-51 se presenta una gráfica con formato de serie temporal de los datos imputados por la técnica de Imágenes Satelitales, donde se puede notar que la gráfica es continua, es decir, no se presentan cortes en la serie temporal por lo tanto se define que la variable ha sido imputada.

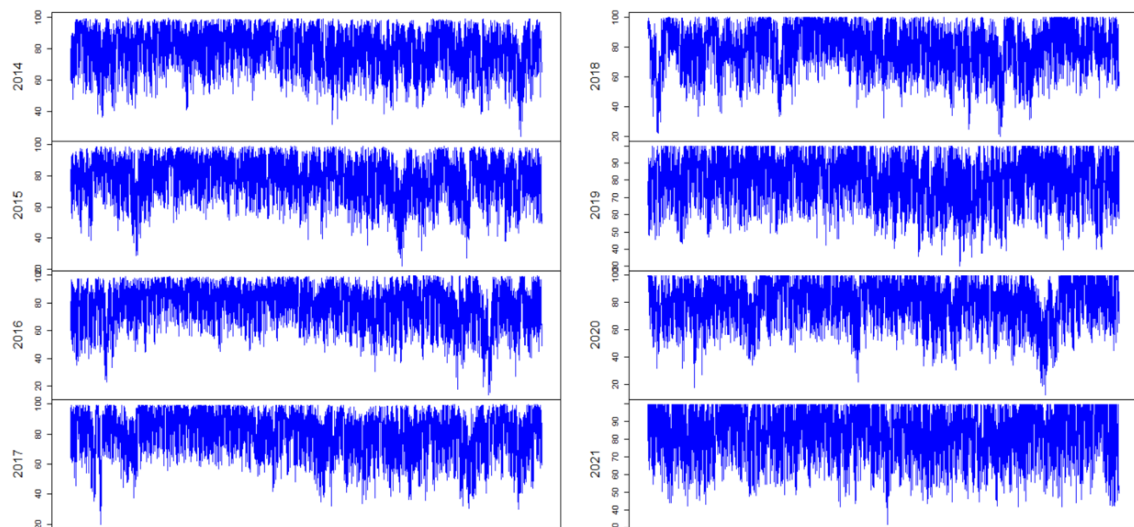


Figura 4-51: Estación San Juan Variable Humedad Relativa Con Datos Imputados.

Realizado por: Altamirano C. y Carrillo P., 2023

Estación San Juan Variable Precipitación

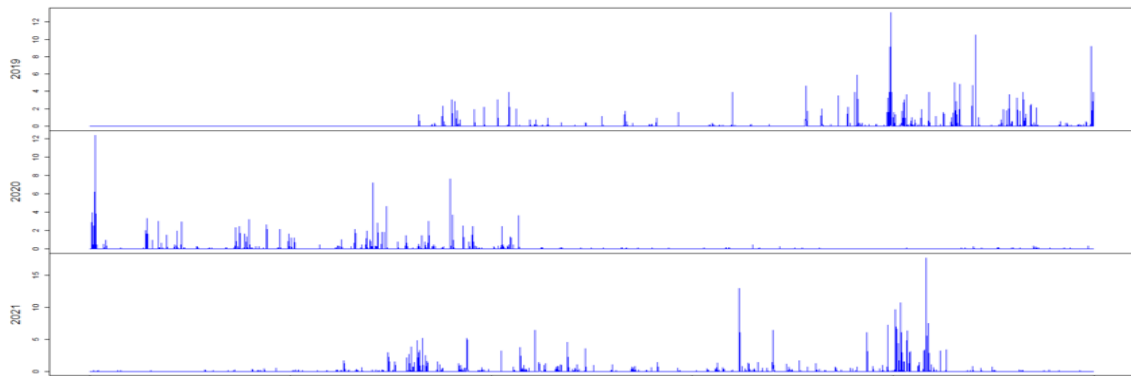


Figura 4.52: Estación San Juan Variable Precipitación Con Datos Imputados

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-52 se puede notar como los datos han sido imputados en su totalidad ya que en las series temporales obtenidas se presentan trazos continuos, pero se puede ver que el comportamiento de las gráficas no llevan un patrón coherente, se puede ver como la precipitación es aproximadamente cero en ciertos segmentos lo cual indica que no existió precipitación lo cual es algo alejado de la realidad, es por eso que se ha efectuado un nuevo análisis tomando como referencia la base de datos proporcionada por el Grupo de Energías Alternativas y Ambiente (GEAA) correspondiente a los datos almacenados por las imágenes satelitales, para obtener nuevas gráficas y así ver como estos datos se comportan en los períodos de tiempo que se ha efectuado el estudio, los resultados de este análisis se presentan en la figura 4-53.

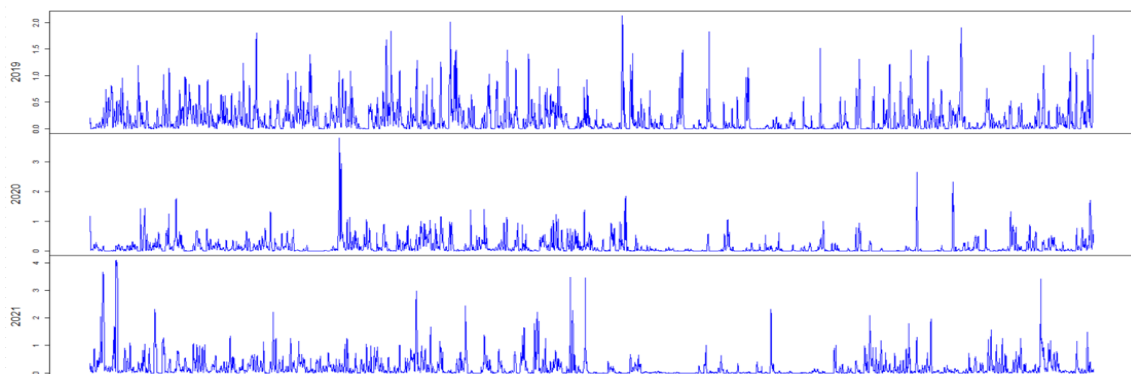


Figura 4-53: Estación San Juan Variable Precipitación Con Datos Imputados Base Imágenes Satelitales

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-53, se puede ver como el comportamiento de las series temporales para cada uno de los años en los que se ha estudiado la variable Precipitación en la estación Cumandá es mucho más acertada ya que se presentan patrones que no están entorno a cero.

4.5. Selección de la mejor técnica de imputación de datos faltantes

Previamente se ha presentado de forma gráfica en la sección 4.3 (Estación Cumandá Imputación de Datos) los resultados obtenidos al aplicar cada una de las técnicas de relleno de datos para la estación Cumandá que es la estación que presenta la mayor cantidad de datos faltantes, mientras que por otra parte a lo largo de la sección 4.4 (Estación San Juan Imputación de Datos) se presentan los resultados de forma gráfica de los resultados obtenidos tras aplicar cada una de las técnicas de relleno de datos sobre la estación San Juan que es la estación que menor cantidad de datos faltantes presenta en el estudio previo.

Para poder determinar cual es la mejor técnica de imputación de datos faltantes para cada una de las variables de estudio (Temperatura Ambiente, Humedad Relativa y Precipitación) se han determinado tres estimadores estadísticos los cuales son, las matrices de correlación entre los datos observados y los datos imputados, el coeficiente de correlación promedio y finalmente los errores cuadráticos medios generados para cada técnica de imputación de datos faltantes.

4.5.1. Matrices de Correlación Estación Cumandá

Variable Temperatura Ambiente

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando la técnica MissForest.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MISSFOREST								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	0.765136014	0.679303047	0.69875369	0.646081997	0.657047906	0.687090465	0.388264903	0.6615682
Imputados 2015	0.69470756	0.75938229	0.700655368	0.644338258	0.636645087	0.682168464	0.459043968	0.632729936
Imputados 2016	0.736781372	0.688738696	0.722567728	0.713919709	0.697925184	0.707882371	0.473897243	0.68546018
Imputados 2017	0.715605063	0.680634505	0.703152705	0.765233975	0.717524273	0.739726622	0.500277767	0.714594494
Imputados 2018	0.708724365	0.652222711	0.682414858	0.673619869	0.759371969	0.725777303	0.411583631	0.674782654
Imputados 2019	0.730966997	0.695271232	0.648436555	0.725295998	0.73705586	0.762606724	0.490901773	0.690819468
Imputados 2020	0.725561337	0.699296655	0.703715797	0.756834076	0.758267449	0.77135109	0.510209971	0.731408881
Imputados 2021	0.66561391	0.63910138	0.626054729	0.656456011	0.66951561	0.693048594	0.151452328	0.731721965

Figura 4-54: Estación Cumandá Coeficientes de Correlación Generados por MissForest.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-54 se puede ver que únicamente para el año 2020 se encuentra por debajo de 0.70 (70%) lo cual indica que los datos imputados tienen un alto grado de relación con los datos originales con lo cual se puede decir que el método de imputación MissForest ha generado buenos resultados para los años 2014, 2015, 2016, 2017, 2019, 2019 y 2021, pero en el año 2020 se puede ver que el coeficiente de correlación entre los datos originales y los datos imputados es de 0.51 (51%) lo cual indica que estos datos tienen un grado de relación medio, esto se debe en gran parte a la cantidad de datos faltantes que se presentan para este año en la estación de Cumandá.

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando la técnica de relleno por medias.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MEDIAS								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.69526615	0.747660158	0.689880016	0.68922645	0.731436553	0.736438413	0.674841195
Imputados 2015	0.695699963	1	0.686851048	0.676252722	0.669544156	0.7051127	0.78852145	0.656865986
Imputados 2016	0.596828376	0.535740345	1	0.426654734	0.497886311	0.568704936	0.816533148	0.490954197
Imputados 2017	0.556434111	0.560051659	0.474354697	1	0.566840862	0.583327646	0.841008777	0.567764769
Imputados 2018	0.648083392	0.635495382	0.60877243	0.645702344	1	0.69276217	0.807870459	0.650245708
Imputados 2019	0.676510638	0.644008518	0.698957217	0.680254975	0.663110874	1	0.776526589	0.64445832
Imputados 2020	0.218287187	0.215152011	0.287403419	0.279926771	0.234769435	0.241970116	1	0.214642999
Imputados 2021	0.645245388	0.629607934	0.624992204	0.640252664	0.671733173	0.680980435	0.782940486	1

Figura 4-55: Estación Cumandá Coeficientes de Correlación Generados Por Medias.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-55 se puede ver que para todos los años se cuenta con correlaciones perfectas es decir, adoptan un valor de 1 (100%) lo que significa que al usar los valores medios de cada uno de los años de estudio para imputar los datos faltantes estos no sufren alteraciones o cambios respecto a los datos originales.

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando el método de la Razón Q .

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR RAZON Q								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	0.772119393	0.718108237	0.711094682	0.679684795	0.702632542	0.749074024	0.72086704	0.669040949
Imputados 2015	0.484723228	0.67427849	0.403642929	0.427143999	0.512315607	0.721940445	0.118591922	0.455610985
Imputados 2016	0.090522524	0.190208139	0.73499391	-0.023463554	-0.013754031	-0.043856625	0.727993978	-0.046043883
Imputados 2017	0.592525909	0.526343075	0.488605281	0.732355044	0.569268917	0.633979742	0.659360576	0.522950141
Imputados 2018	0.580253747	0.598028617	0.531101997	0.576787768	0.736734561	0.69089859	0.466345799	0.601658835
Imputados 2019	0.393781614	0.518412181	0.374070178	0.382642641	0.383904454	0.747661737	0.119139661	0.360968114
Imputados 2020	0.072245799	0.082790536	0.157895024	0.012034111	0.124036351	0.05629237	0.273166937	0.023569031
Imputados 2021	0.636011336	0.639019147	0.625277661	0.674689288	0.62714968	0.688072172	0.805788452	0.720497303

Figura 4-56: Estación Cumandá Coeficientes de Correlación Generados Por Razón Q.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-56 se puede ver que el año 2020 es el que presenta el menor coeficiente de correlación con un valor de 0.27 (27%), para el año 2015 se tiene un coeficiente de correlación de 0.67 (67%) y para el resto de años el coeficiente de correlación se encuentra por sobre el 70 (%) lo que refleja que el grado de relación entre los datos originales y los datos imputados es alto para los años 2014, 2016, 2017, 2018, 2019 y 2021.

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando el método de regresión lineal simple.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR REGRESION LINEAL SIMPLE								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	0.737409929	0.705466925	0.665252752	0.659571667	0.657587244	0.747426323	0.627898647	0.660927281
Imputados 2015	0.666641946	0.776270267	0.656118822	0.663198643	0.641170596	0.702687124	0.588066674	0.608945348
Imputados 2016	0.67197495	0.750796482	0.665453481	0.625332376	0.648229811	0.680796106	0.619898869	0.600514922
Imputados 2017	0.646823527	0.677989455	0.615990713	0.735515367	0.661502336	0.704203643	0.610940695	0.638532816
Imputados 2018	0.651702527	0.678511016	0.642213352	0.645046577	0.773654712	0.73915569	0.635751052	0.666822331
Imputados 2019	0.679254208	0.71779611	0.654772855	0.681999093	0.675714823	0.76841834	0.586018448	0.660385461
Imputados 2020	0.65426909	0.774199125	0.651083684	0.655319745	0.6402239	0.693978292	0.569098593	0.599058783
Imputados 2021	0.638379908	0.653751077	0.619998778	0.667283447	0.67869635	0.704073877	0.651683664	0.724517826

Figura 4-57: Estación Cumandá Coeficientes de Correlación Generados Por Regresión Lineal Simple.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-57 se puede ver que para el año 2020 el coeficiente de correlación presenta un valor de 0.57 (57%) aproximadamente, lo cual indica que la relación entre los datos originales y los imputados es media, además, para el año 2016 el coeficiente de correlación toma un valor de 0.67 (67%) aproximadamente con lo cual se indica que los datos originales y los datos estimados tienen una relación buena, finalmente para los años 2014, 2015, 2017, 2018, 2019 y 2021 los coeficientes de correlación están por encima del (70%) lo que representa que los datos originales y los datos imputados tienen un alto grado de relación entre sí.

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando imágenes satelitales.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR IMÁGENES SATELITÁLES								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	0.765136014	0.647774496	0.69875369	0.602538323	0.618014774	0.687090465	0.388264903	0.634279441
Imputados 2015	0.666255292	0.740322265	0.653370067	0.592823157	0.615945095	0.665786816	0.117393355	0.600654287
Imputados 2016	0.310742781	0.262396246	0.71461813	0.103192125	0.227068871	0.384218736	0.473897243	0.261317699
Imputados 2017	0.168158036	0.175605821	-0.057438873	0.728153272	0.192097312	0.111400178	0.197859206	0.167799587
Imputados 2018	0.383477849	0.359712086	0.245643654	0.362619554	0.659466986	0.407769951	-0.045591935	0.388038309
Imputados 2019	0.489884135	0.441732942	0.592448507	0.354217862	0.466123118	0.742727279	0.218301115	0.467051097
Imputados 2020	-0.270310844	-0.251441534	-0.329478947	-0.211645483	-0.299980138	-0.28395531	0.239028993	-0.28493366
Imputados 2021	0.497468757	0.478272979	0.478253817	0.450447468	0.551438016	0.581393897	0.151452328	0.692240643

Figura 4-58: Estación Cumandá Coeficientes de Correlación Generados por Imágenes Satelitales.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-58 se puede notar que para el año 2020 el coeficiente de correlación es demasiado bajo tomando un valor de 0.24 (24%) aproximadamente el cual indica que la relación entre pares de datos es muy baja, por otra parte para los años 2018 y 2021 el coeficiente de correlación está por debajo del 70 (%) lo que indica que los datos imputados y los datos originales tienen una relación medianamente buena entre ellos, y finalmente para los años 2014, 2015, 2016, 2017, 2019 los coeficientes de correlación supera el 70 (%) dejando así como conclusión que la relación entre los datos originales y los imputados es alta.

Variable Humedad Relativa

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando la técnica MissForest.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MISSFOREST								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.550417161	0.628021348	0.516602015	0.463532536	0.385603776	0.108741097	0.272645165
Imputados 2015	0.550295552	1	0.543452176	0.494351892	0.463041031	0.334707318	0.351829001	0.280377953
Imputados 2016	0.645811439	0.590643058	1	0.56448529	0.49326579	0.395085555	0.336342269	0.272608542
Imputados 2017	0.561655468	0.529841669	0.562494395	1	0.467377534	0.448860469	0.413415529	0.287794149
Imputados 2018	0.448602985	0.445114188	0.451816752	0.392635684	1	0.318094045	-0.0933726	0.235254181
Imputados 2019	0.31953146	0.314703149	0.354317474	0.411555602	0.310839158	1	0.583990793	0.271361264
Imputados 2020	0.258082707	0.473548492	0.379831941	0.512740451	0.216736897	0.664469722	1	0.389712026
Imputados 2021	0.249245574	0.262863591	0.221911995	0.274823069	0.249056504	0.25816712	0.157309935	1

Figura 4-59: Estación Cumandá Coeficientes de Correlación Generados por MissForest.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-59 se puede ver que los coeficientes de correlación generados por el método MissForest para imputación de datos faltantes es del 100% lo que indica que los resultados son muy buenos ya que el grado de relación entre los datos originales y los imputados es perfecto.

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando la técnica de relleno por medias.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MEDIAS								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.546783953	0.628021348	0.511429209	0.457476981	0.385603776	0.108741097	0.27318699
Imputados 2015	0.546125146	1	0.537996263	0.48770517	0.462411538	0.334707318	0.226375837	0.276584846
Imputados 2016	0.506699295	0.420908501	0.999906186	0.310242051	0.344401731	0.324560648	0.334659335	0.209756335
Imputados 2017	0.427893411	0.400026524	0.34919755	1	0.335908909	0.284480543	0.287811958	0.20905003
Imputados 2018	0.431180386	0.436609264	0.430568538	0.381586585	1	0.330724567	0.057775412	0.249417182
Imputados 2019	0.326944931	0.282829051	0.332626692	0.307924232	0.294035669	1	0.195174505	0.183550782
Imputados 2020	0.020875697	0.037252537	0.073977556	0.063517728	0.009124745	0.053261731	0.998564815	0.020776871
Imputados 2021	0.23844225	0.24395782	0.226545032	0.226310033	0.241110633	0.195706885	0.091675889	1

Figura 4-60: Estación Cumandá Coeficientes de Correlación Generados Por Medias.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-60 se puede ver que para todos los años se cuenta con correlaciones perfectas es decir, adoptan un valor de 1 (100%) lo cual representa que los datos originales y los datos imputados tienen un grado de relación casi perfecto ya que para los años 2016 y 2020 los valores no son exactamente 1 pero se aproximan a este.

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando el método de la Razón Q .

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR RAZÓN Q								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.552058606	0.628021348	0.514120693	0.460879511	0.385603776	0.108741097	0.272900113
Imputados 2015	0.414249254	1	0.351058575	0.335034829	0.404452373	0.334707318	0.187008091	0.233376554
Imputados 2016	0.334390334	0.351601949	0.998422693	0.247989403	0.25670091	0.257967414	0.32151899	0.13175941
Imputados 2017	0.240120679	0.334003963	0.19064028	1	0.275287585	0.236135657	0.322937529	0.183507269
Imputados 2018	0.255869279	0.314271166	0.213367699	0.187975281	1	0.246360201	-0.162199231	0.156084783
Imputados 2019	0.076624653	0.078904684	0.121952913	-0.017644718	0.081768833	1	0.246717313	0.053370956
Imputados 2020	-0.120987826	-0.042491446	-0.134189827	-0.119065662	-0.054057237	-0.089027289	0.886016138	-0.037266536
Imputados 2021	0.069325197	0.072400074	0.076341784	0.008255726	0.042371909	0.108395319	0.035414735	1

Figura 4-61: Estación Cumandá Coeficientes de Correlación Generados Por Razón Q.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-61 se puede ver que el año 2020 es el que presenta coeficiente de correlación de 0.89 (89%) el cual es muy bueno ya que indica que la relación entre los datos originales y los datos imputados para este año es alta, por otra parte los años 2014, 2015, 2016, 2017, 2018, 2019 y 2021 presentan coeficientes de correlación de 1 o aproximadamente 1 (100%) lo cual representa que la relación entre los datos faltantes y los originales es perfecta y por tanto los resultados obtenidos son muy buenos.

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando el método de regresión lineal simple.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR REGRESION LINEAL SIMPLE								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.551232231	0.628021348	0.514744113	0.460462063	0.385603776	0.108741097	0.273763501
Imputados 2015	0.546213557	1	0.537912758	0.487779316	0.462534479	0.334707318	0.233807074	0.276969669
Imputados 2016	0.584018381	0.63790656	0.999867148	0.459284361	0.439088514	0.367702431	0.334195777	0.257068176
Imputados 2017	0.509068936	0.568227178	0.481284416	1	0.414447343	0.353813779	0.336786168	0.257669978
Imputados 2018	0.461684174	0.48551395	0.469779189	0.410769007	1	0.342136093	0.073007453	0.259044432
Imputados 2019	0.378952188	0.387546072	0.366621206	0.367437898	0.344081593	1	0.222046408	0.202392123
Imputados 2020	0.484100872	0.91545033	0.473830952	0.453633571	0.412811231	0.319692249	0.998966855	0.24821133
Imputados 2021	0.282986714	0.315400935	0.259071291	0.273646955	0.271003195	0.210421464	0.091617943	1

Figura 4-62: Estación Cumandá Coeficientes de Correlación Generados Por Regresión Lineal Simple.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-62 se puede ver que para todos los años el coeficiente de correlación es 1 o aproximadamente 1, lo cual indica que los resultados obtenidos por este método son bastante aceptables ya que el grado de relación entre pares de datos es casi perfecta.

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando imágenes satelitales.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR IMÁGENES SATELITALES								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	-0.217927169	-0.21656593	-0.31808995	-0.138722262	-0.171429733	-0.231183386	-0.001710927	-0.102889009
Imputados 2015	-0.26767127	-0.19401402	-0.331318598	-0.169405802	-0.19422439	-0.212194737	0.123435373	-0.086241512
Imputados 2016	-0.270814083	-0.235095032	-0.303686278	-0.184121483	-0.206512852	-0.241128663	0.101416412	-0.112131827
Imputados 2017	-0.285345112	-0.259204569	-0.340814589	-0.166909851	-0.201077166	-0.227754814	-0.031468244	-0.085480858
Imputados 2018	-0.289572986	-0.21702186	-0.346149678	-0.176291914	-0.166228047	-0.254160809	-0.055234315	-0.113072797
Imputados 2019	-0.268440896	-0.226721535	-0.33259966	-0.174717364	-0.204347096	-0.184478163	-0.012737195	-0.09797144
Imputados 2020	-0.316668083	-0.248717249	-0.350594268	-0.194971932	-0.226897271	-0.26188521	0.079721436	-0.141129829
Imputados 2021	-0.304715823	-0.259887604	-0.373911819	-0.189977473	-0.227289014	-0.243387963	0.052860684	-0.106805786

Figura 4-63: Estación Cumandá Coeficientes de Correlación Generados por Imágenes Satelitales.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-63 se puede notar que para el año 2020 existe una correlación de 0.08 (8%) la cual es demasiado baja, mientras que para el resto de años las correlaciones son negativas e inferiores al 50% lo que indica que los resultados generados por este método presentan un grado de relación muy bajo respecto a los datos originales.

Variable Precipitación

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando la técnica MissForest.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MISSFOREST			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	0.99999995	-0.013839348	-0.003006396
Imputados 2020	0.003294383	1	-0.014469211
Imputados 2021	-0.002844115	-0.015748965	1

Figura 4-64: Estación Cumandá Coeficientes de Correlación Generados por MissForest.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-64 se puede ver que los coeficientes de correlación generados por el método MissForest para imputación de datos faltantes es del 100% lo que indica que los resultados obtenidos son muy buenos para esta variable ya que el grado de relación entre pares de datos es casi perfecta.

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando la técnica de relleno por medias. En la figura 4-65 se puede ver que para el año 2019 se tiene un coeficiente de correlación de 0.90 (90%) aproximadamente, mientras que para los años 2020 y 2021 los coeficientes de correlación obtenidos son aproximadamente 0 (0%), lo cual deja saber que los resultados generados por la técnica de medias no se relacionan en nada a los datos originales.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MEDIAS			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	0.902261292	-0.008994206	-0.002930872
Imputados 2020	-0.001397944	0.009146221	-3.66091E-05
Imputados 2021	-0.002718228	-0.01311241	0.004776938

Figura 4-65: Estación Cumandá Coeficientes de Correlación Generados Por Medias.

Realizado por: Altamirano C. y Carrillo P. ,2023

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando el método de la Razón Q .

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR RAZÓN Q			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	0.900423727	-0.015026609	-0.00227855
Imputados 2020	-0.002501906	-0.008997368	-0.001627
Imputados 2021	-0.002633885	-0.009864281	0.001494138

Figura 4-66: Estación Cumandá Coeficientes de Correlación Generados Por Razón Q.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-66 se puede ver que el año 2019 es el que presenta coeficiente de correlación más alto con un valor de 0.90 (90%), y para los años 2020 y 2021 los coeficientes de correlación adoptan valores que se aproximan a 0 por lo tanto se concluye que los datos generados por este método no son aceptables.

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando el método de regresión lineal simple.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR REGRESIÓN LINEAL SIMPLE			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	0.615249191	0.04713053	-0.005064462
Imputados 2020	0.002746062	-0.006517103	-0.001656945
Imputados 2021	-0.002982406	-0.009864281	0.001564873

Figura 4-67: Estación Cumandá Coeficientes de Correlación Generados Por Regresión Lineal Simple.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-67 se puede ver que el año 2019 es el que presenta coeficiente de correlación más alto con un valor de 0.62 (62%), y para los años 2020 y 2021 los coeficientes de correlación adoptan valores que se aproximan a 0 por lo tanto se concluye que los datos generados por este método no son confiables.

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando imágenes satelitales.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR IMÁGENES SATELITALES			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	0.999999898	0.003766595	-0.004333089
Imputados 2020	0.002106308	1	0.015959032
Imputados 2021	-0.003128333	-0.015748965	1

Figura 4-68: Estación Cumandá Coeficientes de Correlación Generados por Imágenes Satelitales.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-68 se puede notar que los coeficientes de correlación para los años 2020 y 2021 toman el valor de 1 (100%) y para el año 2019 el coeficiente de correlación se aproxima a 1 (100%) con esto se puede concluir que los datos imputados por imágenes satelitales presentan una relación muy buena con los datos originales presentados en esta variable.

4.5.2. Matrices de Correlación Estación San Juan

Variable Temperatura Ambiente

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando la técnica MissForest.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MISSFOREST								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	0.870499204	0.754750962	0.757639501	0.734408515	0.748678226	0.777391409	0.740857728	0.743280212
Imputados 2015	0.761469431	0.870439788	0.791901534	0.766159155	0.782951583	0.792100159	0.762651905	0.771084191
Imputados 2016	0.751229939	0.791704127	0.88708578	0.779261402	0.761585772	0.797763739	0.783852519	0.779779432
Imputados 2017	0.728029354	0.767508915	0.775792356	0.869052862	0.749208602	0.76912546	0.760862599	0.780268547
Imputados 2018	0.744831131	0.778460099	0.771099699	0.754112815	0.877310695	0.777454497	0.755943024	0.764608785
Imputados 2019	0.769641777	0.788426375	0.785008899	0.772753765	0.776873161	0.863730999	0.772745807	0.762003116
Imputados 2020	0.720708465	0.770327037	0.780082921	0.763300501	0.757460001	0.767257406	0.874101946	0.757792299
Imputados 2021	0.727347905	0.770385255	0.774325215	0.769030178	0.765106829	0.771242389	0.758477124	0.860962262

Figura 4-69: Estación San Juan Coeficientes de Correlación Generados por MissForest.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-69 se puede ver que para el año 2016 se presenta el mayor coeficiente de correlación con un valor de 0.89 (89%) aproximadamente, mientras que para los demás años los coeficientes de correlación se encuentran en el intervalo comprendido entre 0.86 (86%) y 0.88 (88%) con esto se puede definir que el grado de relación entre los datos originales y los datos imputados por el método es alto.

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente

generados por los datos originales y los datos imputados usando la técnica de relleno por medias.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MEDIAS								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.747107715	0.756703575	0.732429403	0.747681705	0.771803038	0.72957295	0.733518594
Imputados 2015	0.747771462	1	0.790218571	0.775628554	0.778667223	0.788023499	0.763170004	0.768070531
Imputados 2016	0.758063929	0.790218571	1	0.778358361	0.770064451	0.797960886	0.780798349	0.778069901
Imputados 2017	0.733175735	0.775628554	0.778358361	1	0.752725472	0.774069817	0.771766479	0.774930925
Imputados 2018	0.744391185	0.772241193	0.760750708	0.747805198	1	0.770625326	0.74993063	0.759448542
Imputados 2019	0.772722116	0.788023499	0.797960886	0.774069817	0.776228716	1	0.774059144	0.765217762
Imputados 2020	0.731078037	0.763170004	0.780798349	0.771766479	0.754188717	0.774059144	1	0.757207947
Imputados 2021	0.735580662	0.768070531	0.778069901	0.774930925	0.764368895	0.765217762	0.757207947	1

Figura 4-70: Estación San Juan Coeficientes de Correlación Generados Por Medias.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-70 se puede ver que para todos los años se cuenta con correlaciones perfectas es decir, adoptan un valor de 1 (100%) lo que representa que los datos imputados tienen una relación perfecta respecto a los datos originales.

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando el método de la Razón Q .

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR RAZÓN Q								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.748218672	0.757937306	0.733495806	0.748394892	0.772766028	0.73109387	0.735555712
Imputados 2015	0.747771462	1	0.790218571	0.775628554	0.778667223	0.788023499	0.763170004	0.768070531
Imputados 2016	0.758063929	0.790218571	1	0.778358361	0.770064451	0.797960886	0.780798349	0.778069901
Imputados 2017	0.733175735	0.775628554	0.778358361	1	0.752725472	0.774069817	0.771766479	0.774930925
Imputados 2018	0.749115305	0.781975923	0.771283467	0.752674935	1	0.77629017	0.754158434	0.764515278
Imputados 2019	0.772722116	0.788023499	0.797960886	0.774069817	0.776228716	1	0.774059144	0.765217762
Imputados 2020	0.731078037	0.763170004	0.780798349	0.771766479	0.754188717	0.774059144	1	0.757207947
Imputados 2021	0.735580662	0.768070531	0.778069901	0.774930925	0.764368895	0.765217762	0.757207947	1

Figura 4-71: Estación San Juan Coeficientes de Correlación Generados Por Razón Q.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-71 se puede ver que para todos los años se cuenta con correlaciones perfectas, es decir, adoptan un valor de 1 (100%) lo que significa que el grado de relación entre los datos originales y los datos imputados por la técnica es excelente.

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando el método de regresión lineal simple.

En la figura 4-72 se puede ver que para el año 2021 el coeficiente de correlación presenta un valor de 0.77 (77 %) aproximadamente, lo cual indica que la relación entre los datos originales y los

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR REGRESIÓN LINEAL SIMPLE								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.748144684	0.757701324	0.733370409	0.74823541	0.772633188	0.730852484	0.735235733
Imputados 2015	0.747771462	1	0.790218571	0.775628554	0.778667223	0.788023499	0.763170004	0.768070531
Imputados 2016	0.758063929	0.790218571	1	0.778358361	0.770064451	0.797960886	0.780798349	0.778069901
Imputados 2017	0.733175735	0.775628554	0.778358361	1	0.752725472	0.774069817	0.771766479	0.774930925
Imputados 2018	0.749013016	0.781198804	0.770260306	0.752725472	1	0.776245652	0.754282407	0.76454037
Imputados 2019	0.772722116	0.788023499	0.797960886	0.774069817	0.776228716	1	0.774059144	0.765217762
Imputados 2020	0.731078037	0.763170004	0.780798349	0.771766479	0.754188717	0.774059144	1	0.757207947
Imputados 2021	0.747771462	1	0.790218571	0.775628554	0.778667223	0.788023499	0.763170004	0.768070531

Figura 4-72: Estación San Juan Coeficientes de Correlación Generados Por Regresión Lineal Simple.

Realizado por: Altamirano C. y Carrillo P. ,2023

imputados es alta, mientras que para el resto de años la correlación presentada entre datos imputados y los datos originales es perfecta ya que se presenta un valor del 100% de relación entre ellos.

- A continuación se presentan los coeficientes de correlación para la variable temperatura ambiente generados por los datos originales y los datos imputados usando la técnica de las imágenes satelitales.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR IMÁGENES SATELITALES								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.746301202	0.755292012	0.731368873	0.745741595	0.770849377	0.728955289	0.732373921
Imputados 2015	0.747771462	1	0.790218571	0.775628554	0.778667223	0.788023499	0.763170004	0.768070531
Imputados 2016	0.758063929	0.790218571	1	0.778358361	0.770064451	0.797960886	0.780798349	0.778069901
Imputados 2017	0.733175735	0.775628554	0.778358361	1	0.752725472	0.774069817	0.771766479	0.774930925
Imputados 2018	0.738208383	0.765738884	0.752620586	0.739735402	1	0.763136498	0.741114263	0.751250548
Imputados 2019	0.772722116	0.788023499	0.797960886	0.774069817	0.776228716	1	0.774059144	0.765217762
Imputados 2020	0.731078037	0.763170004	0.780798349	0.771766479	0.754188717	0.774059144	1	0.757207947
Imputados 2021	0.735580662	0.768070531	0.778069901	0.774930925	0.764368895	0.765217762	0.757207947	1

Figura 4-73: Estación San Juan Coeficientes de Correlación Generados por Imágenes Satelitales.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-73 se puede ver que para todos los años se cuenta con correlaciones perfectas es decir, adoptan un valor de 1 (100%) lo que representa que los datos imputados tienen una relación muy alta respecto a los datos originales.

Variable Humedad Relativa

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando la técnica MissForest. En la figura 4-74 se puede ver que dentro de los coeficientes de correlación generados por los datos imputados mediante el método MissForest el mayor valor se ha producido para el año 2018 con un valor de 0.84 (84%) aproximadamente, mientras que el menor valor se ha producido para

el año 2021 con un valor de 0.81 (81%) aproximadamente, además para los años restantes se puede notar que los valores se encuentran en el intervalo comprendido entre 0.82 (82%) y 0.84 (84%) con lo que se puede resumir que el grado de relación entre los datos originales y los datos imputados es alto generándose así buenos resultados para esta variable.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MISSFOREST								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	0.821678524	0.655063528	0.650038732	0.653720145	0.608460887	0.686441826	0.586624894	0.654273701
Imputados 2015	0.650850543	0.824072245	0.656113499	0.664789271	0.641982024	0.713029903	0.646242933	0.684976448
Imputados 2016	0.65332731	0.668000272	0.833199605	0.657371905	0.598082119	0.687363152	0.636351705	0.676004725
Imputados 2017	0.659777474	0.660210816	0.661934556	0.835285265	0.612150734	0.685790946	0.65778973	0.686294376
Imputados 2018	0.618093104	0.648497336	0.601507379	0.610828159	0.841245831	0.656212082	0.615279024	0.641472933
Imputados 2019	0.69150897	0.694106058	0.668114049	0.689560722	0.659100704	0.826466443	0.617677682	0.696444518
Imputados 2020	0.569121592	0.650999838	0.631290926	0.665504679	0.625253038	0.633252776	0.82788551	0.659853534
Imputados 2021	0.646995286	0.683039732	0.662530692	0.683513786	0.638524775	0.681620666	0.663905061	0.812905833

Figura 4-74: Estación San Juan Coeficientes de Correlación Generados por MissForest.

Realizado por: Altamirano C. y Carrillo P., 2023

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando la técnica de medias.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MEDIAS								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.638435193	0.648268898	0.658538005	0.615466718	0.688616342	0.572204657	0.651319034
Imputados 2015	0.63899193	1	0.659794171	0.672658303	0.635763177	0.706122577	0.648442994	0.676781812
Imputados 2016	0.648984918	0.659794171	1	0.659728172	0.600221856	0.680558036	0.622924886	0.663808488
Imputados 2017	0.65983506	0.672658303	0.659728172	1	0.614281295	0.687335257	0.672576684	0.688229585
Imputados 2018	0.612086743	0.631035014	0.596190067	0.610327657	1	0.645868904	0.616226815	0.633685491
Imputados 2019	0.689313885	0.706122577	0.680558036	0.687335257	0.65077197	1	0.628897344	0.685577872
Imputados 2020	0.573361775	0.648442994	0.622924886	0.672576684	0.620094396	0.628897344	1	0.650475106
Imputados 2021	0.652623857	0.676781812	0.663808488	0.688229585	0.637631614	0.685577872	0.650475106	1

Figura 4-75: Estación San Juan Coeficientes de Correlación Generados por Medias.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-75 se puede ver que para todos los años se cuenta con correlaciones perfectas es decir, adoptan un valor de 1 (100%) lo que significa que los datos resultantes de esta técnica se relacionan a la perfección con los datos originales.

- A continuación se presentan los coeficientes de correlación para la variable humedad generados por los datos originales y los datos imputados usando el método de la Razón Q . En la figura 4-76 se puede ver que para todos los años se cuenta con correlaciones perfectas es decir, adoptan un valor de 1 (100%) lo que representa que el grado de relación entre pares de datos es excelente.

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando regresión lineal simple.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR RAZON Q								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.63963961	0.649470193	0.659979315	0.616103804	0.689518816	0.574287844	0.653175335
Imputados 2015	0.63899193	1	0.659794171	0.672658303	0.635763177	0.706122577	0.648442994	0.676781812
Imputados 2016	0.648984918	0.659794171	1	0.659728172	0.600221856	0.680558036	0.622924886	0.663808488
Imputados 2017	0.65983506	0.672658303	0.659728172	1	0.614281295	0.687335257	0.672576684	0.688229585
Imputados 2018	0.617479136	0.64037227	0.602254394	0.616320413	1	0.65245612	0.620797048	0.637907487
Imputados 2019	0.689313885	0.706122577	0.680558036	0.687335257	0.65077197	1	0.628897344	0.685577872
Imputados 2020	0.573361775	0.648442994	0.622924886	0.672576684	0.620094396	0.628897344	1	0.650475106
Imputados 2021	0.652623857	0.676781812	0.663808488	0.688229585	0.637631614	0.685577872	0.650475106	1

Figura 4-76: Estación San Juan Coeficientes de Correlación Generados Razón Q.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-77 se puede ver que para todos los años el coeficiente de correlación es 1 o aproximadamente 1, lo cual indica que los resultados obtenidos por este método se relacionan a la perfección con los datos originales.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR REGRESION LINEAL SIMPLE								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.639371684	0.649143177	0.659794309	0.615987617	0.689279392	0.573721916	0.652750708
Imputados 2015	0.63899193	1	0.659794171	0.672658303	0.635763177	0.706122577	0.648442994	0.676781812
Imputados 2016	0.648984918	0.659794171	1	0.659728172	0.600221856	0.680558036	0.622924886	0.663808488
Imputados 2017	0.65983506	0.672658303	0.659728172	1	0.614281295	0.687335257	0.672576684	0.688229585
Imputados 2018	0.617067488	0.638589351	0.601440216	0.615204417	1	0.651287896	0.620317729	0.637625694
Imputados 2019	0.689313885	0.706122577	0.680558036	0.687335257	0.65077197	1	0.628897344	0.685577872
Imputados 2020	0.573361775	0.648442994	0.622924886	0.672576684	0.620094396	0.628897344	1	0.650475106
Imputados 2021	0.652623857	0.676781812	0.663808488	0.688229585	0.637631614	0.685577872	0.650475106	1

Figura 4-77: Estación San Juan Coeficientes de Correlación Generados por Regresión Lineal Simple.

Realizado por: Altamirano C. y Carrillo P., 2023

- A continuación se presentan los coeficientes de correlación para la variable humedad relativa generados por los datos originales y los datos imputados usando imágenes satelitales.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR IMÁGENES SATELITALES								
	Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2014	1	0.63598204	0.646138205	0.655047176	0.61355711	0.686610248	0.568949463	0.647751412
Imputados 2015	0.63899193	1	0.659794171	0.672658303	0.635763177	0.706122577	0.648442994	0.676781812
Imputados 2016	0.648984918	0.659794171	1	0.659728172	0.600221856	0.680558036	0.622924886	0.663808488
Imputados 2017	0.65983506	0.672658303	0.659728172	1	0.614281295	0.687335257	0.672576684	0.688229585
Imputados 2018	0.608051333	0.623910391	0.590002476	0.601531945	1	0.635982644	0.607142013	0.625078063
Imputados 2019	0.689313885	0.706122577	0.680558036	0.687335257	0.65077197	1	0.628897344	0.685577872
Imputados 2020	0.573361775	0.648442994	0.622924886	0.672576684	0.620094396	0.628897344	1	0.650475106
Imputados 2021	0.652623857	0.676781812	0.663808488	0.688229585	0.637631614	0.685577872	0.650475106	1

Figura 4-78: Estación San Juan Coeficientes de Correlación Generados por Imágenes Satelitales.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-78 se puede notar que los coeficientes de correlación presentado para todos los años de estudio adoptan el valor de 1 (100%) con lo cual se puede definir que relación existente entre los datos originales y los datos imputados es perfecta.

Variable Precipitación

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando la técnica MissForest.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MISSFOREST			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	0.067092098	-0.010464022	0.01772322
Imputados 2020	-0.009843709	0.538332861	-0.006949025
Imputados 2021	0.10062966	-0.009345298	0.056851029

Figura 4-79: Estación San Juan Coeficientes de Correlación Generados por MissForest.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-79 se puede ver que para el año 2020 el coeficiente de correlación es de 0.54 (54%) mientras que para los años 2019 y 2021 los coeficientes de correlación se aproximan a 0 (0%) con lo cual se define el grado de relación entre los datos originales y los datos imputados es bajo.

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando la técnica de relleno por medias.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR MEDIAS			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	1	-0.011236081	0.012197741
Imputados 2020	-0.011236081	1	-0.009755485
Imputados 2021	0.012197741	-0.009755485	1

Figura 4-80: Estación San Juan Coeficientes de Correlación Generados Por Medias.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-80 se puede ver que los coeficientes de correlación para todos los años de estudio en esta variable toman el valor de 1 (100%) lo cual indica que los resultados de imputación tienen una relación perfecta con los datos originales.

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando el método de la Razón Q .

En la figura 4-81 se puede ver que los coeficientes de correlación para todos los años de estudio en esta variable toman el valor de 1 (100%) con lo cual se define que los datos imputados por este método se relacionan de forma perfecta con los datos originales .

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR RAZÓN Q			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	1	-0.011236081	0.012197741
Imputados 2020	-0.011236081	1	-0.009755485
Imputados 2021	0.012197741	-0.009755485	1

Figura 4-81: Estación San Juan Coeficientes de Correlación Generados Por Razón Q.

Realizado por: Altamirano C. y Carrillo P. ,2023

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando el método de regresión lineal simple.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR REGRESIÓN LINEAL SIMPLE			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	1	-0.011236081	0.012197741
Imputados 2020	-0.011236081	1	-0.009755485
Imputados 2021	0.012197741	-0.009755485	1

Figura 4-82: Estación San Juan Coeficientes de Correlación Generados Por Regresión Lineal Simple.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-82 se puede ver que los coeficientes de correlación toman el valor de 1 (100 %) con lo cual se define que el grado de relación entre los datos originales y los datos imputados es perfecto.

- A continuación se presentan los coeficientes de correlación para la variable precipitación generados por los datos originales y los datos imputados usando imágenes satelitales.

COEFICIENTES DE CORRELACIÓN DATOS FALTANTES VS DATOS IMPUTADOS POR IMÁGENES SATELITÁLES			
	Faltantes 2019	Faltantes 2020	Faltantes 2021
Imputados 2019	0.999999898	0.003766595	-0.004333089
Imputados 2020	0.002106308	1	0.015959032
Imputados 2021	-0.003128333	-0.015748965	1

Figura 4-83: Estación San Juan Coeficientes de Correlación Generados Por Imágenes Satelitales.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-83 se puede notar que los coeficientes de correlación para los años 2020 y 2021 toman el valor de 1 (100 %) y para el año 2019 el coeficiente de correlación se aproxima a 1 (100 %) con esto se puede concluir que los datos imputados por imágenes satelitales presentan una relación perfecta con los datos originales presentados en esta variable.

- Una vez que se han determinado las matrices de correlación generadas por las diferentes técnicas empleadas para la imputación de datos en las estaciones de Cumandá y San Juan se ha visto la necesidad de calcular las correlaciones promedio de cada una de las técnicas para tener un mejor acercamiento a la selección de la mejor técnica de imputación.

4.5.3. Correlación Promedio Estación Cumandá

Variable Temperatura Ambiente

Técnica	MissForest	Medias	Razón Q	Regresión Lineal Simple	Imágenes Satelitales
Correlación	0.7216	1	0.6740	0.7188	0.6602
Correlación (%)	72.160	100	67.398	71.879	66.021

Figura 4-84: Estación Cumandá Coeficientes de Correlación Promedio.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-84 se puede ver cada uno de los coeficientes de correlación promedio generados por cada una de las técnicas de imputación de datos, donde: para MissForest se ha obtenido un coeficiente de correlación de 0.7216 (72.16%), para Medias se tiene un coeficiente de correlación de 1 (100%), para Razón Q se tiene un coeficiente de correlación de 0.6740 (67.40%), para Regresión Lineal Simple se tiene un coeficiente de correlación de 0.7188 (71.88%) y finalmente para Imágenes Satelitales se tiene un coeficiente de correlación de 0.6602 (66.02%). Considerando que para que los resultados sean buenos al momento de la imputación se debe tener un coeficiente de correlación que sea igual o mayor al 70% se obtiene que para la variable Temperatura Ambiente de la estación Cumandá las mejores técnicas de imputación de datos faltantes son MissForest, Medias, y Regresión Lineal Simple.

Variable Humedad Relativa

Técnica	MissForest	Medias	Razón Q	Regresión Lineal Simple	Imágenes Satelitales
Correlación	1	0.9998	0.9856	0.9999	-0.1575
Correlación (%)	100	99.9809	98.555	99.985	-15.754

Figura 4-85: Estación Cumandá Coeficientes de Correlación Promedio.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-85 se puede ver cada uno de los coeficientes de correlación promedio generados por cada una de las técnicas de imputación de datos, donde: para MissForest se ha obtenido un coeficiente de correlación de 1 (100%), para Medias se tiene un coeficiente de correlación de 0.9998 (99.98%), para Razón Q se tiene un coeficiente de correlación de 0.9856 (98.56%), para Regresión Lineal Simple se tiene un coeficiente de correlación de 0.9999 (99.99%) y finalmente para Imágenes Satelitales se tiene un coeficiente de correlación de - 0.1575 (-15.75%). Considerando que para que los resultados sean buenos al momento de la imputación se debe tener un coeficiente de correlación que sea igual o mayor al 70% se tiene por tanto que para la variable Humedad Relativa de la estación Cumandá las mejores técnicas de imputación de datos faltantes son MissForest, Medias, Razón Q y Regresión Lineal Simple.

Variable Precipitación

Técnica	MissForest	Medias	Razón Q	Regresión Lineal Simple	Imágenes Satelitales
Correlación	1	0.3054	0.2976	0.2034	1
Correlación (%)	100	30.54	29.764	20.343	100

Figura 4-86: Estación Cumandá Coeficientes de Correlación Promedio.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-86 se puede ver cada uno de los coeficientes de correlación promedio generados por cada una de las técnicas de imputación de datos, donde: para MissForest se ha obtenido un coeficiente de correlación de 1 (100%), para Medias se tiene un coeficiente de correlación de 0.3054 (30.54%), para Razón Q se tiene un coeficiente de correlación de 0.2976 (29.76%), para Regresión Lineal Simple se tiene un coeficiente de correlación de 0.2034 (20.34%) y finalmente para Imágenes Satelitales se tiene un coeficiente de correlación de 1 (100%).

Considerando que para que los resultados sean buenos al momento de la imputación se debe tener un coeficiente de correlación que sea igual o mayor al 70% se tiene por tanto que para la variable Humedad Relativa de la estación Cumandá las mejores técnicas de imputación de datos faltantes son MissForest e Imágenes Satelitales.

4.5.4. Correlación Promedio Estación San Juan

Variable Temperatura Ambiente

Técnica	MissForest	Medias	Razón Q	Regresión Lineal Simple	Imágenes Satelitales
Correlación	0.8716	1	1	0.9710	1
Correlación (%)	87.165	100	100	97.101	100

Figura 4-87: Estación San Juan Coeficientes de Correlación Promedio.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-87 se puede ver cada uno de los coeficientes de correlación promedio generados por cada una de las técnicas de imputación de datos, donde: para MissForest se ha obtenido un coeficiente de correlación de 0.8716 (87.16%), para Medias se tiene un coeficiente de correlación de 1 (100%), para Razón Q se tiene un coeficiente de correlación de 1 (100%), para Regresión Lineal Simple se tiene un coeficiente de correlación de 0.9710 (97.10%) y finalmente para Imágenes Satelitales se tiene un coeficiente de correlación de 1 (100%).

Considerando que para que los resultados sean buenos al momento de la imputación se debe tener un coeficiente de correlación que sea igual o mayor al 70% se tiene por tanto que para la variable Temperatura Ambiente de la estación San Juan las mejores técnicas de imputación de datos faltantes son MissForest, Medias, Razón Q, Regresión Lineal Simple e Imágenes Satelitales.

Variable Humedad Relativa

Técnica	MissForest	Medias	Razón Q	Regresión Lineal Simple	Imágenes Satelitales
Correlación	0.8278	1	1	1	1
Correlación (%)	82.784	100	100	100	100

Figura 4-88: Estación San Juan Coeficientes de Correlación Promedio.

Realizado por: Altamirano C. y Carrillo P. ,2023

En la figura 4-88 se puede ver cada uno de los coeficientes de correlación promedio generados por cada una de las técnicas de imputación de datos, donde: para MissForest se ha obtenido un coeficiente de correlación de 0.8278 (82.78%), para Medias se tiene un coeficiente de correlación de 1 (100%), para Razón Q se tiene un coeficiente de correlación de 1 (100%), para Regresión Lineal Simple se tiene un coeficiente de correlación de 1 (100%) y finalmente para Imágenes Satelitales se tiene un coeficiente de correlación de 1 (100%). Considerando que para que los resultados sean buenos al momento de la imputación se debe tener un coeficiente de correlación que sea igual o mayor al 70% se tiene por tanto que para la variable Humedad Relativa de la estación San Juan las mejores técnicas de imputación de datos faltantes son MissForest, Medias, Razón Q, Regresión Lineal Simple e Imágenes Satelitales.

Variable Precipitación

Técnica	MissForest	Medias	Razón Q	Regresión Lineal Simple	Imágenes Satelitales
Correlación	0.2208	1	1	1	1
Correlación (%)	22.08	100	100	100	100

Figura 4-89: Estación Cumandá Coeficientes de Correlación Promedio.

Realizado por: Altamirano C. y Carrillo P., 2023

En la figura 4-89 se puede ver cada uno de los coeficientes de correlación promedio generados por cada una de las técnicas de imputación de datos, donde: para MissForest se ha obtenido un coeficiente de correlación de 0.2208 (22.08%), para Medias se tiene un coeficiente de correlación de 1 (100%), para Razón Q se tiene un coeficiente de correlación de 1 (100%), para Regresión Lineal Simple se tiene un coeficiente de correlación de 1 (100%) y finalmente para Imágenes Satelitales se tiene un coeficiente de correlación de 1 (100%). Considerando que para que los resultados sean buenos al momento de la imputación se debe tener un coeficiente de correlación que sea igual o mayor al 70% se tiene por tanto que para la variable Humedad Relativa de la estación San Juan las mejores técnicas de imputación de datos faltantes son Medias, Razón Q, Regresión Lineal Simple e Imágenes Satelitales.

- Una vez que se han determinado las correlaciones promedio entre los datos originales y los datos imputados por cada una de las técnicas o métodos estudiados, se han calculado los Errores Cuadráticos Medios producidos tras el relleno de datos faltantes para así definir la mejor técnica o método de imputación para cada variable de estudio.

4.5.5. Errores Cuadráticos Medios Estación Cumandá

Variable Temperatura Ambiente

ECM	2014	2015	2016	2017	2018	2019	2020	2021	TOTAL
MissForest	10.685	4.627	84.462	85.517	69.686	74.895	13.630	49.868	393.370
Medias	9.418	4.452	206.145	165.626	101.589	119.551	510.016	55.526	1172.323
Razón Q	8.7310	4.7375	238.0504	149.7256	90.6700	128.1720	520.8041	55.9703	1196.8609
Regresión Lineal Simple	12.234	5.130	221.816	157.832	95.474	119.327	489.114	58.466	1159.393
Imágenes Satelitales	12.883	8.295	223.506	176.224	108.240	127.655	543.390	59.284	1259.478

Figura 4-90: Estación Cumandá Errores Cuadráticos Medios.

Realizado por: Altamirano C. y Carrillo P., 2023

Previamente en la sección 4.5.3 (Correlación Promedio Estación Cumandá) se determinó que para la variable Temperatura Ambiente los mejores métodos y técnicas de imputación de datos faltantes son: MissForest, Medias y Regresión Lineal Simple, partiendo de estos resultados en la figura 4-90 se presentan los errores cuadráticos medios generados por cada método o técnica usados para imputar los datos faltantes en la variable Temperatura Ambiente.

Donde se puede ver que el ECM total generado por el método MissForest es de 393.370, el ECM total generado por la técnica de Medias es de 1172.323 y finalmente el ECM total generado por el método de Regresión Lineal Simple es de 1159.393; En base a esta información se define que el método MissForest es el mejor para la imputación de datos faltantes en la variable Temperatura Ambiente.

Variable Humedad Relativa

ECM	2014	2015	2016	2017	2018	2019	2020	2021	TOTAL
MissForest	150.335	0.000	1123.421	1295.240	1162.415	1339.940	5.403	1400.511	6477.266
Medias	138.950	64.607	3058.711	2521.721	1622.242	3330.404	7836.560	2665.269	21238.464
Razón Q	150.9925	64.3087	2918.5070	2572.9665	1651.5524	3369.4444	8509.0645	2774.9570	22011.7930
Regresión Lineal Simple	151.345	63.356	3021.215	2559.185	1737.749	3409.550	8338.526	2745.906	22026.830
Imágenes Satelitales	1111.534	971.259	3537.211	1739.954	1873.377	3362.117	5334.706	2650.879	20581.036

Figura 4-91: Estación Cumandá Errores Cuadráticos Medios.

Realizado por: Altamirano C. y Carrillo P., 2023

Previamente en la sección 4.5.3 (Correlación Promedio Estación Cumandá) se determinó que para la variable Humedad Relativa los mejores métodos y técnicas de imputación de datos faltantes son: MissForest, Medias, Razón Q y Regresión Lineal Simple, partiendo de estos resultados en la figura 4-91 se presentan los errores cuadráticos medios generados por cada método o técnica usados para imputar los datos faltantes en la variable Humedad Relativa, donde se puede ver que el ECM total generado por el método MissForest es de 6477.266, el ECM total generado por la técnica de Medias es de 21238.464, el ECM total generado por la técnica de la Razón Q es de 22011.793 y finalmente el ECM total generado por el método de Regresión Lineal Simple es de 22026.830; En base a esta información se define que el método MissForest es el mejor para la imputación de datos faltantes en la variable Humedad Relativa.

Variable Precipitación

ECM	2019	2020	2021	TOTAL
MissForest	0.0001	0.0000	0.0004	0.0005
Medias	0.044	0.000	1.300	1.345
Razón Q	0.045	0.000	1.304	1.349
Regresión Lineal Simple	0.033	0.076	0.047	0.157
Imágenes Satelitales	0.000	0.000	0.000	0.000

Figura 4-92: Estación Cumandá Errores Cuadráticos Medios.

Realizado por: Altamirano C. y Carrillo P., 2023

Previamente en la sección 4.5.3 (Correlación Promedio Estación Cumandá) se determinó que para

la variable Precipitación los mejores métodos y técnicas de imputación de datos faltantes son: MissForest e Imágenes Satelitales, partiendo de estos resultados en la figura 4-92 se presentan los errores cuadráticos medios generados por cada método o técnica usados para imputar los datos faltantes en la variable Precipitación, , donde se puede ver que el ECM total generado por el método MissForest es de 0.0005 y el ECM total generado por la técnica de Imágenes Satelitales es de 0; En base a esta información se define que la técnica de las Imágenes Satelitales es la mejor para la imputación de datos faltantes en la variable Precipitación.

4.5.6. Errores Cuadráticos Medios Estación San Juan

Variable Temperatura Ambiente

ECM	2014	2015	2016	2017	2018	2019	2020	2021	TOTAL
MissForest	0.298	0.000	0.000	0.000	1.976	0.000	0.000	0.000	2.274
Medias	3.0745	0.0000	0.0000	0.0000	5.1487	0.0000	0.0000	0.0000	8.2231
Razón Q	0.414	0.000	0.000	0.000	1.912	0.000	0.000	0.000	2.326
Regresión Lineal Simple	0.408	0.000	0.000	0.000	1.954	0.000	0.000	0.000	2.362
Imágenes Satelitales	0.407	0.000	0.000	0.000	2.150	0.000	0.000	0.000	2.558

Figura 4-93: Estación San Juan Errores Cuadráticos Medios.

Realizado por: Altamirano C. y Carrillo P. ,2023

Previamente en la sección 4.5.4 (Correlación Promedio Estación San Juan) se determinó que para la variable Temperatura Ambiente los mejores métodos y técnicas de imputación de datos faltantes son: MissForest, Medias, Razón Q, Regresión Lineal Simple e Imágenes Satelitales, partiendo de estos resultados en la figura 4-93 se presentan los errores cuadráticos medios generados por cada método o técnica usados para imputar los datos faltantes en la variable Temperatura Ambiente, donde se puede ver que el ECM total generado por el método MissForest es de 2.274, el ECM total generado por la técnica de Medias es de 8.223, el ECM total generado por el método de la Razón Q es de 2.326, el ECM total generado por el método de Regresión Lineal Simple es de 2.362 y finalmente el ECM total generado por la técnica de Imágenes Satelitales es de 2.558; En base a esta información se define que el método MissForest es el mejor para la imputación de datos faltantes en la variable Temperatura Ambiente.

Variable Humedad Relativa

Previamente en la sección 4.5.4 (Correlación Promedio Estación San Juan) se determinó que para

la variable Humedad Relativa los mejores métodos y técnicas de imputación de datos faltantes son: MissForest, Medias, Razón Q, Regresión Lineal Simple e Imágenes Satelitales, partiendo de estos resultados en la figura 4-94 se presentan los errores cuadráticos medios generados por cada método o técnica usados para imputar los datos faltantes en la variable Temperatura Ambiente, donde se puede ver que el ECM total generado por el método MissForest es de 130.050, el ECM total generado por la técnica de Medias es de 295.7129, el ECM total generado por el método de la Razón Q es de 134.860, el ECM total generado por el método de Regresión Lineal Simple es de 135.361 y finalmente el ECM total generado por la técnica de Imágenes Satelitales es de 160.500; En base a esta información se define que el método MissForest es el mejor para la imputación de datos faltantes en la variable Humedad Relativa.

ECM	2014	2015	2016	2017	2018	2019	2020	2021	TOTAL
MissForest	17.207	0.000	0.000	0.000	112.843	0.000	0.000	0.000	130.050
Medias	90.8029	0.0000	0.0000	0.0000	204.9100	0.0000	0.0000	0.0000	295.7129
Razón Q	16.012	0.000	0.000	0.000	118.848	0.000	0.000	0.000	134.860
Regresión Lineal Simple	15.836	0.000	0.000	0.000	119.525	0.000	0.000	0.000	135.361
Imágenes Satelitales	21.715	0.000	0.000	0.000	138.785	0.000	0.000	0.000	160.500

Figura 4-94: Estación Cumandá Errores Cuadráticos Medios.

Realizado por: Altamirano C. y Carrillo P. ,2023

Variable Precipitación

ECM	2019	2020	2021	TOTAL
MissForest	0.2350	0.0590	0.4511	0.7451
Medias	0.142	0.112	0.000	0.2542
Razón Q	0.123	0.008	0.102	0.2334
Regresión Lineal Simple	0.005	0.009	0.007	0.0210
Imágenes Satelitales	0.000	0.000	0.000	0.0000

Figura 4-95: Estación San Juan Errores Cuadráticos Medios.

Realizado por: Altamirano C. y Carrillo P. ,2023

Previamente en la sección 4.5.4 (Correlación Promedio Estación San Juan) se determinó que para la variable Humedad Relativa los mejores métodos y técnicas de imputación de datos faltantes son: Medias, Razón Q, Regresión Lineal Simple e Imágenes Satelitales, partiendo de estos resultados en la figura 4-95 se presentan los errores cuadráticos medios generados por cada método o técnica usados para imputar los datos faltantes en la variable Temperatura Ambiente, donde se puede ver que el ECM total generado por la técnica de Medias es de 0.2542, el ECM total generado por el método de la Razón Q es de 0.2334, el ECM total generado por el método de

Regresión Lineal Simple es de 0.010 y finalmente el ECM total generado por la técnica de Imágenes Satelitales es de 0; En base a esta información se define que la técnica de Imágenes Satelitales es la mejor para la imputación de datos faltantes en la variable Precipitación.

4.6. Representación gráfica de los resultados

4.6.1. Estación Cumandá

Variable Temperatura Ambiente

En la figura 4-96 se puede ver la representación gráfica de los datos imputados para la variable Temperatura Ambiente en la estación Cumandá en el año 2020 que es el año que presenta mayor cantidad de datos faltantes con un total de 8469 datos, dentro de la gráfica en la primera sección se encuentra el comportamiento de los datos originales en un periodo de tiempo específico, en la segunda gráfica se puede notar como el método ha imputado los datos faltantes completando así la serie y finalmente en el tercer trazo se puede ver el comportamiento general de los datos originales vs los datos imputados.

Variable Humedad Relativa

En la figura 4-97 se puede ver la representación gráfica de los datos imputados para la variable Humedad Relativa en la estación Cumandá en el año 2020 en el cual se presenta la mayor cantidad de datos faltantes con un total de 8581 datos, y se puede notar en la primera sección el comportamiento de los datos observados en un cierto intervalo de tiempo, en la segunda sección se puede ver el comportamiento de los datos imputados contra los datos originales en el intervalo de tiempo determinado en la sección 1, y finalmente en la tercera parte de la gráfica se puede ver el comportamiento general de todos los datos observados versus los datos imputados.

Variable Precipitación

En la figura 4-98 se puede ver la representación gráfica de los datos imputados para la variable Precipitación en la estación Cumandá en el año 2020 el cual presenta la mayor cantidad de datos faltantes con un total de 8469 datos, en la primera sección se presenta el comportamiento de los datos originales, en la segunda sección se puede notar el comportamiento de los datos imputados contra los datos originales en cierto intervalo de tiempo y finalmente en la tercera sección se puede ver el comportamiento general de los datos originales versus los datos imputados en todo el año de estudio.

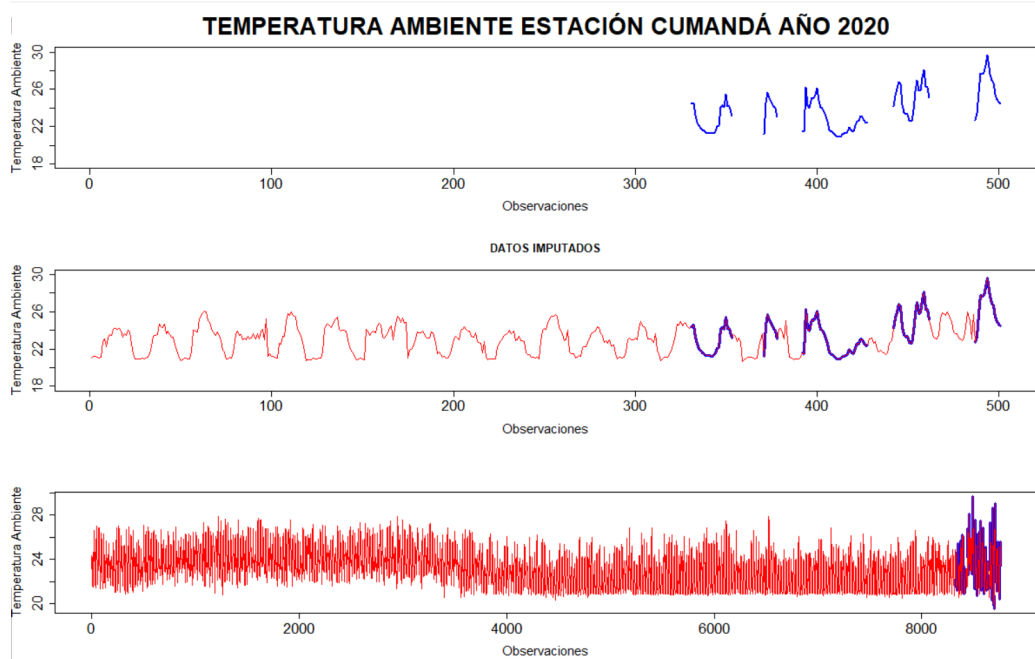


Figura 4-96: Estación Cumandá Datos Imputados Año 2020.

Realizado por: Altamirano C. y Carrillo P. ,2023

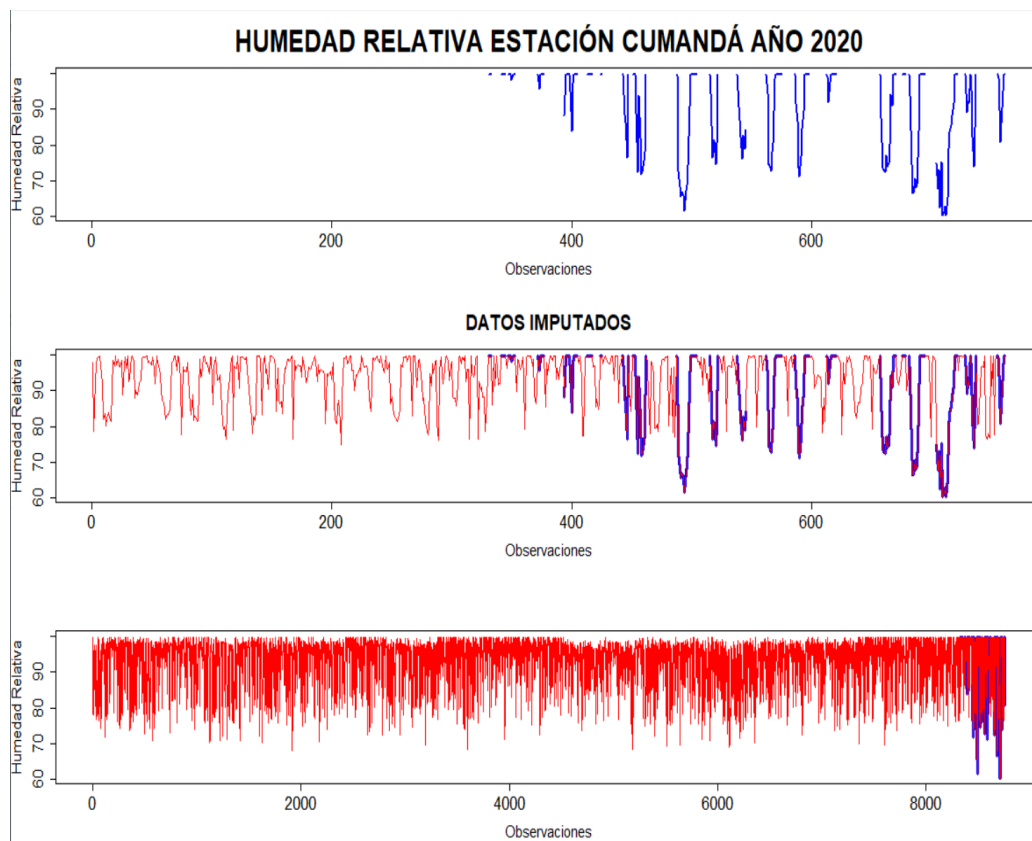


Figura 4-97: Estación Cumandá Datos Imputados Año 2020.

Realizado por: Altamirano C. y Carrillo P. ,2023

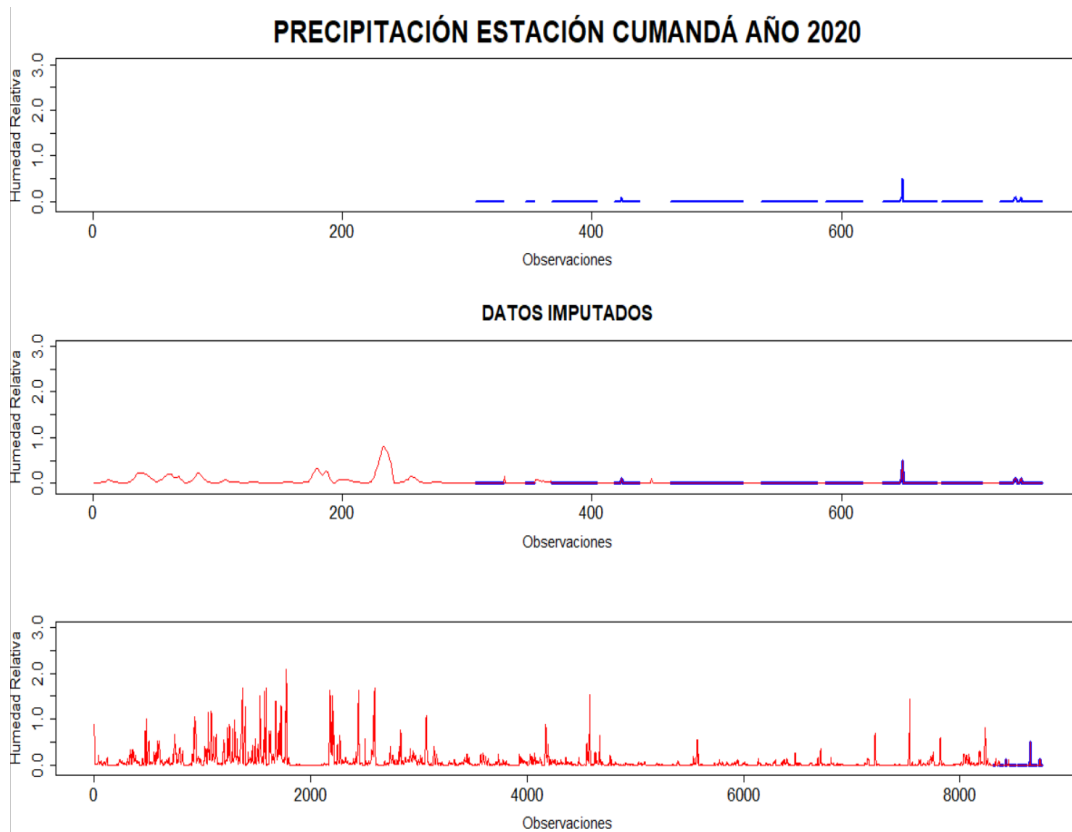


Figura 4-98: Estación Cumandá Datos Imputados Año 2020.

Realizado por: Altamirano C. y Carrillo P., 2023

4.6.2. Estación San Juan

Variable Temperatura Ambiente

En la figura 4-99 se puede ver la representación gráfica de los datos imputados para la variable Temperatura Ambiente en la estación San Juan en el año 2018 que es el año que presenta mayor cantidad de datos faltantes con un total de 157 datos, dentro de la gráfica en la primera sección se encuentra el comportamiento de los datos originales en un periodo de tiempo específico, en la segunda gráfica se puede notar como el método ha imputado los datos faltantes completando así la serie y finalmente en el tercer trazo se puede ver el comportamiento general de los datos originales vs los datos imputados.

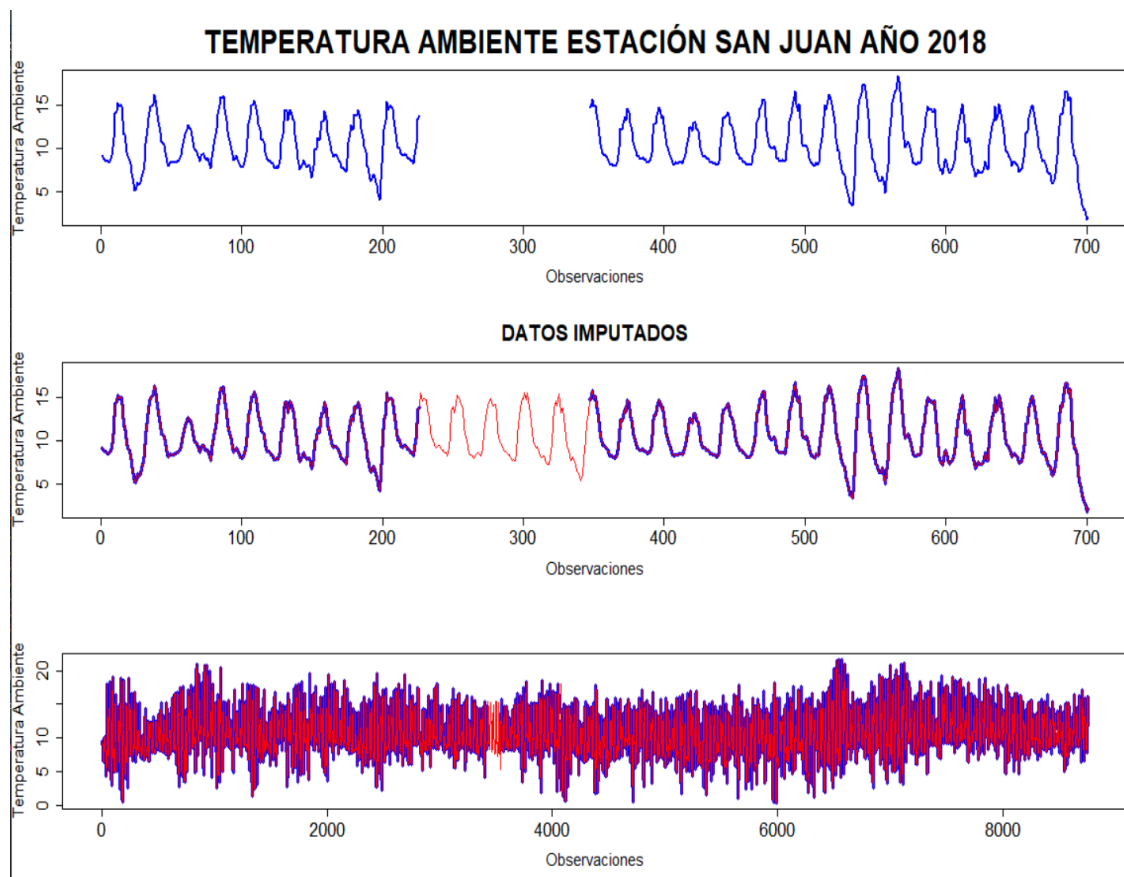


Figura 4-99: Estación Cumandá Datos Imputados Año 2020.

Realizado por: Altamirano C. y Carrillo P., 2023

Variable Humedad Relativa

En la figura 4-100 se puede ver la representación gráfica de los datos imputados para la variable Humedad Relativa en la estación Cumandá en el año 2018 en el cual se presenta la mayor cantidad de datos faltantes con un total de 157 datos, y se puede notar en la primera sección el comportamiento de los datos observados en un cierto intervalo de tiempo, en la segunda sección se puede ver el comportamiento de los datos imputados contra los datos originales en el intervalo de tiempo determinado en la sección 1, y finalmente en la tercera parte de la gráfica se puede ver el comportamiento general de todos los datos observados versus los datos imputados.

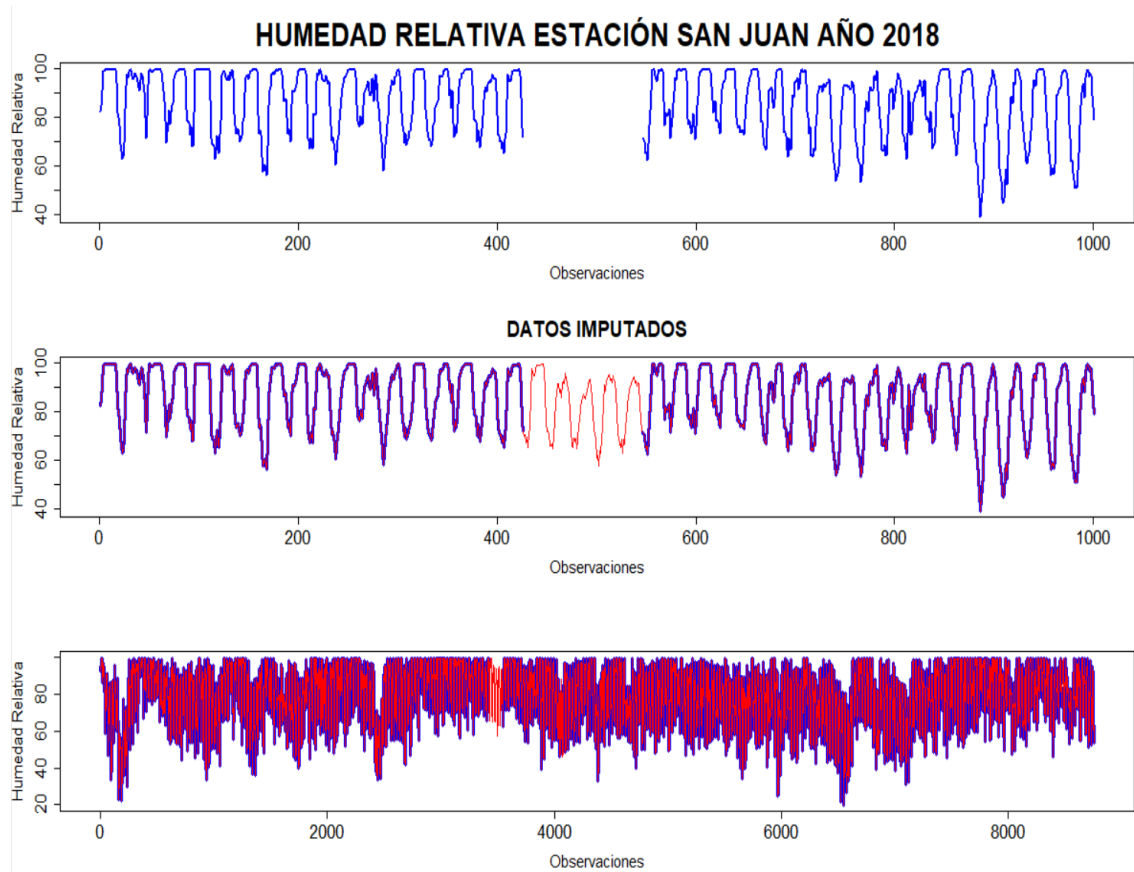


Figura 4-100: Estación San Juan Datos Imputados Año 2020.

Realizado por: Altamirano C. y Carrillo P. ,2023

Variable Precipitación

En la figura 4-101 se puede ver la representación gráfica de los datos imputados para la variable Precipitación en la estación Cumandá en el año 2020 el un total de 24 datos faltantes, en la primera sección se presenta el comportamiento de los datos originales, en la segunda sección se puede notar el comportamiento de los datos imputados contra los datos originales en cierto intervalo de tiempo y finalmente en la tercera sección se puede ver el comportamiento general de los datos originales versus los datos imputados en todo el año de estudio.

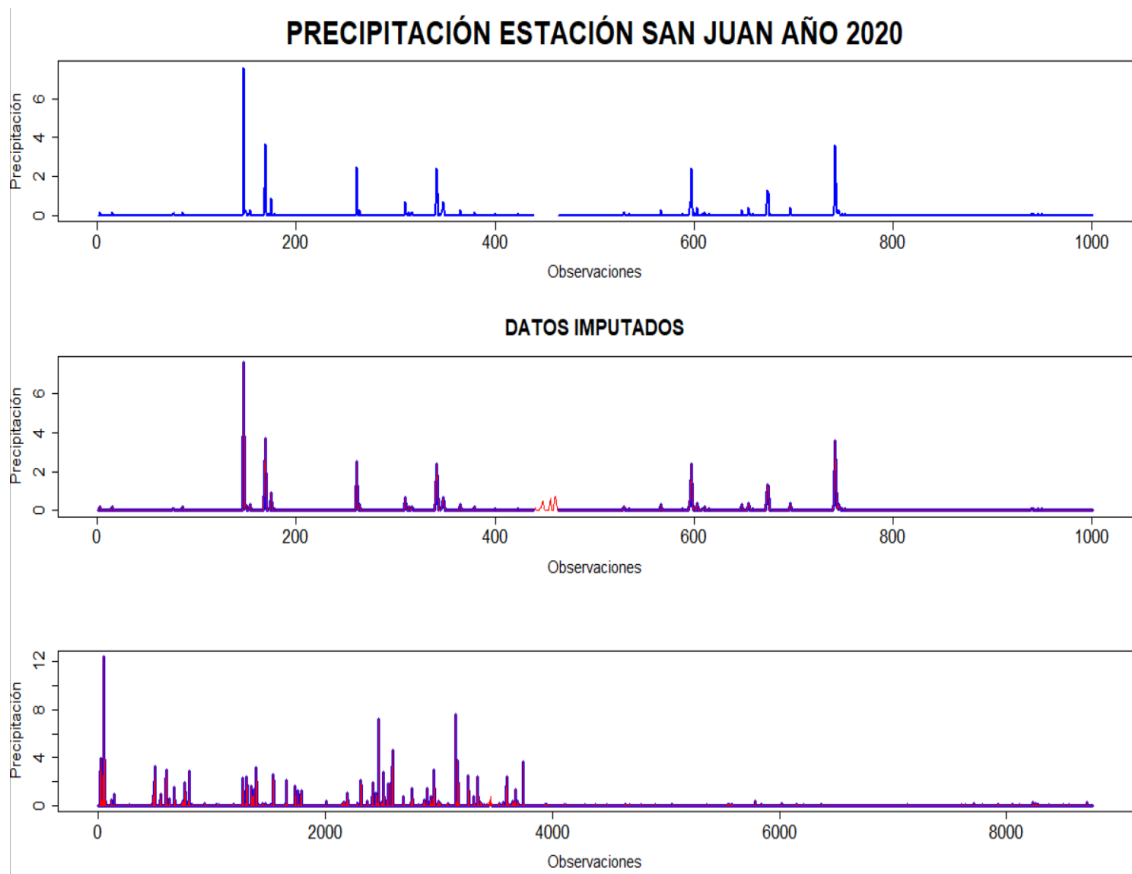


Figura 4-101: Estación San Juan Datos Imputados Año 2020.

Realizado por: Altamirano C. y Carrillo P. ,2023

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- Mediante el análisis estadístico descriptivo se ha determinado cuales son las estaciones que presentan mayor y menor cantidad de datos faltantes, siendo estas la estación Cumandá como la que mayor cantidad de datos faltantes y la estación San Juan como la estación que menor cantidad de datos faltantes presenta.
- Finalizado el relleno de datos faltantes en la estación Cumandá y San Juan se ha determinado de forma gráfica mediante series temporales los comportamientos de cada una de las nuevas bases de datos para validar así que estas han sido imputadas en su totalidad.
- Una vez terminado el proceso experimental mediante los indicadores expuestos en la investigación se ha logrado determinar que para las variables Temperatura Ambiente y Humedad Relativa el método MissForest es el que mejores resultados ha generado, mientras que para la variable Precipitación la técnica de Imágenes Satelitales ha generado mejores resultados.

5.2. Recomendaciones

- Ejecutar mantenimientos de forma periódica a los sensores colocados en las estaciones meteorológicas ubicadas en la provincia de Chimborazo, dándole un énfasis especial a la estación de Cumandá que es la que presenta la mayor cantidad de datos faltantes.
- Realizar un análisis mucho más amplio relacionado a las demás variables con las que cuentan las estaciones meteorológicas ya que en el estudio se ha notado como para Temperatura Ambiente y Humedad Relativa la mejor técnica es usar el comando MISSFOREST de R, mientras que para precipitación la mejor técnica son las imágenes satelitales.
- Automatizar las técnicas que se han determinado como mejores a lo largo de la investigación para así a largo plazo ahorrar recursos al momento de intentar rellenar datos futuros.

BIBLIOGRAFÍA

Ayala, María Fernanda, Villacres David y Tierra Alfonso. “Relación espacio-temporal entre estaciones utilizadas para el relleno de datos de precipitación en Chone, Ecuador”. En: (2018). URL:<http://bdigital2.ula.ve:8080/xmlui/bitstream/handle/654321/1502/Articulo%204.pdf?sequence=1&isAllowed=y> (visitado 09-06-2022).

Carrera-Villacrés, David Vinicio y col. “Relleno de series anuales de datos meteorológicos mediante métodos estadísticos en la zona costera e interandina del Ecuador, y cálculo de la precipitación media”. En: Idesia (Arica) 34.3 (jun. de 2016). Publisher: Universidad de Tarapaca. Facultad de Ciencias agronómicas, págs. 81-90. ISSN: 0718-3429. DOI: [10.4067/S0718-34292016000300010](https://doi.org/10.4067/S0718-34292016000300010). URL: http://www.scielo.cl/scielo.php?script=sci_abstract&pid=S0718-34292016000300010&lng=es&nrm=iso&tlng=es (visitado 01-11-2022).

Haro Rivera, Silvia y col. “DETERMINACION DEL COMPORTAMIENTO METEOROLOGICO DEL VIENTO EN LA PROVINCIA DE CHIMBORAZO, ECUADOR”. En: (ene. de 2020). URL: http://dspace.esPOCH.edu.ec/bitstream/123456789/14577/1/per-n23_v1_04.pdf (visitado 23-10-2022).

Josse, Julie y Husson, Francois (2012) “Handling missing values in exploratory multivariate data analysis methods”. En: (2012). URL: http://www.numdam.org/item/JSFS_2012__153_2_79_0.pdf (visitado 29-10-2022).

Kline, Jack y Milligan, Michael (mayo de 1998). “An Evaluation of Hourly Average Wind-Speed Estimation Techniques”. URL: <https://www.nrel.gov/docs/legosti/fy98/24664.pdf> (visitado 31-10-2022).

Mateus Ferreira, Armando (nov. de 2003). “METODOLOGIAS DE ANALISIS Y IMPUTACION DE DATOS FALTANTES EN SERIES DE VELOCIDAD DEL VIENTO”. URL: https://sgapeio.gal/INFORMEST/VICongreso/artigos/sesion5_05.pdf (visitado 23-10-2022).

Matovelle Carlos, Heras Diego y Solano- Pelaez Jose. “Eficiencia de la Imputacion de Datos Faltantes de Precipitaciones Utilizando Herramientas Computacionales en la Cuenca Hidrografica, Jubones - Ecuador”. es. En: Revista Politécnica 50.2 (ago. de 2022). Number: 2, pags. 23-30. ISSN: 2477-8990. DOI: [10.33333/rp.vol50n2.03](https://doi.org/10.33333/rp.vol50n2.03). URL: https://revistapolitecnica.edu.ec/ojs2/index.php/revista_politecnica2/article/view/1290 (visitado 27-10-2022).

Morgan, Liam (ago. de 2020). “Rpubs - MissForest - Missing Data Imputation”. URL: <https://rpubs.com/lmorgan95/MissForest> (visitado 23-02-2023).

Navarra, Meteo (2023). “Elementos y factores del clima - Meteo Navarra”. URL: <http://meteo.navarra.es/definiciones/elementosfactores.cfm> (visitado 23-01-2023).

Navarro, Javier (jul. de 2017). “Definición de Imagen Satelital”. es-ES. En: Definición ABC. URL: <https://www.definicionabc.com/tecnologia/imagen-satelital.php> (visitado 23-02-2023).

Organización, Meteorológica (mayo de 2016). “Historia de la Organización Meteorológica Internacional”. es. URL: <https://public.wmo.int/es/acerca-de-la-omm/qui%C3%A9nes-somos/historia-de-la-organizaci%C3%B3n-meteorol%C3%B3gica-internacional> (visitado 29-10-2022).

Pérez Natalia, Mullo Héctor y Marcatoma Jessica. ANALISIS DEL CAMBIO CLIMATICO EN UN ECOSISTEMA ALTO ANDINO, RIOBAMBA-ECUADOR. 2020. URL:<http://ceaa.espoeh.edu.ec:8080/revista.perfiles/faces/Articulos/Perfiles23Art1.pdf> (visitado 10-06-2022).

Russi, Jessica y Larrahondo, Eider (2017). “Comparación de métodos de estimación de datos faltantes en series de precipitación diaria en el Valle del Cauca”. URL: <https://bibliotecadigital.univalle.edu.co/bitstream/handle/10893/14419/CB-0573167.pdf?sequence=1> (visitado 10-06-2022).

Serrano Vincenti, Shelia y col. “ANÁLISIS ESTADÍSTICO DE DATOS METEOROLÓGICOS MENSUALES Y DIARIOS PARA LA DETERMINACIÓN DE VARIABILIDAD CLIMÁTICA Y CAMBIO CLIMÁTICO EN EL DISTRITO METROPOLITANO DE QUITO”. es. En: (dic. de 2012). Accepted: 2015-04-20T19:50:40Z. URL: <http://dspace.ups.edu.ec/handle/123456789/8828> (visitado 11-06-2022).

Toainga Oñate, Silvia Daniela (2015). “Dinámica de uso del suelo rural en el periodo 1989 – 2009 y la variabilidad climática en comunidades Altoandinas de la provincia de Chimborazo”. Español. Accepted: 2016-11-29T17:37:00Z Publisher: PUCE. URL: <http://repositorio.puce.edu.ec:80/xmlui/handle/22000/12012> (visitado 10-06-2022).

Toro Trujillo, Ana María y col (ene. de 2018). “Relleno de series diarias de precipitación, temperatura mínima, máxima de la región norte del Urabá Antioqueno | Revista Mexicana de Ciencias Agrícolas”. es-ES.

URL: <https://cienciasagricolas.inifap.gob.mx/index.php/agricolas/article/view/640> (visitado 31-10-2022).

Urrutia Jorge Andrés, Palomino Reiner y Salazar Hever Darío (nov. de 2010). “Metodología para la imputación de datos faltantes en Meteorología”. es. En: Scientia et Technica 3.46. Number: 46, pags. 44-49. ISSN: 2344-7214. DOI: [10.22517/23447214.247](https://doi.org/10.22517/23447214.247).

URL: <https://revistas.utp.edu.co/index.php/revistaciencia/article/view/247> (visitado 23-10-2022).

ANEXOS

date	time	status	2014		2015		2016		2017		2018		2019		2020		2021		
			T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA	status	T_A_MEDIA
01/01/	00:00:00	VALID	22.929	VALID	25.49				VALID	25.89	VALID	23.69						VALID	23.187
01/01/	01:00:00	VALID	22.461	VALID	25.095				VALID	25.012	VALID	23.265						VALID	23.1
01/01/	02:00:00	VALID	22.196	VALID	24.873				VALID	24.543	VALID	23.045						VALID	23.586
01/01/	03:00:00	VALID	22.316	VALID	24.372				VALID	24.42	VALID	22.941						VALID	23.172
01/01/	04:00:00	VALID	22.481	VALID	23.8				VALID	24.52	VALID	22.89						VALID	22.889
01/01/	05:00:00	VALID	21.943	VALID	23.554				VALID	24.177	VALID	22.676						VALID	22.125
01/01/	06:00:00	VALID	21.364	VALID	23.248				VALID	23.781	VALID	22.179						VALID	21.849
01/01/	07:00:00	VALID	21.083	VALID	23.011				VALID	22.654	VALID	21.889						VALID	21.47
01/01/	08:00:00	VALID	20.9	VALID	22.867				VALID	22.343	VALID	21.709						VALID	21.73
01/01/	09:00:00	VALID	20.805	VALID	22.567				VALID	22.379	VALID	21.391						VALID	21.548
01/01/	10:00:00	VALID	20.648	VALID	22.457				VALID	21.907	VALID	20.907						VALID	21.05
01/01/	11:00:00	VALID	20.846	VALID	22.385				VALID	21.65	VALID	20.972						VALID	21.348
01/01/	12:00:00	VALID	20.877	VALID	22.394				VALID	21.635	VALID	20.902						VALID	21.783

Anexo A: Bases de Datos Originales.

Realizado por: Altamirano C. y Carrillo P., 2023

```

1 library(readxl)
2 ##### GRAFICAS #####
3 ##### CUMANDA HUMEDAD RELATIVA 2020 #####
4
5 datos <- read_excel("humedad.xlsx", sheet = 1)
6 attach(datos)
7 colSums(is.na(datos[, 2:9]))
8
9
10 par(mfrow = c(3, 1))
11 plot(datos$DF_2020[8000:8760], col = "blue", xlab = "Observaciones", ylab = "Humedad Relativa", type = "l",
12      lwd = 2, cex.lab = 1.4, cex.axis = 1.7, main = "HUMEDAD RELATIVA ESTACION CUMANDA AÑO 2020", cex.main = 2.5)
13 plot(datos$DF_2020[8000:8760], col = "blue", xlab = "Observaciones", ylab = "Humedad Relativa", type = "l",
14      lwd = 3, cex.lab = 1.4, cex.axis = 1.7, main = "DATOS IMPUTADOS", cex.main = 1.8)
15 lines(datos$DI_2020[8000:8760], col = "red")
16 plot(datos$DF_2020, col = "blue", xlab = "Observaciones", ylab = "Humedad Relativa", type = "l", lwd = 3,
17      cex.lab = 1.4, cex.axis = 1.7)
18 lines(datos$DI_2020, col = "red")
19

```

Anexo B: Obtención de las gráficas de los Datos Imputados en R.

Realizado por: Altamirano C. y Carrillo P., 2023

COEFICIENTES DE CORRELACION DATOS FALTANTES VS DATOS IMPUTADOS POR IMÁGENES SATELITALES											ERRORES CUADRÁTICOS MEDIOS														
Faltantes 2014	Faltantes 2015	Faltantes 2016	Faltantes 2017	Faltantes 2018	Faltantes 2019	Faltantes 2020	Faltantes 2021	Imputados 2014	Imputados 2015	Imputados 2016	Imputados 2017	Imputados 2018	Imputados 2019	Imputados 2020	Imputados 2021	AÑO 2014	AÑO 2015	AÑO 2016	AÑO 2017	AÑO 2018	AÑO 2019	AÑO 2020	AÑO 2021	TOTAL	
Imputados 2014	0.765536094	0.647774496	0.698753569	0.602538323	0.618014774	0.687090465	0.388264903	0.634279441	12.885	8.295	225.506	176.224	108.240	127.655	543.390	59.284	1259.478								
Imputados 2015	0.66625219	0.740322665	0.653370097	0.592823157	0.615945095	0.605786816	0.172933355	0.602654287																	
Imputados 2016	0.510742781	0.262396246	0.714618113	0.103192125	0.227068871	0.384218736	0.473897243	0.261317699																	
Imputados 2017	0.168158036	0.175605821	-0.057438873	0.728153272	0.192097312	0.111400178	0.197859206	0.167799587																	
Imputados 2018	0.382478809	0.359712086	0.245643654	0.262619554	0.803966986	0.402769951	-0.045918315	0.388028308																	
Imputados 2019	0.489884135	0.441732942	0.592448507	0.354217862	0.466123118	0.742727279	0.218301115	0.467051097																	
Imputados 2020	-0.270310844	-0.251441534	-0.329478947	-0.211645483	-0.299980138	-0.28395531	0.239028993	-0.28493366																	
Imputados 2021	0.497468737	0.478272979	0.478253817	0.450447468	0.551438016	0.581393897	0.151452328	0.602340643	0.660211698																

Anexo C: Matrices de Correlación y Errores Cuadráticos Medios.

Realizado por: Altamirano C. y Carrillo P., 2023