



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

**COMPARACIÓN DE TÉCNICAS DE INTERPOLACIÓN ESPACIAL
DE TEMPERATURA AMBIENTE DE LOS MODELOS
ATMOSFÉRICOS CMIP6 Y DE LAS ESTACIONES
METEOROLÓGICAS.**

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERO ESTADÍSTICO

AUTORES:

MALDONADO SIGUENCIA JAMES WILLIAM

YUGSI YUGSI DENNIS ADRIAN

DIRECTORA: ING. CRISTINA ESTEFANÍA RAMOS ARAUJO, MGS.

Riobamba – Ecuador

2024

© 2024, James William Maldonado Siguencia, Dennis Adrian Yugsi Yugsi

Autorizamos 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, Maldonado Siguencia James William y Yugsi Yugsi Dennis Adrian declaramos que el presente Trabajo de Integración Curricular es de nuestra autoría y los resultados de este 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, 28 de mayo de 2024



James William Maldonado Siguencia

030265454-6



Dennis Adrian Yugsi Yugsi

172706484-0

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 INTERPOLACIÓN ESPACIAL DE TEMPERATURA AMBIENTE DE LOS MODELOS ATMOSFÉRICOS CMIP6 Y DE LAS ESTACIONES METEOROLÓGICAS.** realizado por el señor: **MALDONADO SIGUENCIA JAMES WILLIAM** y el señor: **YUGSI YUGSI DENNIS ADRIAN**, 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		2024-05-28
Ing. Cristina Estefania Ramos Araujo, Mgs. DIRECTORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR		2024-05-28
Ing. Natalia Alexandra Pérez Londo, Mgs. ASESORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR		2024-05-28

DEDICATORIA

Este trabajo lo dedico a Dios por brindarme su protección y compañía, a mis padres Cesar y Amable con mucho cariño y mucho amor, ya que sin ellos no lograría mis sueños de obtener mi título académico, ellos siempre confiaron en mí y tuvieron fe y con ganas, dedicación y empeño se logró ese objetivo, a mi hermano Johnny quien me apoya siempre con sus palabras alentadoras y sus grandes sonrisas A mis amigos, compañeros quienes me apoyaron en los buenos y malos momentos en la vida universitaria.

James

Esta tesis va dedicada primeramente a Dios quien ha sido mi guía durante este proceso, en donde su amor y su fidelidad me han acompañado hasta el día de hoy. A mis padres Anita y Héctor, por siempre brindarme su amor y apoyo incondicional en este duro camino, por enseñarme a ser una persona perseverante.

A lo largo de la carrera universitaria he tenido momentos difíciles en donde he querido tirar la toalla, pero en mi mente y en mi corazón siempre ha estado presente el sacrificio y la confianza que mis padres pusieron en mí lo que me ha hecho valorar y me ha dado ánimos y fuerzas para no rendirme hasta cumplir con la meta académica propuesta. A mis hermanos Micaela y Sebastián quienes han estado conmigo en todo momento, brindándome el apoyo moral en los momentos más difíciles de mi carrera.

Sin duda alguna a mi familia quien ha sido mi pilar fundamental y fuente de inspiración para alcanzar este logro académico.

Dennis

AGRADECIMIENTO

Agradezco primero a Dios por darme la fuerza la sabiduría la capacidad el conocimiento para poder realizar cada uno de mis propósitos, por los triunfos y los momentos difíciles, de igual manera a mis padres Cesar y Amable por ser ese pilar fundamental por sus valores y principios ya que siempre me motivaron a seguir esforzándome en el ámbito académico, por su paciencia, palabras de aliento e inspiración en momentos tristes, pero más que nada por su amor, a mis amigos y compañeros por el apoyo incondicional, y por muchos momentos felices en mi vida universitaria. A mis docentes de la carrera de Estadística quienes compartieron su conocimiento para culminar esta etapa importante en mi vida, a pesar de no tener mucho tiempo disponible, me ayudaron me alentaron siempre con una sonrisa. Al grupo de investigación GEAA que me abrieron las puertas en la parte investigativa, y me dieron un impulso, ayudándome al manejo de datos climáticos y de softwares acorde al trabajo. Finalmente agradecer al ingeniero Rommel Palaguachi, David Acurio y la biofísica Isabel Balboa por el apoyo técnico y emocional proporcionado a largo de todo este tiempo.

James

Quiero expresar mi agradecimiento primeramente a Dios por brindarme fuerzas, conocimiento y sabiduría a lo largo de mi carrera universitaria, a mis padres y hermanos por el apoyo incondicional, la fe y la confianza puesta en mi ha sido un pilar fundamental para lograr este objetivo, de igual forma a mis amigos por brindarme su apoyo en todo momento. Agradezco también a mis docentes de la carrera de Estadística quienes han puesto su granito de arena, impartíendome sanos consejos y sobre todo conocimientos esenciales para cumplir este logro académico, también agradezco al grupo de investigación GEEA por apoyarme en el desarrollo de mi investigación.

Dennis

ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS	xi
ÍNDICE DE ECUACIONES	xii
ÍNDICE DE ILUSTRACIONES	xiii
ÍNDICE DE ANEXOS	xvi
RESUMEN	xvii
ABSTRACT	xviii
INTRODUCCIÓN	1
CAPÍTULO I	
1. PROBLEMA DE INVESTIGACIÓN	3
1.1. Antecedentes	3
1.2. Planteamiento del Problema	4
1.3. Problema general de investigación	4
1.4. Problemas específicos de investigación	4
1.5. Objetivos	5
1.5.1. <i>Objetivo general</i>	5
1.5.2. <i>Objetivos específicos</i>	5
1.6. Justificación	5
1.6.1. <i>Justificación Teórica</i>	5
1.6.2. <i>Justificación Metodológica</i>	5
1.6.3. <i>Justificación Práctica</i>	6

CAPÍTULO II

2.	MARCO TEÓRICO	7
2.1.	Bases teóricas	7
2.1.1.	<i>Meteorología</i>	7
2.1.2.	<i>¿Qué es Temperatura?</i>	7
2.1.3.	<i>Geoestadística</i>	7
2.1.4.	<i>Sistema de información geográfica (SIG)</i>	8
2.1.5.	<i>Validación Cruzada</i>	8
2.1.6.	<i>Validación del método de interpolación</i>	8
2.1.7.	<i>Validación del los errores</i>	9
2.1.7.1.	<i>Error Residual</i>	9
2.1.7.2.	<i>Error Cuadrático Medio (RSME)</i>	9
2.1.7.3.	<i>Error Absoluto Medio (MAE)</i>	9
2.1.8.	<i>Coordenadas UTM</i>	9
2.1.9.	<i>System Advisor Model (SAM)</i>	10
2.1.10.	<i>Análisis Geoestadístico</i>	10
2.1.11.	<i>MissForest.</i>	10
2.2.	Bases Conceptuales	11
2.2.1.	<i>Interpolación espacial</i>	11
2.3.	Métodos Determinísticos	12
2.3.1.	<i>Distancia Inversa Ponderada (IDW)</i>	12
2.3.2.	<i>Interpolación Spline</i>	13
2.4.	Métodos Probabilísticos	14
2.4.1.	<i>Interpolación Kriging</i>	14
2.4.2.	<i>Semivariograma</i>	15
2.4.3.	<i>Modelo de Semivariogramas</i>	16

CAPÍTULO III

3.	MARCO METODOLÓGICO	19
3.1.	Tipo de investigación	19
3.2.	Diseño de investigación	19
3.2.1.	<i>Localización del estudio</i>	20

3.2.2.	<i>Población de estudio</i>	20
3.2.3.	<i>Tamaño de la muestra</i>	22
3.2.4.	<i>Método de muestreo</i>	22
3.2.5.	<i>Técnicas de recolección de datos</i>	23
3.3.	Identificación de variables	23
3.3.1.	<i>Variable dependiente</i>	23
3.3.2.	<i>Variables independientes</i>	23
3.3.3.	<i>Operacionalización de variables</i>	23

CAPÍTULO IV

4.	RESULTADOS Y DISCUSIÓN	25
4.1.	Obtención de Datos	25
4.1.1.	<i>Malla de intersección de la provincia de Chimborazo</i>	25
4.1.2.	<i>Obtención de Datos Sam (parte de los modelos CMIP6)</i>	25
4.1.3.	<i>Datos estaciones meteorológicas</i>	27
4.2.	Procesamiento de los datos	27
4.2.1.	<i>Procesamiento de los datos SAM</i>	27
4.3.	Análisis exploratorio de datos	29
4.3.1.	<i>Datos atípicos</i>	29
4.3.2.	<i>Estadísticas Descriptivas</i>	30
4.3.3.	<i>Selección de la muestra de estudio</i>	38
4.3.4.	<i>Estadísticas descriptivas de la muestra de estudio</i>	38
4.3.4.1.	<i>Datos del SAM</i>	39
4.3.4.2.	<i>Datos del Estaciones</i>	40
4.3.4.3.	<i>Transformación Box Cox</i>	41
4.4.	Análisis Geoestadístico	42
4.4.1.	<i>Análisis Geoestadístico datos del SAM y Estaciones</i>	43
4.4.1.1.	<i>Análisis de Tendencia</i>	43
4.4.1.2.	<i>Modelamiento del Semivariograma</i>	48
4.5.	Modelo de interpolación espacial	49
4.5.1.	<i>Kriging – SAM 2015</i>	50
4.5.1.1.	<i>Semivariogramas de Segundo orden</i>	52
4.5.1.2.	<i>Semivariogramas con tendencia constante</i>	53

4.5.2.	<i>IDW (Distancia Inversa Ponderada) Datos del SAM</i>	54
4.5.3.	<i>Spline Datos del SAM</i>	55
4.5.4.	<i>Kriging – Estaciones Meteorológicas 2015</i>	56
4.5.4.1.	<i>Semivariogramas de primer orden</i>	58
4.5.4.2.	<i>Semivariogramas con tendencia constante</i>	59
4.5.5.	<i>IDW (Distancia Inversa Ponderada) Datos Estaciones</i>	60
4.5.6.	<i>Spline Datos Estaciones</i>	61
4.6.	Evaluación de los Métodos	62
4.6.1.	<i>Datos descargados del SAM</i>	62
4.6.2.	<i>Datos de las estaciones</i>	65
4.7.	Mapas de Interpolación	67
4.7.1.	<i>Mapas Datos SAM</i>	67
4.7.2.	<i>Mapas de las Estaciones</i>	74
4.8.	Discusión (Argumentación, sistematización y cuestionamiento de la información obtenida)	77
	CONCLUSIONES	79
	RECOMENDACIONES	80
	BIBLIOGRAFÍA	
	ANEXOS	

ÍNDICE DE TABLAS

Tabla 3-1 : Operacionalización de variables.	23
Tabla 4-1 : Nombre de las Estaciones Meteorológicas.	27
Tabla 4-2 : Puntos con los años faltantes..	28
Tabla 4-3 : Estadísticas descriptivas datos SAM.	35
Tabla 4-4 : Estadísticas descriptivas datos Estaciones.	37
Tabla 4-5 : Estadísticas descriptivas datos Estaciones.	39
Tabla 4-6 : Comparación de los modelos de semivariograma exponencial, gaussiano y esférico.	52
Tabla 4-7 : Comparación de los modelos de semivariograma exponencial, gaussiano y esférico.	53
Tabla 4-8 : Comparación de los modelos de semivariograma exponencial, gaussiano y esférico.	58
Tabla 4-9 : Comparación de los modelos de semivariograma exponencial, gaussiano y esférico.	59
Tabla 4-10 : Comparación con los datos del SAM.	63
Tabla 4-11 : Comparación con los datos de las Estaciones.	65
Tabla 4-12 : Cuadro comparativo de Temperaturas.	74

ÍNDICE DE ECUACIONES

Ecuación 2-1 : Error Residual	9
Ecuación 2-2 : Error Cuadrático Medio (RSME)	9
Ecuación 2-3 : Error Absoluto Medio (MAE)	9
Ecuación 2-4 : Distancia Inversa Ponderada (IDW)	12
Ecuación 2-5 : Distancia inversa entre el punto “i” respecto a los demás puntos “j”	12
Ecuación 2-6 : Interpolación Kriging	15
Ecuación 2-7 : Modelo Esférico	17
Ecuación 2-8 : Condición	17
Ecuación 2-9 : Modelo Gaussiano	18
Ecuación 2-10 : Condición	18
Ecuación 2-11 : Modelo Exponencia	18
Ecuación 2-12 : Condición	18

ÍNDICE DE ILUSTRACIONES

Ilustración 2-1 : Distancia Inversa Ponderada.....	13
Ilustración 2-2 : Spline.....	14
Ilustración 2-3 : Kriging.	15
Ilustración 2-4 : Principales elementos del variograma.	16
Ilustración 2-5 : Modelo de Semivarianza Esférico.....	17
Ilustración 2-6 : Modelo de Semivarianza Gaussiano.....	17
Ilustración 2-7 : Modelo de Semivarianza Exponencial.....	18
Ilustración 3-1 : Desarrollo de la investigación.	19
Ilustración 3-2 : Ubicación en la Epoch.	20
Ilustración 3-3 : Localización de las 11 estaciones meteorológicas en la provincia de Chimborazo.	21
Ilustración 3-4 : Tabla con los nombres y atributos de las 11 estaciones meteorológicas...	21
Ilustración 3-5 : Malla de intersección de la Provincia de Chimborazo.	22
Ilustración 4-1 : Ingreso de coordenadas UTM en el SAM.....	26
Ilustración 4-2 : Comprobación de puntos en Google maps.	26
Ilustración 4-3 : Datos rellenados para el Punto 61-2016.	29
Ilustración 4-4 : Diagrama de caja temperatura 2016-2020.....	30
Ilustración 4-5 : Diagrama de caja datos estaciones temperatura 2014-2020.	30
Ilustración 4-6 : Histograma y curva de la normal temperatura 1998-2001.....	31
Ilustración 4-7 : Histograma y curva de la normal temperatura 2010-2013.....	32
Ilustración 4-8 : Histograma y curva de la normal temperatura 2017-2020.....	32
Ilustración 4-9 : Gráfico de normalidad QQplot 1998-2001.	33
Ilustración 4-10 : Gráfico de normalidad QQplot 2010-2013.....	34
Ilustración 4-11 : Grafico de normalidad QQplot 2017-2020.	34
Ilustración 4-12 : Histograma y curva de la normal 2017-2020.....	36
Ilustración 4-13 : Grafica de normalidad QQplot 2017-2020.	37
Ilustración 4-14 : Diagrama de caja y gráfico de normalidad de la temperatura 2015.....	38
Ilustración 4-15 : Análisis Exploratorio de datos del SAM.....	40

Ilustración 4-16 : Análisis Exploratorio de datos de las Estaciones.....	41
Ilustración 4-17 : Transformación realizada a los datos del SAM.	42
Ilustración 4-18 : (a) Análisis de Tendencia del SAM año 2015 y (b) Análisis de Tendencia de las estaciones año 2015	43
Ilustración 4-19 : Mapa de temperatura en relación con la altura de la provincia de Chimborazo datos SAM 2015.....	44
Ilustración 4-20 : Mapa de temperatura en relación con la altura de la provincia de Chimborazo datos SAM 2020.....	45
Ilustración 4-21 : Mapa de temperatura en relación con la altura de la provincia de Chimborazo Datos Estaciones 2015	46
Ilustración 4-22 : Mapa de temperatura en relación con la altura de la provincia de Chimborazo datos Estaciones 2020.....	48
Ilustración 4-23 : Semivariograma del SAM año 2015.....	49
Ilustración 4-24 : Semivariograma de las Estaciones del año 2015.....	49
Ilustración 4-25 : (a) Localización de las estaciones Meteorológicas y (b) Puntos descargados del SAM de la provincia de Chimborazo.....	50
Ilustración 4-26 : Identificación de los datos de entrenamiento y prueba SAM-2015.....	51
Ilustración 4-27 : Ajuste de los datos del SAM a tres tipos de modelos: (a) exponencial, (b) gaussiano y (c) esférico.....	53
Ilustración 4-28 : Ajuste de los datos del SAM a tres tipos de modelos: (a) exponencial, (b) gaussiano y (c) esférico.....	54
Ilustración 4-29 : Interpolación de IDW con datos de entrenamiento SAM 2015.	55
Ilustración 4-30 : Interpolación con Spline de los datos del SAM 2015.....	56
Ilustración 4-31 : Estaciones meteorológicas de entrenamiento y prueba 2015.	57
Ilustración 4-32 : Ajuste de los datos de las estaciones meteorológicas a tres tipos de modelos: (a) exponencial, (b) gaussiano y (c) esférico.	59
Ilustración 4-33 : Tres tipos de modelos de semivariograma ajustados a los datos: (a) exponencial, (b) gaussiano y (c) esférico.	60
Ilustración 4-34 : Interpolación IDW con los datos de las estaciones meteorológicas 2015.	61
Ilustración 4-35 : Interpolación Spline con los datos de las estaciones meteorológicas 2015.....	62
Ilustración 4-36 : Interpolación de Kriging con un semivariograma Exponencial de los datos del SAM 2015.....	64
Ilustración 4-37 : Interpolación de kriging con un semivariograma Gaussiano para los datos de las estaciones.....	66

Ilustración 4-38 : Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2015.	68
Ilustración 4-39 : Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2016.	69
Ilustración 4-40 : Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2017.	70
Ilustración 4-41 : Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2018.	71
Ilustración 4-42 : Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2019.	72
Ilustración 4-43 : Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2020.	73
Ilustración 4-44 : Interpolación de Kriging 2015-2018.	75
Ilustración 4-45 : Interpolación de Kriging 2019-2020.	76

ÍNDICE DE ANEXOS

ANEXO A: CÓDIGOS UTILIZADOS EN EL SOFTWARE ESTADÍSTICO RSTUDIO

ANEXO B: AVAL DE LA INVESTIGACIÓN

RESUMEN

La interpolación es esencial para estimar datos en áreas sin información. Este estudio aborda la escasez de datos en zonas sin estaciones meteorológicas en la provincia de Chimborazo, lo cual dificulta entender el comportamiento de la variable temperatura. Por ende, el objetivo del presente trabajo de titulación fue comparar técnicas de interpolación espacial de datos de temperatura ambiente de los modelos atmosféricos CMIP6 y de las estaciones meteorológicas. La metodología fue cuantitativa, con un diseño no experimental de tipo transversal. Se descargaron datos del SAM (523 puntos) periodo 1998-2020, detectando y completando información faltante con missForest en RStudio. Por otra parte, los datos de temperatura de las estaciones meteorológicas periodo 2014-2020 fueron proporcionados por el Grupo de Energías Alternativas y Ambiente (GEAA). Se empleo un muestreo por conveniencia, determinando que el año 2015 muestra mayor confiabilidad y precisión debido a la menor cantidad de datos faltantes. Las estadísticas muestran un incremento en la media y la mediana a lo largo de tiempo. Las pruebas de normalidad de Kolmogorov-Smirnov y Shapiro-Wilks, indicaron que los datos del SAM no siguen una distribución normal a diferencia de los datos de las estaciones. Luego se implementaron las técnicas de interpolación IDW, Spline y Kriging, y se evaluaron usando métricas de validación. El método que mejor se ajustó y presentó menores errores fue Kriging, para los datos del SAM el Kriging Universal con semivariograma exponencial y tendencia constante presentó un EM de -0.01, SEM de 1.64 y RSME de 1.10, mientras que para los datos de las estaciones meteorológicas el método de Kriging Ordinario con semivariograma gaussiano y tendencia constante presentó un EM de -0.16, SEM de 2.74 y RSME de 2.93, el estudio demuestra que la interpolación Kriging es la más efectiva, ya que utiliza la correlación espacial y el análisis de tendencia.

Palabras clave: <INTERPOLACIÓN ESPACIAL>, <SPLINE>, <KRIGING>, <SAM>, <DISTANCIA INVERSA PONDERADA (IDW)>, <MODELOS ATMOSFÉRICOS>.

0781-DBRA-UPT-2024



SUMMARY

Interpolation is essential to estimate data in areas without information. This study addresses the scarcity of data in areas without meteorological stations in the province of Chimborazo, which makes it difficult to understand the behavior of the temperature variable. Therefore, the objective of this degree work was to compare spatial interpolation techniques of ambient temperature data from CMIP6 atmospheric models and weather stations. The methodology was quantitative, with a non-experimental cross-sectional design. Data were downloaded from the SAM (523 points) period 1998-2020, detecting and completing missing information with missForest in RStudio. On the other hand, temperature data from meteorological stations period 2014-2020 were provided by the Alternative Energies and Environment Group (GEAA). Convenience sampling was used, determining that 2015 shows higher reliability and accuracy due to fewer missing data. Statistics show an increase in the mean and median over time. Kolmogorov-Smirnov and Shapiro-Wilks normality tests, indicated that the MAR data do not follow a normal distribution unlike the station data. IDW, Spline and Kriging interpolation techniques were then implemented and evaluated using validation metrics. The best fitting method that presented lower errors was Kriging, for the SAM data the Universal Kriging with exponential semivariogram and constant trend presented an EM of -0.01, SEM of 1.64 and RSME of 1.10, while for weather station data the Ordinary Kriging method with Gaussian semivariogram and constant trend presented an EM of -0.16, SEM of 2.74 and RSME of 2.93, the study shows that the Kriging interpolation is the most effective, since it uses spatial correlation and trend analysis.

Keywords: <SPATIAL INTERPOLATION>, <SPLINE>, <KRIGING>, <SAM>, <INVERSE DISTANCE WEIGHTED (IDW)>, <ATMOSPHERIC MODELS>.



Edgar Mesias Jaramillo Moyano

0603497397

INTRODUCCIÓN

En el año 2015 la agencia National Oceanic and Atmospheric Administration (NOAA) ha indicado que la temperatura media global ha variado $1,33^{\circ}\text{C}$ respecto a la media del siglo pasado. De la misma manera, según Aguilar la temperatura media mundial de los océanos se ha incrementado $0,74^{\circ}\text{C}$, siendo esto corroborado por Tim Wheeler mediante varios estudios realizados a lo largo del tiempo (Aguilar, 2020, p.3).

La Organización Meteorológica Mundial (OMM) ha establecido que las variables meteorológicas que definen el comportamiento climático de una región se pueden describir mediante: temperatura, precipitación, dirección del viento, velocidad del viento y radiación solar. La disponibilidad de información climatológica durante un periodo prologado de tiempo es crucial para mejorar los pronósticos y toma de decisiones. Por lo tanto, estos datos ayudan analizar fenómenos como el cambio climático, permitiendo la formulación de medidas para mitigar los impactos negativos (Martínez, Paredes y Valdez Peña 2022).

De acuerdo con los hallazgos realizados por (Reyes 2017, p.2), los métodos de interpolación espacial probabilísticos como el Kriging Ordinario utilizando semivariogramas Exponencial y Gaussiano proporcionan resultados óptimos para las variables precipitación y temperatura dentro de la provincia del Azuay. Asimismo (Andrades et al. 2020), reporta que los métodos de interpolación con criterios estadísticos determina la autocorrelación espacial y tendencia a través de un semivariograma, estos modelos Esférico, Exponencial y Gaussiano se ajustan mejor dependiendo de la distribución de los datos (Mohd Aziz et al. 2019, p.5).

En la investigación realizada por (Velásquez et al. 2018) ,para interpolar datos de precipitación y temperatura en las regiones Andina, Pacífica y Caribe de la república de Colombia, se realizó un geoprocesamiento. Posteriormente se aplicó el método Distancia Inversa Ponderada (IDW) obteniendo superficies interpoladas para 258 estaciones meteorológicas en la variable temperatura.

Se ha estudiado el comportamiento climático del Ecuador mediante las componentes: temperatura y precipitación, descubriendo variaciones temporales y espaciales. En la región Andina se ha determinado dos estaciones: días húmedos y días secos, indicando cambios de temperatura paulatinamente a través del tiempo, mientras que en la Amazonía los días húmedos son considerados durante la mayor parte del año (Pérez, Mullo y Marcatoma 2020, p.2).

La limitada disponibilidad de información meteorológica espacial y temporal en la provincia de Chimborazo es un problema debido a la falta de recursos económicos, técnicos, logísticos y administrativos para el mantenimiento y la instalación de las estaciones meteorológicas automáticas. Además, uno de los principales problemas en el análisis de fenómenos meteorológicos es la escasez de información en puntos no muestreados. En este contexto los métodos de interpolación determinísticos y probabilísticos ayudan a estimar datos en regiones faltantes con la influencia de estaciones meteorológicas cercanas (Reyes 2017), (Martínez, Paredes y Valdez Peña 2022).

Es importante indicar que la interpolación espacial es un cálculo matemático, que permite estimar valores desconocidos a partir de valores conocidos en el espacio, partiendo de dos coordenadas X e Y, generando la tercera coordenada Z y dando como resultado una función de probabilidad con valores proyectados. Por lo tanto, los métodos de interpolación son: deterministas y geoestadísticos, en donde no existe preferencia sobre qué método utilizar, ya esto se debe a la gran cantidad de criterios de selección (Beltran 2020).

De manera general, el método de interpolación más utilizado es el IDW que se define como el promedio ponderado de los valores de ubicación adyacentes a la ubicación prevista (Diaz et al. 2023). Mientras que el Spline suaviza una superficie continua al reducir al mínimo su curvatura (Mayorga et al. 2022), utilizando métodos de regresión matemática, en los cuales prevalece el ajuste de los mínimos cuadrados (Chang 2019). Así mismo, el método de Kriging utiliza semivariogramas para cuantificar la estructura espacial de los datos, ya que los valores más cercanos a un punto conocido tienen mayor peso y este disminuye a medida que se aleja del punto de interés (Chang 2019).

Es primordial tener en cuenta que los modelos de predicción espacial mencionados anteriormente como IDW, Kriging y Spline, buscan aproximar superficies continuas a partir de un número discreto de puntos de medición. Existe controversia respecto al mejor método, por lo que se debe analizar las variantes de los diferentes modelos, la distribución espacial de las observaciones, las particularidades del campo de estudio, tanto la naturaleza como el periodo temporal y espacial de resolución de las variables atmosféricas. (Velásquez et al. 2018).

En resumen, la investigación aborda la comparación de métodos de interpolación espacial Distancia Inversa Ponderada (IDW), Spline, Kriging Universal y Ordinario, con el propósito de identificar el modelo más efectivo. Para este estudio se utilizan datos de la variable temperatura de las 11 estaciones meteorológicas en un periodo 2014-2020. También se obtuvo datos de 523 puntos dentro de la provincia de Chimborazo del System Advisor Model (SAM) parte de los modelos CMIP6, desde 1998-2020.

CAPÍTULO I

1. PROBLEMA DE INVESTIGACIÓN

1.1. Antecedentes

En la antigüedad una forma de representar información física natural era mediante métodos cartográficos manual. Los mapas o cartas analógicas eran portadores de datos y variables geográficas, con capacidades limitadas de procesamiento de información. Esto conduce a problemas en la gestión, análisis y actualización de la información espacial. Por ello, a partir de 1960, se desarrolló un nuevo método de procesamiento de información geográfica con la ayuda de programas informáticos. Fue así como los sistemas de información geográfica surgieron paulatinamente con el avance tecnológico (Chang 2019).

Los mapas climáticos en diferentes dominios presentan una serie de problemas relacionados con la expansión espacial de los datos. Esto se debe a la baja densidad de estaciones meteorológicas y a la falta de información meteorológica, lo que supone una limitación importante y en algunos casos decisiva para las actividades agrícolas y medioambientales. Este problema es aún más grave en los países del Tercer Mundo, donde las redes de estaciones de observación son escasas y es imposible realizar un mapa adecuado (Chang 2019).

Uno de los problemas de la cartografía es la relación con el carácter puntal de los datos y la heterogeneidad de la distribución de estaciones meteorológicas. Y una de las mejores soluciones son los métodos de interpolación. Básicamente, la interpolación es una técnica estadística (matemática) que se utiliza para obtener información de puntos no muestreados. Sin embargo, aún no se ha identificado un enfoque general debido a factores como el tamaño del área de estudio, las variables topográficas y la escala temporal considerada (diaria, mensual, anual) (Cifuentes Carvajal 2016). Las técnicas de interpolación espacial para estimar variables climáticas, en un sistema bidimensional algunos patrones no son detectados, por ello, se ha encontrado en la bibliografía científica diferentes técnicas para representar el comportamiento climático, la resolución temporal de los datos, el espacio variable adicional y si hay información útil adicional que se puede incluir en el proceso de interpolación (Berndt y Haberlandt 2018).

1.2. Planteamiento del Problema

La pregunta del estudio surge debido a que el manejo y mantenimiento de las estaciones meteorológicas puede utilizar muchos recursos económicos. De la misma forma, sobre la falta de datos en las zonas donde no existen estaciones meteorológicas en la provincia de Chimborazo. Se han encontrado diferentes técnicas para abordar este problema: Distancias Ponderadas Inversas (IDW), Spline, y Kriging. Los resultados que ofrecen estos métodos pueden variar, si la técnica brinda estimaciones que no difieren de la variable observada se denominan técnica precisa, de otra manera, técnica imprecisa.

El clima en Ecuador está influenciado por la latitud, longitud, así también como el relieve que dan origen a las regiones naturales (costa, sierra y oriente). De la misma forma la Cordillera de los Andes se presenta como una barrera ante la presencia de corriente oceánicas en la Región Andina (Pourrut, 1987). Estas características se ven reflejadas en el comportamiento de la temperatura y la precipitación que presentan variaciones en las diferentes regiones del país. La distribución temporal de las precipitaciones en el Ecuador da lugar a dos estaciones bien diferenciadas, una lluviosa y otra seca, con excepción de la Amazonía en donde llueve todo el año (Pérez, Mullo y Marcatoma 2020).

La temperatura tiene una relación con la altitud, según el Instituto Geofísico del Ecuador, en la provincia de Chimborazo la altura varía desde los 135 m.s.n.m hasta los 6310 m.s.n.m que es la cumbre del volcán Chimborazo, un verdadero termómetro del clima (Rivera Vásquez 2020). Por lo tanto, comprender el comportamiento de la variable temperatura ambiente es complejo debido a la baja densidad de observaciones, este estudio busca comparar diferentes técnicas de interpolación espacial, de la variable temperatura ambiente de los modelos atmosféricos CMIP6 y de las estaciones meteorológicas.

1.3. Problema general de investigación

¿Qué tan eficientes son las técnicas de interpolación espacial de la variable temperatura ambiente?

1.4. Problemas específicos de investigación

- ¿Existe una diferencia significativa entre las técnicas de interpolación espacial de la variable temperatura ambiente?
- ¿Cuál de las técnicas de interpolación espacial es la más precisa?

1.5. Objetivos

1.5.1. *Objetivo general*

- Comparar las técnicas de interpolación espacial de datos de temperatura ambiente de los modelos atmosféricos CMIP6 y las estaciones meteorológicas.

1.5.2. *Objetivos específicos*

- Realizar un análisis estadístico descriptivo de los datos de la variable temperatura.
- Seleccionar las técnicas de interpolación espacial encontradas en la bibliografía para la variable temperatura según la estimación del error.
- Comparar las técnicas de interpolación espacial en base a las métricas estadísticas.
- Implementar la mejor técnica de interpolación espacial en el conjunto de datos de temperatura.

1.6. Justificación

1.6.1. *Justificación Teórica*

La necesidad de este trabajo de integración curricular surge respecto a la variable temperatura ambiente, teniendo en cuenta que el comportamiento de esta variable es complejo. No se ha realizado un análisis comparativo de dicha variable que afirme específicamente que técnica de interpolación espacial ofrece resultados de zonas no cubiertas a partir de zonas cubiertas.

1.6.2. *Justificación Metodológica*

Para la presente investigación, la información de la variable meteorológica se descargará del *System Advisor Model* (SAM) en 523 puntos, estos puntos hacen referencia a la distribución espacial de la provincia del Chimborazo, con una distancia de 4 km respectivamente entre cada punto considerando que cada uno posee información desde el año de 1998 al 2020. Los diferentes métodos de interpolación espacial serán implementados en el software de ARCGIS.

1.6.3. *Justificación Práctica*

El presente trabajo de Integración Curricular tiene como finalidad comparar diferentes técnicas de interpolación espacial de la variable Temperatura de la red de estaciones meteorológicas del Grupo de Investigación de Energías Alternativas y Ambiente (GEAA) de la ESPOCH mediante los métodos determinísticos Distancia Inversa Ponderada (IDW) y Spline, y probabilísticos Kriging. Los resultados de interpolación espacial dentro de la provincia de Chimborazo para la variable temperatura beneficiaran al GEAA y algunas otras instituciones y organizaciones relacionadas con el ámbito meteorológico.

CAPÍTULO II

2. MARCO TEÓRICO

2.1. Bases teóricas

2.1.1. *Meteorología*

Es la ciencia del clima que se enfoca en el estudio de la atmósfera y su estructura, patrones eventos y fenómenos, su interés es comprender cómo cambia el clima, la temperatura, la humedad, la energía eólica, el viento, la energía que pueden cambiar a corto plazo (Arizala y Colt 2020).

2.1.2. *¿Qué es Temperatura?*

Para definir la temperatura, primero debemos entender el término calor; El calor es el resultado de la energía liberada por el movimiento de las partículas que componen toda la materia. Estas partículas crecen más rápido y con más violencia, lo que hace que el calor aumente a una temperatura que mide el calor de todo el material (Arizala y Colt 2020).

Cantidad física en el estudio de la termodinámica, la temperatura también se puede confundir con otro término denominado calor. Coloquialmente, la temperatura es una medida de cuán “fría” o “caliente” es una sustancia, por ello, se deduce que cuanto más “calor” tiene un objeto, mayor es su temperatura, y viceversa. Sin embargo, dado que muchas propiedades de las sustancias cambian a medida que varía su temperatura, es relativamente fácil medirlas para crear un concepto claro y razonable de ellas (Leal 2020).

2.1.3. *Geoestadística*

La geoestadística es una rama de la estadística especializada en el análisis y modelado de la variabilidad espacial de fenómenos que tienen un componente espacial, es decir, que se pueden obtener utilizando coordenadas de datos vinculados en ubicaciones específicas. Esta rama se centra en el análisis y modelado de variables de información espacial a partir de un conjunto de métodos destinados a estimar valores desconocidos a partir de valores conocidos (Beltran 2020).

En general, esta información se procesa a través de medidas descriptivas de: tendencia central, localización, dispersión o variabilidad, análisis de correlación y regresión. Las geoestadísticas para

el análisis de variables utilizan métodos de interpolación que identifican métodos específicos para cada zona en función de la variable de interés estudiada, lo que les permite mejorar la representación espacial de las variables climáticas (Reyes 2017).

2.1.4. Sistema de información geográfica (SIG)

El Instituto de Geografía Militar define SIG como un sistema de información geográfica a los resultados de aplicar tecnologías de la información (TI) hasta la gestión de la información geográfica (GI), que se está convirtiendo en una herramienta útil en este sentido. Además, la posibilidad de preparar la información de forma didáctica mejora el análisis de resultados (Reyes 2017).

2.1.5. Validación Cruzada

La validación cruzada es una técnica de validación de modelos utilizada para evaluar lo que hace el algoritmo. El análisis estadístico se puede generalizar a un conjunto de datos independiente el cual se realiza con la ayuda de particiones de un conjunto de datos de la siguiente manera: un subconjunto para entrenar el algoritmo y el resto de los datos para la prueba. Cada ronda de validación cruzada implica clasificar aleatoriamente el conjunto de datos original en un conjunto de entrenamiento y un conjunto de prueba (Rincon, D’Giano y Risk 2018).

2.1.6. Validación del método de interpolación

Existe una variedad de técnicas para determinar la confiabilidad de los métodos utilizados y la precisión para la toma de decisiones. El más utilizado es el de la validación cruzada, consistente en extraer uno de los valores de la variable en estudio y calcular su semivariograma de los datos, en donde con el valor extraído se evalúa y se estudia de igual forma (Reyes 2017).

Una buena solución para aliviar muchos de estos problemas es usar la validación cruzada que se utiliza en todo tipo de investigación estadística para probar la confiabilidad de un grupo de prueba antes de aplicar el análisis a un grupo problemático. Básicamente, es una forma de establecer opciones, modificarlas y usarlas más tarde para configuraciones desconocidas. Evitaremos realizar alteraciones al cambiar parámetros para que nuestros textos sin errores se clasifiquen correctamente y aprendemos qué opciones son las mejores para poder implementarlas sin cambiar nuestro conjunto de problemas.

Un estudio de Adisoma (1993) encontró que el uso del método Jackknife complementa el uso tradicional del kriging. Esto se debe a que se ha demostrado que supera a otras técnicas (Reyes 2017).

2.1.7. Validación del los errores

Al evaluar los métodos de interpolación aplicados en la variable en estudio y además sugerir el método que dé el menor error de estimación. Se utilizan ecuaciones matemáticas para cuantificar y validar un método en específico (Chang 2019).

2.1.7.1. Error Residual

Es la diferencia entre los valores observados y valores predichos (Chang 2019).

$$ER = (OBSERVADOS - PREDICHOS) \quad (2-1)$$

2.1.7.2. Error Cuadrático Medio (RSME)

El error cuadrático medio es la diferencia entre el valor predicho y el valor real. Estas diferencias se deben elevar al cuadrado, calcular el promedio de todos los valores y obtener la raíz cuadrada. RMSE mide el tamaño del error (Chang 2019).

$$RMSE = (N^{-1} \sum_{i=1}^N (P_i - O_i)^2)^{0.5} \quad (2-2)$$

Donde:

P_i =valor predicho

O_i =valor real

2.1.7.3. Error Absoluto Medio (MAE)

El error absoluto medio determina qué tan cerca están las predicciones de los datos observados y es el promedio de la diferencia absoluta entre los valores predichos y observados (Chang 2019).

$$MAE = N^{-1} \sum_{i=1}^N |P_i - O_i| \quad (2-3)$$

P_i =valor predicho

O_i =valor observado

2.1.8. Coordenadas UTM

El sistema de coordenadas Universal Transverse Mercator (UTM) es un sistema que se obtiene especificando la proyección cartográfica Transverse Mercator (TM). El sistema UTM se originó

en la década de 1940 desarrollado por el ejército de los EE. UU. Satisfaciendo las necesidades de pronóstico global con fines militares. Los requisitos de proyección son minimizar la distorsión del azimut y mantener la distorsión de escala dentro de los límites especificados. La proyección cartográfica Transverse Mercator (TM) es una representación conforme de un elipsoide y es superior a otros tipos de proyecciones cartográficas utilizadas, con la excepción de la proyección conforme utilizada para representar regiones polares, que utiliza la proyección estereográfica polar (Frau et al. 2012).

2.1.9. *System Advisor Model (SAM)*

El Modelo Solar Advisor (SAM) es un modelo financiero y de rendimiento diseñado para ayudar a los investigadores en la toma de decisiones en el campo de las energías renovables (es decir, eólica y solar). desarrollado originalmente por el Laboratorio Nacional de Energía Renovable en colaboración con otros laboratorios en 2005. El Laboratorio Nacional de Energía Renovable (NREL) publicó su primera versión en agosto del 2007 y para el año 2010 el software pasó a llamarse "Modelo System Advisor" para reflejar la inclusión de tecnologías fuera del sistema solar (Shahzad 2022).

2.1.10. *Análisis Geoestadístico*

La aplicación de la geoestadística es una herramienta muy importante para el análisis de datos gráficos. La evaluación de la ubicación, la variabilidad, la forma, la distribución y el cálculo de las métricas puede determinar si ciertos supuestos necesarios para aplicar la teoría geoestadística son correctos o determinar qué procedimientos de pronóstico son más útiles (Cifuentes Carvajal 2016).

2.1.11. *MissForest.*

El método de missForest, que es una técnica de imputación no paramétrica diseñada específicamente para abordar datos de tipo mixto, corresponde a un método que trata de manera simultánea tanto para variables continuas como categóricas. En donde el algoritmo realiza un proceso repetitivo para llenar los valores faltantes hasta que se alcance un criterio de finalización, aplicando la menor cantidad de iteraciones respecto a los datos (Rozo Ángel 2023, pág.7).

Uno de los problemas que se presentan durante el análisis de información meteorológica es la ausencia de datos ya sea en el SAM que es desarrollado por el Laboratorio Nacional de Energías Renovables (NREL) o de las estaciones meteorológicas existentes en las áreas de estudio, ya que al ser recolectados y guardados automáticamente por algún sistema automatizado, provocando fallos

ya sea por la falta de energía o por algún tipo de configuración del mecanismo (Reyes 2017).

2.2. Bases Conceptuales

Se han utilizado diversos modelos de estimación espacial como distancia inversa ponderada (IDW), Kriging ordinario y universal los mismos que permiten generar superficies continuas de datos climáticos a partir de un número discreto de puntos de medición. Estos métodos tratan de generar mapas climáticos en zonas de interés determinadas a través de mediciones en puntos específicos, incluyendo en algunos casos variables topográficas o geográficas (Beltran, 2020).

- **Deterministas:** Crea superficies que son continuas a través de un grado de similitud o suavidad. Esta categoría incluye: métodos como el IDW y Spline (Beltran, 2020).
- **Geoestadísticos:** Crea una superficie continúa basada en las propiedades estadísticas de los datos originarios utilizando métodos como Kriging (Beltran, 2020).

2.2.1. Interpolación espacial

La interpolación espacial es un cálculo matemático que te permite calcular valores desconocidos los cuales se pueden calcular en el espacio dibujando una línea, curva o área a partir del valor y el cual se debe conocer, siempre que esté dentro de los límites de los puntos de medición. Entonces se debe saber que existen varios tipos de interpolaciones que se dividen en dos grandes grupos, deterministas y probabilísticos (Vicencio, 2018).

La interpolación es una técnica de geoestadística basada en los principios de autocorrelación y dependencia espacial, que predice valores de celda en las ubicaciones de puntos no muestreados y mide el grado de correlación entre objetos cercanos y lejanos. Esto indica que los objetos o eventos cercanos tienen propiedades similares y que es menos probable que los objetos distantes tengan propiedades similares, esto se llama correlación espacial y es una base importante de los métodos de interpolación.

Actualmente se conocen diferentes métodos de interpolación con una serie de usos y aplicaciones específicas que profundizan en función de la naturaleza de los datos y el interés del estudio. Los más populares son la Ponderación de Distancia Inversa (IDW), Spline, Kriging, etc (Pedreros, 2021).

2.3. Métodos Determinísticos

Las técnicas deterministas estiman los valores de las variables cuyas entradas producen ciertas salidas únicas. Su enfoque se basa en la relación entre variables independientes para un comportamiento dado y una o más variables dependientes (modelización de la realidad), es decir, cuantas más variables involucradas, más complejo se vuelve el proceso de estimación (Chang, 2019).

2.3.1. Distancia Inversa Ponderada (IDW)

Con la interpolación IDW, los puntos de demostración se ponderan en el proceso de interpolación de modo que la influencia de un punto en relación con otros puntos disminuye a medida que aumenta la distancia desde el punto desconocido creado (Beltran, 2020).

También determina los valores en diferentes celdas utilizando secuencias de puntos de muestra. Este peso es una función de la distancia entre el punto de entrada y la ubicación de las bandejas de salida, es decir, cuanto mayor es la distancia, menos efecto tiene la celda en el valor de salida (Pedreros, 2021).

Este método detecta que la similitud es mayor que el valor de los puntos que están cerca entre sí y puede reducirse, pero la distancia entre diferentes puntos conocidos aumenta. Luego, para estimar el valor 'Z' del punto desconocido, calculamos el estadístico promedio ponderado de los puntos circundantes, que se obtiene como la distancia correspondiente entre cada punto conocido y el punto desconocido (Chang, 2019).

Expresión matemática:

$$Z_i = \frac{\sum_j (Z_j * W_{ij})}{\sum_j W_{ij}} \quad (2-4)$$

$$W_{ij} = \frac{1}{D_{ij}^a} \quad (2-5)$$

Donde:

Z_i = Valor estimado en el punto "i", media ponderada de valores conocidos.

Z_j = Valores conocidos de la variable en todos los puntos "j".

W_{ij} Distancia inversa entre el punto "i" respecto a los demás puntos "j".

D_{ij}^a Distancia entre el punto "i" y los demás puntos "j" conocido.

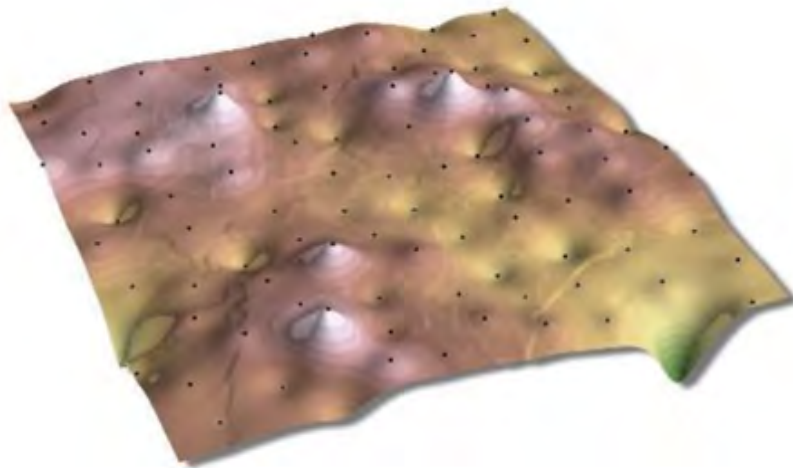


Ilustración 2–1: Distancia Inversa Ponderada.

Fuente: Universidad Nacional de Colombia.

Realizado por: (Pedreros 2021).

2.3.2. Interpolación Spline

El Spline emplea técnicas de interpolación para estimar valores mediante funciones matemáticas minimizando la curvatura general de una superficie, dando como resultado una superficie lisa que atraviesa la entrada al punto de interés (Atehortúa 2021).

Esta es una técnica determinista donde la superficie Z se puede ajustar usando métodos de regresión matemática, basados en una variable conocida (Z) y se adapta un ajuste de mínimos cuadrados, es decir, el valor con la menor desviación o variación. Los métodos de regresión calculan múltiples parámetros ajustados por ecuaciones lineales, cuadráticas, cúbicas o de orden superior. Sin embargo, la confianza de ajuste siempre se deriva de la suma de los valores conocidos en cada punto disponible. Otro parámetro para considerar es el coeficiente de acoplamiento, pues cuanto más cerca esté de uno (1), mejor será la dirección y menor la desviación (Chang, 2019).

Esta interpolación además de predecir picos o valles utiliza funciones matemáticas que ayudan a minimizar la curvatura de la superficie; el resultado es una superficie lisa que se extiende exactamente sobre los puntos de entrada. Además, de que predice picos o valles, es la mejor manera de representar las superficies de fenómenos ligeramente variables como la temperatura (Pedreros, 2021).

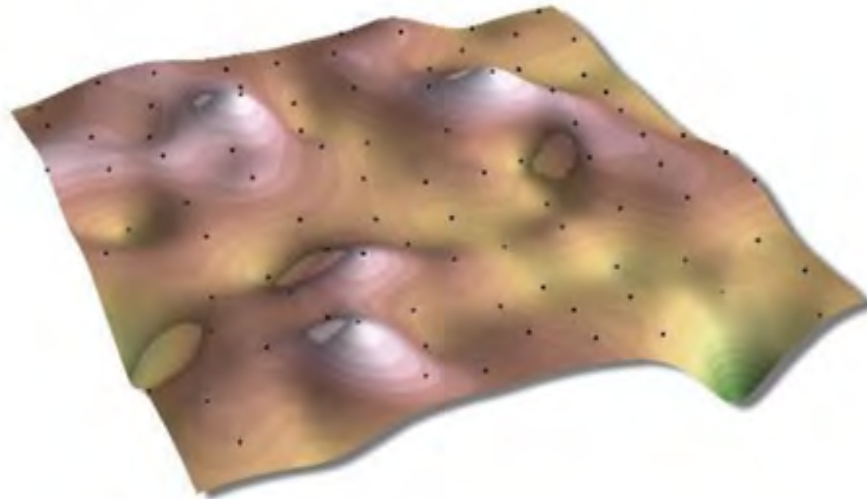


Ilustración 2–2: Spline.

Fuente: Universidad Nacional de Colombia.

Realizado por: (Pedreros 2021).

2.4. Métodos Probabilísticos

Los métodos probabilísticos proporcionan estimaciones probabilísticas de la calidad del ajuste en función de la correlación espacial de los datos geográficos. Esto permite pronosticar para: áreas o volúmenes más grandes e incluir datos en el cálculo para verificar la precisión del ajuste (Chang, 2019).

2.4.1. *Interpolación Kriging*

Es un método geoestadístico para producir una ubicación aproximada de un rango de puntos con valores altos. La base de este método es la información sobre la naturaleza espacial del evento en relación con su altura, y de esta manera se puede ver si se refleja más en la variable que se está estudiando. El método Kriging se basa en modelos que usan autocorrelación, con este método se puede proporcionar la precisión del pronóstico promedio excluyendo el pronóstico. Kriging se utiliza principalmente cuando cada elemento de los datos tiene un borde de ataque o una longitud (Atehortúa 2021).

Kriging asumió que la distancia entre los puntos de muestra reflejaba una relación espacial que podría usarse para describir la dinámica de la superficie, ajustando los recursos a una cantidad específica de puntos o a todos los puntos dentro de un radio específico determinando el valor de salida de cada ubicación (Pedreros, 2021).

El método Kriging es similar a IDW, puesto que pondera los valores de umbral medidos para calcular una predicción de ubicación desconocida. Al tener una fórmula general de ambos interpoladores se determina como una suma ponderada de los datos (Beltran 2020).

Expresión matemática:

$$Z(s_o) = \sum_{i=1}^N \lambda_i Z(S_i) \quad (2-6)$$

Donde:

$Z(S_i)$ = El valor medido en la ubicación n.º i.

λ_i = Un peso desconocido para el valor medido en la ubicación n.º i.

s_o = La ubicación de la predicción y.

N = El número de valores medidos.

En kriging ordinario, el peso, λ_i , depende de qué tan bien se ajuste el modelo a los puntos de medición, la distancia desde la ubicación esperada y la relación espacial entre las mediciones alrededor de la ubicación esperada (Beltran 2020).

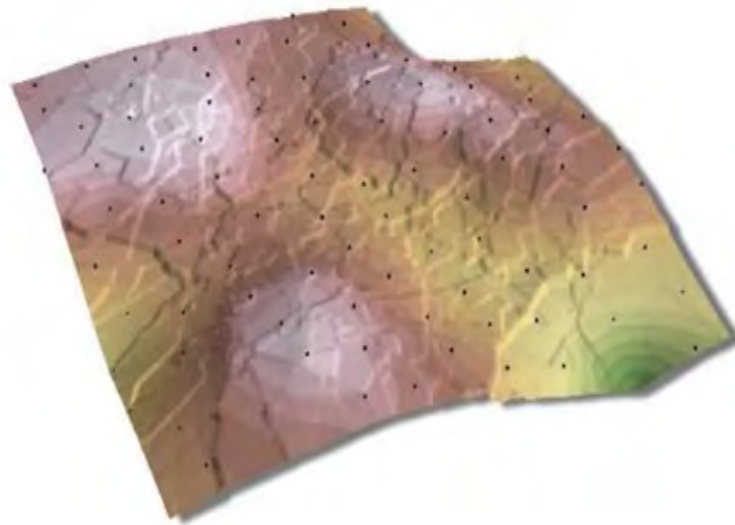


Ilustración 2–3: Kriging.

Fuente: Universidad Nacional de Colombia.

Realizado por: (Pedreros 2021).

2.4.2. *Semivariograma*

Un variograma o semivariograma es una herramienta geográfica importante para resumir información, ya que, es una función creciente que alcanza un límite llamado meseta igual a la desviación de muestreo. Se alcanza una meseta en un valor de h , llamado rango, que define

la banda de influencia alrededor del punto donde la autocorrelación es cero. En el análisis geoestadístico, primero se construye un semivariograma experimental y, a través de él, se construye un semivariograma teórico (modelo adjunto) para obtener una función matemática exacta (Reyes 2017).

Parámetros definidos en los semivariogramas

Nugget: También conocido como efecto pepita, atribuyéndolo al error de medición de una de coordenadas.

Sill: También conocido como la varianza asintótica del proceso y es el valor en donde la distancia se vuelve constante.

Rango: Representa la distancia por la cual los valores de las variables ya no se correlacionan.

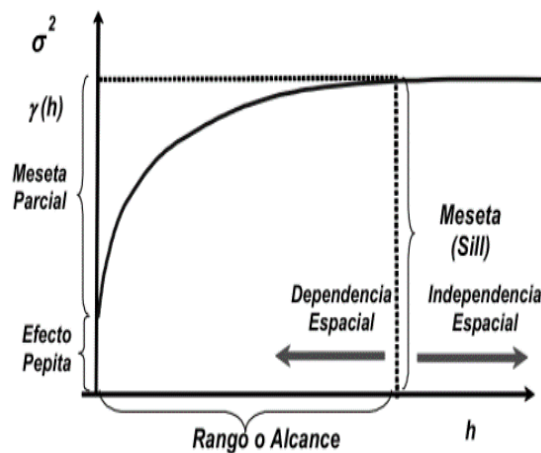


Ilustración 2–4: Principales elementos del variograma.

Fuente: Universidad de Guayaquil.

Realizado por: (Chang 2019).

2.4.3. Modelo de Semivariogramas

Según (Reyes 2017). describe los modelos de semivariograma teóricos más empleados:

- **Modelo esférico**

En este modelo, el crecimiento es rápido cerca del punto de partida, cuanto más avanza, más se ralentiza; cero en longitud en la región de incremento.

SPHERICAL

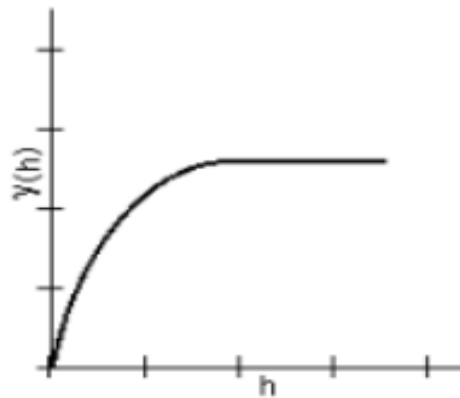


Ilustración 2-5: Modelo de Semivarianza Esférico.

Fuente: Universidad Distrital Francisco José de Caldas.

Realizado por: (Beltran 2020).

$$y(h) = c_0 + c \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) \quad 0 < h \leq a \quad (2-7)$$

$$y(h) = c_0 + c; h > a \quad (2-8)$$

- **Modelo gaussiano**

En el modelo gaussiano, la dependencia espacial disminuye a medida que la distancia se aproxima al infinito. Tiene una forma parabólica cercana a la original.

GAUSSIAN

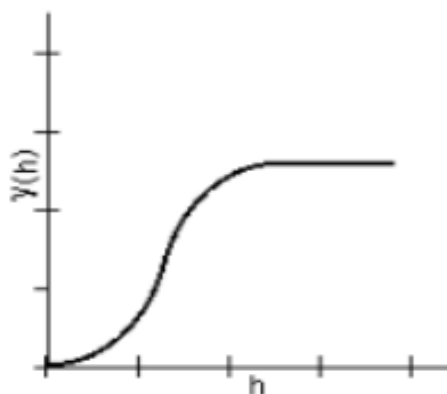


Ilustración 2-6: Modelo de Semivarianza Gaussiano.

Fuente: Universidad Distrital Francisco José de Caldas.

Realizado por: (Beltran 2020).

$$y(h) = c_0 + c \left(1 - \exp\left(\frac{h^2}{r^2}\right) \right) h > 0 \quad (2-9)$$

$$y(h) = 0 \quad (2-10)$$

- **Modelo exponencial**

La dependencia espacial aumenta exponencialmente con la distancia entre puntos conocidos.

EXPONENTIAL

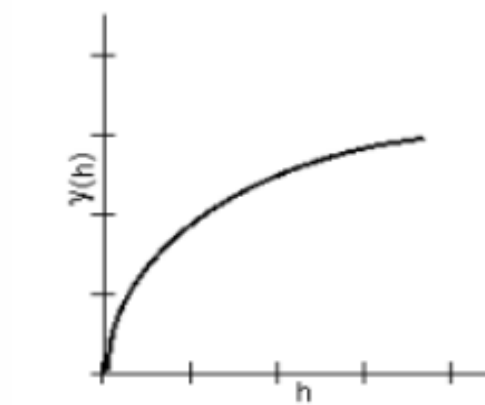


Ilustración 2-7: Modelo de Semivarianza Exponencial.

Fuente: Universidad Distrital Francisco José de Caldas.

Realizado por: (Beltran 2020).

$$y(h) = c_0 + c \left(1 - \exp\left(\frac{-h}{r}\right) \right) h > 0 \quad (2-11)$$

$$y(h) = 0 \quad (2-12)$$

CAPÍTULO III

3. MARCO METODOLÓGICO

3.1. Tipo de investigación

El tipo de investigación del presente trabajo es **cuantitativa**, debido a que se realiza una serie de procesos secuenciales para estudiar el comportamiento de la variable temperatura ambiente en la provincia de Chimborazo, en cuanto a la manipulación de las variables es **no experimental** dado que la base de datos lo proporciona el GEAA. Por el nivel de profundización es **explicativa**, ya que se realiza un análisis detallado de los métodos. De acuerdo con el tipo de inferencia, se presenta una investigación **deductiva** por que se comparan técnicas de interpolación espacial para determinar el método que mejor se ajusta. Por consiguiente, la investigación es **aplicada** ya que se implementa para las estaciones meteorológicas en el periodo 2014-2020 así como para los datos del SAM en el periodo 1998-2020. Además, es una investigación **transversal** porque analiza a la variable temperatura.

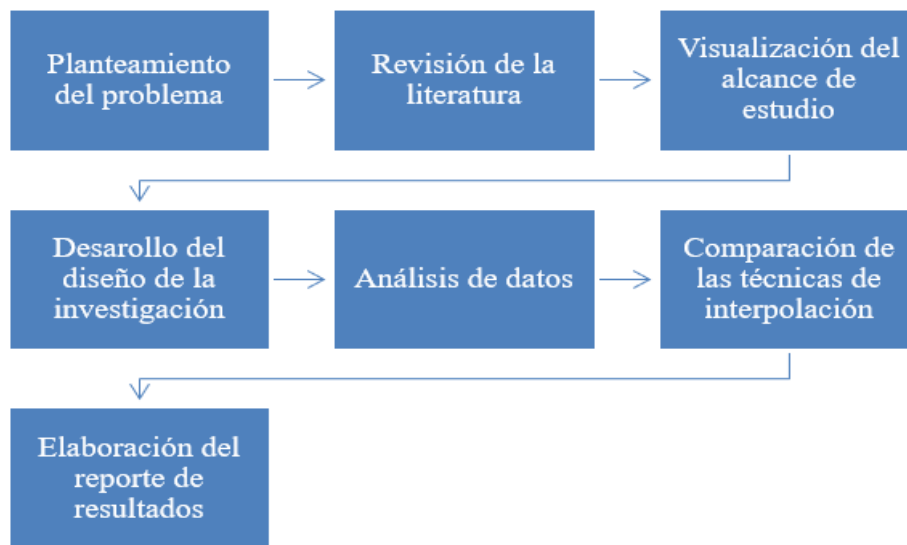


Ilustración 3–1: Desarrollo de la investigación.

Realizado por: Maldonado J., y Yugsi D., 2024.

3.2. Diseño de investigación

El diseño de la investigación es de tipo no experimental debido a que los datos de la variable se registran en las distintas estaciones meteorológicas, así como en los datos de salida del SAM. Por lo tanto, la investigación es cuantitativa ya que se va a comparar técnicas de interpolación espacial.

3.2.1. Localización del estudio

El proyecto de investigación planteado se lo desarrollará en el grupo de investigación GEAA, en la ciudad de Riobamba, en las calles: Av. Pedro Vicente y Entrada 1 ESPOCH - Riobamba.

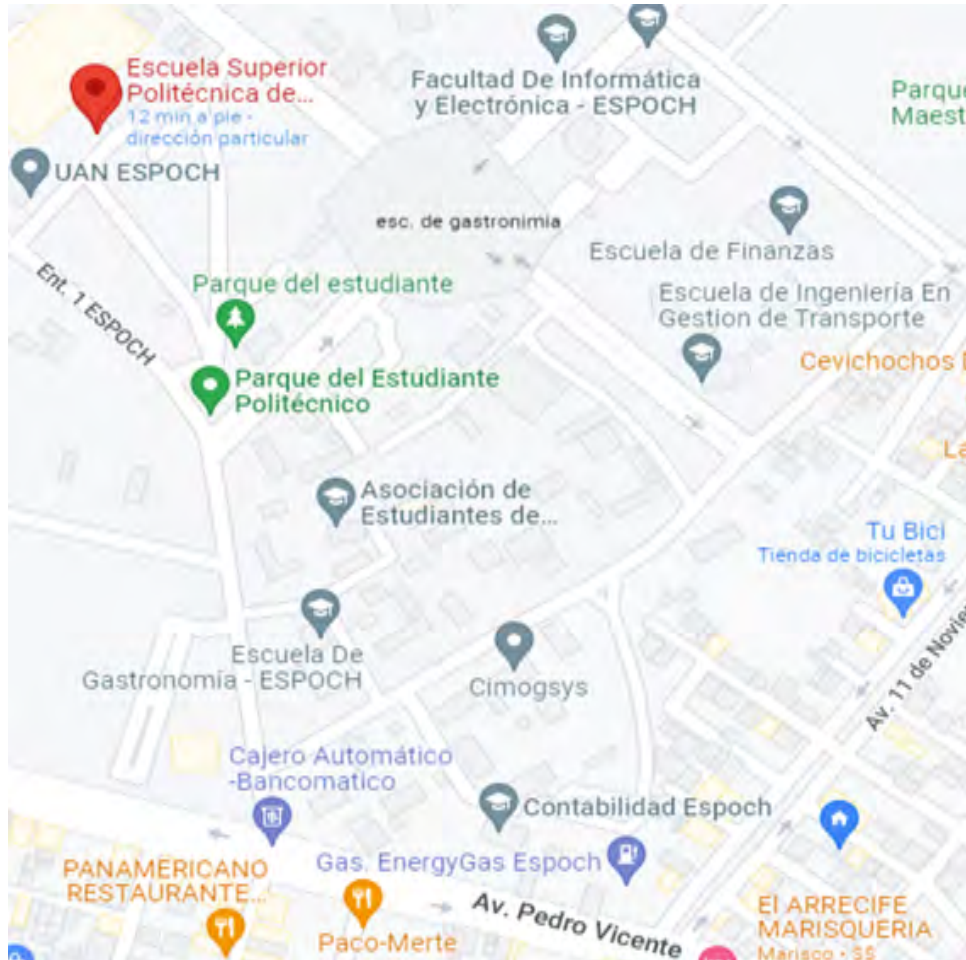


Ilustración 3–2: Ubicación en la Espoch.

Realizado por: Maldonado J., y Yugsi D., 2024.

3.2.2. Población de estudio

Para el proyecto de investigación la población corresponde a los datos de temperatura de las 11 estaciones meteorológicas en el periodo 2014-2020, y los datos del System Advisor Model (SAM) a una resolución 4km por 4km en el periodo 1998-2020 para la provincia del Chimborazo.

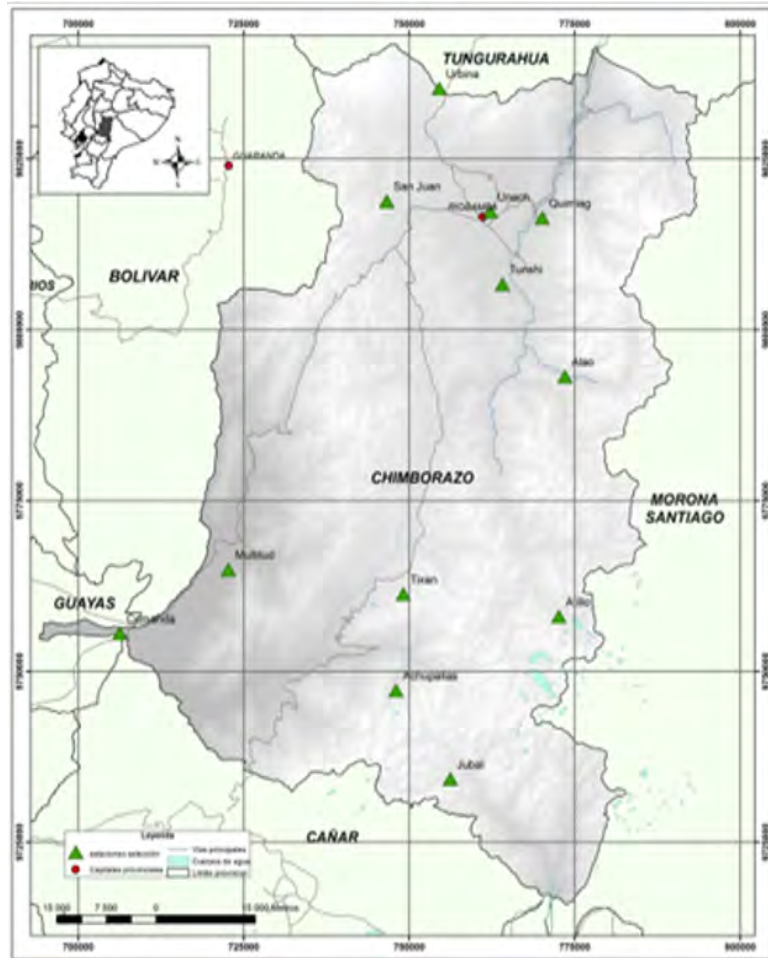


Ilustración 3–3: Localización de las 11 estaciones meteorológicas en la provincia de Chimborazo.

Fuente: GEAA.

Realizado por: GEAA.

No	Nombre de la estación	Coordenadas (UTM-WGS84 Zona 17S)		Altura (m.s.n.m)
		E	N	
1	Alao	773499	9793173	3064
2	Atillo	772610	9758048	3467
3	EsPOCH	758398	9816965	2754
4	Matus	777564	9827892	2471
5	Multitud	722699	9764910	1483
6	Quimiag	770084	9816393	2709
7	San Juan	746596	9818849	3232
8	Tixan	749103	9761332	3546
9	Tunshi	764087	9806678	2840
10	Cumandá	706262	9755579	331
11	Urbina	754533	9835326	3642

Ilustración 3–4: Tabla con los nombres y atributos de las 11 estaciones meteorológicas.

Fuente: GEAA.

Realizado por: GEAA.

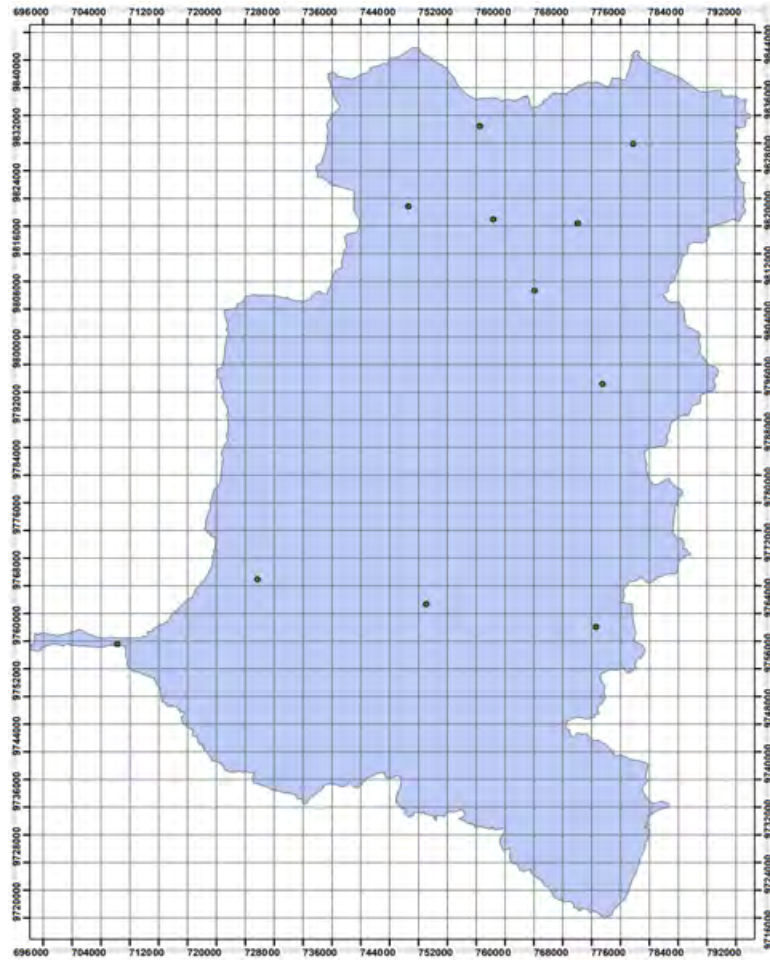


Ilustración 3-5: Malla de intersección de la Provincia de Chimborazo.

Realizado por: Maldonado J., y Yugsi D., 2024.

3.2.3. *Tamaño de la muestra*

Para determinar el método óptimo se toma como referencia los datos del SAM y de las estaciones meteorológicas del año 2015, dado que los equipos y los instrumentos de medición iniciaban su funcionamiento y contaban con los respectivos certificados de calibración de fábrica. Además, en un análisis previo, el año en estudio reportó menor porcentaje de datos faltantes.

3.2.4. *Método de muestreo*

Se realiza un muestreo no probabilístico por conveniencia.

3.2.5. *Técnicas de recolección de datos*

No se ha empleado ninguna técnica de recolección de datos debido a que los datos han sido recopilados y proporcionados por el GEAA - ESPOCH.

3.3. Identificación de variables

3.3.1. *Variable dependiente*

- Promedio Temperatura

3.3.2. *Variables independientes*

- Latitud
- Longitud
- Tiempo

3.3.3. *Operacionalización de variables*

Tabla 3–1: Operacionalización de variables.

VARIABLE	CONCEPTO	INDICADOR	INSTRUMENTO
Temperatura	La temperatura es una cantidad física que indica la energía interna de un objeto o ambiente.	Grados Celsius.	Sensor de temperatura y humedad.
Latitud	Es la distancia medida en grados hacia el norte o hacia el sur desde la línea ecuatorial, se utiliza para especificar la ubicación de un punto en la superficie terrestre en relación con los polos norte y sur.	Latitud Grados minutos y segundos	

Longitud	Es una medida angular de la ubicación de un punto en la superficie de la Tierra medida en grados hacia el este o el oeste desde el meridiano de Greenwich, utilizada para especificar la posición este-oeste de un lugar en el planeta.	Longitud Grados minutos y segundos	
Tiempo	El tiempo es la duración de las cosas cambiantes período, hora, día, semana, siglo, etc.	Horas Dias	

Realizado por: Maldonado J., y Yugsi D., 2024.

CAPÍTULO IV

4. MARCO DE ANALISIS E INTERPRETACION DE RESULTADOS

4.1. Obtención de Datos

4.1.1. *Malla de intersección de la provincia de Chimborazo*

En la provincia de Chimborazo, se ha implementado un diseño de sistema de coordenadas UTM formando una grilla de 4 km ubicado en la zona 17S. Mediante este diseño se identificó 523 puntos o nodos de la malla como se indica en la ilustración 3-5. Estos puntos fueron etiquetados estratégicamente para recopilar datos de temperatura promedio anual en el periodo establecido en la investigación.

Se ha considerado temperatura promedio debido a los cambios en los patrones que se analizan con la tendencia central debido a que proporcionan información del estado de la atmósfera a largo plazo, tradicionalmente definido como clima (Garreaud 2011). Esta medida proporciona información relevante lo que permite evaluar el impacto del cambio climático.

4.1.2. *Obtención de Datos Sam (parte de los modelos CMIP6)*

Los puntos se descargaron del programa SAM (parte de los modelos CMIP6), donde se creó un proyecto en el servidor net (<https://sam.nrel.gov/>) como se visualiza en la ilustración 4-1, los datos se obtienen puntualmente ingresando longitud y latitud en coordenadas geográficas de cada nodo de la malla dentro de la provincia de Chimborazo.

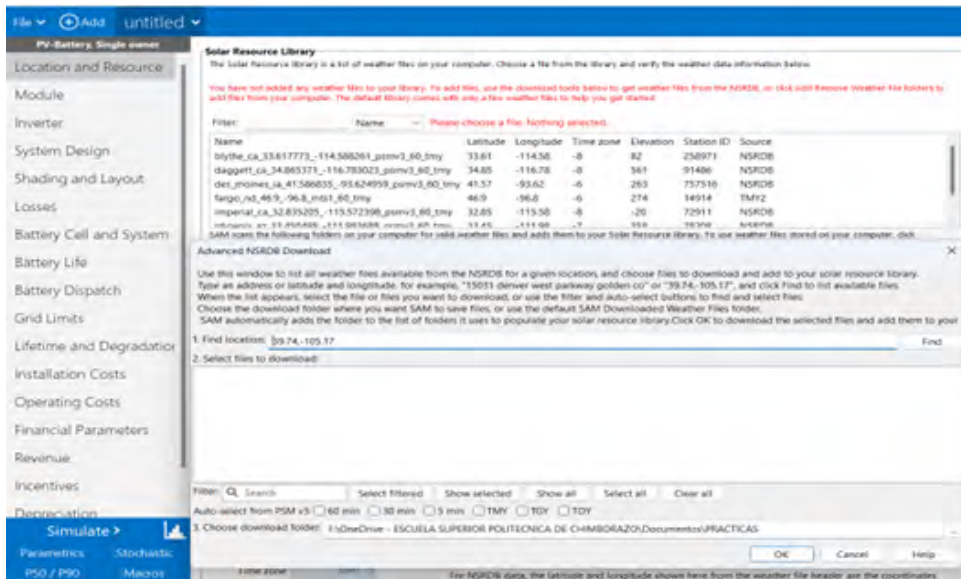


Ilustración 4-1: Ingreso de coordenadas UTM en el SAM.

Realizado por: Maldonado J., y Yugsi D., 2024.

Posteriormente, se contrasto las ubicaciones utilizando la página de Google Maps se observa en la ilustración 4-2.



Ilustración 4-2: Comprobación de puntos en Google maps.

Realizado por: Maldonado J., y Yugsi D., 2024.

Se obtuvo 17 variables meteorológicas y 17520 observaciones en cada año dentro de los 523 puntos en un formato csv, además, se calculó el promedio anual de la variable temperatura mediante tablas dinámicas proceso automatizado utilizando las macros en Excel.

4.1.3. Datos estaciones meteorológicas

Los datos de temperatura se han obtenido de las 11 estaciones meteorológicas monitoreadas por el GEAA. La información de las bases de datos son preprocesadas, homogeneizadas, y la reducción de la variabilidad generada por errores en la toma de mediciones mediante el uso de la herramienta Climatol (Guijarro Pastor 2004).

A continuación, se detallan los nombres de las diferentes estaciones junto con la parroquia y el cantón al que pertenecen, así como su respectiva altura (metros) y coordenadas UTM como indica la tabla 4-1.

Tabla 4-1: Nombre de las Estaciones Meteorológicas.

N	Estaciones M	Parroquia	Cantón	X	Y	Altura
1	Alao	Pungala	Riobamba	773491	9793166	3064
2	Atillo	Cebadas	Guamote	772589	9758044	3467
3	Cumandá	Cumandá	Cumandá	706262	9755579	331
4	EsPOCH	Lizarzaburu	Riobamba	758398	9816965	2754
5	Matus	Matus	Penipe	777564	9827892	2471
6	Multitud	Multitud	Alausí	725681	9764904	1483
7	Quimiag	Quimiag	Riobamba	770082	9816406	2718
8	San Juan	San Juna	Riobamba	746640	9818903	3244
9	Tixán	Tixán	Alausí	749115	9761350	3546
10	Tunshi	Licto	Riobamba	764074	9806681	2840
11	San Isidro de Patulú	San Isidro de Patulú	Chocavid Central	754579	9835357	3467

Realizado por: Maldonado J., y Yugsi D., 2024..

4.2. Procesamiento de los datos

4.2.1. Procesamiento de los datos SAM

Se ha encontrado limitaciones dentro de los datos del SAM por que no proporcionan información completa en diferentes años. Por ende, fue necesario llevar a cabo una imputación de datos para rellenar faltantes en diferentes puntos como indica la tabla 4-2.

Tabla 4-2: Puntos con los años faltantes.

Punto	Año	Porcentaje
61	2016	0.0019 %
66	2015	0.0019 %
143	2013	0.0019 %
380	2008	0.0019 %
381	2013	0.0019 %
450	2009	0.0019 %
496	2014	0.0019 %
160	2009	0.0019 %
182	2018	0.0019 %
184	2007	0.0019 %
240	2015, 2016, 2017, 2018, 2019	0.0096 %
244	2019	0.0019 %
317	2005	0.0019 %
482	1998	0.0019 %
513	2000, 2001	0.0038 %
100	2017	0.0019 %

Realizado por: Maldonado J., y Yugsi D., 2024.

Se ha empleado la técnica de missForest para el relleno de datos faltantes en diferentes puntos del SAM con el lenguaje de programación RStudio completando en todos los años como indica la ilustración 4-3. Además, la imputación de temperaturas tiene un buen desempeño si se tiene datos de estaciones vecinas, por lo contrario, para precipitación tiene un resultado mucho más incierto (Schmidt et al. 2015).

	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DECIEMBRE	P_GENERAL
P61-1998	9.23185484	9.4389881	8.65188172	9.84652778	8.72849462	7.37222222	6.53427419	6.87567204	7.29861111	8.03561828	8.17361111	8.06787634	8.2713027
P61-1999	7.81451613	7.07068452	7.72580645	7.77013889	7.4563172	7.04166667	6.40188172	6.12163979	7.00069444	7.25537634	7.74236111	7.84072581	7.27015076
P61-2000	7.43145161	7.72767857	7.85416667	7.89375	7.65456989	7.26458333	6.93615591	7.08736559	7.53819444	8.61962366	8.5625	8.91599462	7.79050286
P61-2001	7.87231183	8.44122024	8.22916667	8.69791667	8.4139785	6.98472222	6.69758065	6.73454301	7.44791667	8.54569893	8.44791667	8.88172043	7.94955771
P61-2002	8.53965054	8.63244048	8.17137097	8.50486111	8.46774194	6.92222222	7.13642473	6.66061828	7.54236111	8.15322581	8.19583333	9.00739247	7.99451192
P61-2003	8.99798387	9.03125	8.40591398	8.48055556	8.13642473	7.25347222	6.82930108	7.2244624	7.9875	8.53091398	8.52430556	8.38306452	8.14859431
P61-2004	8.83870968	8.5796131	8.48521505	8.48958333	8.16465054	7.00555556	6.84408602	6.62701613	7.28888889	8.3125	8.47638889	8.44018817	7.96269961
P61-2005	8.86962366	9.53199405	9.1780914	8.97777778	8.81384409	8.66944444	8.40188172	8.63575269	9.45277778	8.90994624	8.56458333	8.20362903	8.85077885
P61-2006	8.50403226	8.56026786	8.35416667	8.56736111	7.93884409	7.24652778	7.04099462	7.48185484	8.15902778	9.1922043	9.27430556	9.2264785	8.29550545
P61-2007	10.0456989	9.0483631	8.97513441	9.15625	9.31451613	7.74513889	8.40658602	6.68615591	7.81458333	8	8.21458333	7.75	8.51308417
P61-2008	8.0625	7.66741071	7.49596774	7.85972222	7.72446237	7.13819444	6.60013441	7.0625	7.54722222	8.06586022	8.12569444	8.12231183	7.62266505
P61-2009	8.20564516	8.13020833	8.22110215	8.38125	8.28427419	8.01666667	7.83467742	8.07795699	8.46736111	8.90389785	9.02430556	8.95228495	8.3749692
P61-2010	9.37836022	9.58184524	9.22379032	9.10486111	9.03158602	7.45972222	7.78293011	7.46505376	7.97152778	8.25067204	7.69583333	7.62298387	8.38076384
P61-2011	7.79637097	7.85044643	7.9233871	8.1	8.20564516	7.74375	7.11491936	7.43884409	7.68402778	7.92607527	7.96597222	8.14650538	7.82466198
P61-2012	8.00873856	7.53869048	7.93413979	7.96944444	7.73991936	7.22013889	6.93145161	6.77822581	7.45069444	8.17405914	8.59444444	8.2909462	7.71924496
P61-2013	9.10819893	8.06324405	8.59677419	8.66111111	8.09206989	6.99513889	6.78494624	7.26209677	7.74722222	8.22715054	8.19166667	8.49731183	8.01891094
P61-2014	8.49731183	8.41666667	8.33266129	8.86180556	8.6344086	7.62708333	7.2973118	7.11693548	8.03611111	8.25537634	8.60555556	8.56653226	8.18751493
P61-2015	8.47849462	8.49776786	8.33467742	8.29652778	8.40456989	7.22777778	7.43010753	7.77352151	7.91458333	8.90389785	9.1875	9.02956989	8.28991629
P61-2016	8.66700806	8.56731399	8.30590726	8.53440972	8.24888441	7.31790972	6.95209677	7.23730511	7.84138194	8.41462366	8.46420833	8.62742608	8.09820625
P61-2017	8.06922043	8.06845238	8.14919355	8.60625	8.26276882	7.76736111	6.5094086	7.47446237	7.79722222	8.73185484	8.35069444	8.56048387	8.02894772
P61-2018	8.8452957	9.67782738	9.24038979	8.62534722	8.88736559	7.87625	7.79307796	7.55201613	8.55402778	9.31922043	9.67201389	8.90591398	8.74572882
P61-2019	9.21008065	9.76919643	9.40530914	9.67270833	9.08548387	8.36083333	7.61377688	7.67560484	8.60465278	9.01646505	9.55409722	9.7375672	8.97548131
P61-2020	9.72318548	10.1231399	10.1188172	9.77076389	9.53669355	8.48194444	7.94153226	8.56263441	8.94659722	9.59448925	9.48666667	9.60900538	9.32462247

Ilustración 4-3: Datos rellenados para el Punto 61-2016.

Realizado por: Maldonado J., y Yugsi D., 2024.

Por otra parte, los datos de las estaciones no fueron sometidos a ningún proceso, porque han sido tratados a priori por el GEAA.

4.3. Análisis exploratorio de datos

4.3.1. Datos atípicos

Se ha observado presencia de datos sospechosos de la variable temperatura del SAM periodo 1998-2020 como se aprecia en la ilustración 4-4 esto podría deberse a la ubicación geográfica de la provincia de Chimborazo, teniendo zonas muy altas y frías con temperaturas de hasta -2 °C y otras muy calientes de hasta 26 °C.

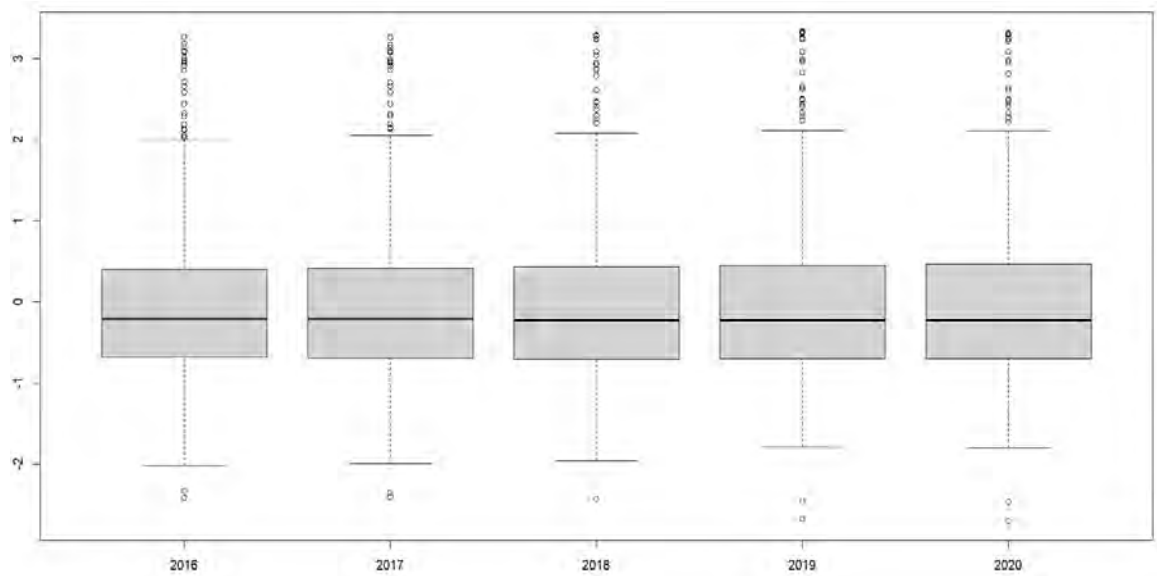


Ilustración 4-4: Diagrama de caja temperatura 2016-2020.

Realizado por: Maldonado J., y Yugsi D., 2024.

Se ha realizado el diagrama de caja de las estaciones meteorológicas periodo 2014-2020 donde se observa que el comportamiento de la variable temperatura en la estación de Cumandá es diferente debido a que pertenece a la región costa como se puede observar en la ilustración 4-5.

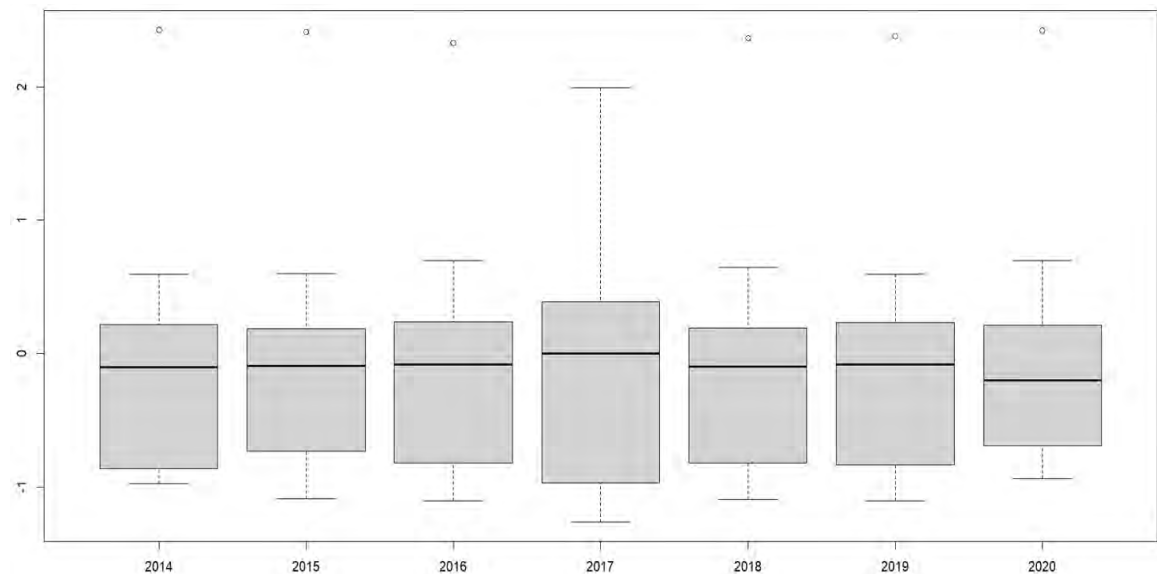


Ilustración 4-5: Diagrama de caja datos estaciones temperatura 2014-2020.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.3.2. Estadísticas Descriptivas

Se realizó un análisis estadístico descriptivo para todos los años descargados del SAM. En donde se ha observado que el comportamiento de los datos en los histogramas de la variable temperatura

no se aproximan a una distribución normal como se observa en la ilustración 4-6, 4-7, 4-8.

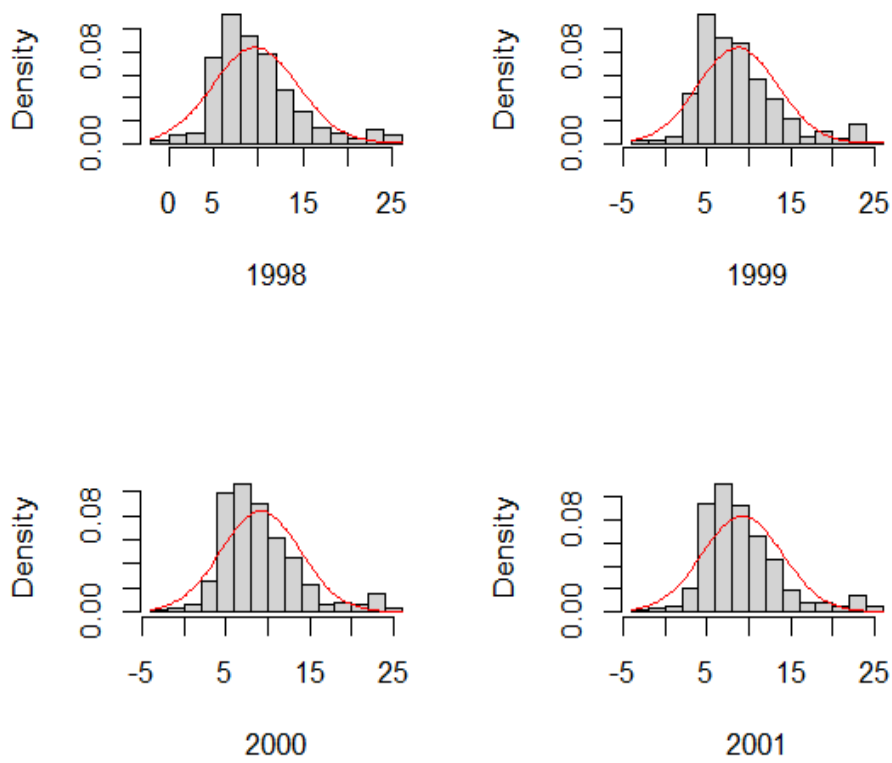


Ilustración 4-6: Histograma y curva de la normal temperatura 1998-2001.

Realizado por: Maldonado J., y Yugsi D., 2024.

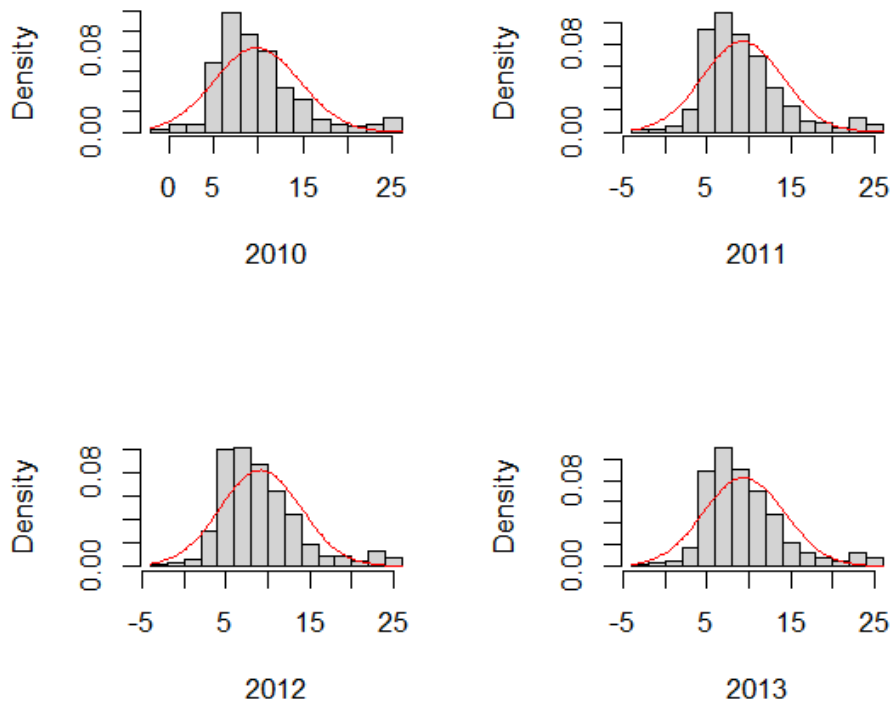


Ilustración 4-7: Histograma y curva de la normal temperatura 2010-2013.

Realizado por: Maldonado J., y Yugsi D., 2024.

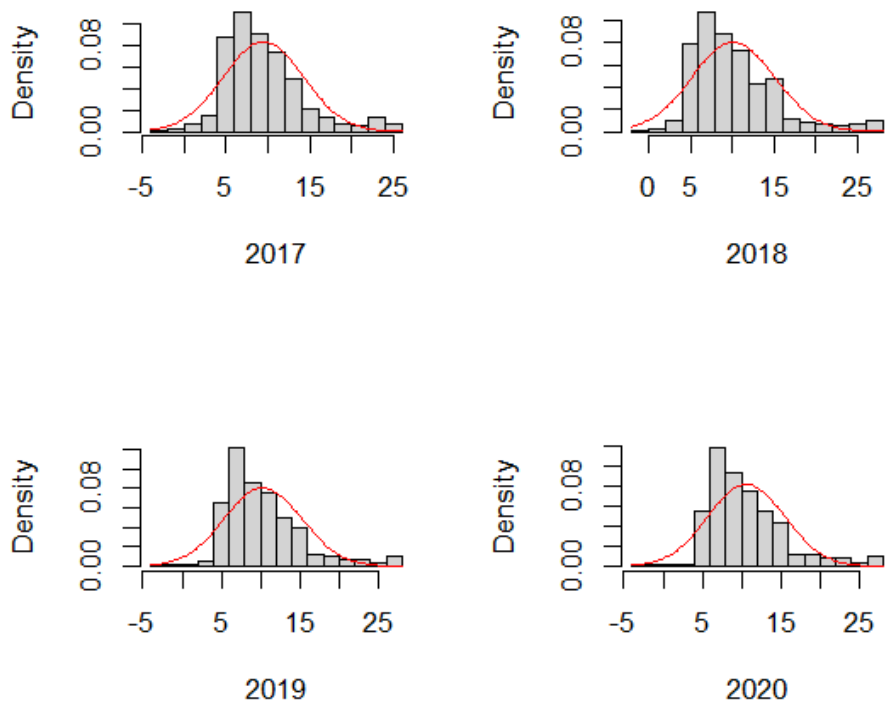


Ilustración 4-8: Histograma y curva de la normal temperatura 2017-2020.

Realizado por: Maldonado J., y Yugsi D., 2024.

Mediante un test de Normalidad de Kolmogórov-Smirnov con un nivel de confianza de 95 % donde

se rechaza la hipótesis nula y se concluye que los datos del SAM no provienen de una distribución normal se reitera en las ilustraciones de QQ Plot , como se observa en la ilustración 4-9, 4-10, 4-11.

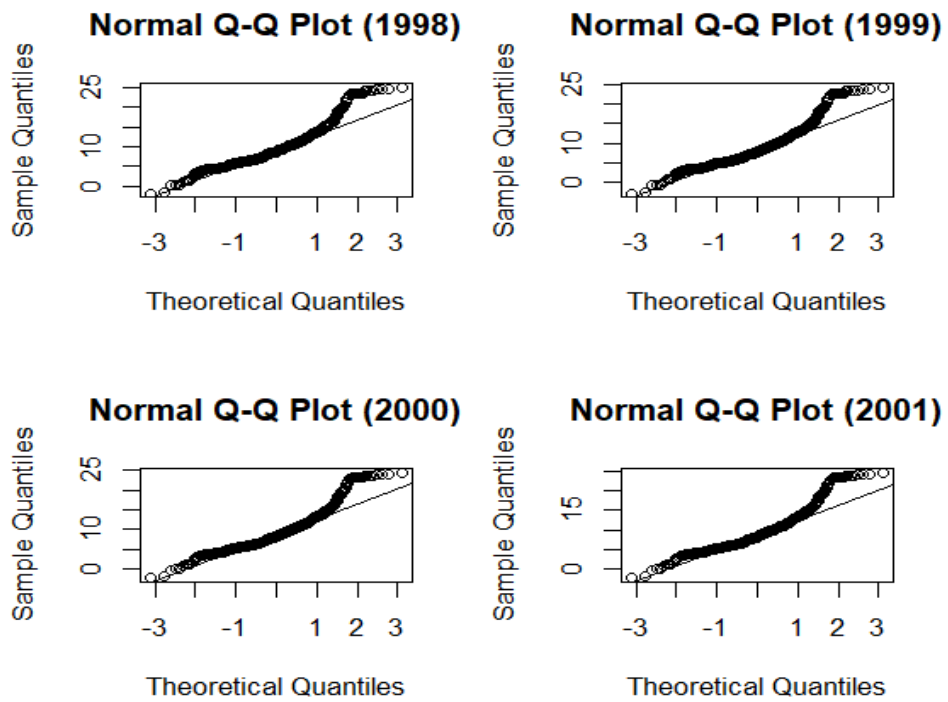


Ilustración 4-9: Gráfico de normalidad QQplot 1998-2001.

Realizado por: Maldonado J., y Yugsi D., 2024.

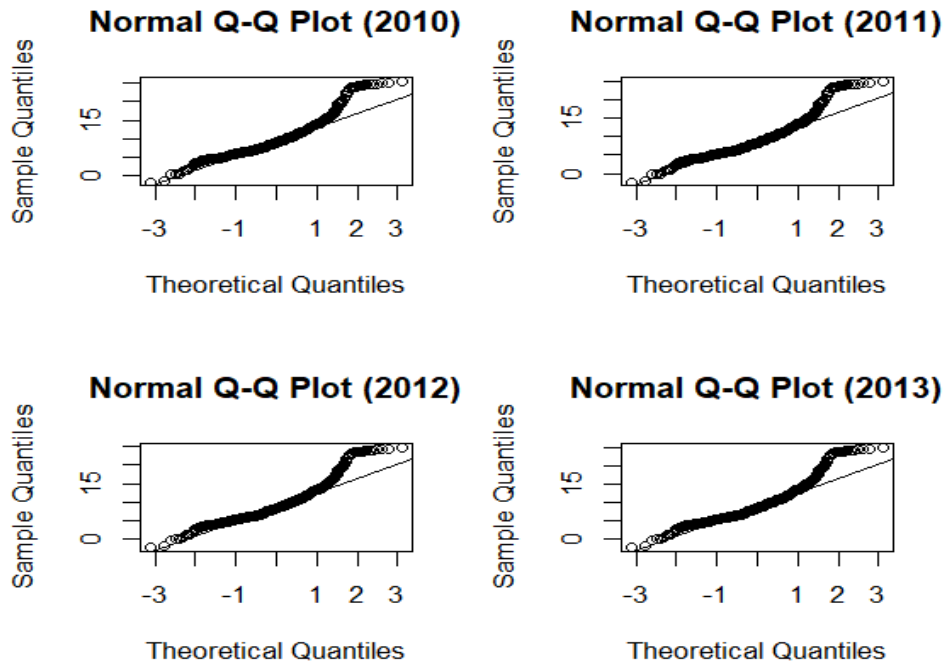


Ilustración 4-10: Gráfico de normalidad QQplot 2010-2013.

Realizado por: Maldonado J., y Yugsi D., 2024.

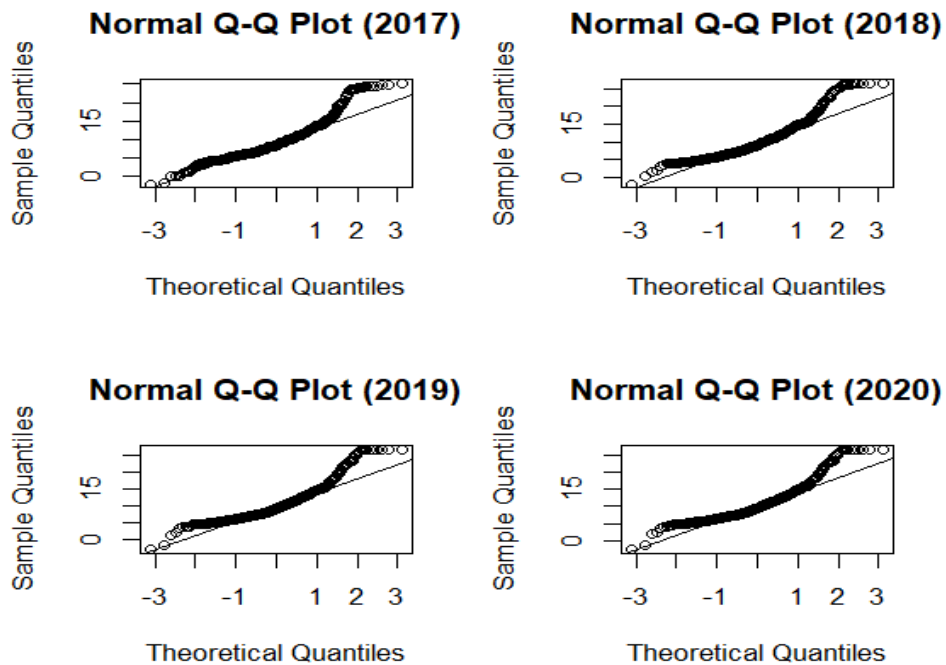


Ilustración 4-11: Grafico de normalidad QQplot 2017-2020.

Realizado por: Maldonado J., y Yugsi D., 2024.

Se realizo un análisis descriptivo de los datos del SAM para el periodo 1998-2020, calculado la

media, mediana, desviación estándar, curtosis, y sesgo. Se observó que la media y la mediana eran casi idénticas debido a la ubicación de los puntos, no había mucha variación de los datos entre los diferentes años. Además, se encontró que los datos tenían una curtosis positiva y un sesgo derecho, lo que indica que la distribución era leptocúrtica y asimétrica hacia la derecha, con una cola más larga en el lado derecho de la media indicado en la tabla 4-3.

Tabla 4–3: Estadísticas descriptivas datos SAM.

SAM	Media	Mediana	Std.Dev	Kurtosis	Sesgo
1998	9.685	8.664	4.749	1.536	1.127
1999	8.711	7.697	4.746	1.554	1.142
2000	9.165	8.123	4.741	1.545	1.143
2001	9.253	8.216	4.774	1.593	1.163
2002	9.336	8.301	4.838	1.627	1.169
2003	9.508	8.502	4.851	1.623	1.165
2004	9.356	8.345	4.840	1.639	1.173
2005	10.283	9.205	4.785	1.618	1.161
2006	9.749	8.712	4.828	1.983	1.182
2007	9.970	8.928	4.805	1.658	1.175
2008	9.079	8.046	4.773	1.576	1.150
2009	9.775	8.745	4.851	1.657	1.177
2010	9.858	8.861	4.788	1.621	1.158
2011	9.300	8.313	4.805	1.608	1.159
2012	9.159	8.141	4.832	1.644	1.175
2013	9.449	8.441	4.828	1.618	1.163
2014	9.669	8.630	4.897	1.730	1.200
2015	9.744	8.653	4.954	1.750	1.209
2016	9.989	8.985	4.868	1.646	1.170
2017	9.486	8.485	4.812	1.605	1.159
2018	10.163	9.046	4.986	1.639	1.224
2019	10.339	9.200	4.953	1.871	1.275
2020	10.629	9.536	4.886	1.793	1.249

Realizado por: Maldonado J., y Yugsi D., 2024.

Mientras que para los datos de temperatura promedio de las **estaciones meteorológicas** entre 2014 y 2020 se observó a simple vista en los histogramas que los datos no se ajustan a una distribución

normal, como se observa en la ilustración 4-12, implicando aplicar un test de normalidad para corroborar.

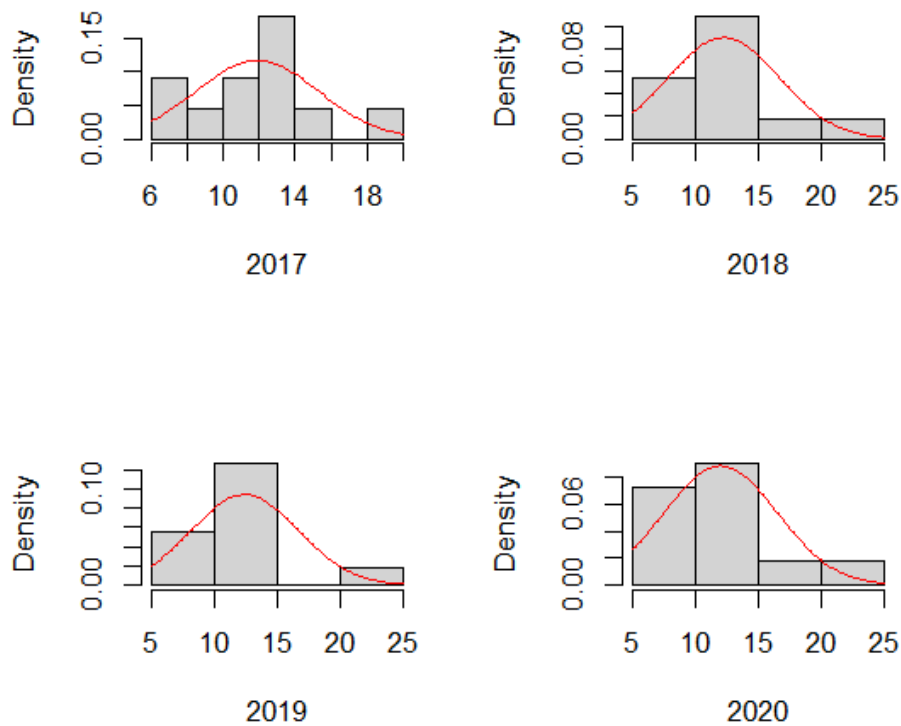


Ilustración 4–12: Histograma y curva de la normal 2017-2020.

Realizado por: Maldonado J., y Yugsi D., 2024.

De la misma manera mediante un test de Normalidad de Kolmogórov-Smirnov con un nivel de confianza de 95 % donde no se rechaza la hipótesis nula y se concluye que los datos de las estaciones si provienen de una distribución normal debido ha que se considerado datos anuales como se reitera en las ilustraciones de QQ Plot , como se observa en la ilustración 4-13

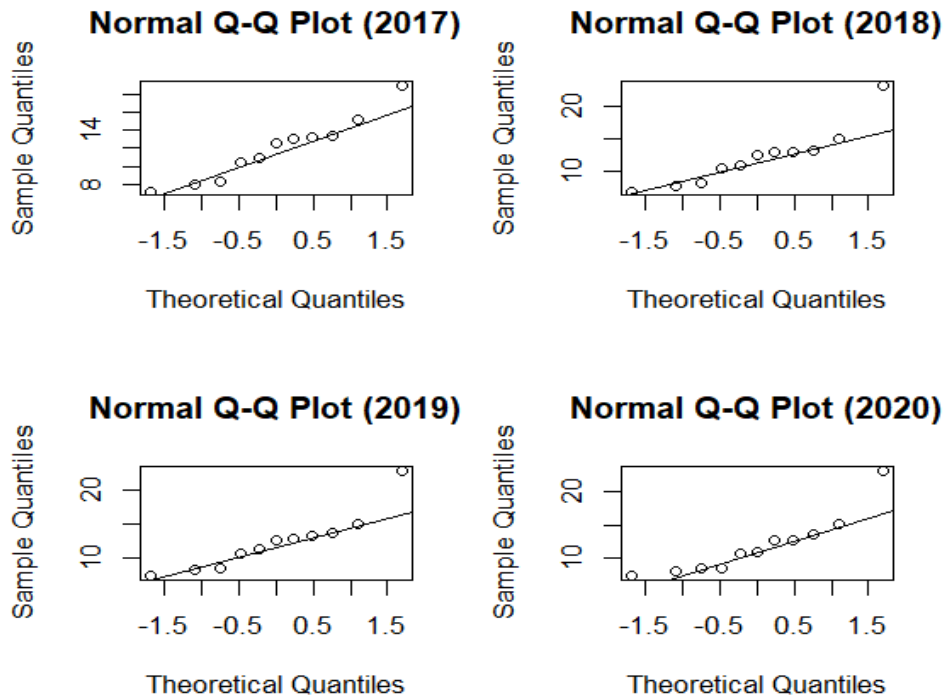


Ilustración 4-13: Grafica de normalidad QQplot 2017-2020.

Realizado por: Maldonado J., y Yugsi D., 2024.

De igual manera, se realizó el mismo proceso para los datos de las estaciones meteorológicas con un cambio en el 2017 la curtosis es negativa y el sesgo positivo, indicando una distribución platicúrtica y asimetría hacia la derecha, como indica la tabla 4-4.

Tabla 4-4: Estadísticas descriptivas datos Estaciones.

Estaciones	Media	Mediana	Std.Dev	Kurtosis	Sesgo
2014	12.488	11.982	4.985	0.704	1.200
2015	12.483	12.749	4.503	0.937	1.181
2016	12.737	13.006	4.285	0.469	0.994
2017	11.900	12.588	3.431	-0.773	0.379
2018	12.267	12.501	4.423	0.677	1.079
2019	12.441	12.688	4.248	0.739	1.091
2020	12.024	11.068	4.512	0.779	1.214

Realizado por: Maldonado J., y Yugsi D., 2024.

4.3.3. Selección de la muestra de estudio

4.3.4. Estadísticas descriptivas de la muestra de estudio

Para la identificación de atípicos de los datos descargados del SAM y de las estaciones meteorológicas se seleccionó una muestra siendo este el año 2015 por su mayor confiabilidad, precisión y menor cantidad de datos faltantes. A pesar de la identificación de anomalías ilustración 4-14, no se procedió a su eliminación debido a que no se presentaron tanto en las estaciones meteorológicas como en el SAM y también debido a la ubicación geográfica de la provincia de Chimborazo.

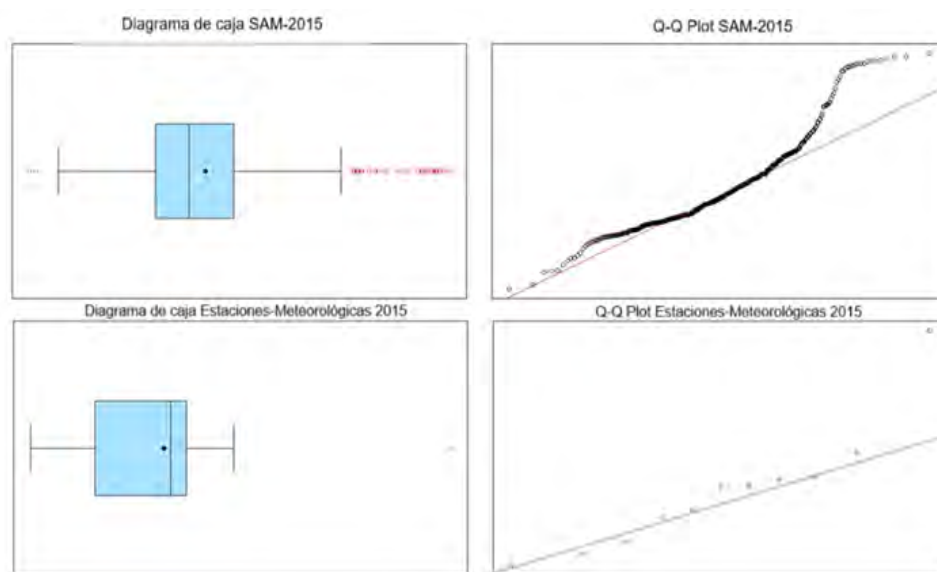


Ilustración 4-14: Diagrama de caja y gráfico de normalidad de la temperatura 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

Se ha observado que el cantón Cumandá presenta características distintivas en comparación con otros cantones, posiblemente debido a su transición de la región Sierra a Costa. Este fenómeno también se reflejó en los datos de las estaciones en el cantón Cumandá, donde se registró una temperatura media de 23.80 °C.

Las estadísticas descriptivas para este año en particular como indica la tabla 4-5, proporcionaron una visión completa de la distribución de los datos, la media y la media indican una ligera asimetría que se corrobora con la curtosis y el sesgo positivo, teniendo una forma leptocúrtica y asimetría hacia la derecha, pero no sucede lo mismo con las estaciones meteorológicas ya que el sesgo no es tan grande, esto se debe a la baja cantidad de estaciones.

Tabla 4-5: Estadísticas descriptivas datos Estaciones.

Estadísticas	SAM 2015	Estaciones Meteorológicas 2015
Media	9.744	12.483
Mediana	8.653	12.749
Std. Dev	4.954	4.503
kurtosis	1.750	9.937
Sesgo	1.209	1.181

Realizado por: Maldonado J., y Yúgsi D., 2024.

4.3.4.1. Datos del SAM

Histograma: Para el año 2015, como se observa en la tabla 4-5 la medida de sesgo o Skewness tuvo un valor de 1.2009, mientras que la medida de forma Kurtosis registró un valor de 1.750, por ende, se concluye que la distribución es leptocúrtica con asimetría hacia la derecha.

Prueba de Normalidad: Se utilizó la prueba de normalidad **Kolmogórov-Smirnov** se observó en el histograma una ligera tendencia a una distribución normal, pero al analizar el gráfico de normalidad (QQ plot) se evidencia que los datos no siguen una distribución normal, dado que el p-valor es menor que el nivel de significancia de 0.05, por ende, se rechaza la hipótesis nula, indicando que los datos no provienen de una distribución normal reflejado en la ilustración 4-15.

QQ Plot: Muestra que en el eje horizontal se encuentran los cuantiles de la distribución normal y en el eje vertical los cuantiles de la variable temperatura. Se evidencia que los puntos no se alinean correctamente a lo largo de la recta diagonal y se asume que los datos no están distribuidos normalmente.

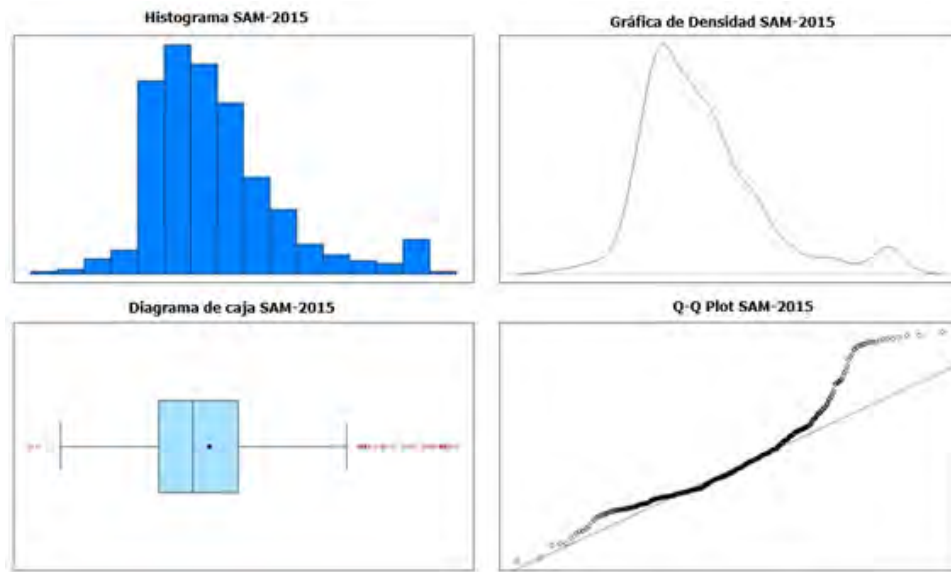


Ilustración 4–15: Análisis Exploratorio de datos del SAM.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.3.4.2. Datos del Estaciones

Histograma: Para el año 2015 se obtuvo medidas de centrado (media y mediana) similares, con diferencias inferiores a 1. En cuanto a las medidas de forma, como se muestra en la tabla 4-5 la medida de sesgo o Skewness arrojó un valor de 1,181 y la Kurtosis presentó un valor de 0,937, concluyendo que la distribución es leptocúrtica y asimetría a la derecha.

Prueba de Normalidad: Se utilizó la prueba de normalidad **Shapiro-Wilks** se observa en el histograma una leve tendencia a una distribución normal, pero al analizar el gráfico de normalidad (QQ plot), se evidencia que los datos si provienen de una distribución normal, dado que el p-valor es mayor al nivel de significancia de 0.05, no se rechaza la hipótesis nula, indicando que los datos si provienen de una distribución normal reflejado en la ilustración 4-16.

QQ Plot: Se muestra que en el eje horizontal se encuentran los cuantiles de la distribución normal y en el eje vertical los cuantiles de la variable temperatura. Se ha observado que los puntos de datos cerca de las colas no caen exactamente a lo largo de la recta diagonal, pero la mayor parte de puntos se alinean correctamente a lo largo de la recta y se asume que los datos parecen estar distribuidos normalmente.

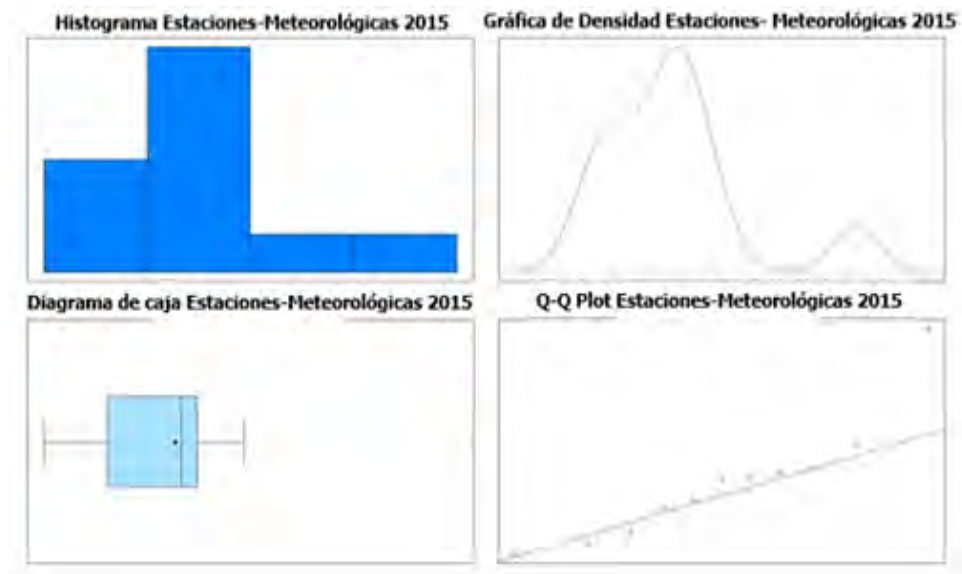


Ilustración 4–16: Análisis Exploratorio de datos de las Estaciones.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.3.4.3. Transformación Box Cox

Dado a la naturaleza de los datos y a los cambios drásticos que producen las transformaciones de tipo logarítmica, exponencial o $1/x$, se optó por una transformación que no altere mucho la forma original de los datos: la transformación box cox para evaluar el efecto de los datos, al examinar el gráfico de normalidad (QQ plot), los datos no siguen una distribución normal.

Se utilizó la prueba de normalidad **Kolmogorov Smirnov**, dado que el p-valor es menor que el nivel de significancia de 0.05, se rechaza la hipótesis nula, indicando que los datos no provienen de una distribución normal como se indica en la ilustración 4-17.

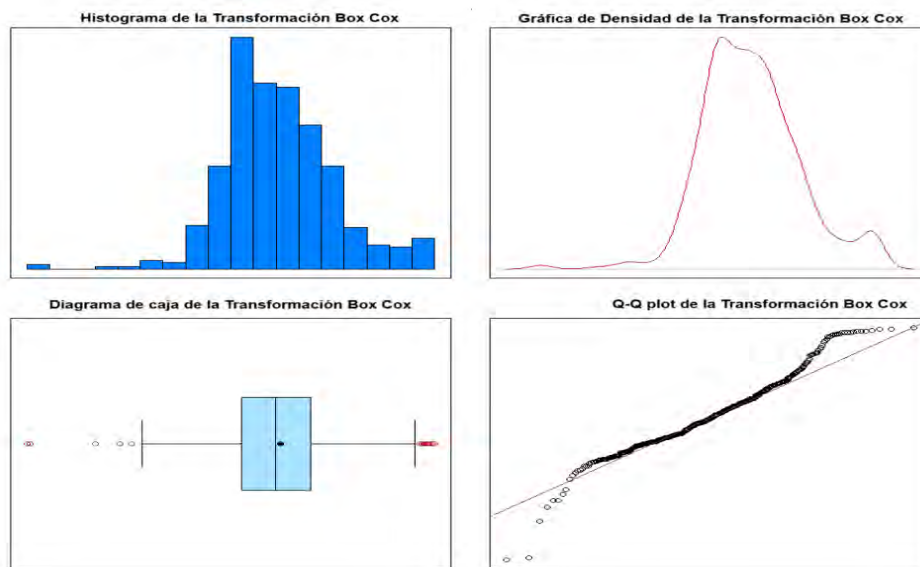


Ilustración 4–17: Transformación realizada a los datos del SAM.

Realizado por: Maldonado J., y Yugsi D., 2024.

Tras realizar un análisis descriptivo y comprobar mediante inferencia que los datos del SAM transformados no siguen una distribución normal, como se muestra en la ilustración 4-17, se aplicó el método de Kriging Universal según lo recomendado por (Saltos 2016).

4.4. Análisis Geoestadístico

Se ha considerado los siguientes pasos para la interpolación espacial.

1. Análisis exploratorio de datos
2. Muestreo Aleatorio
3. Realizar el modelo de interpolación
4. Validar el Modelo

4.4.1. Análisis Geoestadístico datos del SAM y Estaciones

4.4.1.1. Análisis de Tendencia

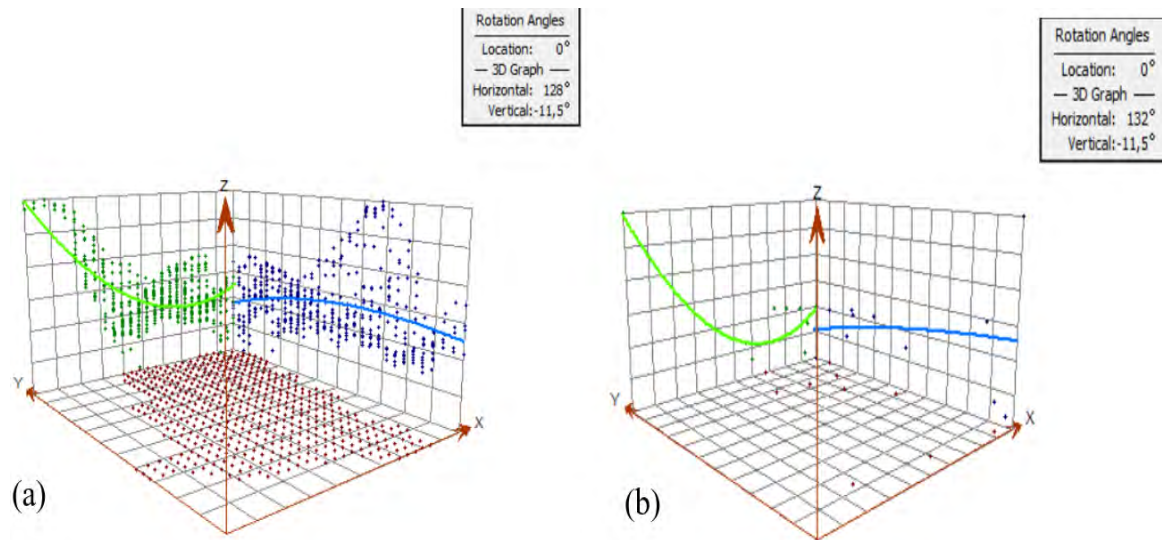
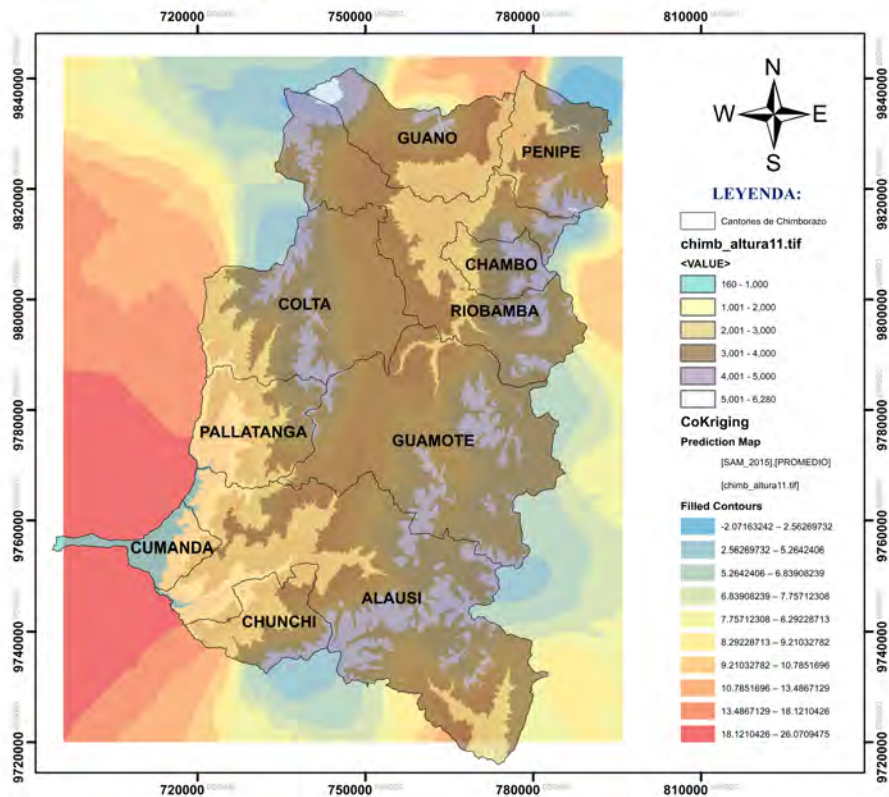


Ilustración 4-18: (a) Análisis de Tendencia del SAM año 2015 y (b) Análisis de Tendencia de las estaciones año 2015

Realizado por: Maldonado J., y Yugsi D., 2024.

La ilustración 4-18 corresponde a los datos del SAM del año 2015 mostrando una parábola de color verde en el plano Y, Z la cual representa un polinomio de segundo grado, puesto que la parábola sigue una curva creciente, lo que representa la transición de la sierra-costa, experimentando un aumento de temperatura a medida que desciende a la región costa, esta tendencia es característica en la cordillera occidental. Mientras que, en el plano X, Z se observó una línea de color azul que describe el comportamiento de un polinomio de primer grado, esta tendencia hace referencia a la cordillera oriental en donde la altura varía entre los 3000 a 4000 metros sobre el nivel del mar y en esta región la temperatura es relativamente constante como se visualiza en la Ilustración 4-19.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Sigüencia
Dennis Adrian Yugsi Yugsi



Ilustración 4–19: Mapa de temperatura en relación con la altura de la provincia de Chimborazo datos SAM 2015

Realizado por: Maldonado J., y Yugsi D., 2024.

Esta ilustración presenta el Cokriging de los datos del SAM año 2015 relacionando las variables altura y temperatura, se observó que a menor altura existe una mayor temperatura, esto se lo puede corroborar en el cantón Cumandá, el cual se encuentra a una altura que ronda los 1000 metros sobre el nivel del mar y para el año 2015 obtuvo una temperatura de 18 hasta 26 °C, y también se observó que a mayor altura existe una menor temperatura y esto se lo puede corroborar en el cantón Riobamba, el cual se encuentra a una altura de 3000 a 4000 metros sobre el nivel del mar y

para el mismo año obtuvo una temperatura de 10 hasta 13 °C. En el caso del Cokriging referente a los datos del SAM año 2020 se observó que de igual forma existe la relación de altura con la temperatura como se visualiza en la ilustración 4-20.

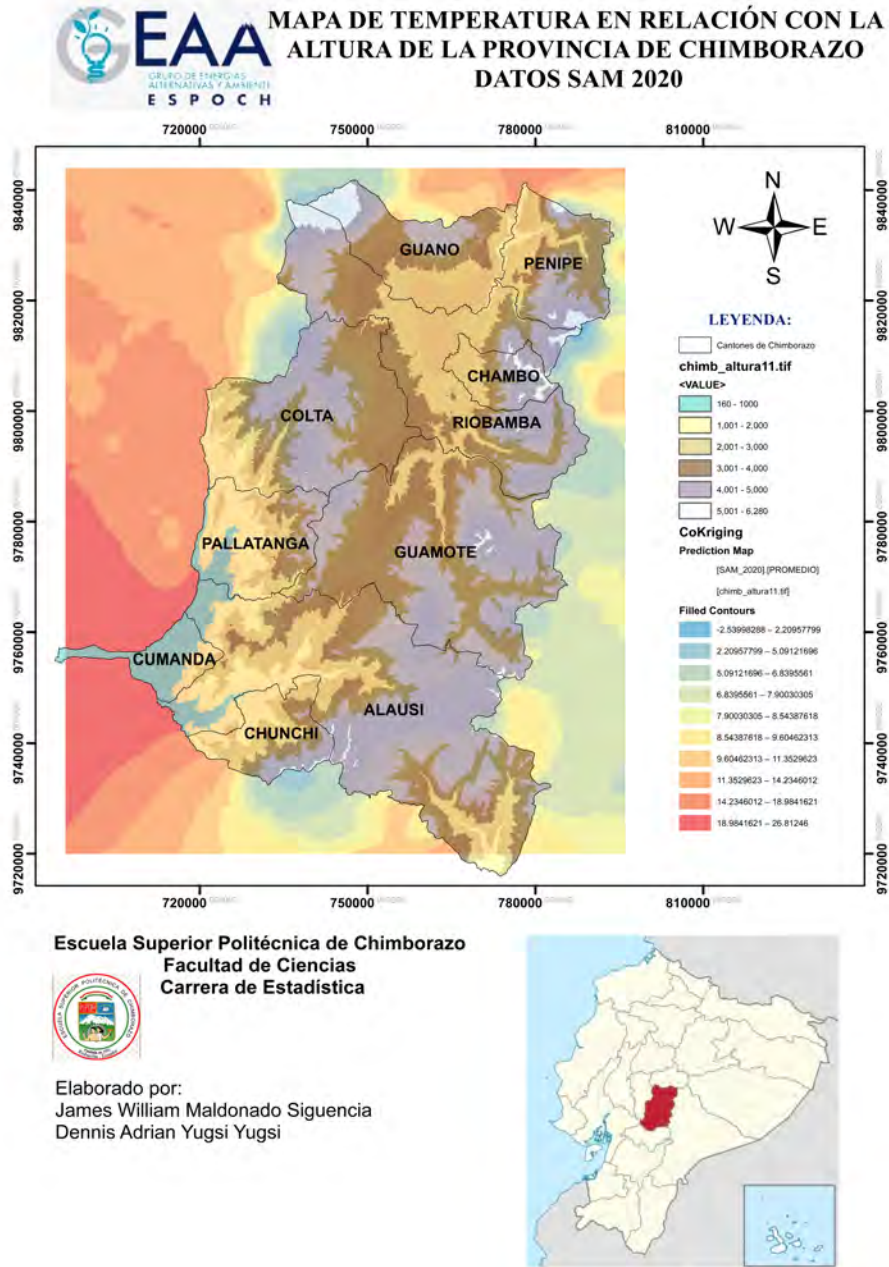


Ilustración 4–20: Mapa de temperatura en relación con la altura de la provincia de Chimborazo datos SAM 2020

Realizado por: Maldonado J., y Yugsi D., 2024.

A pesar de contar con 11 estaciones meteorológicas, se observó que el patrón de comportamiento y tendencia son similares a los datos del SAM como se visualiza en la Ilustración 30. Esta ilustración presenta los datos de las estaciones del año 2015 mostrando una parábola de color verde en el

plano Y, Z que representa un polinomio de segundo grado, esta tendencia es característica de la cordillera occidental. Mientras que, en el plano X, Z se observó una línea de color azul que describe el comportamiento de un polinomio de primer grado, esta tendencia hace referencia a la cordillera oriental en donde la altura varía entre los 3000 a 4000 metros sobre el nivel del mar como se visualiza en la Ilustración 4-21.

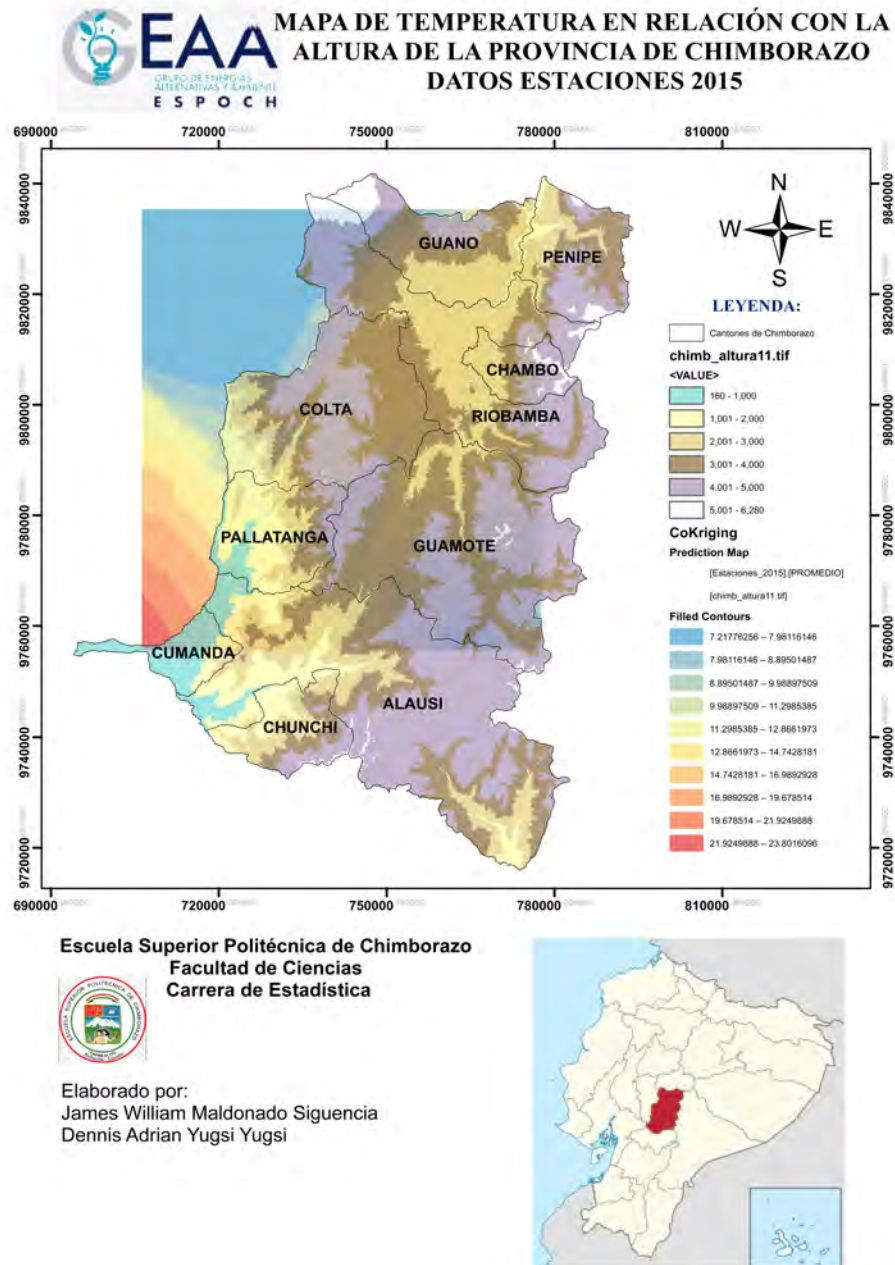
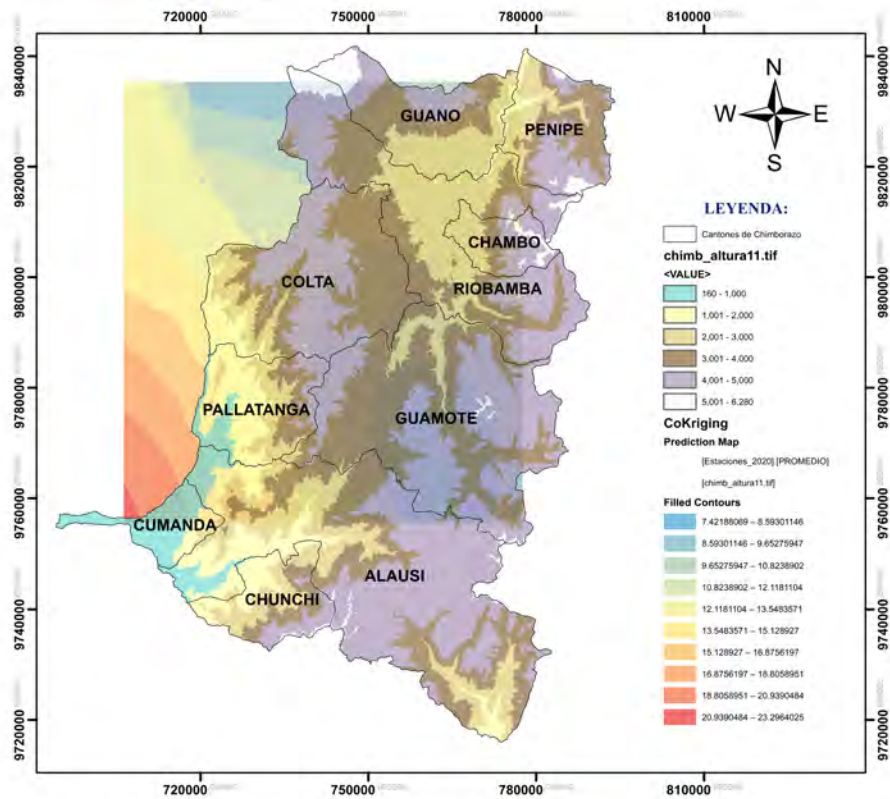


Ilustración 4-21: Mapa de temperatura en relación con la altura de la provincia de Chimborazo Datos Estaciones 2015

Realizado por: Maldonado J., y Yugsi D., 2024.

Esta ilustración presenta el Cokriging de las Estaciones año 2015 relacionando las variables altura

y temperatura, se observó que a menor altura existe una mayor temperatura, esto se lo puede corroborar en el cantón Cumandá, el cual se encuentra a una altura que ronda los 1000 metros sobre el nivel del mar y para el año 2015 obtuvo una temperatura de 21 hasta 23 °C, y también se observó que a mayor altura existe una menor temperatura y esto se lo puede corroborar en el cantón Riobamba, el cual se encuentra a una altura de 3000 a 4000 metros sobre el nivel del mar y para el mismo año obtuvo una temperatura de 11 hasta 12 °C. En el caso del Cokriging referente a los datos del SAM año 2020 se observó que de igual forma existe la relación de altura con la temperatura como se visualiza en la ilustración 4-22.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Sigüencia
Dennis Adrian Yugsi Yugsi



Ilustración 4–22: Mapa de temperatura en relación con la altura de la provincia de Chimborazo datos Estaciones 2020

Realizado por: Maldonado J., y Yugsi D., 2024.

4.4.1.2. Modelamiento del Semivariograma

En la ilustración 4-23 se evidencia la autocorrelación espacial de las temperaturas, se examinó la nube de puntos del semivariograma, cada punto representa un par de ubicaciones recalcando que las más cercanas entre si deben ser más parecidas y mostrando un comportamiento alejado del eje X e Y, señalando la existencia de datos a temperaturas altas por el cambio de región.

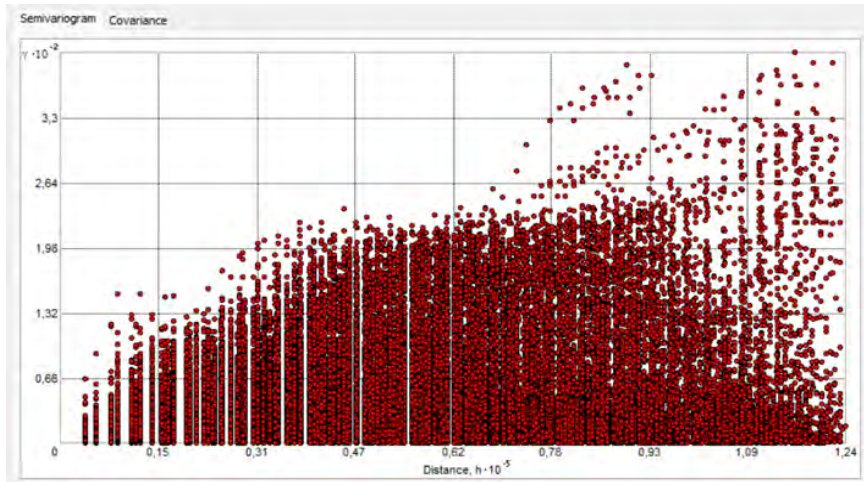


Ilustración 4–23: Semivariograma del SAM año 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

En la ilustración 4-24 se evidencia la autocorrelación espacial de las temperaturas, se examinó la nube de puntos del semivariograma, cada punto representa un par de ubicaciones recalando que las más cercanas entre si deben ser más parecidas y mostrando un comportamiento alejado del eje X e Y, esto debido a temperaturas altas y bajas por la localización de las estaciones a nivel de la provincia de Chimborazo.

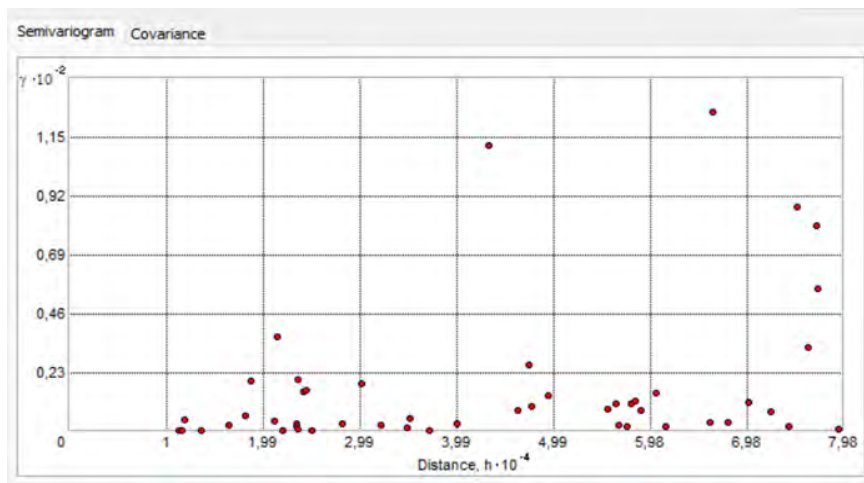


Ilustración 4–24: Semivariograma de las Estaciones del año 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5. Modelo de interpolación espacial

Para llevar a cabo esta investigación, se emplearon técnicas de interpolación espacial como el Kriging Universal que manejan datos que no provienen de una distribución normal, y Ordinario cuando los datos provienen de una distribución normal. Además, se implementó la técnica de

interpolación de la distancia inversa ponderada (IDW), la cual otorga mayor peso a los puntos cercanos y menor peso a los puntos más distantes. Por último, se usó la interpolación Spline, que desarrolla curvas suaves para aproximar un conjunto de datos.

En la ilustración 4-25 se observa la ubicación de las estaciones meteorológicas y los puntos descargados del SAM dentro de la provincia de Chimborazo. Para la aplicación de las diferentes técnicas de interpolación se dividió a los datos en dos subconjuntos de: entrenamiento (80%) y prueba (20%); con el porcentaje respectivo de prueba se llevó a cabo la validación cruzada.

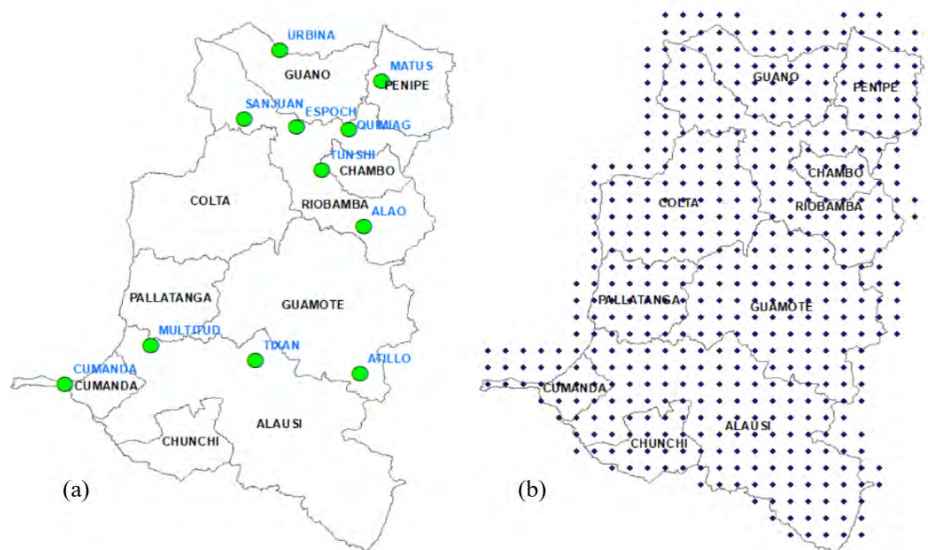
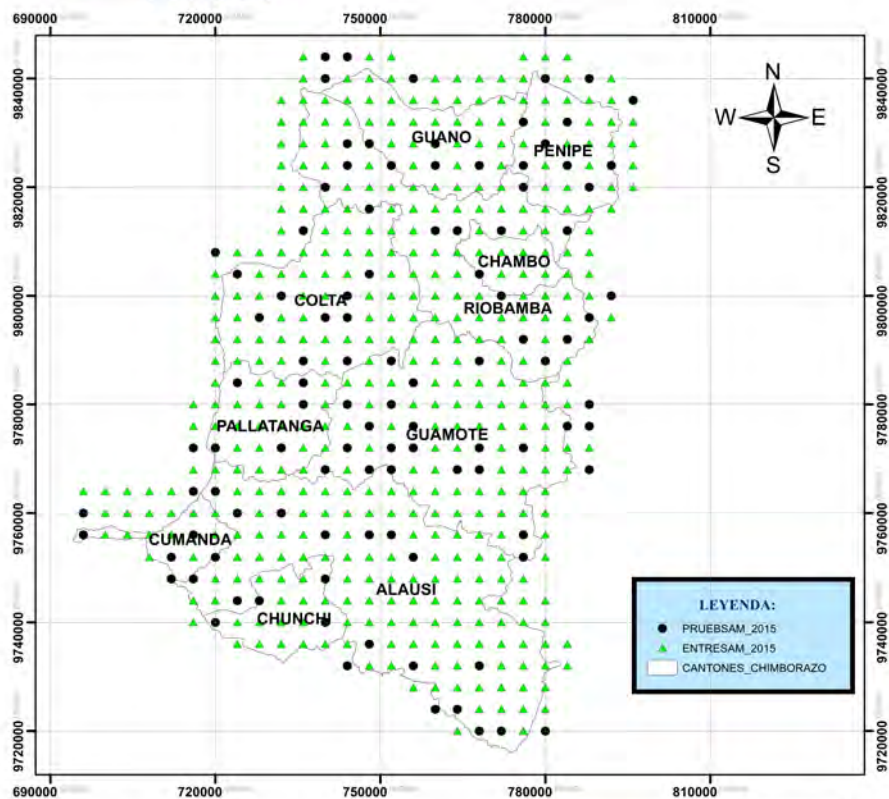


Ilustración 4-25: (a) Localización de las estaciones Meteorológicas y (b) Puntos descargados del SAM de la provincia de Chimborazo.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5.1. *Kriging – SAM 2015*

Para aplicar el método geoestadístico Kriging Universal, no es un requisito que los datos provengan de una distribución normal.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4–26: Identificación de los datos de entrenamiento y prueba SAM-2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

Se trabajó de dos formas con tendencia de segundo grado y sin tendencia, empleando 3 tipos de semivariogramas: gaussiano, exponencial y esférico. El primer paso es establecer un semivariograma basado en los datos.

4.5.1.1. *Semivariogramas de Segundo orden*

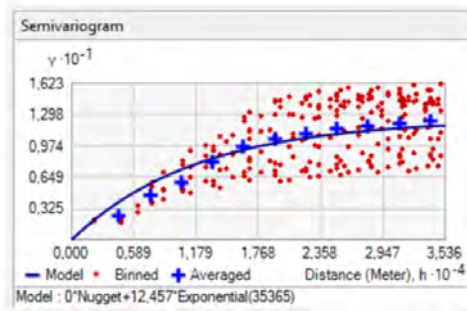
El semivariograma consta de dos partes: experimental y modelo. El variograma experimental se obtiene calculando la varianza de cada punto del conjunto con respecto a todos los demás puntos y trazando la varianza en función de la distancia entre puntos. Luego de calcular el variograma experimental, el siguiente paso es definir el variograma modelo que es una función matemática simple que se utiliza para modelar la tendencia de los variogramas experimentales (Mohd Aziz et al. 2019). Los resultados de los tres modelos de semivariogramas se muestran en la tabla 4-6 comparando el valor de pepita, umbral parcial y el rango.

Tabla 4-6: Comparación de los modelos de semivariograma exponencial, gaussiano y esférico.

	Exponencial	Gaussiano	Esférico
Pepita	0	1.0841	0
Umbral	12.4572	10.8093	12.0726
Rango	35.364	22.147	28.838

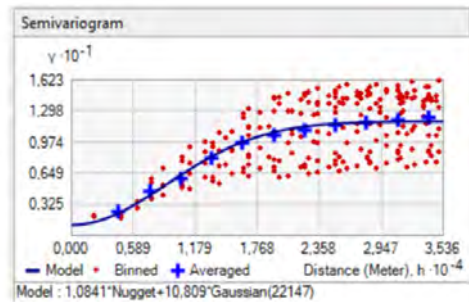
Realizado por: Maldonado J., y Yugsi D., 2024.

En la ilustración 4-27 se observan los modelos teóricos de los semivariogramas con tendencia de grado dos que se ajustan a los promedios, es decir, entre más próximos estén el margen de error es mínimo.



(a)

(b)



(c)

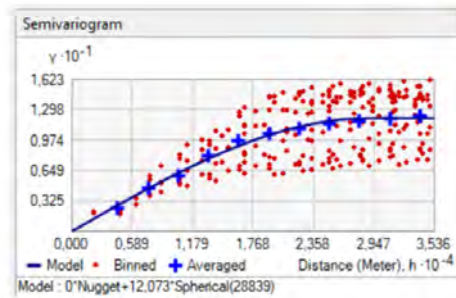


Ilustración 4–27: Ajuste de los datos del SAM a tres tipos de modelos: (a) exponencial, (b) gaussiano y (c) esférico.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5.1.2. Semivariogramas con tendencia constante

La Tabla 4-7 contiene los resultados de cada semivariograma comparando el valor de pepita, el umbral parcial y el rango. Se realizó una serie de análisis adicionales para decidir qué modelo es correcto utilizando un esquema de validación cruzada.

Tabla 4–7: Comparación de los modelos de semivariograma exponencial, gaussiano y esférico.

	Exponencial	Gaussiano	Esférico
Pepita	0	6.0741	1.2174
Umbral	41.4226	31.5889	38.7448
Rango	159.298	86.219	112.167

Realizado por: Maldonado J., y Yugsi D., 2024.

En la ilustración 4-28 se observan los modelos teóricos con tendencia constante de los semivariogramas que se ajustan a los promedios, es decir, entre más próximos estén el margen de error es mínimo.

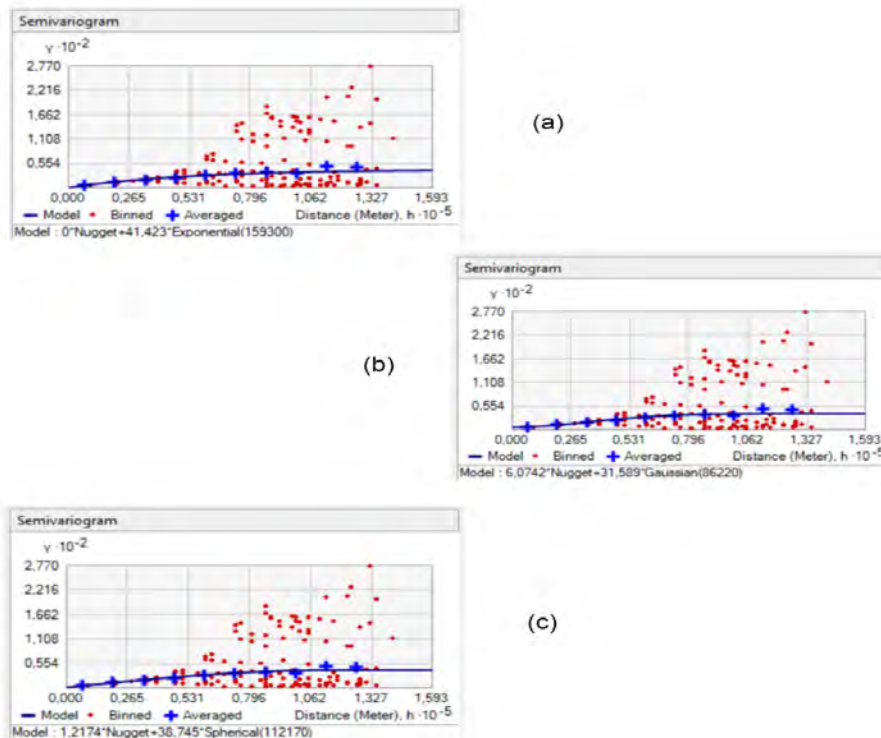
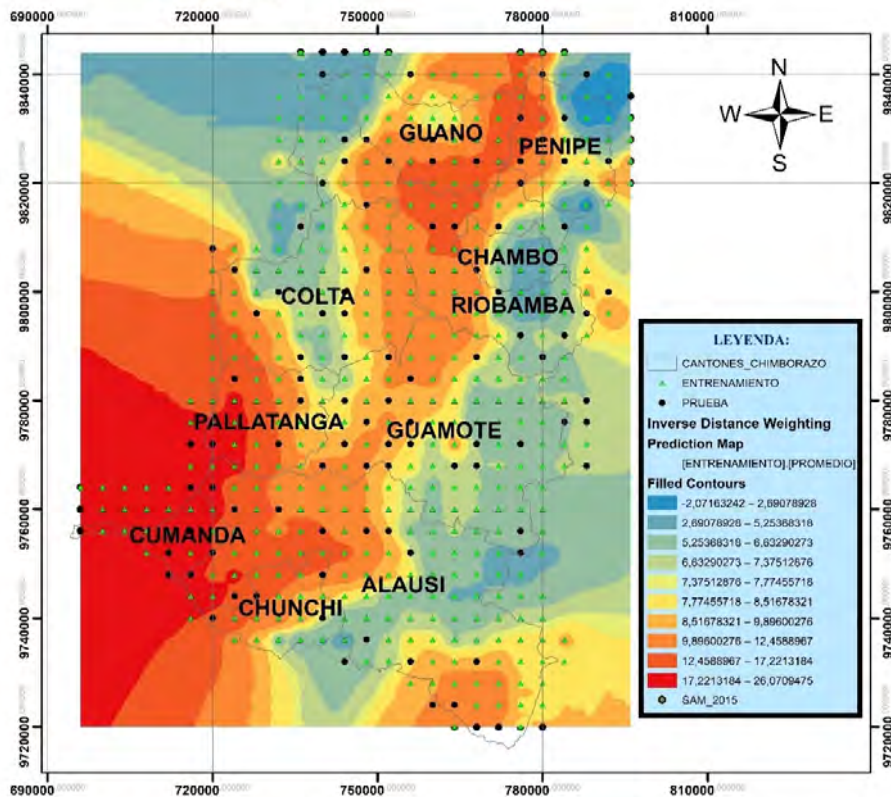


Ilustración 4-28: Ajuste de los datos del SAM a tres tipos de modelos: (a) exponencial, (b) gaussiano y (c) esférico.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5.2. IDW (Distancia Inversa Ponderada) Datos del SAM

Es un método que estima promedios a partir de los valores de los puntos muestreados, es decir, se asigna más peso a los puntos más cercanos y menos peso a los más lejanos y se realizó la interpolación IDW con el conjunto de datos de entrenamiento del SAM correspondiente al año 2015, posteriormente una vez realizado la interpolación se generó una capa ráster en función de los datos que se utilizaron indicado en la ilustración 4-29.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi

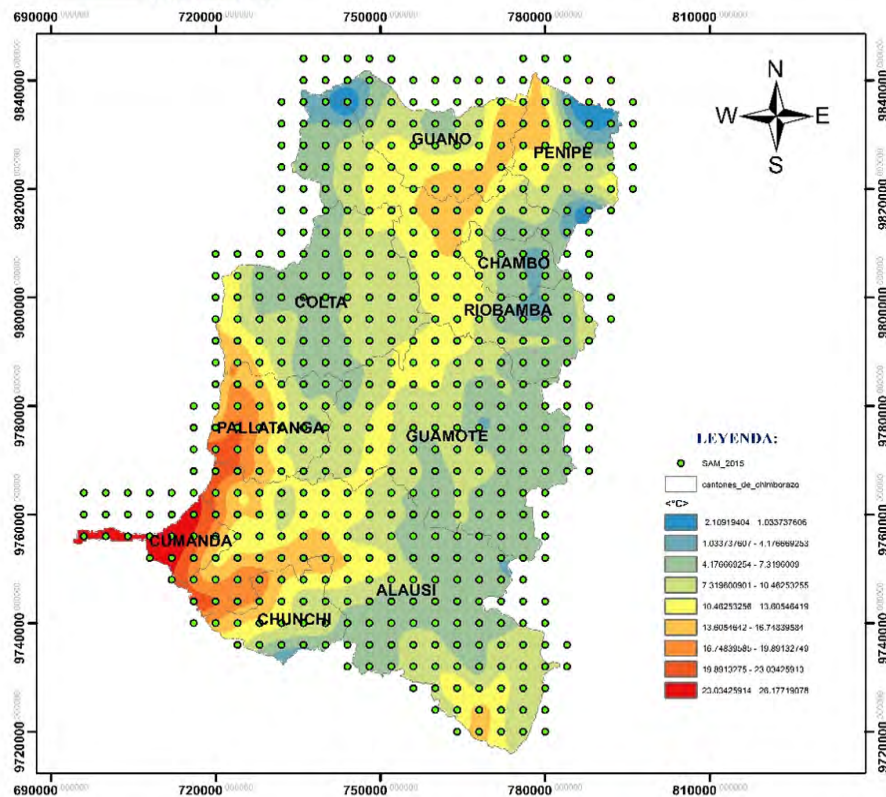


Ilustración 4–29: Interpolación de IDW con datos de entrenamiento SAM 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5.3. *Spline Datos del SAM*

El Spline se caracteriza por su curva suave que se emplea para aproximar o interpolar un conjunto de datos, se realizó la interpolación indicando que en el valle interandino la temperatura oscila entre los 10°C y 13°C incrementando en el cantón Cumandá por el cambio de región como se observa en la ilustración 4-30.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Sigüencia
Dennis Adrian Yugsi Yugsi

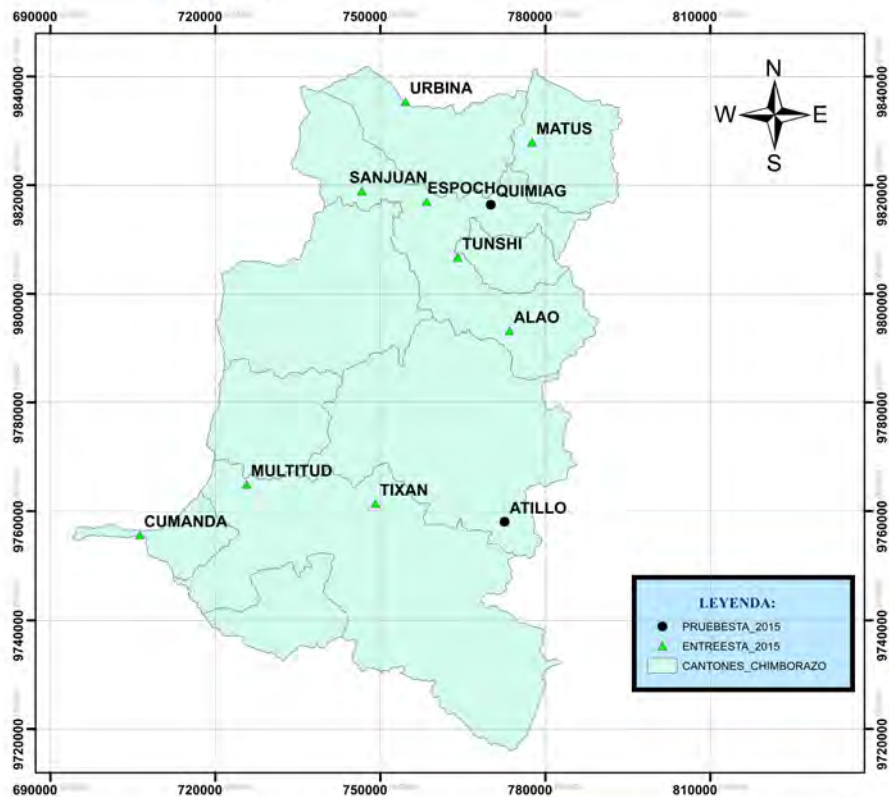


Ilustración 4–30: Interpolación con Spline de los datos del SAM 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5.4. Kriging – Estaciones Meteorológicas 2015

Los métodos geoestadísticos son eficientes cuando se implementan datos que siguen una distribución normal. Por ende, el método geoestadístico Kriging Ordinario brindará una alta exactitud cuando los datos cumplan el supuesto de normalidad (Rohma 2023) como indica la ilustración 4-31.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4–31: Estaciones meteorológicas de entrenamiento y prueba 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

Se trabajó de dos formas con tendencia de primer grado y sin tendencia, utilizando 3 tipos de semivariogramas: gaussiano, exponencial y esférico. Además, cabe recalcar que para realizar la interpolación de igual manera se trabajó con datos de entrenamiento y de validación.

4.5.4.1. *Semivariogramas de primer orden*

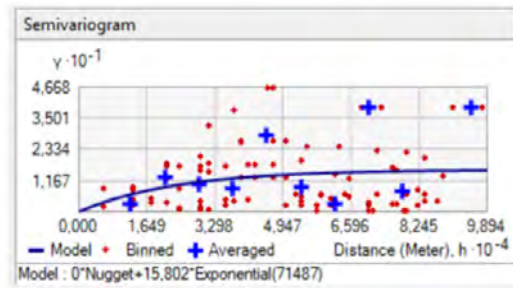
La Tabla 4-8 indica los valores del modelo de cada semivariograma de la comparación de: pepita, el umbral parcial y el rango.

Tabla 4-8: Comparación de los modelos de semivariograma exponencial, gaussiano y esférico.

	Exponencial	Gaussiano	Esférico
Pepita	0	7.6704	4.8817
Umbral	15.8015	7.6771	10.2327
Rango	71.486	71.486	71.486

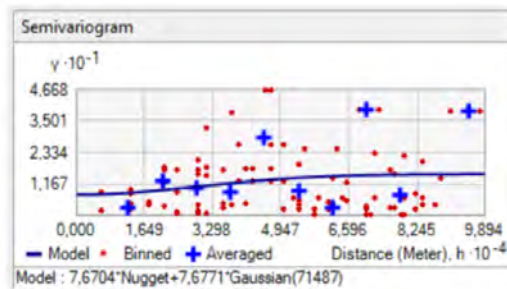
Realizado por: Maldonado J., y Yugsi D., 2024.

En la ilustración 4-32 se presentó el modelo teórico con tendencia de primer orden de los semivariogramas notándose una aproximación a los promedios, lo cual indica que entre más próximos estén el error es menor, caso contrario el error se incrementa.



(a)

(b)



(c)

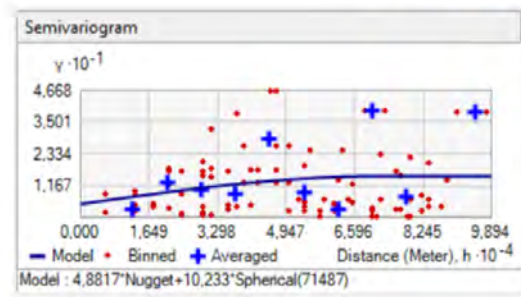


Ilustración 4–32: Ajuste de los datos de las estaciones meteorológicas a tres tipos de modelos: (a) exponencial, (b) gaussiano y (c) esférico.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5.4.2. Semivariogramas con tendencia constante

La Tabla 4-9 indica los valores del modelo de cada semivariograma de la comparación de: pepita, el umbral parcial y el rango.

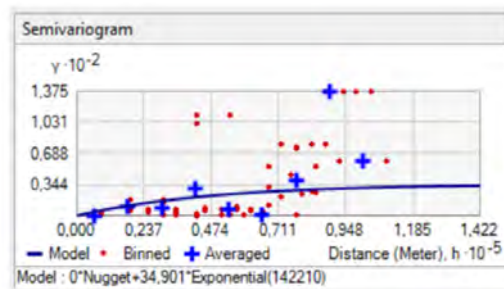
Tabla 4–9: Comparación de los modelos de semivariograma exponencial, gaussiano y esférico.

	Exponencial	Gaussiano	Esférico
Pepita	0	0.9491	0
Umbral	34.9009	32.3501	44.5779
Rango	142.211	71.486	142.211

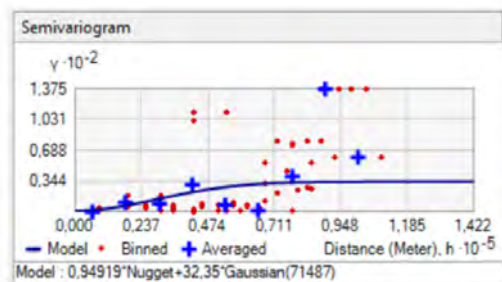
Realizado por: Maldonado J., y Yugsi D., 2024.

En la ilustración 4-33 se presentó el modelo teórico con tendencia constante de los semivariogramas

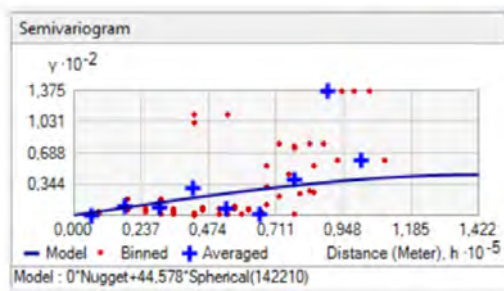
notándose una aproximación a los promedios, lo cual indica que entre más próximos estén el error es menor, caso contrario el error se incrementa.



(a)



(b)



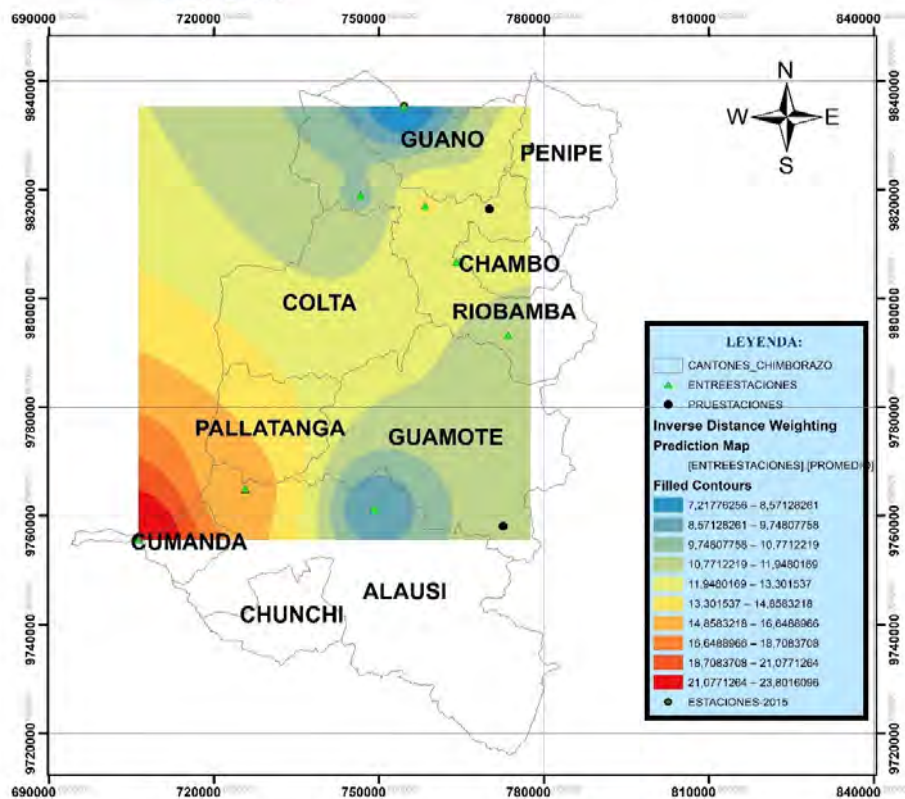
(c)

Ilustración 4–33: Tres tipos de modelos de semivariograma ajustados a los datos:
(a) exponencial, (b) gaussiano y (c) esférico.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5.5. IDW (*Distancia Inversa Ponderada*) Datos Estaciones

La interpolación del método IDW con el conjunto de datos de las estaciones meteorológicas correspondiente al año 2015, una vez realizado la interpolación se generó una capa raster en función de los datos que se utilizaron indicado en la ilustración 4-34.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Sigüencia
Dennis Adrian Yugsi Yugsi

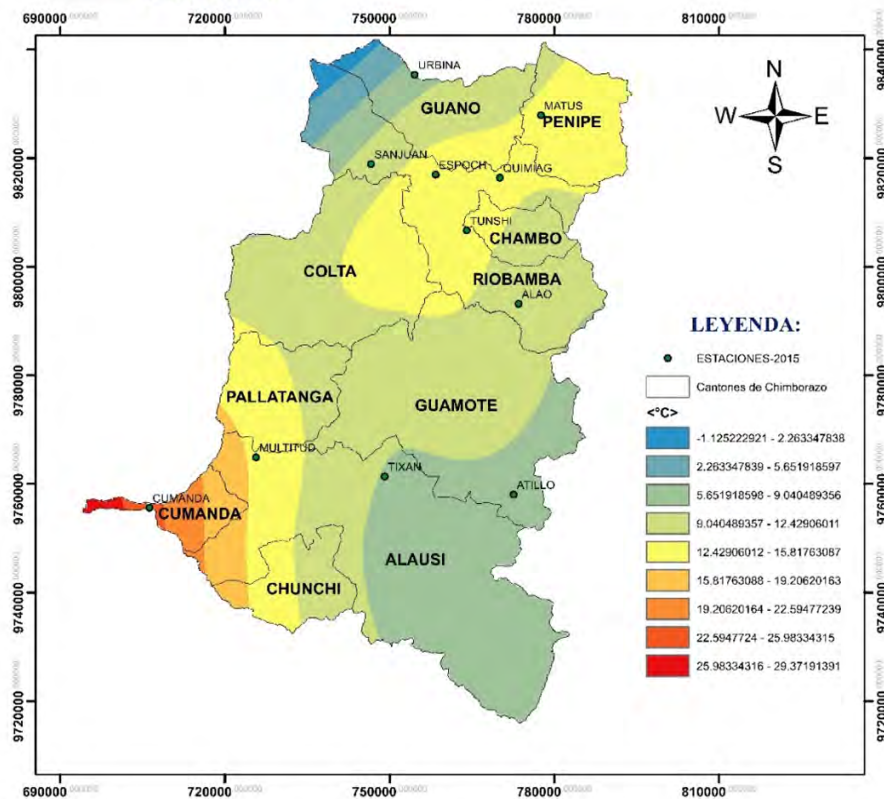


Ilustración 4-34: Interpolación IDW con los datos de las estaciones meteorológicas 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.5.6. Spline Datos Estaciones

Se realizó la interpolación Spline con los datos de las estaciones meteorológicas del año 2015. En la ilustración 4-35 se detectó temperaturas altas en el cantón Cumandá, por otra parte, para el noroeste del cantón Guano y Riobamba se observaron bajas temperaturas.



Escuela Superior Politécnica de Chimborazo
 Facultad de Ciencias
 Carrera de Estadística



Elaborado por:
 James William Maldonado Sigüencia
 Dennis Adrian Yugsi Yugsi



Ilustración 4–35: Interpolación Spline con los datos de las estaciones meteorológicas 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.6. Evaluación de los Métodos

4.6.1. Datos descargados del SAM

Se llevó a cabo validación cruzada, que consistió en la partición de los datos en un 80% para el entrenamiento y un 20% para la prueba. Esta metodología contribuye significativamente a la

obtención de estimaciones más robustas.

Para la evaluación de estos métodos se empleó diferentes criterios de validación tales como: el error medio (EM), raíz del error cuadrático medio (RMSE) y el error medio estandarizado (SEM). Los resultados se presentan en la tabla 4-10, donde se observó que el Kriging Universal con tendencia de segundo grado y semivariograma exponencial demostró el mejor desempeño con un EM de 0.0307, RMSE de 1.1727 y SEM de 2.0107, sin embargo, para el semivariograma Gaussiano presento un EM de 0.0586, RMSE de 1.2536 y SEM de 1.3086; mientras que con semivariograma Esférico un EM de 0.0306, RMSE de 1.1656 y SEM de 1.5759.

En el caso del Kriging Universal con tendencia constante, utilizando un semivariograma exponencial se obtuvo un EM de -0.0115, RMSE de 1.1000 y SEM de 1.6464, con semivariograma Gaussiano, un EM de 0.0098, RMSE de 1.8527 y SEM de 2.5582, mientras que con semivariograma Esférico, un EM de -0.0105, RMSE de 1.2120 y SEM de 1.8246.

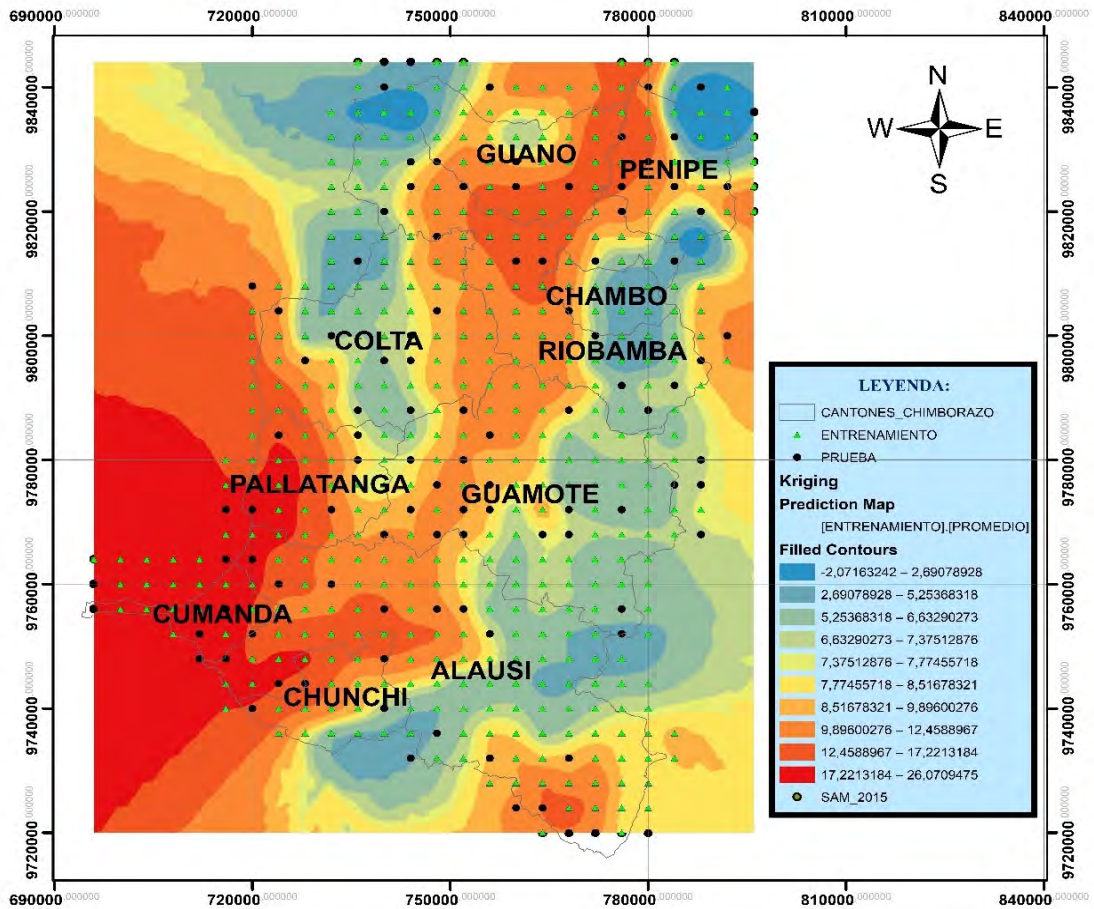
Posteriormente, utilizando la interpolación IDW, se registraron valores de EM de -0.0612, RMSE de 1.4934, finalmente para la interpolación con el método del Spline se utilizó el software estadístico R-Studio, presentando EM de -0.1132, RMSE de 11.0286 y SEM de -0.0229.

Tabla 4–10: Comparación con los datos del SAM.

Variable	Método	EM			RMSE	EM estandarizada
Temperatura	Kriging Tendencia segundo grado	Exponencial	0.0307	1.1727	2.0107	
		Gaussiano	0.0586	1.2536	1.3086	
		Esférico	0.0306	1.1656	1.5759	
	Constante	Exponencial	-0.0115	1.1000	1.6464	
		Gaussiano	0.0098	1.8527	2.5582	
		Esférico	-0.0105	1.2120	1.8246	
	IDW	-0.0612			1.4934	-
SPLINE	-0.1132			11.0286	-0.0229	

Realizado por: Maldonado J., y Yugsi D., 2024.

Al analizar los errores del Kriging Universal con los datos descargados (SAM), en la ilustración 4-36 se observó que el semivariograma exponencial del Kriging con tendencia constante mostro el mejor rendimiento con un error medio EM -0.0115, raíz del error cuadrático medio RMSE de 1.1000 y el error medio estandarizado SEM. de 1.6464. Con estos resultados, se procedió a trabajar con los siguientes años.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4-36: Interpolación de Kriging con un semivariograma Exponencial de los datos del SAM 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.6.2. Datos de las estaciones

Para la evaluación de estos métodos se empleó diferentes criterios de validación tales como: el error medio (EM), raíz del error cuadrático medio (RMSE) y el error medio estandarizado (SEM). Los resultados se presentan en la tabla 4-11, donde se observó que el Kriging Ordinario con tendencia de primer grado y semivariograma exponencial demostró el mejor desempeño con un EM de -0.1244, RMSE de 5.0644 y SEM de 3.3965, sin embargo, para el semivariograma Gaussiano presento un EM de -0.1954, RMSE de 5.2126 y SEM de 3.4852; mientras que con semivariograma Esférico un EM de -0.1756, RMSE de 5.1008 y SEM de 3.4448.

En el caso del Kriging Ordinario con tendencia constante, utilizando un semivariograma exponencial se obtuvo un EM de -0.0919, RMSE de 4.3171 y SEM de 3.9278, con semivariograma Gaussiano, un EM de -0.1658, RMSE de 2.9346 y SEM de 2.7368, mientras que con semivariograma Esférico, un EM de -0.0969, RMSE de 4.0853 y SEM de 3.4077.

Posteriormente, utilizando la interpolación IDW, se registraron valores de EM de -0.0908, RMSE de 4.4661, finalmente para la interpolación con el método del Spline se utilizó el software estadístico R-Studio, presentando EM de 0.7051, RMSE de 19.5328 y SEM de 0.1423.

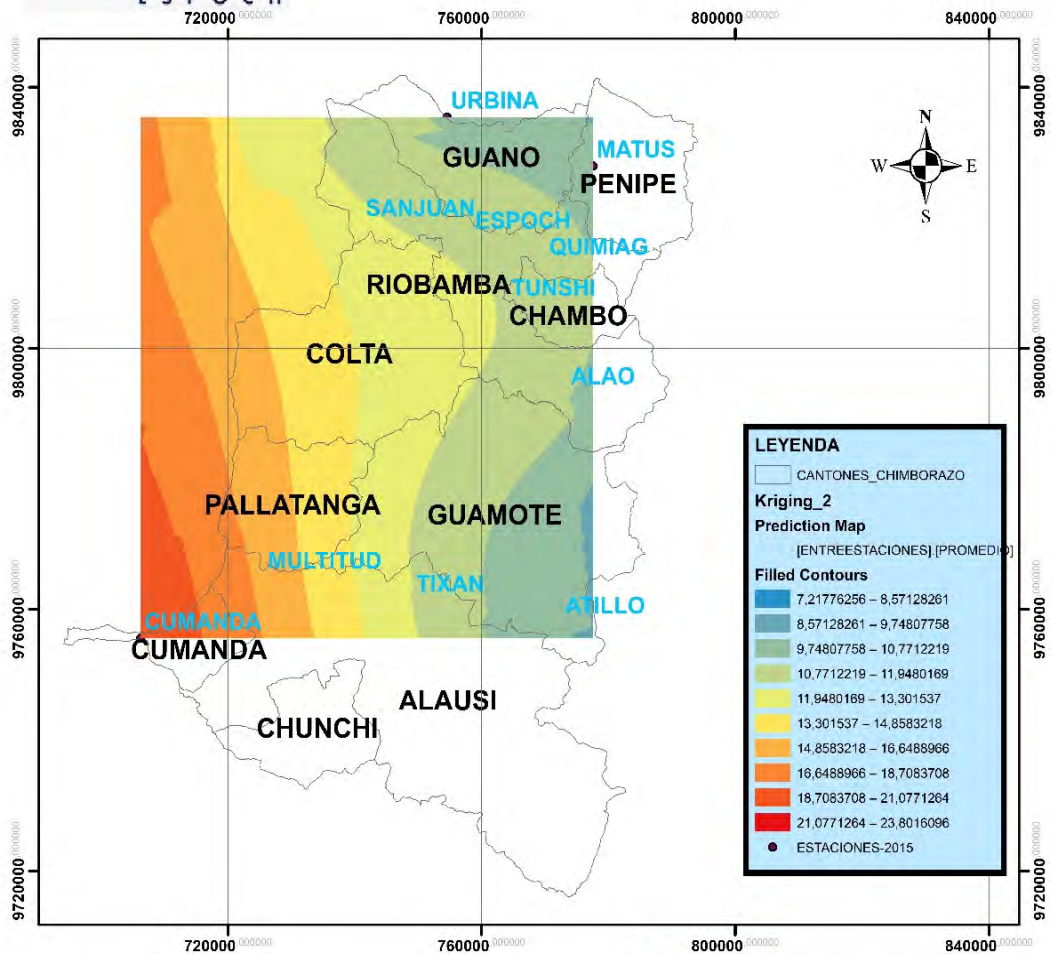
Tabla 4–11: Comparación con los datos de las Estaciones.

Variable	Método	EM			RMSE	EM estandarizada
Temperatura	Kriging	Tendencia segundo grado	Exponencial	-0.1244	5.0644	3.3965
			Gaussiano	-0.1954	5.2126	3.4852
			Esférico	-0.1756	5.1008	3.4448
	Constante	Exponencial	-0.0919	4.3171	3.9278	
		Gaussiano	-0.1658	2.9346	2.7368	
		Esférico	-0.0969	4.0853	3.4077	
	IDW	-0.0908			4.4661	-
Spline	0.7051			19.5328	0.1423	

Realizado por: Maldonado J., y Yugsi D., 2024.

Al analizar los errores del Kriging Ordinario con los datos de las estaciones Meteorológicas, en la ilustración 4-37 se observó que el semivariograma Gaussiano del Kriging con tendencia constante mostro el mejor rendimiento con un error medio EM de -0.1658, raíz del error cuadrático medio RMSE de 2.9346 y el error medio estandarizado SEM. de 2.7368. Con estos resultados, se procedió a trabajar con los siguientes años.

MAPAS DE TEMPERATURA



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4–37: Interpolación de kriging con un semivariograma Gaussiano para los datos de las estaciones.

Realizado por: Maldonado J., y Yugsi D., 2024.

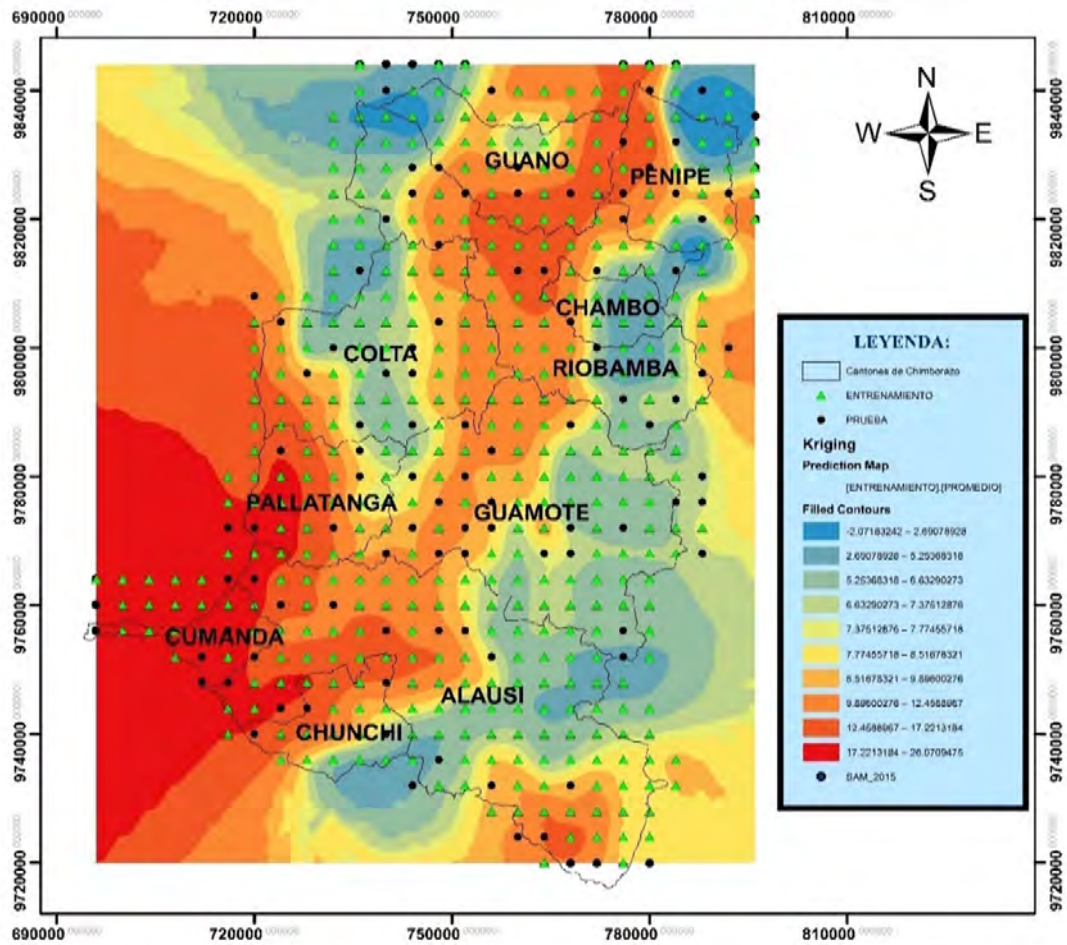
Después de llevar a cabo un análisis detallado para el año 2015 y determinar el método de interpolación más adecuado, se procedió a realizar el mismo estudio, para los años subsiguientes,

abarcando el periodo 2016 - 2020.

4.7. Mapas de Interpolación

4.7.1. *Mapas Datos SAM*

En las ilustraciones 4-38, 4-39, 4-40, 4-41, 4-42, 4-43 se presentan los mapas de interpolación de Kriging Universal con el semivariograma exponencial con tendencia constante de la variable temperatura de los datos del SAM-2015 perteneciente a los modelos CMIP6.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística

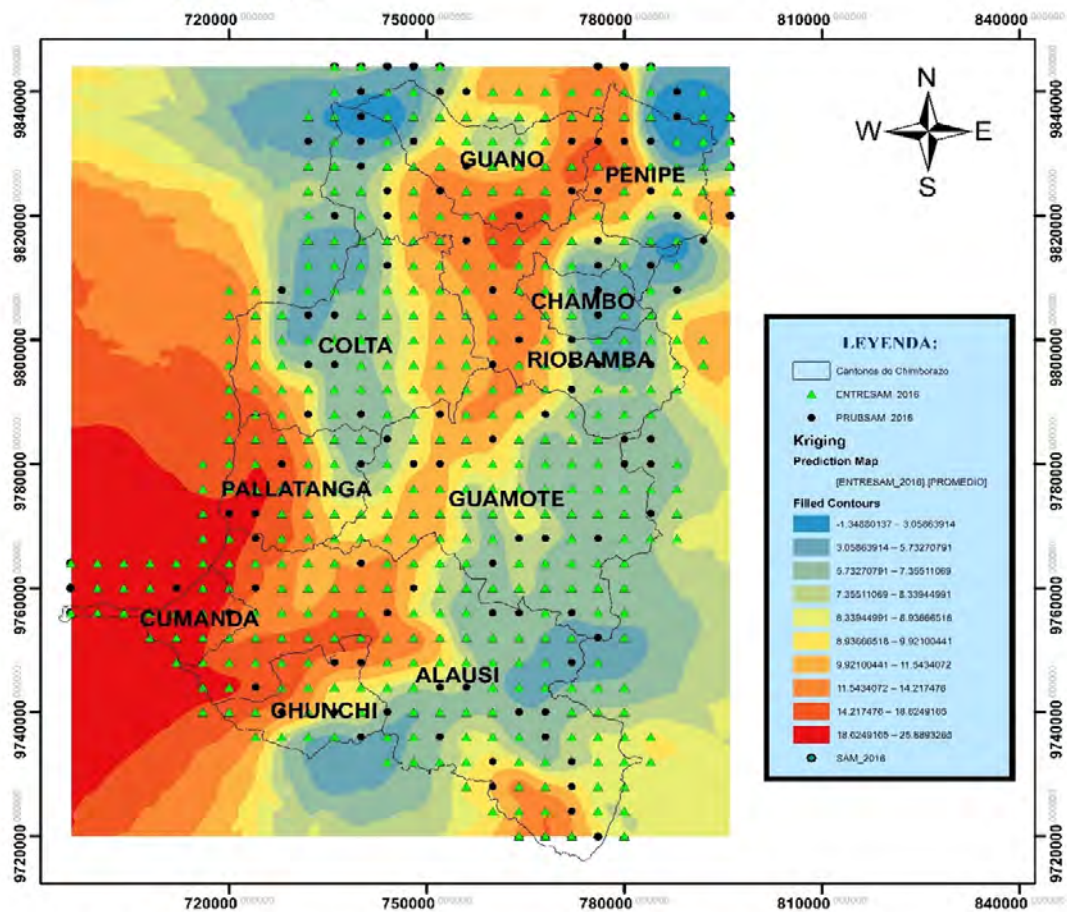


Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4-38: Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2015.

Realizado por: Maldonado J., y Yugsi D., 2024.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística

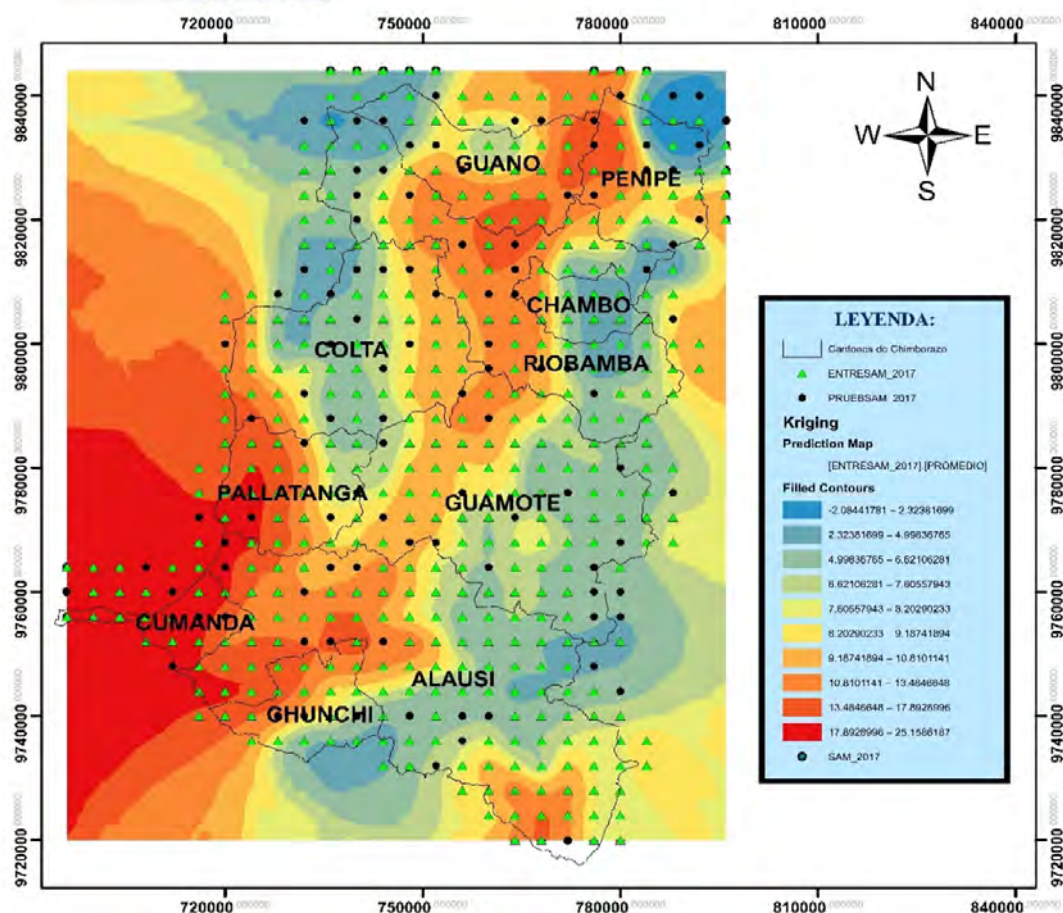


Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4-39: Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2016.

Realizado por: Maldonado J., y Yugsi D., 2024.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística

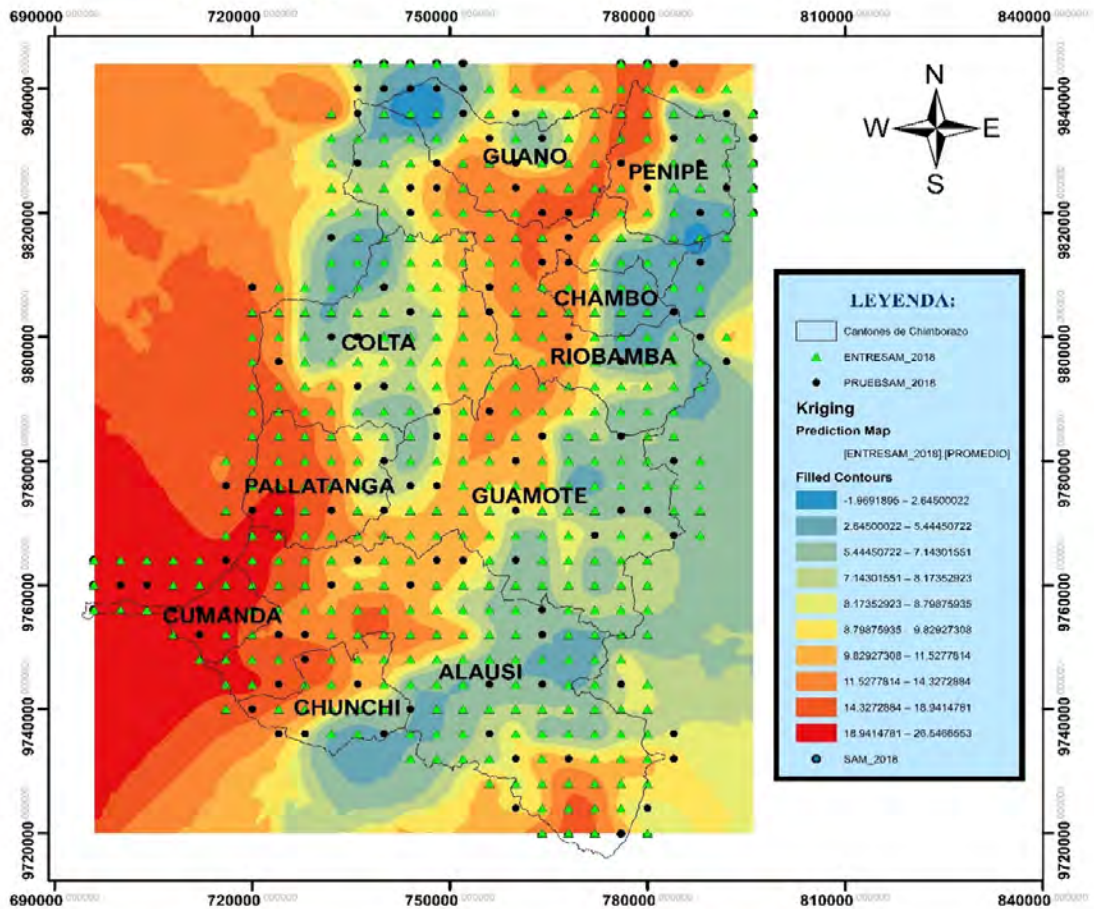


Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4-40: Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2017.

Realizado por: Maldonado J., y Yugsi D., 2024.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística

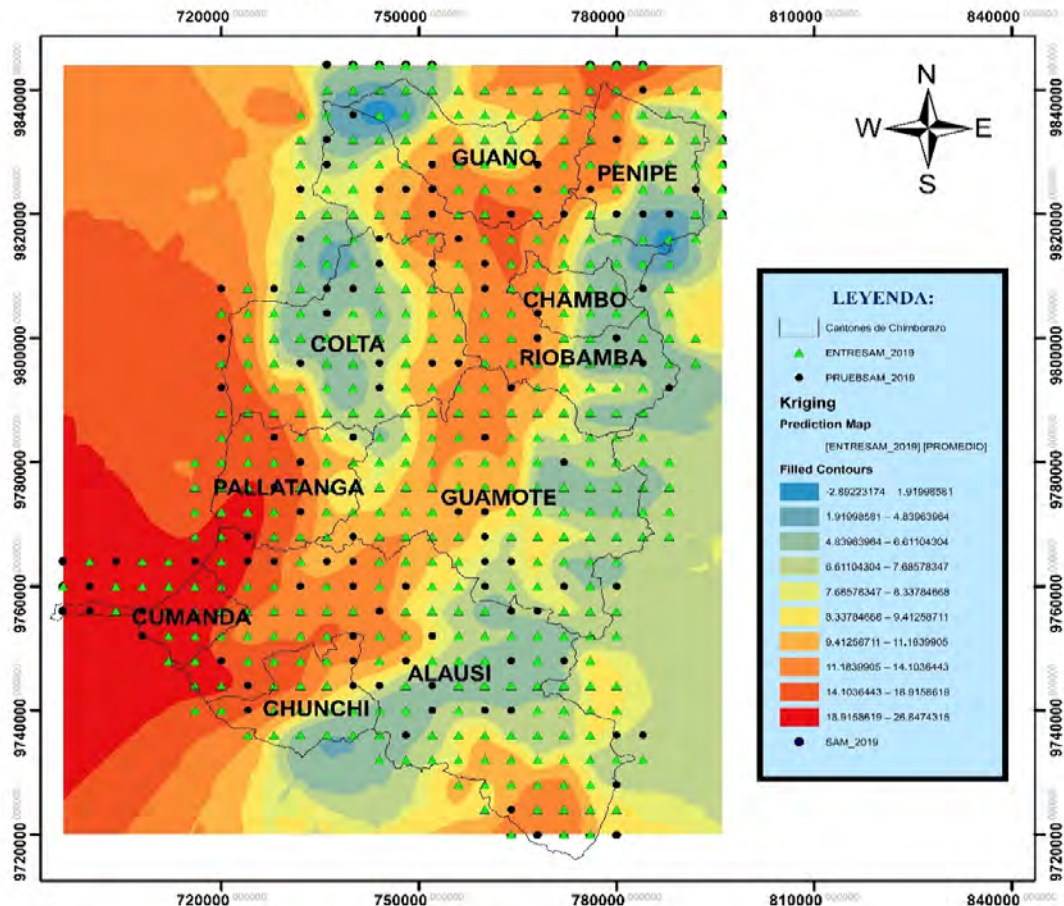


Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4-41: Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2018.

Realizado por: Maldonado J., y Yugsi D., 2024.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística

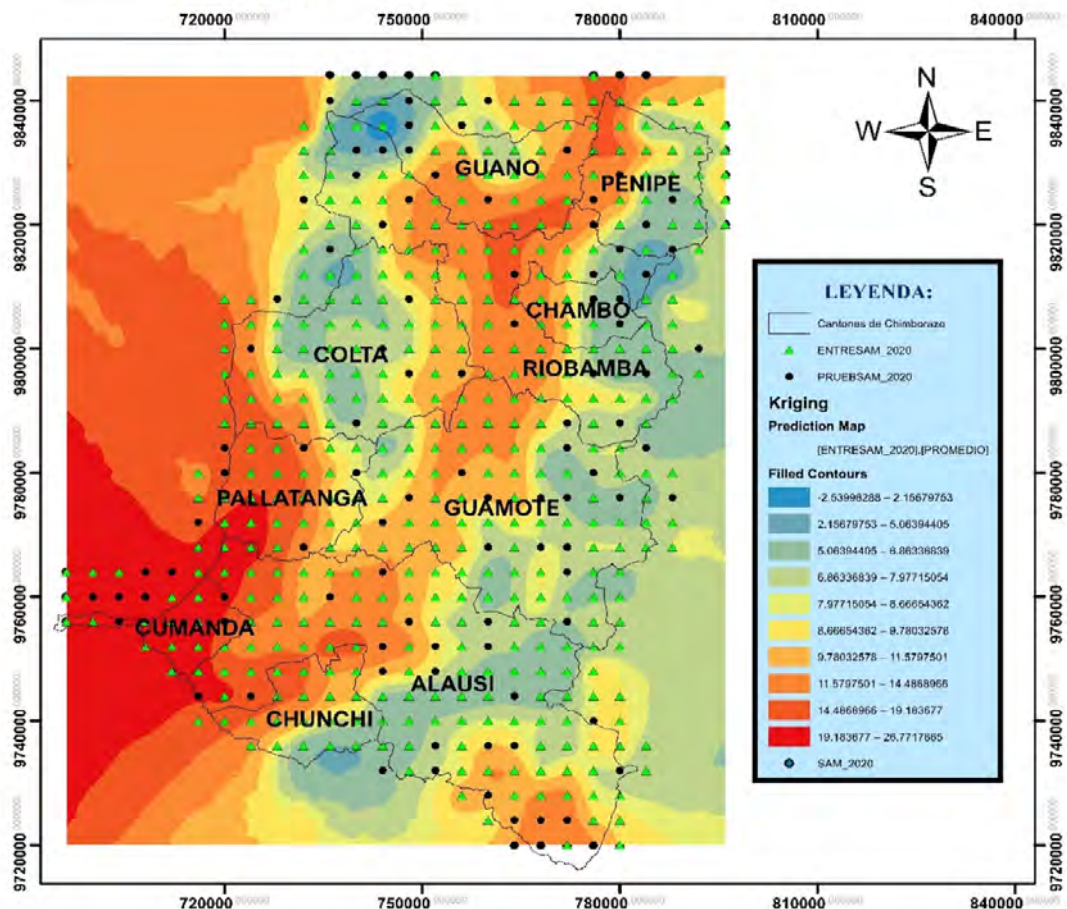


Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4-42: Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2019.

Realizado por: Maldonado J., y Yugsi D., 2024.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Siguencia
Dennis Adrian Yugsi Yugsi



Ilustración 4–43: Mapa de Temperatura de Kriging Universal con semivariograma exponencial de los datos del SAM 2020.

Realizado por: Maldonado J., y Yugsi D., 2024.

Se compararon las temperaturas durante un período de cinco años (2015-2020) utilizando los datos del SAM. Se aplicó la prueba no paramétrica de de Wilcoxon, para determinar si existe diferencia

significativa en la variable de estudio entre los cantones: Riobamba y Alausí; con un nivel de significancia de 0.05. A un nivel de confianza del 95 % se rechazó la hipótesis nula (H0), lo que indica que hay una diferencia significativa de temperaturas en los años analizados.

Según la tabla 4-12, la temperatura promedio del cantón Alausí aumento 0.63°C entre el 2015 y el 2020, pasando de 10.08°C a 10.70°C. En el mismo periodo, el cantón Riobamba experimento un incremento mayor de 0.98°C, de 8.53°C a 9.51°C. Este aumento térmico podría deberse a factores como el crecimiento poblacional, la disminución de las zonas verdes y el cambio climático.

Tabla 4–12: Cuadro comparativo de Temperaturas.

Cantón Alausí promedio 2015	Cantón Alausí promedio 2020	Diferencia
10.07634	10.7031	0.6268
Cantón Riobamba promedio 2015	Cantón Riobamba promedio 2020	Diferencia
8.5306	9.5111	0.9806

Realizado por: Maldonado J., y Yugsi D., 2024.

4.7.2. Mapas de las Estaciones

En las ilustraciones 4-44, 4-45 se presentan los mapas de interpolación de Kriging Ordinario con el semivariograma gaussiano con tendencia constante de la variable temperatura promedio de los datos de las estaciones meteorológicas dentro de la provincia de Chimborazo.

MAPAS DE TEMPERATURA

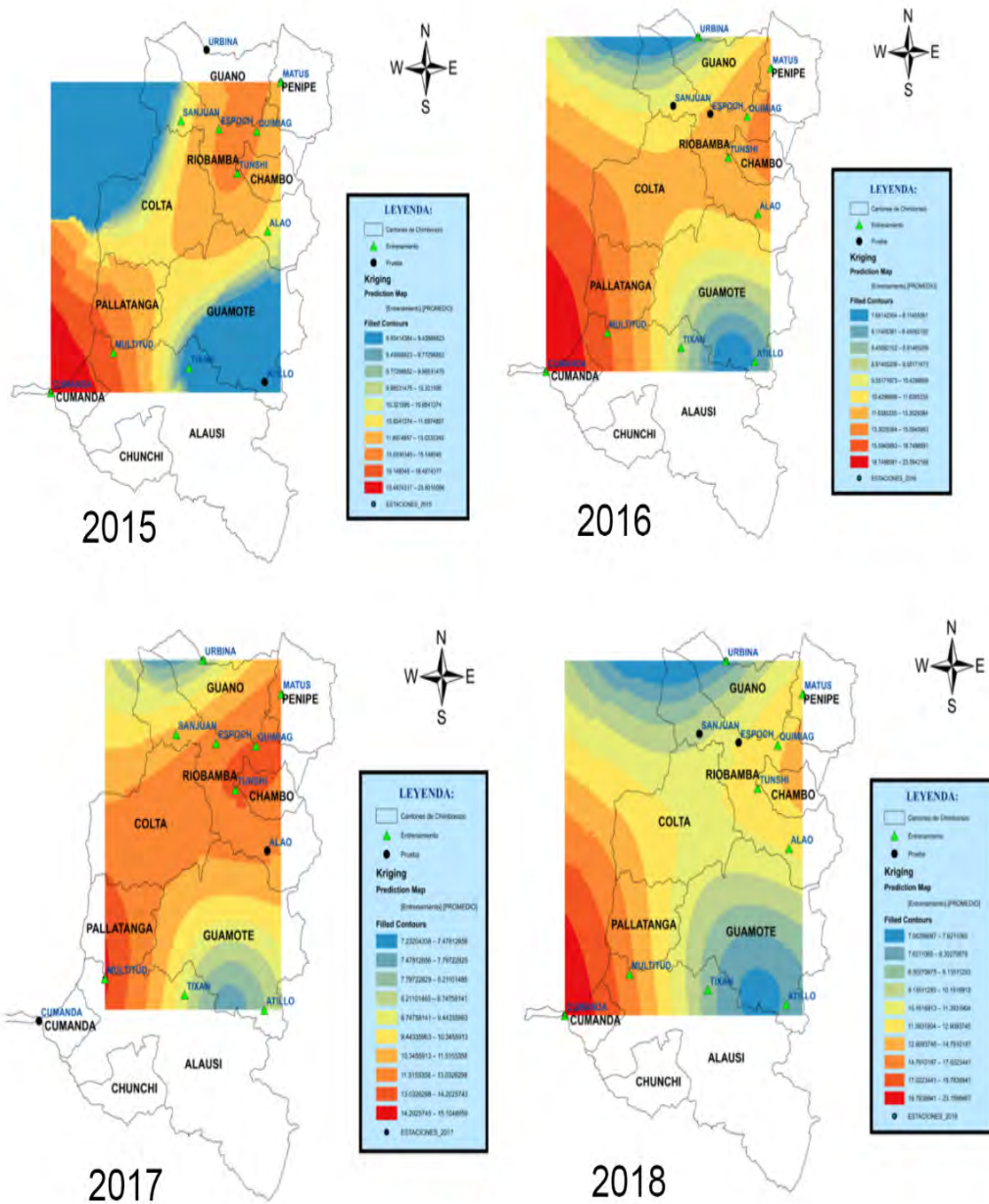
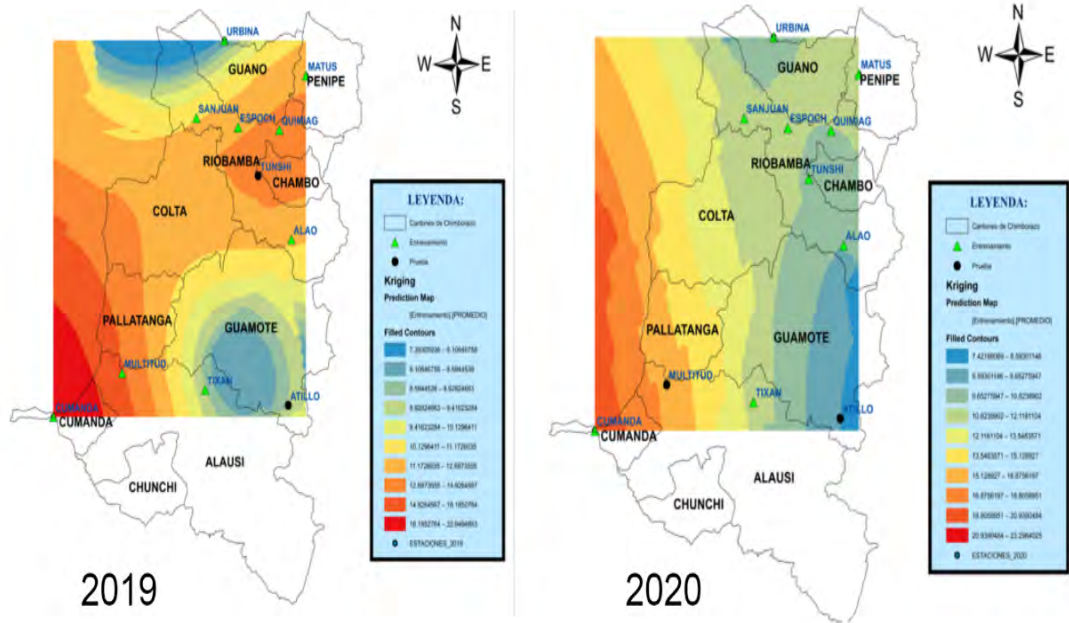


Ilustración 4-44: Interpolación de Kriging 2015-2018.

Realizado por: Maldonado J., y Yugsi D., 2024.



Escuela Superior Politécnica de Chimborazo
Facultad de Ciencias
Carrera de Estadística



Elaborado por:
James William Maldonado Sigüencia
Dennis Adrian Yugsi Yugsi



Ilustración 4-45: Interpolación de Kriging 2019-2020.

Realizado por: Maldonado J., y Yugsi D., 2024.

4.8. Discusión (Argumentación, sistematización y cuestionamiento de la información obtenida)

Se comparo las diferentes técnicas de interpolación espacial para la variable temperatura ambiente de los modelos atmosféricos CMIP6 y de las estaciones meteorológicas. Los métodos de interpolación determinísticos como el IDW y Spline determinan un valor para cada celda en la cuadrícula de la malla de datos utilizando exclusivamente las características físicas de las muestras de elevación.; lo que implica la creación de superficies basadas en los puntos medidos y en fórmulas matemáticas (Chirinos y Mallqui 2016) , mientras que los métodos de interpolación probabilísticos como el Kriging emplea un número predeterminado de puntos utilizando modelos de autocorrelación en la densidad o a su vez considera todos los puntos dentro de un radio determinado para calcular el valor correspondiente en cada ubicación (Chirinos y Mallqui 2016), este método reduce los errores tanto para los datos de las estaciones meteorológicas como para modelos atmosféricos.

Se buscaron métodos más robustos para trabajar con los datos del SAM pertenecientes a los modelos CMIP6 como lo sugiere (Reyes 2017), en el caso de no cumplir normalidad, se lleva a cabo transformaciones en los datos como: Box-Cox, raíz cuadrada, logaritmo, $1/x$, entre otras. (<https://aguaysig.com/analisis-geoestadistico-con-arcgis-parte-2-analisis-exploratorio-de-los-datos/>). Para representar el comportamiento de la variable temperatura, siguiendo la recomendación de la teoría de (Saltos 2016), quien sostiene que no es necesario que los datos provengan de una distribución normal al aplicar el método kriging universal. En términos de los métodos de interpolación, de acuerdo con el estudio de (Mohd Aziz et al. 2019) identificó que los semivariogramas más apropiados para ajustarse a los datos de temperatura son: Esférico, Gaussiano y Exponencial. Además, según (Chang 2019), la importancia de elegir el mejor método de interpolación espacial considerando como norma general los estadísticos de error, favoreciendo aquellos que exhiban los valores más bajos en relación con el error cuadrático medio (RMSE) y el error medio absoluto (MAE).

Para los datos de las estaciones estudiadas se optó por utilizar kriging ordinario, ya que este método permite trabajar con datos provenientes de una distribución normal. Sin embargo, en la provincia de Chimborazo existen pocas estaciones y los errores en las interpretaciones de las temperaturas anuales podrían ser mayores, como se refleja en la tabla de comparaciones (Tabla 4-9). Cabe mencionar que como establece (Vicencio 2018), la ubicación de las estaciones emerge como un factor crucial a tener en cuenta, ya que están influenciadas por las características del terreno, presentando variaciones en elevación y distancias extensas.

En la provincia de Chimborazo, en los cantones de Riobamba y Alausí, las temperaturas registradas para el año 2015 revelaron cifras de 8.5306 °C y 10.0763°C. Para el año 2020, se evidenció crecimientos de 9.5111°C y 10.7031°C respectivamente. Este incremento, de 0.9806°C y 0.6268°C, son valores que coinciden con lo que reporta Bustamante para la provincia de Chimborazo, existe un incremento de la variable que contrasta con los modelos globales que indican un incremento generalizado (Bustamante 2017).

El incremento de la temperatura es de 0.9806°C en el cantón Riobamba y de 0.6268°C en el cantón Alausí como se observa de forma espacial como menciona Bustamante. Los cambios de temperatura tendrán impactos directos en la distribución de los ecosistemas naturales restantes, la productividad de los cultivos y los patrones de distribución de las especies que transmiten enfermedades tropicales, lo que podría aumentar su distribución en altitud bajo las nuevas condiciones (Bustamante 2017).

CONCLUSIONES

Se realizó un análisis descriptivo de los datos de temperatura del SAM y de las estaciones, identificando media, mediana y moda. Se visualizó el comportamiento de los datos con histogramas (distribución leptocúrtica y asimétrica a la derecha), se visualizó diagramas de caja y se aplicaron pruebas de normalidad de Kolmogorov-Smirnov y Shapiro-Wilks corroborando con los gráficos Q-Q Plot, determinando que los datos del SAM no provienen de una distribución normal a diferencia con los datos de las estaciones en donde si proviene de una distribución normal con una frecuencia anual.

Se seleccionaron diferentes métodos de interpolación espacial para la variable temperatura: dos métodos determinísticos Distancia Inversa Ponderada (IDW) y Spline, basándose en valores medidos más cercanos determinando la suavidad de la superficie resultante y un probabilístico Kriging ya sea el universal para datos que no provienen de una distribución normal y el ordinario para datos que, si provienen de una distribución normal.

Se comparó las técnicas de interpolación espacial IDW, Splines y Kriging para los datos del SAM y de las estaciones meteorológicas. El método óptimo y el que mejor se ajusta para interpolar datos de la variable temperatura y el que presenta menores errores fue el Kriging en el caso de los datos del SAM Kriging Universal con semivariograma exponencial y tendencia constante presentó un EM de -0.01, SEM de 1.64 y RSME de 1.10 para el SAM, mientras que para los datos de las estaciones meteorológicas el método de Kriging Ordinario con semivariograma gaussiano y tendencia constante presentó un EM de -0.16, SEM de 2.74 y RSME de 2.93.

Una vez determinada la técnica de interpolación espacial Kriging como la más efectiva tomando en cuenta el semivariograma y la tendencia específica tanto para los datos del SAM como para las estaciones meteorológicas en el año 2015, se replicó esta técnica para los siguientes años y se elaboró un atlas digital que incluye todos los mapas desarrollados, los cuales están disponibles en el repositorio del GEA, de acceso libre.

RECOMENDACIONES

Para investigaciones futuras, se sugiere trabajar con la mediana para estudiar la temperatura de una zona determinada. Además, es recomendable realizar un análisis exploratorio de datos exhaustivo, ya que esto permite identificar datos atípicos que puedan mejorar el ajuste de los datos y explicar mejor su comportamiento en los mapas de interpolación.

Para futuras investigaciones, se recomienda utilizar una mayor cantidad de datos provenientes de las estaciones meteorológicas. Esto permitirá obtener una representación más precisa en los mapas de interpolación, que al momento de realizar la interpolación se genere una capa raster que cubra de manera completa la zona de interés, mejorando así la precisión y la calidad de los resultados.

Se recomienda utilizar la prueba de normalidad Kolmogórov-Smirnov para muestras grandes y para muestras menores a 50 lo más factible es aplicar la prueba de normalidad Shapiro-Wilks, esto para verificar si los datos a trabajar provienen o no de una distribución normal y posteriormente a futuro determinar que técnicas de interpolación se ajustan a nuestros datos.

Para futuras investigaciones relacionadas con diversas técnicas de interpolación espacial es importante tener un conocimiento previo acerca del manejo del programa ArcGIS, específicamente en la comprensión de la herramienta denominada Geostatistical Analyst, la cual es la base para realizar un estudio con mapas de interpolación.

BIBLIOGRAFÍA

1. **AGUILAR, H.C., 2020.** *Vulnerabilidad y gestión del riesgo de desastres frente al cambio climático en Piura, Perú.* Semestre Económico, vol. 23, no. 54, ISSN 2248-4345. DOI 10.22395/seec.v23n54a5.
2. **ANDRADES, J., CUESTA, L., CAMARGO, C., LÓPEZ, J., TORRES, H. y OSORIO, A., 2020.** *Propuesta metodológica para la construcción y selección de modelos digitales de elevación de alta precisión.* Colombia Forestal, vol. 23, no. 2, ISSN 0120-0739. DOI 10.14483/2256201x.15155.
3. **ARIZALA, K.O.E. y COLT, T.F.R., 2020.** *Diseño de un modelo predictivo basado en técnicas de Machine Learning que permita determinar la temperatura usando los datos de una mini estación meteorológica en la ciudad de Guayaquil. [en línea]. bachelorThesis. S.l.:Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería en Sistemas Computacionales. [consulta: 11 enero 2023]. Disponible en: <http://repositorio.ug.edu.ec/handle/redug/48892>.*
4. **ATEHORTÚA, D.A.P., 2021.** *Diseño de mapas de información de radiación solar en el departamento de Caldas a través de métodos de interpolación y herramientas de SIG.* En: Accepted: 2021-08-18T20:05:29Z, Universidad Católica de Manizales [en línea], [consulta: 17 diciembre 2022]. Disponible en: <https://repositorio.ucm.edu.co/jspui/handle/10839/3288>.
5. **BELTRAN, P.A., 2020.** *Comparación de técnicas de interpolación espacial (IDW y Kriging) utilizando medidas de campo eléctrico en la Comuna Dos de la ciudad de San José de Cúcuta.* En: Accepted: 2020-07-31T19:10:43Z [en línea], [consulta: 17 diciembre 2022]. Disponible en: <http://repository.udistrital.edu.co/handle/11349/24828>.
6. **BERNDT, C. y HABERLANDT, U., 2018** *Spatial interpolation of climate variables in Northern Germany—Influence of temporal resolution and network density.* *Journal of Hydrology: Regional Studies*, vol. 15, ISSN 2214-5818. DOI 10.1016/j.ejrh.2018.02.002.
7. **BUSTAMANTE, D.P., 2017.** *ESCENARIO DE CAMBIO CLIMÁTICO A NIVEL DE SUBCUENCAS HIDROGRÁFICAS PARA EL AÑO 2050 DE LA PROVINCIA DE CHIMBORAZO- ECUADOR.* La Granja, vol. 26, no. 2, ISSN 1390-8596, 1390-3799. DOI 10.17163/lgr.n26.2017.02.

8. **CHANG, J.K.C.L., 2019.** *Métodos de interpolación para mapas de temperatura del aire en la ciudad de Guayaquil.* bachelorThesis. S.l.: Facultad de Ciencias Naturales. Universidad de Guayaquil. [consulta: 8 enero 2023]. Disponible en: <http://repositorio.ug.edu.ec/handle/redug/39633>.
9. **CHIRINOS, L.R. y MALLQUI, C.G., 2016.** *Comparación entre Interpoladores Espaciales en el Estudio de Distribución de Partículas Sedimentables Insolubles en la Cuenca Atmosférica de Lima y Callao.* Información tecnológica, vol. 27, no. 4, ISSN 0718-0764. DOI 10.4067/S0718-07642016000400012.
10. **CIFUENTES CARVAJAL, A., 2016.** *Evaluación de diferentes métodos de interpolación para la variable precipitación en el departamento de Caldas – Colombia.* En: Accepted: 2016-06-24T21:15:21Z [en línea], [consulta: 3 abril 2024]. Disponible en: <https://ridum.umanizales.edu.co/handle/20.500.12746/2652>.
11. **DIAZ, R.E.V., PACHECO, J.R.G., HENAO, R.G., RAMÍREZ, V.M.V., ORTIZ, W.A.W. y FLOREZ, C.V.F., 2023.** *Metodologías de interpolación y predicción espacial para el análisis de las propiedades físicas del suelo en la hoya del río Suárez (Colombia).* Siembra, vol. 10, no. 1, ISSN 2477-8850. DOI 10.29166/siembra.v10i1.4118.
12. **FRAU, C.M., PINO, L.M., ROJAS, Y.O. y HERNÁNDEZ, Y.M., 2012.** *Modelamiento geométrico del cambio de coordenadas UTM causadas por mudanza de referencial geodésico: Caso Sirgas - Chile.* Boletim de Ciências Geodésicas, vol. 18, no. 4, ISSN 1982-2170. DOI 10.1590/S1982-21702012000400005.
13. **GARREAUD, R.D., 2011.** *Cambio Climático: Bases Físicas e Impactos en Chile.* , no. 93, Disponible en: <https://bibliotecadigital.ciren.cl/handle/20.500.13082/29289>
14. **GUIJARRO PASTOR, J.A., 2004.** *CLIMATOL: software libre para la depuración y homogeneización de datos climatológicos.* S.l.: Asociación Española de Climatología. [consulta: 10 enero 2024]. ISBN 978-84-8102-384-8. Disponible en: <https://repositorio.aemet.es/handle/20.500.11765/9036>.
15. **LEAL, R.G., 2020.** *1 ¿Qué es la Temperatura?,* Disponible en: <https://dcb.ingenieria.unam.mx/wp-content/themes/tempera-child/Publicaciones/Naturalis/Boletines/bfyq35.pdf>.

16. **MANTILLA SALTOS, G.F., 2016.** *Comparación de metodologías estadísticas para interpolar la precipitación en el Ecuador* [en línea]. bachelorThesis. S.l.: ESPOL.FCNM. [consulta: 3 abril 2024]. Disponible en: <http://www.dspace.espol.edu.ec/handle/123456789/45357>.
17. **MARTÍNEZ, Z., PAREDES, J. y VALDEZ PEÑA, S., 2022.** *Estimación de datos faltantes de temperatura combinando IDW y una serie truncada de Fourier.*, vol. 22, ISSN 1578-8768.
18. **MAYORGA, J.L.A., MONTES, M.V.C., ESPINOZA, W.A.B., ANZUETO, W.A.P. y SOTOJ, E.M.P., 2022.** *ANÁLISIS COMPARATIVO DE MODELOS DE INTERPOLACIÓN DE LA PRECIPITACIÓN EN GUATEMALA (1981-2019).*, Revista de la Escuela de Estudios de Postgrado, no. 1, ISSN 2706-9508.
19. **MOHD AZIZ, M.K.B., YUSOF, F., MOHD DAUD, Z., YUSOP, Z. y KASNO, M.A., 2019.** *Comparison of Semivariogram Models in Rain Gauge Network Design.* *MATEMATIKA*, ISSN 0127-9602, 0127-8274. DOI 10.11113/matematika.v35.n2.1155.
20. **PEDREROS, J.G.G., 2021.** *Modelo de distribución espacial del desequilibrio de radiación terrestre medido por medio de radiómetros puestos en órbita*, Trabajo de grado - Maestría. S.l.: Universidad Nacional de Colombia. [consulta: 17 diciembre 2022]. Disponible en: <https://repositorio.unal.edu.co/handle/unal/80077>.
21. **PÉREZ, N.A., MULLO, H.S. y MARCATOMA, J.A., 2020.** *Análisis del cambio climático en un ecosistema alto andino, Riobamba-Ecuador.* *Perfiles*, vol. 1, no. 23, ISSN 2477-9105.
22. **Pourrut, Pierre. 1987.** *CLIMA DEL ECUADOR* Quito : s.n., 1987.
23. **REYES, S.T.C., 2017.** *UTILIZACIÓN DE HERRAMIENTAS GEOESTADÍSTICAS PARA LA GENERACIÓN DE INFORMACIÓN ESPACIAL METEOROLÓGICA EN LA PROVINCIA DEL AZUAY.* 2017. bachelorThesis. S.l.: s.n. Disponible en: <http://dspace.ucuenca.edu.ec/handle/123456789/28419>.
24. **RINCON, A.Q., D'GIANO, C. y RISK, M., 2018.** *Epileptic seizure prediction using Pearson's product-moment correlation coefficient of a linear classifier from generalized Gaussian modeling.* En: arXiv:2006.01359 [stat], Neurología Argentina, vol. 10, no. 4, ISSN 18530028. DOI 10.1016/j.neuarg.2018.06.004.

25. **RIVERA VÁSQUEZ, J.I., 2020.** *La malnutrición infantil en Chimborazo: una mirada multidimensional*. En: Accepted: 2021-01-26T17:46:51Z [en línea]. workingPaper. S.I.: Quito, EC: Universidad Andina Simón Bolívar, Sede Ecuador. [consulta: 3 abril 2024]. Disponible en: <http://repositorio.uasb.edu.ec/handle/10644/7818>
26. **ROHMA, N.N., 2023.** *PERBANDINGAN PENDUGAAN METODE ORDINARY KRIGING DAN METODE ORDINARY KRIGING DENGAN TEKNIK JACKKNIFE*. *MAP (Mathematics and Applications) Journal*, vol. 4, no. 2, ISSN 2721-1185. DOI 10.15548/map.v4i2.4736.
27. **ROZO ÁNGEL, V., 2023.** *Revelando a los presuntos responsables “desconocidos” de asesinatos selectivos en el conflicto armado colombiano*. En: Accepted: 2023-11-16T21:15:26Z [en línea], [consulta: 3 abril 2024]. Disponible en: <http://repositorio.udesa.edu.ar/jspui/handle/10908/23406>.
28. **SCHMIDT, F., BERT, F., PODESTÁ, G., VEIGA, H., HERRERA, N., SKANSI, M. de los M., CLAUS, F. y BASUALDO, A., 2015.** *Evaluación de métodos para la imputación de registros faltantes*. En: Accepted: 2018-11-22T13:38:18Z [en línea], [consulta: 19 diciembre 2023]. Disponible en: <http://repositorio.smn.gob.ar/handle/20.500.12160/849>.
29. **SHAHZAD, U., 2022.** *Critical Analysis For Solar Cell Models Using System Advisor Model Simulations*. *Journal of Electrical Engineering, Electronics, Control and Computer Science*, vol. 9, no. 1, ISSN 2668-8476.
30. **VELÁSQUEZ, D.F.A., CARRILLO, G.A.A., BARBOSA, E.O.R., LATORRE, D.A.G. y MALDONADO, F.E.M., 2018.** *INTERPOLACION REGNIE PARA LLUVIA Y TEMPERATURA EN LAS REGIONES ANDINA, CARIBE Y PACÍFICA DE COLOMBIA*. *Colombia Forestal*, vol. 21, no. 1, DOI <https://doi.org/10.14483/2256201X.11601>.
31. **VICENCIO, Á.E.G., 2018.** *Análisis comparativo de metodologías de interpolación espacial en la distribución de precipitaciones y temperaturas de la Región de La Araucanía*. [en línea], Disponible en: <http://cybertesis.uach.cl/tesis/uach/2018/fcg934a/doc/fcg934a.pdf>.

0781-DBRA-UPT-2024



ANEXOS

Anexo A: Códigos utilizados en el software estadístico Rstudio

```
1 library(readxl)
2 TOTAL_PUNTOS_TO <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/TOTAL PUNTOS.xlsx")
3     sheet = "TODOS")
4
5 ESTACIONES_PROMEDIOS <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/ESTACIONES P
6     sheet = "TODOS")
7 View(ESTACIONES_PROMEDIOS)
8 View(TOTAL_PUNTOS_TO)
9
10 library(MVN)
11
12 # Al trabajar con varios años se los selecciona
13 # en partes explicadas a continuación.
14
15 total<- TOTAL_PUNTOS_TO[,4:12] #4:12, 13:21, 22:26 así consecutivamente
16 summary(total)
17 View(total)
18
19 total <- ESTACIONES_PROMEDIOS[,7:10]
20 summary(total)
21
22 # ESTADÍSTICAS DESCRIPTIVAS
23
24 Result <- mvn(data=total, mvnTest="royston", univariatePlot="box")
25 Result<- mvn(data=total, mvnTest="royston", univariatePlot="histogram")
26 Result<-mvn(data=total, mvnTest = "royston", univariatePlot = "qqplot")
27
28 #####
29 ks.test(total$`2016`, "pnorm", mean = mean(total$`2016`)
30     , sd = sd(total$`2016`))
31 ks.test(total$`2017`, "pnorm", mean = mean(total$`2017`)
32     , sd = sd(total$`2017`))
33 ks.test(total$`2018`, "pnorm", mean = mean(total$`2018`)
34     , sd = sd(total$`2018`))
35 ks.test(total$`2019`, "pnorm", mean = mean(total$`2019`))
```

```

36     , sd = sd(total$`2019`)
37 ks.test(total$`2020`, "pnorm", mean = mean(total$`2020`)
38     , sd = sd(total$`2020`)
39
40 # test de normalidad univariante
41
42 ##Test de Lilliefors (correccion de Kolmogorov)
43
44 result <- mvn(data = total, mvnTest = "royston",
45     univariateTest = "Lillie", desc = TRUE)
46 result
47
48 ## Test de Anderson Darling
49 result <- mvn(data = total, mvnTest = "royston",
50     univariateTest = "AD", desc = TRUE)
51 result
52
53 ##### Test de normalidad multivariante
54
55 total<- TOTAL_PUNTOS_TO[,4:26]
56 head(total)
57
58 #### Prueba de Mardia con MVN.
59
60 # test de normalidad
61 # test de Mardia en MVN
62 result <- mvn(data = total, mvnTest = "mardia")
63 result$multivariateNormality
64
65 ##### Seleccion de la muestra del ano 2015 #####
66
67 library(readxl)
68 TOTAL_PUNTOS <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/TOTAL PUNTOS.xlsx",
69     sheet = "2015")
70
71 ESTACIONES_PROMEDIOS <-read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/ESTACIONES PR
72     sheet = "2015")
73 #View(ESTACIONES_PROMEDIOS)
74 #View(TOTAL_PUNTOS)

```



```

75 Datos_SAM_ano_2015<-TOTAL_PUNTOS$PROMEDIO
76 estaciones_2015<-ESTACIONES_PROMEDIOS$PROMEDIO
77 #View(Datos_SAM_ano_2015)
78 #View(estaciones_2015)
79
80 ##### datos atipicos y detecccion de datos atipicos
81 PASWR::EDA(Datos_SAM_ano_2015)
82 PASWR::EDA(estaciones_2015)
83
84 # Prueba de normalidad para la muestra
85 # Shapiro Test 50 DATOS
86 shapiro.test(estaciones_2015)
87
88 # kolgomorov
89 ks.test(Datos_SAM_ano_2015, "pnorm", mean = mean(Datos_SAM_ano_2015),
90         sd = sd(Datos_SAM_ano_2015))
91
92 ##### TRASFORMACION BOXCOX #####
93 library(fpp)
94 library(trend)
95 library(dplyr)
96 ### Eliminando los datos negativos
97
98 ati <- c(27,55,38)
99 TOTAL_sin_atipicos<-slice(TOTAL_PUNTOS,-ati)
100 View(TOTAL_sin_atipicos)
101
102 SAM_ano_2015 <- TOTAL_sin_atipicos$PROMEDIO
103
104 lambda <- BoxCox.lambda(SAM_ano_2015, method = c("loglik")
105                       ,lower = -5, upper = 5)
106 lambda
107 trans <- BoxCox(SAM_ano_2015, lambda)
108 shapiro.test(trans)
109 ks.test(trans, "pnorm", mean = mean(trans), sd = sd(trans))
110 PASWR::EDA(trans)
111
112 ##### DATOS DESCARGADOS SAM #####
113 library(splines)

```

```

114 library(readxl)
115
116 datos <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/TOTAL PUNTOS.xlsx",
117                    sheet = "2015")
118 # Numero de observaciones
119 n <- nrow(datos)
120 # Proporción para el conjunto de prueba
121 prop_prueba <- 0.2
122 # Tamaño del conjunto de prueba
123 tamaño_prueba <- round(n * prop_prueba)
124 # Tamaño del conjunto de entrenamiento
125 tamaño_entrenamiento <- n - tamaño_prueba
126 # Inicializar vectores para almacenar los errores
127 errores_me <- c()
128 errores_mse <- c()
129 # Realizar la validación cruzada
130 for (i in 1:(n - tamaño_prueba)) {
131   # Índices para el conjunto de prueba
132   índice_prueba <- (i + 1):(i + tamaño_prueba)
133
134   # Índices para el conjunto de entrenamiento
135   índice_entrenamiento <- setdiff(1:n, índice_prueba)
136
137   # Datos de entrenamiento y prueba
138   datos_entrenamiento <- datos[índice_entrenamiento, ]
139   datos_prueba <- datos[índice_prueba, ]
140
141   # Ajustar el modelo de spline a los datos de entrenamiento
142   modelo_spline <- smooth.spline(x = datos_entrenamiento$X,
143                                 y = datos_entrenamiento$PROMEDIO)
144   # Predecir en los datos de prueba
145   predicciones <- predict(modelo_spline, x = datos_prueba$X)$y
146   # Calcular el error cuadrático medio y almacenar
147   error_mse <- mean((predicciones - datos_prueba$PROMEDIO)^2)
148   errores_mse <- c(errores_mse, error_mse)
149
150   # Calcular el error medio y almacenar
151   error_me <- mean(predicciones - datos_prueba$PROMEDIO)
152   errores_me <- c(errores_me, error_me)

```

```

153 }
154 # Calcular el error promedio
155 error_promedio_mse <- meanerrores_mse)
156 error_promedio_me <- meanerrores_me)
157 # Calcular la desviacion estandar del valor real
158 desviacion_estandar_real <- sd(datos$PROMEDIO)
159 # Calcular el error medio estandarizado
160 error_promedio_mse_estandarizado <- error_promedio_me / desviacion_estandar_real
161 # Imprimir resultados
162 cat("Error Cuadratico Medio Promedio (RMSE):",
163     error_promedio_mse, "\n")
164 cat("Error Medio Promedio (ME):",
165     error_promedio_me, "\n")
166 cat("Error Medio Estandarizado (MSE Estandarizado):",
167     error_promedio_mse_estandarizado, "\n")
168
169 ##### DATOS ESTACIONES #####
170 library(readxl)
171
172 # Cargar los datos desde el archivo Excel
173 ESTACIONES_PROMEDIOS <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/ESTACIONES P
174                                     sheet = "2015")
175 # Numero total de observaciones
176 n1 <- nrow(ESTACIONES_PROMEDIOS)
177 # Proporción para el conjunto de prueba
178 prop_prueba <- 0.2
179 # Tamaño del conjunto de prueba
180 tamaño_prueba <- round(n1 * prop_prueba)
181 # Inicializar vectores para almacenar los errores
182 errores_me <- c()
183 errores_mse <- c()
184 # Realizar la validación cruzada
185 for (i in 1:(n1 - tamaño_prueba)) {
186     # Índices para el conjunto de prueba
187     índice_prueba <- (i + 1):(i + tamaño_prueba)
188
189     # Índices para el conjunto de entrenamiento
190     índice_entrenamiento <- setdiff(1:n1, índice_prueba)
191

```

```

192 # Datos de entrenamiento y prueba
193 datos_entrenamiento <- datos[indice_entrenamiento, ]
194 datos_prueba <- datos[indice_prueba, ]
195
196 # Ajustar el modelo de spline a los datos de entrenamiento
197 modelo_spline <- smooth.spline(x = datos_entrenamiento$X,
198                               y = datos_entrenamiento$PROMEDIO)
199
200 # Predecir en los datos de prueba
201 predicciones <- predict(modelo_spline, x = datos_prueba$X)$y
202
203 # Calcular el error cuadratico medio y almacenar
204 error_mse <- mean((predicciones - datos_prueba$PROMEDIO)^2)
205 errores_mse <- c(errores_mse, error_mse)
206
207 # Calcular el error medio y almacenar
208 error_me <- mean(predicciones - datos_prueba$PROMEDIO)
209 errores_me <- c(errores_me, error_me)
210 }
211 # Calcular el error promedio
212 error_promedio_mse <- mean(errores_mse)
213 error_promedio_me <- mean(errores_me)
214 # Calcular la desviacion estandar del valor real
215 desviacion_estandar_real <- sd(datos$PROMEDIO)
216 # Calcular el error medio estandarizado
217 error_promedio_mse_estandarizado <- error_promedio_me / desviacion_estandar_real
218 # Imprimir resultados
219 cat("Error Cuadratico Medio Promedio (RMSE):",
220     error_promedio_mse, "\n")
221 cat("Error Medio Promedio (ME):",
222     error_promedio_me, "\n")
223 cat("Error Medio Estandarizado (MSE Estandarizado):",
224     error_promedio_mse_estandarizado, "\n")
225
226 ##### CANTON RIOBAMBA #####
227 CANTONRIOBAMBA15 <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/TOTAL PUNTOS.xls",
228                               sheet = "CANTON-RIOBAMBA2015")
229
230 CANTONRIOBAMBA20 <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/TOTAL PUNTOS.xls",

```

```

231         sheet = "CANTON-RIOBAMBA2020")
232 CR2015<-CANTONRIOBAMBA15$PROMEDIO
233 CR2020<-CANTONRIOBAMBA20$PROMEDIO
234 ## Diferencias entre ambas
235 dif<-(CR2015-CR2020);dif
236 # Observar las diferencias
237 data<-data.frame(CR2015,CR2020,dif);data
238 # Rangos de diferencias
239 ragdid<-rank(abs(dif[dif != 0])) # ignora las direncias que sean 0
240 # Rangos
241 rangosfe<-c(ragdid[1:4],0, ragdid[5])# diferencias de las categorias
242
243 # Compilar toda la informacion
244 data<-data.frame(CR2015,CR2020,signo=sign(dif),
245                 dif=abs(dif),rangos=ragdid);data
246
247 aumento <- data.frame(mean(CR2015),mean(CR2020),
248                       diferencia=abs(mean(CR2015)-mean(CR2020)))
249 aumento
250 ## Sumar los rangos positivos
251 sumapositivos<-sum(data[data$signo==1,'rangos'])
252 sumanegativos<-sum(data[data$signo==-1,'rangos'])
253 # Exporta el valor mas maximo o mas grnade hallado
254 w<-max(c(sumanegativos,sumapositivos))
255 # Prueba de Wilcoxon
256 wilcox.test(CR2015,CR2020,paired = T,
257             exact = T,correct = F,coef.int = 0.95)
258 #####CANTON ALAUSI #####
259 CANTONALAUSI15 <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/TOTAL PUNTOS.xlsx"
260                             sheet = "CANTON-ALAUSI2015")
261
262 CANTONALAUSI20 <- read_excel("~/OCTAVO SEMESTRE/PUNTOS CORRECTOS/TOTAL PUNTOS.xlsx"
263                             sheet = "CANTON-ALAUSI2020")
264 CA2015<-CANTONALAUSI15$PROMEDIO
265 CA2020<-CANTONALAUSI20$PROMEDIO
266 ## Diferencias entre ambas
267 dif<-(CA2015-CA2020);dif
268 # Observar las diferencias
269 data<-data.frame(CA2015,CA2020,dif);data

```

```

270 # Rangos de diferencias
271 ragdid<-rank(abs(dif[dif != 0])) # ignora las diferencias que sean 0
272 # Rangos
273 rangosfe<-c(ragdid[1:4],0, ragdid[5])# diferencias de las categorias
274
275 # Compilar toda la informacion
276 data<-data.frame(CA2015,CA2020,signo=sign(dif),
277                 dif=abs(dif),rangos=ragdid);data
278
279 aumento <- data.frame(mean(CA2015),mean(CA2020)
280                       ,diferencia=abs(mean(CA2015)-mean(CA2020)))
281 aumento
282
283 ## Sumar los rangos positivos
284 sumapositivos<-sum(data[data$signo==1,'rangos'])
285 sumanegativos<-sum(data[data$signo==-1,'rangos'])
286 # Exporta el valor mas maximo o mas grande hallado
287 w<-max(c(sumanegativos,sumapositivos))
288 # Prueba de Wilcoxon
289 wilcox.test(CA2015,CA2020,paired = T,
290            exact = T,correct = F,coef.int = 0.95)

```

Anexo B: Aval de la Investigación



AVAL

EL DIRECTOR DEL GRUPO DE ENERGÍAS ALTERNATIVAS Y AMBIENTE
(GEAA)

DOCTOR CELSO GUILLERMO RECALDE MORENO, SUSCRIBE

AVALA:

Que el señor **James William Maldonado Siguencia** con CI: 030265454-6 y el señor **Dennis Adrian Yugsi Yugsi** con CI: 172706484-0, estudiantes de la Carrera de Estadística se encuentran desarrollando su trabajo de integración Curricular con el tema denominado: **"Comparación de técnicas de interpolación espacial de temperatura ambiente de los modelos atmosféricos CMIP6 y de las estaciones meteorológicas"** en el GRUPO DE ENERGÍAS ALTERNATIVAS Y AMBIENTE PERTENECIENTE A LA FACULTAD DE CIENCIAS EN LA ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO. Cuya tutora es la Ing. Cristina Ramos y Miembro del Tribunal la Ing. Natalia Pérez.

Se extiende el presente documento a solicitud del interesado, para los fines que estime convenientes.

Riobamba, 25 de enero del 2023



CELSO
GUILLERMO
RECALDE
MORENO
Dr. Celso Guillermo Recalde
DIRECTOR GEAA-FAC. CIENCIAS

Firmado
digitalmente por
CELSO GUILLERMO
RECALDE MORENO



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
CERTIFICADO DE CUMPLIMIENTO DE LA GUÍA PARA
NORMALIZACIÓN DE TRABAJOS DE FIN DE GRADO

Fecha de entrega: 11/07/2024

INFORMACIÓN DEL AUTOR
Nombres – Apellidos: James William Maldonado Siguenca Dennis Adrian Yugsi Yugsi
INFORMACIÓN INSTITUCIONAL
Facultad: Ciencias
Carrera: Estadística
Título a optar: Ingeniero Estadístico
<div style="text-align: center;"> Ing. Cristina Estefania Ramos Araujo, Mgs. Director del Trabajo de Titulación</div> <div style="text-align: center;"> Ing. Natalia Alexandra Perez Londo, Mgs. Asesor del Trabajo de Titulación</div>