



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA DE INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

**“PREDICCIÓN DE LA TEMPERATURA AMBIENTAL
MEDIANTE MODELOS ESTADÍSTICOS FUNCIONALES DE
LAS ESTACIONES MONITOREADAS POR EL GEAA EN LA
PROVINCIA DE CHIMBORAZO (2014-2019)”**

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERA EN ESTADÍSTICA INFORMÁTICA

AUTORA: ROXANA GABRIELA MARIÑO CAJO

DIRECTORA: Ing. JOHANNA ENÍTH AGUILAR REYES MSc.

Riobamba - Ecuador

2021

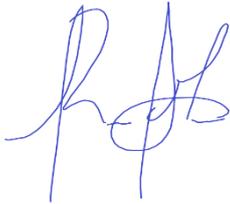
© 2021, Roxana Gabriela Mariño Cajo

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

Yo, Roxana Gabriela Mariño Cajo, declaro que el presente trabajo de integración curricular es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo 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 Julio del 2021



Roxana Gabriela Mariño Cajo

070509469-6

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA DE INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

El Tribunal del Trabajo de Integración Curricular certifica que: El Trabajo de Integración Curricular: Tipo: Proyecto de Investigación, “**PREDICCIÓN DE LA TEMPERATURA AMBIENTAL MEDIANTE MODELOS ESTADÍSTICOS FUNCIONALES DE LAS ESTACIONES MONITOREADAS POR EL GEAA EN LA PROVINCIA DE CHIMBORAZO (2014-2019)**”, realizado por la señorita: **ROXANA GABRIELA MARIÑO CAJO**, 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

Dr. Celso Guillermo Recalde PhD.
PRESIDENTE DEL TRIBUNAL

 Firmado digitalmente
por CELSO GUILLERMO
RECALDE MORENO
Fecha: 2021.08.04
20:17:40 -05'00'

28 de Julio de 2021

Ing. Johanna Eníth Aguilar Reyes MSc.
**DIRECTORA DEL TRABAJO DE
TITULACIÓN**

 Firmado electrónicamente por:
**JOHANNA ENITH
AGUILAR REYES**

28 de Julio de 2021

Ing. Nancy Chariguamán Maurisaca MSc.
MIEMBRO DEL TRIBUNAL

 NANCY
ELIZABETH
CHARIGUAMAN
MAURISACA
CHECI SERIALNUMBER=0603612458,
SN=CHARIGUAMAN MAURISACA,
CN=NANCY ELIZABETH
CHARIGUAMAN MAURISACA,
OU=NANCY ELIZABETH,
E=HE.CB@live.com, S=CHIMBORAZO,
L=ROSBAMBIA, O=Certificado de Clase
2 de Persona Física EC (FIRMA)

28 de Julio de 2021

DEDICATORIA

El presente trabajo lo dedico a Dios, quien me ha guiado y cuidado para cumplir con cada una de las metas propuestas. A mi familia, quienes han sido el pilar fundamental, dándome fortaleza y ánimos para nunca darme por vencida, A mis madres Manuela Cajo y Rosita García por apoyarme para poder alcanzar esta meta, a mis hermanos, sobrinas y a mi Padre Manuel Mariño que me cuida desde el cielo.

Gabriela Mariño

AGRADECIMIENTOS

Mi profundo agradecimiento a Dios por darme fuerzas para levantarme y seguir luchando para cumplir mis sueños, a mi familia quienes confiaron en mí y me brindaron el apoyo necesario para poder culminar con éxito esta etapa, a los docentes por sus enseñanzas, compromiso y dedicación al impartir sus conocimientos, a la Ing. Johanna Aguilar, Ing. Nancy Chariguamán, quienes me guiaron para poder terminar este trabajo , a mis amigos que fueron un apoyo para mi vida y con quienes compartí momentos de mucha alegría y al grupo de Energías Alternativas y Ambiente (GEAA) por facilitarme las bases de datos de las estaciones meteorológicas para que pueda utilizarlas en este estudio.

Gabriela Mariño

TABLA DE CONTENIDOS

ÍNDICE DE TABLAS.....	x
ÍNDICE DE FIGURAS.....	xi
ÍNDICE DE GRÁFICOS.....	xii
ÍNDICE DE ANEXOS.....	xiv
RESUMEN.....	xv
SUMMARY.....	xvi
INTRODUCCIÓN.....	1

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL.....	2
1.1. Antecedentes.....	2
1.2. Planteamiento del Problema.....	3
1.3. Formulación del Problema.....	3
1.4. Justificación.....	3
1.4.1. Justificación Teórica.....	3
1.5. Objetivos.....	4
1.5.1. Objetivos Generales.....	4
1.5.2. Objetivos Específicos.....	4
1.6. Bases Teóricas.....	4
1.6.1. Análisis funcional.....	4
1.6.2. Variable Funcional.....	5
1.6.3. ¿Que son los datos funcionales?.....	5
1.6.4. ¿Qué es un base funcional?.....	5
1.6.4.1. Bases de Fourier.....	5
1.6.4.2. Bases de Wavelets.....	6
1.6.4.3. Bases B-splines.....	6
1.6.4.4. Bases de potencias y exponenciales.....	7
1.6.4.5. Bases polinomiales.....	8
1.6.5. Suavizamiento de la función.....	8
1.6.5.1. Suavizar datos funcionales por mínimos cuadrados.....	8
1.6.5.2. Suavizar datos funcionales por mínimos cuadrados ponderados.....	9
1.6.5.3. Mínimos cuadrados localizados suavizados mediante kernels.....	10
1.6.6. Profundidad Funcional.....	11
1.6.6.1. Profundidad Fraiman-Muniz (FM).....	11

1.6.6.2.	<i>Profundidad modal (MD)</i>	12
1.6.6.3.	<i>Profundidad de proyección aleatoria (RPD)</i>	12
1.6.7.	Análisis exploratorio funcional	12
1.6.7.1.	<i>Media funcional</i>	12
1.6.7.2.	<i>Varianza funcional</i>	13
1.6.7.3.	<i>Covarianza funcional</i>	13
1.6.7.4.	<i>Correlación Funcional</i>	13
1.6.8.	Análisis funcional en componentes principales	13
1.6.9.	Regresión Funcional del Componente Principal	14
1.6.9.1.	<i>Análisis Robusto de Componentes Principales</i>	16
1.6.10.	Técnicas de Predicción de series de tiempo univariadas y multivariadas	17
1.6.11.	Método de pronósticos de intervalos	18
1.6.12.	Medidas de Evaluación de pronósticos	19
1.6.12.1.	<i>Medidas dependientes de la escala</i>	19
1.6.12.2.	<i>Medidas basadas en porcentaje</i>	20
1.6.13.	Medidas de Precisión del Pronóstico	21
1.6.13.1.	<i>Coficiente U de Theil</i>	21
1.6.13.2.	<i>Test de Diebold -Mariano</i>	21
1.7.	Bases Conceptuales	21
1.7.1.	<i>¿Qué es la meteorología?</i>	21
1.7.2.	<i>Organización Meteorológica Mundial (OMM)</i>	21
1.7.3.	<i>¿Calidad de los datos?</i>	22
1.7.4.	<i>Tipos de Errores existentes en bases de datos</i>	22
1.7.5.	<i>Valores Umbrales</i>	22
1.7.6.	<i>¿Qué es la temperatura del aire?</i>	23
1.7.7.	<i>Naturaleza Cíclica de los cambios de la temperatura</i>	23
1.7.9.	Método de Imputación Random Forest	24
1.7.10.	Ventajas de Random Forest	24
1.7.11.	Función rfImpute	25
1.7.12.	Software Estadístico R	25

CAPÍTULO II

2.	MARCO METODOLÓGICO	26
2.1.	Tipo y Diseño de Investigación	26
2.2.	Diseño de la Investigación	26
2.3.	Localización del Proyecto	26

2.4.	Población de estudio	27
2.5.	Tamaño de la Muestra	28
2.6.	Método de Muestreo	28
2.7.	Recolección de la Información	28
2.8.	Análisis Estadístico	28
2.9.	Operacionalización de las variables	28

CAPÍTULO III

3.	RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS	29
3.1.	Base de datos	29
3.2.	Validación de los datos	29
3.3.	Base funcional óptima	32
3.4.	Detección de datos Atípicos	34
3.5.	Análisis exploratorio de datos funcionales	37
3.5.1.	<i>Media y Desviación estándar Funcional</i>	37
3.5.1.1.	<i>Estación Alao</i>	37
3.5.1.2.	<i>Estación Cumandá</i>	37
3.5.1.3.	<i>Estación Atillo</i>	38
3.5.1.4.	<i>Estación Multitud</i>	39
3.5.1.5.	<i>Estación Quimiag</i>	39
3.5.1.6.	<i>Estación San Juan</i>	40
3.5.1.7.	<i>Estación Tixán</i>	41
3.5.1.8.	<i>Estación Tunshi</i>	41
3.5.1.9.	<i>Estación Urbina</i>	42
3.5.1.10.	<i>Estación Espoch</i>	43
3.5.1.11.	<i>Estación Matus</i>	43
3.5.2.	<i>Varianza Funcional</i>	44
3.6.	Modelización de datos funcionales (Ajuste del Modelo)	48
3.6.1.	<i>Componentes Principales de Estación Alao</i>	49
3.6.2.	<i>Componentes Principales de Estación Atillo</i>	50
3.6.3.	<i>Componentes Principales de Estación Cumandá</i>	51
3.6.4.	<i>Componentes Principales de Estación Espoch</i>	52
3.6.5.	<i>Componentes Principales de Estación Matus</i>	53
3.6.6.	<i>Componentes Principales de Estación Multitud</i>	54
3.6.7.	<i>Componentes Principales de Estación Quimiag</i>	55
3.6.8.	<i>Componentes Principales de Estación San Juan</i>	56

3.6.9.	<i>Componentes Principales de Estación Tixán</i>	57
3.6.10.	<i>Componentes Principales de Estación Tunshi</i>	58
3.6.11.	<i>Componentes Principales de Estación Urbina</i>	59
3.7.	Pronósticos de Temperatura Ambiental	60
3.7.1.	<i>Pronósticos de Estación Alao</i>	60
3.7.2.	<i>Pronósticos de Estación Atillo</i>	61
3.7.3.	<i>Pronósticos de Estación Cumandá</i>	61
3.7.4.	<i>Pronósticos de Estación Espoch</i>	62
3.7.5.	<i>Pronósticos de Estación Matus</i>	62
3.7.6.	<i>Pronósticos de Estación Multitud</i>	63
3.7.7.	<i>Pronósticos de Estación Quimiag</i>	63
3.7.8.	<i>Pronósticos de Estación San Juan</i>	64
3.7.9.	<i>Pronósticos de Estación Tixán</i>	64
3.7.10.	<i>Pronósticos de Estación Tunshi</i>	65
3.7.11.	<i>Pronósticos de Estación Urbina</i>	65
3.8.	Criterios de Evaluación de los Modelos	66
3.9.	Análisis de Precisión de los Modelos	67
3.9.1.	<i>Coficiente U de Theil</i>	67
3.9.2.	<i>Test de Diebold-Mariano</i>	67
3.10.	Pronósticos de la Temperatura Ambiental	68
3.10.1.	<i>Valores Pronosticados de Alao</i>	69
3.10.2.	<i>Valores Pronosticados de Atillo</i>	69
3.10.3.	<i>Valores Pronosticados de Cumandá</i>	70
3.10.4.	<i>Valores Pronosticados de Espoch</i>	70
3.10.5.	<i>Valores Pronosticados de Matus</i>	71
3.10.6.	<i>Valores Pronosticados de Multitud</i>	71
3.10.7.	<i>Valores Pronosticados de Quimiag</i>	72
3.10.8.	<i>Valores Pronosticados de San Juan</i>	72
3.10.9.	<i>Valores Pronosticados de Tixán</i>	73
3.10.10.	<i>Valores Pronosticados de Tunshi</i>	73
3.10.11.	<i>Valores Pronosticados de Urbina</i>	74
	CONCLUSIONES	75
	RECOMENDACIONES	76
	GLOSARIO	
	BIBLIOGRAFÍA	
	ANEXOS	

ÍNDICE DE TABLAS

Tabla 1-1:	Valores umbrales de la temperatura del aire	23
Tabla 1-2:	Operacionalización de las Variables.....	28
Tabla 1-3:	Porcentaje de Faltantes por año de cada estación.....	29
Tabla 2-3:	Reemplazo de valores INVALID por NA	30
Tabla 3-3:	MSE de suavizados con Fourier y B-splines	33
Tabla 4-3:	Valores de componentes Principales de Alao.....	49
Tabla 5-3:	Valores de componentes Principales de Atillo	50
Tabla 6-3:	Valores de componentes Principales de Cumandá	51
Tabla 7-3:	Valores de componentes Principales de Espoch.....	52
Tabla 8-3:	Valores de componentes Principales de Matus	53
Tabla 9-3:	Valores de componentes Principales de Multitud	54
Tabla 10-3:	Valores de componentes Principales de Quimiag	55
Tabla 11-3:	Valores de componentes Principales San Juan.....	56
Tabla 12-3:	Valores de componentes Principales de Tixán	57
Tabla 13-3:	Valores de componentes Principales de Tunshi	58
Tabla 14-3:	Valores de componentes Principales de Urbina	59
Tabla 15-3:	Criterios de evaluación del modelo	66
Tabla 16-3:	Coeficiente U de Theil.....	67
Tabla 17-3:	Test de Diebold-Mariano.....	67
Tabla 18-3:	Valores Pronosticados de Alao.....	69
Tabla 19-3:	Valores Pronosticados de Atillo	69
Tabla 20-3:	Valores Pronosticados de Cumandá	70
Tabla 21-3:	Valores Pronosticados de Espoch.....	70
Tabla 22-3:	Valores Pronosticados de Matus	71
Tabla 23-3:	Valores Pronosticados de Multitud.....	71
Tabla 24-3:	Valores Pronosticados de Quimiag.....	72
Tabla 25-3:	Valores Pronosticados de San Juan	72
Tabla 26-3:	Valores Pronosticados de Tixán	73
Tabla 27-3:	Valores Pronosticados de Tunshi	73
Tabla 28-3:	Valores Pronosticados de Urbina	74

ÍNDICE DE FIGURAS

Figura 1-1: Bases B-splines	6
Figura 2-1: Curvas típicas de radiación y temperatura del aire	24
Figura 1-2: Coordenadas de las Estaciones del GEAA	26
Figura 2-2: Mapa de Ubicación de Estaciones Meteorológicas del GEAA.....	27

ÍNDICE DE GRÁFICOS

Gráfico 1-2:	Mapa de Ubicación de Estaciones Meteorológicas del GEAA	27
Gráfico 1-3:	Validación de Información prueba de tolerancia.....	29
Gráfico 2-3:	Validación de Información prueba de coherencia	30
Gráfico 3-3:	Faltantes de la Estación Atillo.....	31
Gráfico 4-3:	Faltantes de Estación Atillo según los días	31
Gráfico 5-3:	Datos Imputados de la Estación Atillo	32
Gráfico 6-3:	Suavizado con diferentes bases y número de bases.....	33
Gráfico 7-3:	Errores con 15 B-splines	34
Gráfico 8-3:	Base Optima 15 B-splines	34
Gráfico 9-3:	Curvas atípicas de las Estaciones suavizadas 1	35
Gráfico 11-3:	Media y Desviación Funcional Alao	37
Gráfico 12-3:	Media y Desviación Funcional Cumandá.....	38
Gráfico 13-3:	Media y Desviación Funcional Atillo.....	38
Gráfico 14-3:	Media y Desviación Funcional Multitud.....	39
Gráfico 15-3:	Media y Desviación Funcional Quimiag	40
Gráfico 16-3:	Media y Desviación Funcional San Juan.....	40
Gráfico 17-3:	Media y Desviación Funcional Tixán.....	41
Gráfico 18-3:	Media y Desviación Funcional Tunshi.....	42
Gráfico 19-3:	Media y Desviación Funcional Urbina.....	42
Gráfico 20-3:	Media y Desviación Funcional Espoch	43
Gráfico 21-3:	Media y Desviación Funcional Matus	44
Gráfico 22-3:	Varianza Funcional Espoch, Matus.....	44
Gráfico 23-3:	Varianza Funcional Alao, Cumandá, Atillo	45
Gráfico 24-3:	Varianza Funcional Multitud, Quimiag, San Juan	46
Gráfico 25-3:	Varianza Funcional Tixán, Tunshi, Urbina	47
Gráfico 26-3:	Datos particionados para modelización y validación	48
Gráfico 27-3:	Componentes Principales Alao.....	49
Gráfico 28-3:	Componentes Principales Atillo	50
Gráfico 29-3:	Componentes Principales Cumandá.....	51
Gráfico 30-3:	Componentes Principales Espoch.....	52
Gráfico 31-3:	Componentes Principales Matus	53
Gráfico 32-3:	Componentes Principales Multitud	54
Gráfico 33-3:	Componentes Principales Quimiag	55
Gráfico 34-3:	Componentes Principales San Juan.....	56
Gráfico 35-3:	Componentes Principales Tixán	57

Gráfico 36-3:	Componentes Principales Tunshi	58
Gráfico 37-3:	Componentes Principales Urbina	59
Gráfico 38-3:	Pronósticos Alao.....	60
Gráfico 39-3:	Pronósticos Atillo	61
Gráfico 40-3:	Pronósticos Cumandá.....	61
Gráfico 41-3:	Pronósticos Espoch.....	62
Gráfico 42-3:	Pronósticos Matus	62
Gráfico 43-3:	Pronósticos Multitud	63
Gráfico 44-3:	Pronósticos Quimiag	63
Gráfico 45-3:	Pronósticos San Juan	64
Gráfico 46-3:	Pronósticos Tixán	64
Gráfico 47-3:	Pronósticos Tunshi	65
Gráfico 48-3:	Pronósticos Urbina	65

ÍNDICE DE ANEXOS

- ANEXO A:** Imputación de datos Faltantes
- ANEXO B:** Lectura de datos
- ANEXO C:** Base Funcional y Suavizado
- ANEXO D:** Cálculo de datos atípicos
- ANEXO E:** Análisis Descriptivo Funcional
- ANEXO F:** Modelado y Pronóstico
- ANEXO G:** Errores de Pronóstico
- ANEXO H:** Errores al escoger la base
- ANEXO I:** Estaciones graficadas por hora y años

RESUMEN

El objetivo del estudio fué predecir la temperatura ambiental, mediante modelos estadísticos funcionales de las estaciones de Alao, Atillo, Tunshi, Tixán, Matus, Quimiag, Urbina, San Juan, Espoch, Multitud, Cumandá, haciendo uso de los datos de temperatura ambiental almacenados desde el año 2014 al 2019 por el Grupo de Energías Alternativas y Ambiente (GEAA). Las bases fueron analizadas, ordenadas y espaciadas para realizar imputaciones mediante Random Forest; se utilizó técnicas de datos funcionales como: suavizados B-plines, detección de curvas atípicas, análisis exploratorio funcional, medias funcionales, desviación estándar funcional y varianza funcional. El modelado de los datos se realizó por medio de regresión del componente principal funcional, para descomponer una serie temporal de funciones en un conjunto de componentes principales funcionales y sus puntuaciones asociadas, además se utilizaron series de tiempo univariante (ARIMA) y multivariante (VAR) para el pronóstico de los coeficientes. La validación de los modelos se analizaron los errores por medio de medidas dependiente de escala y medidas basadas en porcentaje, mientras que, la precisión del modelo fué calculado mediante el coeficiente U the Theil y el Test de Diebold-Mariano. Se concluye que existe menores errores, al utilizar el modelo de regresión del componente principal funcional con coeficientes pronosticados por series de tiempo multivariante. Se recomienda realizar análisis de datos funcionales de forma continua a la variable temperatura ambiental u otras variables meteorológicas para conocer su comportamiento a través del tiempo.

Palabras clave: <ANÁLISIS DE DATOS FUNCIONALES>, <SERIES DE TIEMPO FUNCIONALES>, <PRONÓSTICOS FUNCIONALES>, <PAQUETE FTSA>, <TEMPERATURA AMBIENTAL>, <COMPONENTE PRINCIPAL FUNCIONAL>.

LEONARDO
FABIO MEDINA
NUSTE

Firmado digitalmente por
LEONARDO FABIO MEDINA
NUSTE
Fecha: 2021.08.04 17:29:04
-05'00'



1466-DBRA-UTP-2021

SUMMARY

The objective of the study was to forecast the ambient temperature, through functional statistical models of the stations of Alao, Atillo, Tunshi, Tixán, Matus, Quimiag, Urbina, San Juan, Espoch, Multitud, Cumandá, using stored ambient temperature data. from 2014 to 2019 by the Alternative Energy and Environment Group (GEAA). The databases were analyzed, ordered and spaced to make imputations using Random Forest; functional data techniques such as: smoothing B-plines, detection of atypical curves, functional exploratory analysis, functional means, functional standard deviation and functional variance were used. Data modeling was performed by means of functional principal component regression, to decompose a time series of functions into a set of functional principal components and their associated scores, in addition univariate (ARIMA) and multivariate (VAR) time series were used for forecasting coefficients. For the validation of the models, errors were analyzed by means of scale-dependent measures and percentage-based measures, while the precision of the model was calculated using the Theil's U coefficient and the Diebold-Mariano test. It is concluded that there are fewer errors, when using the regression model of the functional principal component with coefficients predicted by multivariate time series. It is recommended to perform functional data analysis continuously at the ambient temperature variable or other meteorological variables to know its behavior over time.

Keywords: <FUNCTIONAL DATA ANALYSIS>, <FUNCTIONAL TIME SERIES>, <FUNCTIONAL FORECASTS>, <FTSA PACKAGE>, <AMBIENT TEMPERATURE>, <MAIN FUNCTIONAL COMPONENT>.

INTRODUCCIÓN

El estudio de variables meteorológicas data desde la antigüedad, desde entonces ya se tenía constancia de cambios producidos en la atmósfera, analizados mediante variables como temperatura, radiación solar, precipitación, velocidad del viento, entre otras, las mismas que son analizadas hoy en día para la previsión del tiempo, alertas meteorológicas y estudios relacionados al clima en áreas como: actividades agropecuarias , actividades aéreas, climatológicas, seguimientos de incendios, entre otros. Al analizar la fluctuación de variables meteorológicas en un periodo de tiempo permitirá planificar actividades como en construcción, turismo, industria, textil, recursos hídricos (IDEAM, 2018, p.23).

La temperatura ambiental o también conocida como temperatura del aire es una de las magnitudes más utilizadas para describir el estado de la atmósfera, principalmente en el sector agrícola, siendo el factor responsable de la variabilidad anual de la producción, la predicción de esta variable permite ejecutar estrategias de riego, selección de cultivos, condiciones ambientales como temperatura del aire altas o bajas, con el fin de planificar estrategias para el desarrollo de actividades(Campetella y Cerne, 2011, p.12).

El análisis de datos Funcionales es un método utilizado generalmente en observaciones que provienen de procesos estocásticos, es decir que dependen del tiempo, siendo resultados de la evolución temporal de una variable en estudio, el cual consiste en hacer que los datos que son tratados como vectores se conviertan en curvas o funciones, analizando y tratando a esas curvas por medio métodos estadísticos como: regresión del componente principal funcional , utilizado para el ajuste del modelo y series de tiempo univariadas y multivariadas utilizadas para el pronóstico de coeficientes(Shang, 2013, pp.2-3).

El objetivo de la investigación es predecir la temperatura ambiental de las 11 estaciones monitoreadas por el GEAA, como: Cumandá, Espoch, Multitud, Matus, Quimiag, San Juan, Tunshi, Tixán, Urbina, Alao, Atillo, las mismas que pertenecen a la Provincia del Chimborazo, con el fin de utilizar el modelo que pronostique la temperatura ambiental con el menor error posible. El modelado de los datos se lo realiza por medio de regresión del componente principal funcional, con coeficientes pronosticados por series de tiempo multivariante, utilizando modelos VAR y coeficientes pronosticados por series de tiempo univariante, utilizando modelos ARIMA. Para la evaluación de los modelos, se analizan los errores de pronóstico, permitiendo escoger el modelo basado en el menor error presente, las medidas utilizadas en esta investigación son medidas dependientes de escala como: MSE, RMSE, MAE, MDAE, las medidas basadas en porcentaje como: MAPE, MDAPE, RMSPE y las medidas de precisión como: Coeficiente U de Theil y el Test de Diebold-Mariano.

CAPÍTULO I

1. MARCO TEÓRICO REFERENCIAL

1.1. Antecedentes

Las variables meteorológicas han sido objeto de estudio por su gran importancia en diferentes campos e investigaciones como en: “análisis estadístico de datos meteorológicos mensuales y diarios para la determinación de variabilidad climática y cambio climático en el Distrito Metropolitano de Quito "(Serrano Vincenti et al., 2012: p.20), cuyo objetivo es conocer el cambio del clima y cuanto ha variado en los últimos años, en la ciudad de Quito. Se utiliza variables meteorológicas como temperaturas mensuales mínimas, máximas y precipitación.

Existe otro tipo de estudios, donde se utiliza Análisis de datos Funcionales, el cual engloba una gran variedad de métodos estadísticos, superficies o cualquier otra función que varía de forma continua (Escabias et al., 2012, p.6).El estudio denominado "Análisis de Datos Funcionales Aplicado al Riesgo Financiero " utiliza técnicas funcionales para el análisis, detectando datos atípicos, y construyendo modelos para poder clasificarlos (Castillo et al., 2017: pp.5-7).

Uno de los métodos del análisis funcional es la predicción mediante modelos funcionales para conocer características y comportamientos de variables meteorológicas, generando un sin número de investigaciones como: "Aplicaciones en meteorología de modelos funcionales de predicción en componentes principales" (Carreño Madinabeitia, 2013, p.8), cuyo objetivo es mejorar las predicciones, mediante modelos en componentes principales funcionales.

En otro estudio denominado como: “Conjuntos de tasas de mortalidad para predecir las tasas de mortalidad por cáncer en España" (Gonzalez, 2013, p.8),se genera modelos de regresión por componentes principales ,utilizando métodos de predicción , métodos de suavización y un análisis completo donde se predice tasas de mortalidad.

1.2. Planteamiento del Problema

El avance tecnológico e industrial ha traído consigo un sin número de factores que influyen en la variación de la temperatura, como la polución producida por los aerosoles, cambios de la órbita de la tierra, gases de efecto invernadero (CO₂), deforestaciones, causando incrementos en la misma, provocando en algunos casos grandes pérdidas económicas, al no prever sobre esos cambios, es por ello que surge la necesidad de conocer su comportamiento para generar un modelo funcional que permita realizar predicciones, permitiendo así planificar estrategias tempranas para realizar actividades.

1.3. Formulación del Problema

¿Cuál de los modelos de predicción con datos funcionales ayuda a predecir con el menor error posible?

1.4. Justificación

1.4.1. Justificación Teórica

Las series de tiempo se suelen analizar con herramientas estadísticas basadas en ciertos supuestos como estacionariedad, aleatoriedad, entre otras, que por lo general no reflejan la variabilidad de las trayectorias o fluctuaciones propia de este tipo de variables, además por la presencia de la gran cantidad de datos existente, es necesario expresar la información como funciones definidas sobre un dominio temporal continuo. Una de las técnicas que se usa en la actualidad es el análisis funcional, porque permite analizar grandes cantidades de datos, mediante funciones, analizando la fluctuación de variables a través del tiempo, siendo esta técnica necesaria para conocer cómo se comporta la temperatura ambiental por la gran cantidad de datos almacenados en los últimos 6 años de las estaciones meteorológicas de la Provincia de Chimborazo, con el propósito de generar modelos estadísticos funcionales para la predicción de la temperatura ambiental hasta su máxima capacidad temporal, mediante regresión del componente principal funcional, generando nuevas líneas bases de la utilización de este método, el mismo que puede ser replicado para otras variables.

1.5. Objetivos

1.5.1. *Objetivos Generales*

Predecir la temperatura ambiental mediante modelos estadísticos funcionales de las estaciones monitoreadas por el GEAA en la provincia de Chimborazo (2014-2019).

1.5.2. *Objetivos Específicos*

- Validar la información proporcionada según como lo establece la OMM.
- Establecer el formato requerido de las bases de datos para realizar el estudio.
- Realizar un análisis exploratorio de datos funcional de la variable temperatura ambiental.
- Identificar modelos de datos funcionales para predecir la variable temperatura ambiental.
- Evaluar los modelos con el fin de seleccionar aquel que presente menor error posible.
- Mostrar las predicciones de la temperatura ambiental hasta su máxima capacidad temporal.

1.6. Bases Teóricas

1.6.1. *Análisis funcional*

Una variable funcional es aquella cuyos valores del tiempo se caracterizan porque se evalúan en cualquier momento en el dominio, a diferencia de la forma discreta, en la que originalmente se midieron u observaron. Por tanto, Los datos funcionales son representaciones de curvas $\{x_1(t), \dots, x_n(t)\}$ con $t \in T$. Cada curva se puede observar en diferentes momentos de su argumento t como $x_i = (x_i(t_o), \dots, x_i(t_{mi}))'$ o el conjunto de tiempos $t_o, \dots, t_{mi}, i = 1, \dots, n$ y estos no son necesariamente los mismos para cada curva.

Se busca construir una forma funcional de las curvas para evaluarlas en cualquier momento t . Este método asume que las curvas o trayectorias de las muestras son de dimensión finito, representado en base de funciones $\{\phi_1(t), \dots, \phi_p(t)\}$ y expresado como:

$$x_i(t) = \sum_{j=1}^p a_{ij} \phi_j(t) \quad i = 1, \dots, n \quad (1)$$

La forma funcional de las curvas se determina cuando se conocen los coeficientes base $a_i = (a_{i1}, \dots, a_{ip})'$. Estos se pueden obtener de las observaciones discretas por mínimos cuadrados o

por interpolación, dependiendo de las características de las curvas y las observaciones, se pueden utilizar varias clases de bases (Escabias y Mar, 2008, p.1).

1.6.2. Variable Funcional

Cuando X una variable aleatoria toma valores en el espacio funcional ε , ya sea normado o seminormado, se representa como variable funcional". Generalmente los datos funcionales se presentan discretizados sobre un conjunto de puntos $\{t_j\}, j = 1, \dots, d$, con $d > 0$, no necesariamente equidistantes (Castillo et al, 2017: p.2).

1.6.3. ¿Que son los datos funcionales?

Son un conjunto de datos $\{x_1(t), \dots, x_n(t)\}$, de la observación de n variables funcionales X_1, \dots, X_n idénticamente distribuidas. Cuando se empieza a trabajar con datos funcionales, se deberá determinar el espacio funcional con su respectiva métrica que se representará los datos.

Un espacio funcional utilizado comúnmente, es el espacio \mathbb{R}^n , cuando el área entre curvas puede proporcionar información relevante acerca del comportamiento de los datos, es usual recurrir al espacio funcional L_2 (el espacio de Hilbert de todas las funciones integrables cuadradas sobre un cierto intervalo) (Castillo et al, 2017: p.3). Para esto, suponiendo un dominio:

$$L^2 = \{X: T \rightarrow \mathbb{R}, \text{ tal que } \int X^2(t) dt < \infty\} \quad (2)$$

1.6.4. ¿Qué es un base funcional?

Son varias funciones independientes $\{\phi_p(t)\}$, tales que, pueden ser aproximadas mediante combinaciones lineales. De esta forma, la observación funcional puede aproximarse como: $x_i(t) = \sum_{j=1}^p a_{ij} \phi_j(t)$. Estas bases son útiles para obtener trayectorias de funciones, guardando la información presente en los datos (Carreño, 2013, p.20).

Existe muchas bases usadas en datos funcionales, entre ellas, las bases de Fourier para datos periódicos, B-splines, bases Wavelets, bases exponenciales y bases polinomiales. (Ginzo, 2008, p.42).

1.6.4.1. Bases de Fourier

Las bases de Fourier son usadas cuando se desea estimar trayectorias regulares o periódicas, son muy útiles para aproximar trayectorias muy estables, pero no son apropiadas si existe discontinuidad en trayectorias (Ginzo, 2008, p.43). La serie está definida por:

$$f(t) = c_0 + c_1 \text{sen}(wt) + c_2 \text{cos}(wt) + 3\text{sen}(2wt) + 4\text{cos}(2wt) + \dots \quad (3)$$

donde $T = [0, T]$ define el periodo de oscilación y $w = \frac{2\pi}{T}$ está formada por las siguientes funciones ortonormales:

$$\phi_0(t) = \frac{1}{\sqrt{T}} \quad (4)$$

$$\phi_{2r-1}(t) = \frac{1}{\sqrt{\frac{T}{2}}} \text{sen}(rwt) \quad (5)$$

$$\phi_{2r}(t) = \frac{1}{\sqrt{\frac{T}{2}}} \text{cos}(rwt) \quad (6)$$

1.6.4.2. Bases de Wavelets

Este tipo de bases son utilizadas cuando se desea representar datos funcionales, las mismas que son obtenidas por medio de dilataciones y cambios de una wavelet madre propia ϕ (Castro, 2019, p.9). Estas bases no afirman que los datos que se presentan son periódicos, como en las bases de Fourier. (Carreño, 2013, p.24).

$$\phi_{kj}(t) = 2^{k/2} \phi(2^k t - j) \quad (7)$$

1.6.4.3. Bases B-splines

Los B-splines son curvas polinómicas, que se diferencian según el orden que se presenten. La base B-splines de grado p (orden $p+1$) es una función de grado p , presentando derivadas continuas hasta $p-1$ en los intervalos extremos. Están representados como combinaciones lineales de B-splines que presentan un mismo grado de partición de dominio (Carreño, 2013, p.22).

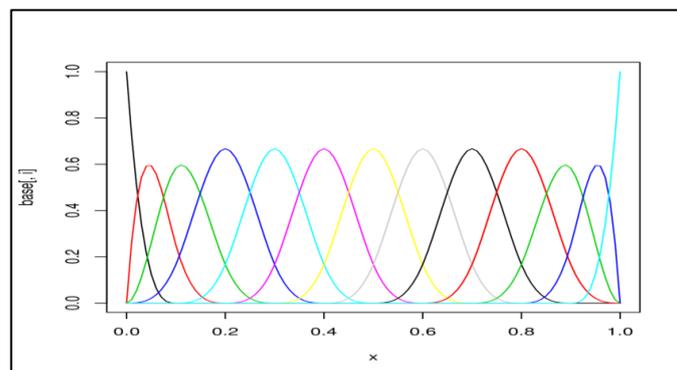


Figura 1-1: Bases B-splines

Fuente: Martínez, 2015, p.55.

Este tipo de bases cumplen con algunas características como:

- "Consiste en $p + 1$ trozos de polinomios de orden p que se unen en p nodos internos.
- Las derivadas hasta el orden $p - 1$ son continuas en los puntos de unión.
- El B-splines es positivo en el dominio expandido por $p + 2$ nodos y 0 en el resto.
- Excepto en los extremos, se solapa con $2p$ trozos de polinomios de sus vecinos.
- Para cada valor de x , $p + 1$ B-splines son no nulos.
- Para construir un B-splines de grado p son necesarios $p + 2$ nodos".

Se denota una partición de nodos en un intervalo T , su partición extendida es $t_{-3} < t_{-2} < t_{-1} < t_0 \dots < t_{m+1} < t_{m+2} < t_{m+3}$ la base B-splines $p + 1$ de grado p está definido de la siguiente forma:

$$B_{j,1}(t) = \begin{cases} 1 & t_{j-2} < t < t_{j-1} \\ 0 & \text{en otro caso} \end{cases} \quad (8)$$

con $j = -1, 0, 1, \dots, m + 4$

$$B_{j,p}(t) = \frac{t-t_{j-2}}{t_{j+p-2}-t_{j-2}} B_{j,p}(t) + \frac{t_{j+p-1}-t}{t_{j+p-1}-t_{j-2}} B_{j+1,p}(t) \quad (9)$$

Con $p = 1, 2, \dots$ y $j = -1, 0, 1, 2, \dots, m - p + 4$, si $p = 4$ estas bases son B-splines cúbicos, las cuales son utilizadas para aproximar curvas regulares, de esta manera este tipo de funciones son denotados como: $B_{j,4}(t) = B_j(t)$ con $j = -1, 0, 1, 2, \dots, m + 1$ (Escudero, 2016, pp.14-15).

1.6.4.4. Bases de potencias y exponenciales

Generalmente las bases de potencias se forman con monomios y son de la siguiente forma:

$$1, t, t^2, \dots, t^k, \dots \quad k = 1, 2, \dots \quad (10)$$

Las bases exponenciales se forman de la siguiente forma:

$$e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_j t} \quad j = 1, 2, \dots \quad (11)$$

En donde los parámetros λ_j son todos distintos, en ocasiones $\lambda_1 = 0$. Las ecuaciones diferenciales lineales con coeficientes constantes tienen como soluciones expansiones en términos de bases exponenciales (González, 2016, p.14).

1.6.4.5. Bases polinomiales

Las bases polinomiales son de la forma $\phi_j(t) = (t - \omega)^k, k = 0, \dots, K$, donde ω es un parámetro que se elige en el centro del intervalo de aproximación (Millán, 2017, p.31). Para evitar errores en los cálculos, es necesario escoger un buen parámetro. Uno de los problemas que se presenta en este tipo de bases es que no existe un buen ajuste en las colas.

1.6.5. Suavizamiento de la función

La forma funcional del dato $x(t)$, tiene ciertos puntos esparcidos en el dominio T, pero se observan los valores asociados a $x(t_1), \dots, x(t_m)$ para alguno $t_1, \dots, t_m \in T$, es decir en primera instancia se debe recuperar la función $X(s)$ para funciones puntuales, suavizar la función, utilizando la estructura de espacio vectorial de $L_2(T)$ para recrear la función con bases funcionales, obteniendo la forma funcional de las trayectorias muestrales mediante aproximaciones de los coeficientes básicos (Correal, 2017, pp.11-12).

1.6.5.1. Suavizar datos funcionales por mínimos cuadrados

La técnica de mínimos cuadrados es una optimización matemática, cuyo fin es obtener una función continua, que, bajo un criterio de mínimo de error cuadrático se pueda aproximar de forma correcta a los datos. (González, 2016, pp. 11-12).

Se tiene $y_i = x_{t_i} + \varepsilon$ que representa observaciones de una única curva, para poder estimar:

$$x(t) \approx \sum_{j=1}^k c_j \phi_j(t) \quad (12)$$

por medio de mínimos cuadrados, minimizando los errores cuadráticos de:

$$SEE = \sum_{i=1}^n (y_i - x(t_i))^2 = \sum_{i=1}^n (y_i - \Phi(t_i)c)^2 \quad (13)$$

Se expresa en terminos de bases funcionales, con errores independientes e idénticamente distribuidos:

$$x_i(t_k) = \sum_{p=1}^p a_{ip} \phi_p(t_k) \quad (14)$$

Con $k = 0, \dots, K$ $i = 1, \dots, N$ y expresando en forma matricial tenemos,

$$x_i = \Phi_i a_i \quad (15)$$

Donde $a_i = (a_{i1}, \dots, a_{ip})'$ y $\Phi_i = (\phi(t_k))_{K \times P}$ y x_i que es un vector de datos discretos como $x_i = (x_{i1}, \dots, x_{ik})'$, siendo la expresión básica a_{i1} coeficientes, aproximados por método de mínimo cuadrados y viene expresado de forma matricial:

$$ECM(x_i|a_i) = (x_i - \Phi_i a_i)'(x_i - \Phi_i a_i) \quad (16)$$

Derivando en base a a_i se obtiene:

$$2\Phi_i' a_i - 2\Phi_i' x_i = 0 \quad (17)$$

El estimador de mínimos cuadrados, despejando el coeficiente

$$\hat{a}_i = (\Phi_i' \Phi_i)^{-1} \Phi_i' x_i \quad (18)$$

Calculando el vector \hat{x}_i

$$\hat{x}_i = \Phi_i \hat{a}_i = \Phi_i (\Phi_i' \Phi_i)^{-1} \Phi_i' x_i \quad (19)$$

Finalmente se representa de la siguiente manera la curva aproximada y ajustada:

$$\hat{x}_i(t) = \hat{a}_i' \Phi(t) \quad (20)$$

1.6.5.2. Suavizar datos funcionales por mínimos cuadrados ponderados

Se utiliza mínimos cuadrados ponderados cuando no presentan errores idénticamente distribuidos y varianza dispersa, es decir no es constante. Para mejorar este problema se debe incluir un valor significativo, considerado como peso a los residuos, para realizar este proceso se utiliza un criterio de mínimos cuadrados. (Millán, 2017, p.33).

$$ECM(x_i|a_i) = (x_i - \Phi_i a_i)' W (x_i - \Phi_i a_i) \quad (21)$$

Al derivar la expresión y despejar el valor de a , se obtienen el estimador correspondiente a mínimos cuadrado ponderados. Cuando se desconocer Σ_e^{-1} , se asuma covarianzas de errores iguales a cero.

$$\hat{a}_i = (\Phi_i' W \Phi_i)^{-1} \Phi_i' W x_i \quad (22)$$

1.6.5.3. Mínimos cuadrados localizados suavizados mediante kernels

Los mínimos cuadrado localizados, mediante kernel, permiten suavizar los datos con observaciones cercanas (Millán, 2017, pp.35-37). Está representado el estimador de la función de la siguiente forma :

$$x(t) = \sum_{j=1}^p w_j y_j \quad (23)$$

Siendo así los valores de los pesos w_j grandes para valores muestrales t_j próximo a t (Aguilera, 2009, p.21). Existen algunos ejemplos de kernel como:

- Uniforme:

$$Kern(u) = \begin{cases} 0.5 & \text{si } |u| \leq 1 \\ 0, & \text{en otro caso} \end{cases} \quad (24)$$

- Cuadrático:

$$Kern(u) = \begin{cases} 0.75(1 - u)^2 & \text{si } |u| \leq 1 \\ 0, & \text{en otro caso} \end{cases} \quad (25)$$

- Gaussiano:

$$Kern(u) = (2\pi)^{-1/2} e^{(-u^2/2)} \quad (26)$$

Donde los pesos son:

$$w_j(t) = Kern\left(\frac{t_j - t}{h}\right) \quad (27)$$

Cuando se presentan valores relativamente pequeños en h , reciben mayor peso los valores más cercanos a t . Este valor de h es considerado como ancho de banda. Existen estimaciones como combinaciones lineales de observaciones locales, utilizando el criterio de mínimos cuadrados, despejando valores de \hat{a}_i , se obtiene el estimador:

$$x(t) = \sum_{j=1}^p S_j(t) y_j \quad (28)$$

$$SEE = \sum_{i=1}^p w_j(t) [x_i - \sum_{i=1}^K a_i \Phi_i(t_i)]^2 \quad (29)$$

$$w_j(t) = Kern\left(\frac{t_j - t}{h}\right) \quad (30)$$

En términos matriciales

$$ECM(x_i | a_i) = (x_i - \Phi_i a_i)' W(t) (x_i - \Phi_i a_i) \quad (31)$$

$$\hat{a}_i = (\Phi_i' W(t) \Phi_i)^{-1} \Phi_i' W(t) x_i \quad (32)$$

$$\hat{x}(t) = \sum_{k=1}^K \hat{a}_k \Phi_k(t) \quad (33)$$

$$\hat{x}(t) = \sum_{j=1}^p S_j(t) y_j \quad (34)$$

Siendo los valores S_j pertenecen al vector

$$S(t) = W(t) \Phi_i [\Phi_i' W(t) \Phi_i]^{-1} \phi_j(t) \quad (35)$$

1.6.6. Profundidad Funcional

Las medidas de profundidad son utilizadas para detección de outliers y cálculos estadísticos descriptivos funcionales, a continuación se describe las medidas más utilizadas (Estrada, 2014, p.4).

1.6.6.1. Profundidad Fraiman-Muniz (FM)

Según (Bárceñas, 2014, p. 52) la profundidad Fraiman-Muniz se define como " $F_{n,t}(x_i(t))$ " la distribución empírica acumulada de los valores de las curvas $x_1(t), \dots, x_n(t)$ en un instante de tiempo $t \in [a, b]$, entonces la profundidad funcional de Fraiman y Muniz de una curva x_i con respecto al conjunto x_1, \dots, x_n y viene dada por:

$$FM_n(x_i) = \int_a^b D_n(x_i(t)) dt \quad (36)$$

Donde $D_n(x_i(t))$ es la profundidad de $x_i(t)$ una curva univariada dada por:

$$D_n(x_i(t)) = 1 - \left| \frac{1}{2} - F_{n,t}(x_i(t)) \right| \quad (37)$$

Para cada $1 \leq i \leq n$

1.6.6.2. Profundidad modal (MD)

Esta profundidad se caracteriza por que es la curva más densa y el conjunto de curvas x_1, \dots, x_n se representa de la siguiente manera:

$$MD_n(x_i, h) = \sum_{k=1}^n K\left(\frac{\|x_i - x_k\|}{h}\right) \quad (38)$$

Donde K es una función kernel y h representa el ancho de banda (Sanchez y Rodriguez, 2018, p.13).

1.6.6.3. Profundidad de proyección aleatoria (RPD)

Este tipo de profundidad se caracteriza por que hace uso de las derivadas para poder construir una curva funcional de forma aleatoria sobre un punto R^2 y está representada por:

$$RPD_n(x_i) = \frac{1}{p} \sum_{p=1}^p D_n(\langle v_p, x_i \rangle, \langle v_p, x_i' \rangle) \quad (39)$$

Donde D_n denotada en R^2 del punto $(\langle v_p, x_i \rangle, \langle v_p, x_i' \rangle) \in R^2$

1.6.7. Análisis exploratorio funcional

En el análisis exploratorio funcional se debe considerar la media funcional, varianza funcional, desviación estándar funcional, entre otros. Esto ayuda a observar la forma de los datos, analizar la dispersión y variabilidad presente. Se las definen por la naturaleza de los datos del caso de estudio (Navarro Pérez, 2004, p.15).

1.6.7.1. Media funcional

La media funcional se la define como la función $x(t)$ o también como el valor medio de la función $x(t)$ y se la representa como:

$$\bar{x} = \frac{1}{T} \langle x, 1 \rangle = \frac{1}{\int_0^T 1(t)^2 dt} * \int_0^T x(t) * 1(t) dt \quad (40)$$

Donde la función $1(t) = 1 \forall t \in R$

1.6.7.2. Varianza funcional

Una muestra de funciones $x_1(t), x_2(t), \dots, x_n(t)$, de una función aleatoria (t) , la cual está definida en el intervalo $[0, T]$, se representa la función de varianza muestral de (t) como:

$$\text{var}(x(t)) = \frac{1}{n-1} \sum_{i=1}^n [x_i(t) - \bar{x}(t)]^2 \quad (41)$$

1.6.7.3. Covarianza funcional

Una muestra de funciones $x_1(t), x_2(t), \dots, x_n(t)$, de una función aleatoria (t) , la cual está definida en el intervalo $[0, T]$, se representa la función de covarianza muestral de (t) entre t_1 y t_2 como:

$$\text{cov}(x(t_1), x(t_2)) = \frac{1}{n-1} \sum_{i=1}^n (x_i(t_1) - \bar{x}(t_1))(x_i(t_2) - \bar{x}(t_2)) \quad (42)$$

1.6.7.4. Correlación Funcional

Una muestra de funciones $x_1(t), x_2(t), \dots, x_n(t)$, de una función aleatoria (t) , la cual está definida en el intervalo $[0, T]$ (Taguacundo, 2019, p.21). Se representa la función de correlación muestral de (t) entre t_1 y t_2 como:

$$\text{cor}(x(t_1), x(t_2)) = \frac{\text{cov}_x(t_1, t_2)}{\sqrt{\text{var}_x(t_1)\text{var}_x(t_2)}} \quad (43)$$

1.6.8. Análisis funcional en componentes principales

Análisis en componentes principales (ACP), es sin lugar a duda uno de los métodos usados para la reducción de la dimensionalidad, disminuyendo así el número de variables. Cuando se trata de componentes principales funcionales (AFCP) se empieza a trabajar con funciones, espacios funcionales, entre otros. EL objetivo del AFCP es el mismo que el del ACP, el cual consiste en reducir la dimensionalidad de las variables (Chávez et al , 2015, p.16).

Entonces se podría decir que $x_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, p$, donde p representa el número de variables (Pérez, 2016, p. 3). Las componentes presentan valores que se denominan peso $\alpha_{11}, \alpha_{21}, \dots, \alpha_{p1}$ que maximizan lo siguiente:

$$\frac{1}{n-1} \sum_{i=1}^n \left[\sum_{j=1}^p \alpha_{j1} x_{ij} \right]^2 \quad (44)$$

En donde $\sum_{j=1}^p \alpha_{j1}^2 = 1$. Cuando se presenta en forma funcional se reemplaza los vectores de pesos, por funciones $\alpha_1(t)$ y las sumatorias por integrales, entonces $\alpha_1(t)$ cuando es continuo tenemos:

$$\frac{1}{n-1} \sum_{i=1}^n [\int \alpha_1(t) x_i(t) dt]^2 \quad (45)$$

La componente principal funcional se representa como:

$$\alpha_1(t) = \max \|\alpha\|^2 = \frac{1}{n-1} \sum_{i=1}^n [\int \alpha_1(t) x_i(t) dt]^2 \quad (46)$$

1.6.9. Regresión Funcional del Componente Principal

Sea $(f_t: t \in Z)$ una serie de tiempo funcional estacionaria arbitraria, con observaciones f_t elementos del espacio $H = L^2(I)$, equipado con el producto interior $\langle a, b \rangle = \int_I a(x)b(x)dx$, donde x representa un continuo dentro de un rango de soporte de función denotado por I . Cada función es una función integrable cuadrada que satisface $\|f_t\|^2 = \int_I f_t^2(x)dx < \infty$. La notación $f \in L_H^p(\Omega, A, P)$ se utiliza para indicar $p > 0, E(\|f\|^p) < \infty$. Cuando $p = 1, f(x)$ tiene la curva media $\mu(x) = E[X(x)]$; cuando $p = 2$, el operador de covarianza $K(s, x) = Cov[f(s), f(x)] = E\{[f(s) - \mu(s)][f(x) - \mu(x)]\}$ y está definido por:

$$K(s, x) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(x), \quad s, x, \in I \quad (47)$$

Donde $\phi_k(x)$ denota k el componente principal ortonormal y λ_k denota el k valor propio, basado en la separabilidad del espacio de Hilbert. La expansión de un proceso estocástico f_t se puede expresar como:

$$f_t(x) = u(x) + \sum_{k=1}^{\infty} \beta_k \phi_k(x) \quad (48)$$

Donde las puntuaciones, también llamados coeficientes de los componentes principales β_k están dadas por la proyección $[f(x) - \mu(x)]$ en la dirección de la k^{th} función propia ϕ_k , i. e., $\beta_k = \langle f - \mu, \phi_k \rangle$

Se detalla cada uno de los pasos:

- El primer paso es suavizar la temperatura ambiental usando regresión splines.

- Una vez que $f_t(x)$ es visualizada como una función continua, haciendo énfasis a la hora x que es vista como variable continua. En lo que sigue $y_t(x_i)$ denotará la temperatura ambiental para la hora $x_i \in [x_1, \dots, x_p]$ en el día t . Se representa $\sigma_t(x_i)$ como desviación estándar del error observacional, que valora el valor de error presente en la variación de la temperatura ambiental en el día t , entonces el conjunto $\{x_i, y_t(x_i)\}$ observaciones satisface:

$$y_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \quad (49)$$

Con $i = 1, \dots, p$; $t = 1, \dots, n$

- A nivel de población, un proceso estocástico denotado por f se puede descomponer en la función media y la suma de las multiplicaciones de componentes principales funcionales ortogonales y las puntuaciones de componentes principales (Shang, 2013, p.2). Puede expresarse como:

$$f_t(x) = u(x) + \sum_{k=1}^{\infty} \beta_k \phi_k(x) \quad (50)$$

- Donde μ es la función media de la población no observable, B_k es la k -ésima puntuación de los componentes principales y ϕ_k es la k -ésima componente principal funcional de la población. En la práctica solo se puede observar n realizaciones de evaluadas en un intervalo $x \in [0, \tau]$, denotado por $f_t(x)$ para $t = 1, 2, \dots, n$. A nivel de muestra la descomposición del componente principal funcional se puede escribir como:

$$f_t(x) = \bar{f}_t(x) + \sum_{k=1}^K \hat{\beta}_{t,k} \hat{\phi}_k(x) + \hat{\varepsilon}_t(x) \quad (51)$$

Donde:

- " $\bar{f}_t(x)$ es la función media, representando el promedio de temperatura ambiental de las horas a través de los días.
- $\hat{\phi}_1(x), \dots, \hat{\phi}_k(x)$ denotan las funciones básicas ortonormales, i.e, conforman el conjunto de las primeras k componentes principales funcionales, con $k < n$;
- $\hat{\beta}_{t,1}, \dots, \hat{\beta}_{t,k}$ es un conjunto de puntuaciones incorreladas de las componentes principales funcionales seleccionadas
- $\hat{\varepsilon}_t(x) \sim N(0, \sigma_x^2)$ denota la función residual, de media cero".

El valor de K es escogido como un valor representativo de la variabilidad explicada por todos los componentes, entonces K es:

$$K = \underset{K: K \geq 1}{\operatorname{argmin}} \left\{ \frac{\sum_{k=1}^K \hat{\lambda}_k}{\sum_{k=1}^{\infty} \hat{\lambda}_k} 1\{\hat{\lambda}_k > 0\} \geq \delta \right\} \quad (52)$$

Fijando k la secuencia $\hat{\beta}_{1,k}, \hat{\beta}_{2,k}, \hat{\beta}_{n,k}$ son realizaciones de una serie temporal $\{\beta_{t,k}\}_t$. Para poder obtener las predicciones de estos coeficientes para h años más adelante, se puede usar modelos de series de tiempo univariantes o multivariantes.

Entonces $f(x) = [f_1(x), f_2(x), \dots, f_n(x)]^T$ que son datos observados y el conjunto de las componentes principales funcionales $B = [\hat{\phi}_1(x), \hat{\phi}_2(x), \dots, \hat{\phi}_k(x)]^T$, las predicciones siguientes de h de $y_{n+h}(x)$ se consiguen como:

$$y_{n+h|n}(x) = \mathbf{E}[y_{n+h}(x)|f(x), B] = \bar{f}_t(x) + \sum_{k=1}^K \hat{\beta}_{n+h|n,k} \hat{\phi}_k(x) \quad (53)$$

Donde $\hat{\beta}_{n+h|n,k}$ denota la predicción de h pasos delante de $\hat{\beta}_{n+h,k}$ usando series de tiempo univariante o multivariante (González, 2013, pp.13-15).

1.6.9.1. Análisis Robusto de Componentes Principales

En presencia de valores atípicos, el operador de covarianza $K(s, x) = \operatorname{Cov}[f(s), f(x)]$ puede no ser robusto contra valores atípicos y los resultados de los componentes principales funcionales estimados pueden ser erróneos. Para evitar ese problema se utiliza el FPCA robusto, del algoritmo de Hyndman y Ullah, el cual consiste en extraer componentes principales funcionales robustos reduciendo el efecto de los valores atípicos (Shang, 2017, p.11-15).

Se calcula el error cuadrático integrado de la curva

$$v_i = \int_a [\bar{f}_t(x) - \sum_{k=1}^K \hat{\beta}_{t,k} \hat{\phi}_k(x)]^2 dx, \quad \text{para } t = 1, \dots, n, \quad (54)$$

Donde las curvas periféricas tienden a tener grandes valores de v_i pesos, luego se le asigna $v_i < s + \lambda\sqrt{s}$ donde $\lambda > 0$ representa un parámetro de ajuste para controlar la cantidad de robustez y s es la media de $\{v_1, \dots, v_n\}$, según (Hyndman & Ullah, 2007, p.6) cuando $\lambda = 2.33$ la eficiencia de $\Phi(2.33/\sqrt{2}) = 95\%$, es decir que el 5% se trata como atípico.

1.6.10. Técnicas de Predicción de series de tiempo univariadas y multivariadas

Para obtener $\hat{\beta}_{n+h|n,k}$ utilizando una técnica de predicción se series de tiempo univariadas, es decir un modelo ARIMA, capaz de modelar series de tiempo no estacionarias. Las puntuaciones de los componentes principales tienen de forma general la forma:

$$(1 - \varphi_1^k B - \dots - \varphi_p^k B^p)(1 - B)^{d^k} \beta_{t,k} = \alpha + (1 + \theta_1^k B + \dots + \theta_q^k B^q) w_{t,k}, \quad (55)$$

Donde α representa el intercepto, $(\varphi_1^k, \dots, \varphi_p^k)$ los coeficientes asociados con el componente autorregresivo, $(\theta_1^k, \dots, \theta_q^k)$ son los coeficientes asociados con el componente de media móvil, B denota el operador de retroceso, d^k es el operador de diferencia para k^{th} puntaje del componente principal, $\beta_{t,k}$ representa el k^{th} puntaje estimado de componentes principales para el t^{th} observación y $w_{t,k}$, es el error. El modelo ARIMA óptimo se selecciona en función de un criterio de información.

El modelado de series de tiempo univariante puede ocasionar pérdida de información en la dependencia de la matriz de puntuaciones del componente principal, es por ello que se ha implementado técnicas multivariantes para dar solución a este problema.

El modelo de series de tiempo multivariante comúnmente utilizado es el modelo VAR, es fácil de estimar por medio de mínimos cuadrados ordinarios (MCO), máximo verosimilitud o método bayesiano.

La serie de tiempo multivariante de puntuaciones de componentes principales $\beta_k = [\beta_{1,k}, \dots, \beta_{n,k}]^T$ sigue un VAR, modelo de orden ϑ , VAR (ϑ), si:

$$\beta_k = \phi_0 + \sum_{v=1}^{\vartheta} \phi_v \beta_{k-v} + a_k \quad (56)$$

Donde ϕ_0 es un vector constante n -dimensional vector constante, ϕ_v son $n \times n$ matrices para $v > 0$ y $\phi_v \neq 0$, a_k es un conjunto de vectores de error aleatorio idénticamente distribuido (iid) con media cero y una matriz de covarianza definida, que cuenta con todos los valores propios positivos y ϑ puede ser determinado por un criterio de información.

El VAR (ϑ) puede escribirse como: El VAR (1)

$$\beta_k = B^T x_k + a_k, \quad k = \vartheta + 1, \dots, n \quad (57)$$

Con $x_k = (1, \beta_{k-1}^T, \dots, \beta_{k-\vartheta}^T)^T$ y $B = (\phi_0^T, \phi_1^T, \dots, \phi_{\vartheta}^T)^T$ contiene todos los coeficientes de regresión desconocidos. $X = (x_{\vartheta+1}, \dots, x_n)^T$ contiene los valores de las variables explicativas y

$Y = (\beta_{\vartheta+1}, \dots, \beta_n)^T$ como la matriz de variable respuesta. Los coeficientes de regresión desconocidos pueden ser estimados por MCO

$$\hat{B}_{MCO} = (X^T X)^{-1} X^T Y \quad (58)$$

h pronósticos de puntajes de componentes principales y curva de pronósticos son:

$$\hat{\beta}_{n+h|n} = \hat{B}_{MCO} (x_{n+h}) \quad (59)$$

$$y_{n+h|n}(x) = \hat{\beta}_{n+h|n,k} \hat{\phi}_k(x) \quad (60)$$

Donde:

$$\hat{\phi}_k = [\phi_1(x), \dots, \phi_k(x)] \quad (61)$$

Un modelo VAR está restringido a series de tiempo multivariadas estacionarias, aunque un modelo de corrección de errores vectoriales puede tratar con series de tiempo multivariadas no estacionarias. El número de parámetros sigue siendo comparativamente mayor que aquellos del modelo de serie de tiempo univariante, y esto puede dificultar la precisión de la estimación de parámetros, especialmente para un tamaño de muestra pequeño.

1.6.11. Método de pronósticos de intervalos

Los intervalos de predicción son una herramienta que ayuda a medir la incertidumbre probabilística asociada con pronósticos puntuales, es importante proporcionar pronósticos de intervalo para poder evaluar la incertidumbre futura para la planificación de diferentes estrategias. Los errores para las puntuaciones estimadas de los componentes principales, están dadas por:

$$\hat{w}_{j,k} = \hat{\beta}_{n-j+1,k} - \hat{\beta}_{n-j+1|n-j,k} \quad \text{para } j = 1, \dots, n - K \quad (62)$$

Donde $\hat{\beta}_{n-j+1,k}$ denota el k^{th} puntaje estimado del componente principal por día $n - j + 1$; $\hat{\beta}_{n-j+1|n-j,k}$ que es un pronóstico de un paso adelante obtenido por una de las técnicas de series univariante o multivariante, K representa el número de componentes principales, mientras que $n - K$ se observa como el número total de errores de puntuación de componentes principales en la muestra. Cuando el primer K de la descomposición del primer componente principal se aproxima a los datos, los residuos del modelo presentan un ruido aleatorio y estos valores residuales se los representa como: $\hat{\epsilon}_{n+1}(x)$. Sumando estas fuentes de errores se obtiene:

$$\hat{y}_{n+1|n}^b(x) = \hat{f}_t(x) + \sum_{k=1}^K \hat{\beta}_{n+1|n,k}^b \hat{\Phi}_k(x) + \hat{e}_{n+1}^b(x) \quad (63)$$

1.6.12. Medidas de Evaluación de pronósticos

1.6.12.1. Medidas dependientes de la escala

- *MSE (Mean Square Error)*

Se lo conoce como error cuadrático medio, mide el promedio de los errores al cuadrado, se lo calcula con la siguiente fórmula:

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (64)$$

Donde:

n representa el número de muestras, mientras que \hat{y}_t son los valores estimados.

- *RMSE (Root Mean Square Error)*

Se lo conoce como raíz del error cuadrático medio y mide la raíz promedio de los errores al cuadrado, se lo calcula con la siguiente fórmula:

$$MSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (65)$$

Donde:

n representa el número de muestras, mientras que \hat{y}_t son los valores estimados.

- *MAE (Mean Absolute Error)*

Permite calcular el error absoluto medio y se obtiene los valores con la siguiente formula:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (66)$$

- *MdAE (Median Absolute Error)*

Permite calcular la diferencia de la mediana absoluta entre dos valores, se la obtiene mediante la fórmula siguiente:

$$MdAE = \text{Mediana}(|y_t - \hat{y}_t|) \quad \text{para } t = 1, \dots, n \quad (67)$$

1.6.12.2. *Medidas basadas en porcentaje*

▪ *MAPE (Mean Absolute Percentage Error)*

Representa el error porcentual absoluto medio, mide el tamaño de error en terminos porcentuales y se lo calcula en base a la siguiente fórmula:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (68)$$

Donde:

n representa el número de muestras, mientras que \hat{y}_t son los valores estimados.

▪ *MDAPE (Median Absolute Percentage Error)*

Se denomina error porcentual absoluto de la mediana y mide el tamaño de error en terminos porcentuales y se lo calcula en base a la siguiente formula:

$$MAPE = \text{Mediana} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (69)$$

Donde:

n representa el número de muestras, mientras que \hat{y}_t son los valores estimados.

▪ *RMSPE (Root Mean Square Percentage Error)*

Representa la raíz del error porcentual cuadrático medio y se lo calcula con la siguiente formula:

$$RMSPE = \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{y_t - \hat{y}_t}{y_t} \right)^2} \quad (70)$$

Donde:

n representa el número de muestras, mientras que \hat{y}_t son los valores estimados.

▪ *RMDSPE (Root Median Square Percentage Error)*

Se denomina raíz del error porcentual cuadrático de la mediana y se lo calcula con la siguiente formula:

$$RMDSPE = \sqrt{\text{Median} \left(\left(\frac{y_t - \hat{y}_t}{y_t} \right)^2 \right)} \quad (71)$$

1.6.13. Medidas de Precisión del Pronóstico

1.6.13.1. Coeficiente U de Theil

El valor del coeficiente varía entre 0 y 1, este coeficiente permite observar que tan buena es la precisión del modelo, cuando este valor es próximo a 0 el modelo presenta una buena precisión, a diferencia de valores cercanos a 1, el modelo no es certero y poco confiable. (Vélez, 2016, pp.54-56),

$$U = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n (y_t^2) + \frac{1}{n} \sum_{t=1}^n (\hat{y}_t^2)}} \quad (72)$$

1.6.13.2. Test de Diebold -Mariano

Este tipo de test , es una medida de precisión , que mediante planteamiento de hipótesis según el investigador puede observar la precisión del modelo ,presenta un test de igualdad entre dos conjuntos de datos (Acurio y Pilco, 2019, pp.45-55).

1.7. Bases Conceptuales

1.7.1. ¿Qué es la meteorología?

Es una ciencia que estudia los fenómenos de la atmosfera y sus propiedades en un tiempo y lugar específico, describe condiciones climáticas y su evolución en un tiempo determinado. Cuando se estudia la atmósfera implica una serie de magnitudes o variables meteorológicas a observar, dentro de las cuales se tiene temperatura, presión atmosférica, humedad, entre otras, las mismas que varían según el espacio geográfico y el tiempo (Rodríguez et al., 2004, p.6).

1.7.2. Organización Meteorológica Mundial (OMM)

La OMM fundada en 1873 es el portavoz autorizado del sistema de las Naciones Unidas sobre el tiempo, el clima y el agua, el mismo que contribuye de forma significativa al desarrollo sostenible, reducción de pérdida de vidas y bienes, que se da por los desastres naturales, respecto al tiempo, clima y agua. Esta organización se encarga de proteger al medio ambiente para las presentes y futuras generaciones, proporcionando pronósticos y alertas tempranas para salvar vidas, reducir daños económicos, materiales y medioambientales. Por medio de la OMM se realizan actividades

de investigación con el objetivo de elaborar modelos informáticos que representen la mayor exactitud posible de predicciones meteorológicas (OMM, 2009, pp.3-5).

1.7.3. ¿Calidad de los datos?

El control de calidad de los datos es importante para verificar si el valor de un dato es representativo de la medición que se ha efectuado y no se ha visto alterado por factores que no se relacionen con el mismo, es por ello que las observaciones deben ser sometidos a un control de calidad adecuado que permita garantizar e identificar que la hora y los elementos observados sean coherentes, este procedimiento se deberá realizar antes de analizar los datos, mediante técnicas estadísticas, para así evitar errores de interpretación de resultados (Organización Meteorológica Mundial, 2011, p.52).

Los problemas que se presentan generalmente en controlar la calidad de los datos, es la cantidad de observaciones, haciendo que este proceso se vuelva tedioso, es por ello que gracias al avance de la tecnología existen softwares estadísticos que permiten mejorar el proceso de validación, identificando los valores válidos e inválidos que se presenten en una base de datos (Organización Meteorológica Mundial, 2011, p.53).

1.7.4. Tipos de Errores existentes en bases de datos

Existen diversos tipos de errores presentes en una base de datos con observaciones de variables meteorológicas como: fecha incorrecta, instrumentos de medición mal utilizados o en mal estado, al momento del proceso de transmisión de datos, introducción de datos por teclado, validación, modificación y manipulación de los formatos de datos (Organización Meteorológica Mundial, 2011, p.53). Estos son algunas de los errores que se presentan de manera habitual, pero que deben solucionarse mediante procedimientos automáticos o semiautomáticos, evitando así la presencia de estos errores, mediante pruebas de formato (corrección de fechas), completitud de datos, pruebas de coherencia (variación de los elementos) y pruebas de tolerancia (valores umbrales) (Organización Meteorológica Mundial, 2011, pp.54-55).

1.7.5. Valores Umbrales

Los valores umbrales sirven para verificar si los valores de datos instantáneos (promedio de un minuto o suma) están dentro de los límites aceptados, los mismos que dependen de diferentes parámetros meteorológicos y de las condiciones climatológicas de un sitio o de una temporada (Zahumenský, 2004, p.6).

Tabla 1-1: Valores umbrales de la temperatura del aire

Variable	Valores Umbrales	
	Mínimo	Máximo
Temperatura Ambiental	-80	60

Fuente: Zahumenský,2004.

Realizado por: Mariño Gabriela ,2021.

1.7.6. ¿Qué es la temperatura del aire?

La temperatura es una de las variables que se usa para explicar la atmosfera, la misma que cambia en el transcurso del día y noche, dependiente del lugar, su altitud y latitud, es decir de la ubicación geográfica. La temperatura es una magnitud que se relaciona con la rapidez del movimiento de las partículas que constituyen la materia, es decir que a mayor agitación que se presente, mayor será la temperatura. El instrumento utilizado para evaluar la temperatura se llama termómetro de mercurio y su unidad de medida está dada en escala de Celsius, expresada en grados centígrados(°C) y en escala de Fahrenheit(°F) (Rodríguez et al., 2004, pp.15-16).

1.7.7. Naturaleza Cíclica de los cambios de la temperatura

Se produce un cambio en la temperatura en la diferencia de la energía que ingresa es decir la solar, y la energía saliente, la que se encuentra en el suelo, conocida como terrestre, cuando existe un aumento considerable de energía entrante a comparación de la saliente, hace que se presente valores positivos. Es así como se va aumentando la temperatura del aire de manera indirecta , al incrementar la temperatura de la superficie terrestre (Andrades y Muñoz, 2012, p.19). En esta situación la superficie terrestre va aumentando de temperatura, pero cuando se presente al revés, es decir mayor energía terrestre y menor energía solar, la temperatura del aire disminuye. A continuación, se observa como varia la temperatura del aire de forma diaria, alcanzando temperaturas elevadas en el intervalo de 11 a 16 horas y luego empieza a disminuir (Andrades y Muñoz, 2012, p.20).

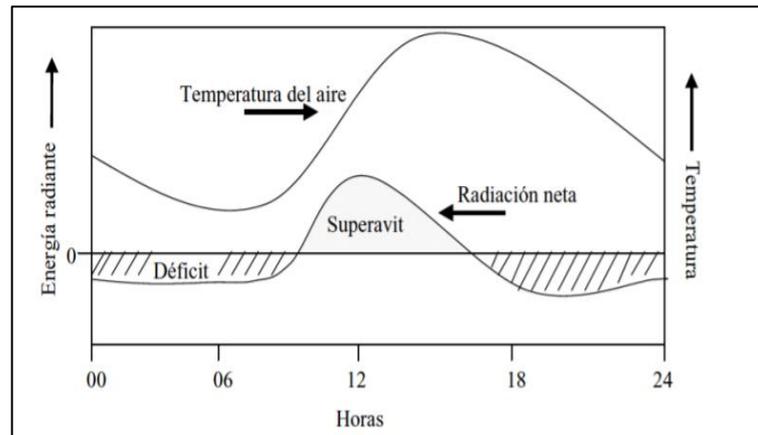


Figura 2-1: Curvas típicas de radiación y temperatura del aire

Fuente: Andrades , Muñez.2012.

1.7.8. Aparatos de medida

- Termómetros de dilatación. - Pueden ser termómetros de mercurio o de alcohol ayudando a observar las temperaturas máximas y mínimas.
- Termógrafos. -Requieren calibraciones, son muy pocos precisos.
- Termómetros de infrarrojos. – Estos sensores es utilizado con mucha frecuencia , empleados en áreas de cultivo para registrar su máxima longitud (Andrades y Muñez , 2012, p.24).

1.7.9. Método de Imputación Random Forest

Existen muchos métodos de imputacion de datos, que permiten completar esos datos faltantes. En la actualidad se han incrementado métodos, mediante arboles aleatorios denominado así por su nombre en inglés como Random Forest, en donde se genera arboles de forma aleatorio, utilizando el método de bootstrapping con muestras con reemplazamiento. El valor de cada nodo se lo escoge aleatoriamente, y siendo $m < p$, donde p representa el número de variables utilizadas para la distribución. El algoritmo permite predecir los valores promediando las predicciones de los árboles(Campos, 2017, pp.34-35).

1.7.10. Ventajas de Random Forest

El método tiene varias ventajas como:

- Es certero para un conjunto de datos extenso.
- Imputa grandes cantidades de valores perdidos, manteniendo la similitud con los datos reales.
- Realiza calificación mediante la relación entre las variables.
- Permite localizar valores atípicos.

1.7.11. Función rfImpute

RfImpute, utiliza el propio método del random forest para hacer las imputaciones. A partir del resultado de un random forest se calcula una medida de similitud o de proximidad entre dos individuos computando el número de árboles del random forest que clasifican en el mismo nodo final a ambos. La función comienza haciendo una primera imputación en base a la moda para variables categóricas o la media para las continuas.

1.7.12. Software Estadístico R

El software R es una de las herramientas que más se utiliza para hacer análisis de datos, además es flexible, gratuito y de código abierto, con varias librerías y paquetes que facilitan el análisis y el cálculo para el proceso de los datos, agilitando los resultados, siendo estos más precisos. Uno de los paquetes incorporados por R permite analizar datos mediante funciones, calculando bases funcionales, estadísticos descriptivos funcionales, cálculo de componentes principales funcionales, entre otros (Santana y Farfán, 2014, p.7).

CAPÍTULO II

2. MARCO METODOLÓGICO

2.1. Tipo y Diseño de Investigación

Por el método de investigación es de tipo cuantitativa porque se estudia fenómenos observables a través de técnicas estadísticas, matemáticas y computacionales, según el objetivo es aplicada , es decir que se realizara un análisis de los datos en el software R, según el nivel de profundización en el objeto de estudio es descriptivo , es decir analiza un conjunto de datos meteorológicos en un intervalo de tiempo , el mismo que permite describir el comportamiento de los datos a través de los años, según la manipulación de las variables es no experimental porque no se efectuará ningún experimento para la obtención de los datos, siendo recolectadas de fuente secundaria, Según el periodo temporal es longitudinal , es decir se estudia una única variable, analizándola en un tiempo de 6 años.

2.2. Diseño de la Investigación

La toma de los datos fué realizada por el Grupo de Energías Alternativas y Ambiente (GEAA), por medio de instrumentos de medición en salidas de campos.

2.3. Localización del Proyecto

La investigación es desarrollada en la provincia de Chimborazo, en las siguientes estaciones:

N	Estaciones	Parroquia	Cantón	X	Y
1	Cumandá	Cumandá	Cumandá	706262	9755579
2	San Juan	San Juan	Riobamba	746640	9818903
3	Tixán	Tixán	Alausí	749115	9761350
4	Tunshi	Licto	Riobamba	764074	9806681
5	Urbina	San Andrés	Guano	754579	9835357
6	Alao	Pungala	Riobamba	773491	9793166
7	Atillo	Cebadas	Guamote	772589	9758044
8	Epoch	Lizarzaburo	Riobamba	758398	9816965
9	Matus	Matus	Penipe	777564	9827892
10	Multitud	Multitud	Alausí	725681	9764904
11	Quimiag	Quimiag	Riobamba	770082	9816406

Figura 1-2: Coordenadas de las Estaciones del GEAA

Fuente: GEAA,2019.

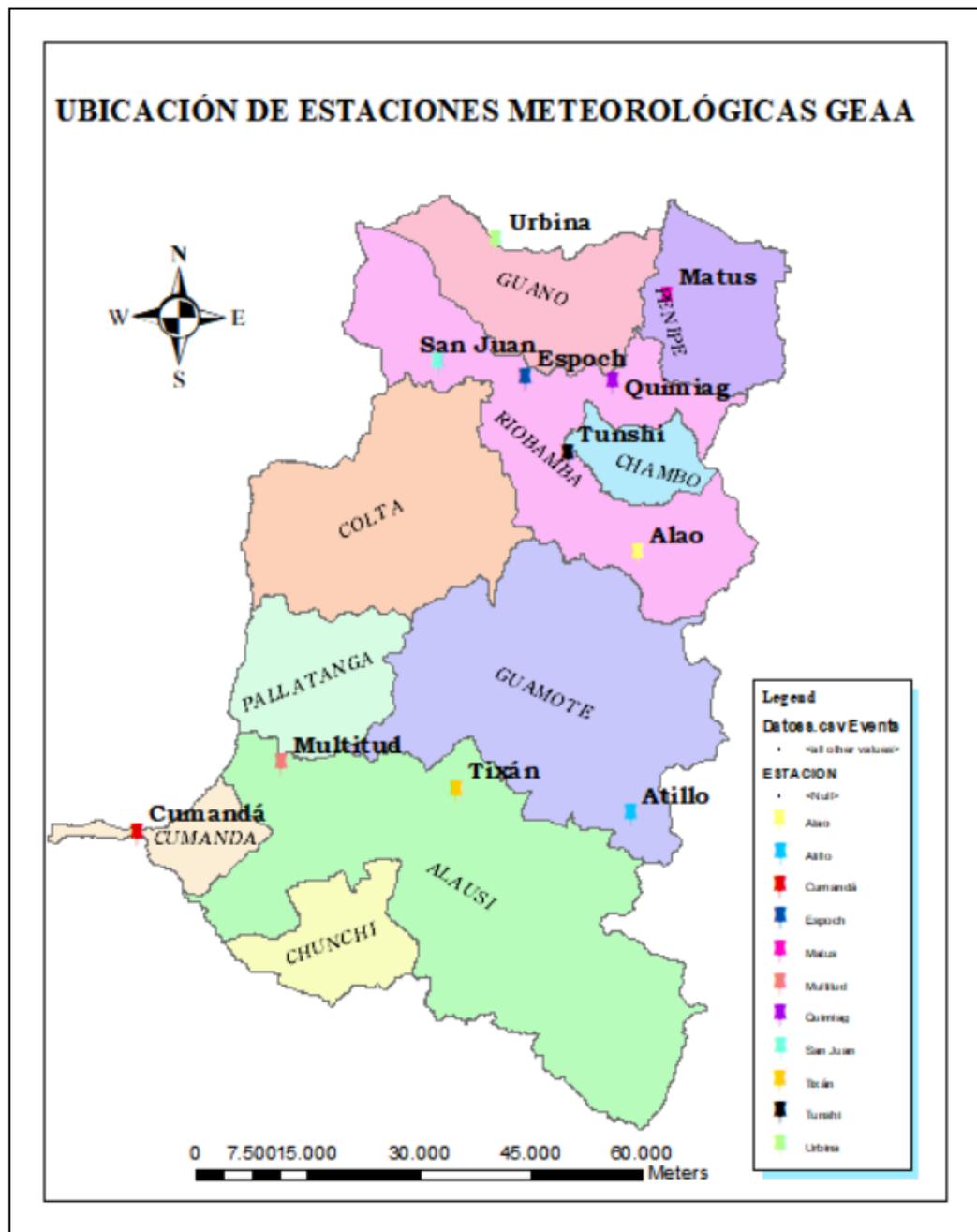


Figura 2-2: Mapa de Ubicación de Estaciones Meteorológicas del GEAA

Realizado por: Mariño Gabriela, 2021.

2.4. Población de estudio

La población de estudio comprende las bases de datos del año 2014 al 2019 de las estaciones meteorológicas del GEAA: Alao, Atillo, Cumandá, EsPOCH, Matus, Multitud, Quimiag, San Juan, Tixán, Tunshi y Urbina de la variable temperatura ambiental, con un total de 445165 valores registrados de temperatura ambiental.

2.5. Tamaño de la Muestra

Se trabajó con toda la población de estudio

2.6. Método de Muestreo

No se realizó ningún tipo de muestreo para la toma de los datos

2.7. Recolección de la Información

La variable temperatura ambiental es tomada mediante un termómetro de mercurio, cuya información es almacenada por el Grupo de Energías Alternativas y Ambiente de la ESPOCH.

2.8. Análisis Estadístico

Se utiliza los paquetes para el análisis de datos funcionales que se encuentran en el Software R, haciendo uso funciones que permitan imputar los datos, analizar estadísticas descriptivas funcionales, modelado de datos por medio de regresión del componente principal funcional de pronóstico con coeficientes de series de tiempo univariante y con coeficientes de series de tiempo multivariante.

2.9. Operacionalización de las variables

Tabla 1-2: Operacionalización de las Variables

Variable	Definición	Unidad de medida	Escala
Temperatura ambiental	La temperatura indica en valores numéricos el nivel de energía interna que se encuentra en un lugar en ese momento.	Grados Centígrado	Intervalo

Fuente: GEAA,2019

Realizado por: Mariño Gabriela.,2021.

CAPÍTULO III

3. RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS

3.1. Base de datos

Los datos de la Temperatura Ambiental de las 11 estaciones meteorológicas del 2014 al 2019, fueron proporcionados por el GEAA y contienen las mediciones por hora de la variable temperatura ambiental, dando un total de 8760 datos recopilados por año, a excepción del 2016 que presenta una cantidad de 8784 datos por ser año bisiesto. En la siguiente tabla se observa el porcentaje de faltantes de cada Estación.

Tabla 1-3: Porcentaje de Faltantes por año de cada estación

ESTACIONES	AÑOS					
	2014	2015	2016	2017	2018	2019
ALAO	0,8%	0,1%	0,1%	0,3%	0,6%	5,8%
ATILLO	6,70%	0,1%	0,1%	4,6%	26,1%	67,9%
CUMANDA	1,8%	0,8%	39,9%	29,8%	19,2%	26,6%
ESPOCH	24,7%	2,0%	0,1%	4,6%	19,2%	8,0%
MATUS	90,7%	1,94%	0,1%	0,05%	5,3%	31,8%
MULTITUD	5,5%	6,3%	0,1%	18,3%	0,6%	5,4%
QUIMIAG	0,2%	0,1%	0,1%	0,2%	81,0%	8,6%
SAN JUAN	0,3%	0,1%	0,1%	0,3%	1,9%	5,5%
TIXÁN	3,3%	42,0%	0,1%	0,3%	11,1%	5,5%
TUNSHI	2,7%	0,1%	7,2%	20,4%	28,2%	44,6%
URBINA	0,4%	0,1%	0,1%	0,0%	47,7%	18,7%

Fuente: GEAA,2019.

Realizado por: Mariño Gabriela,2021.

3.2. Validación de los datos

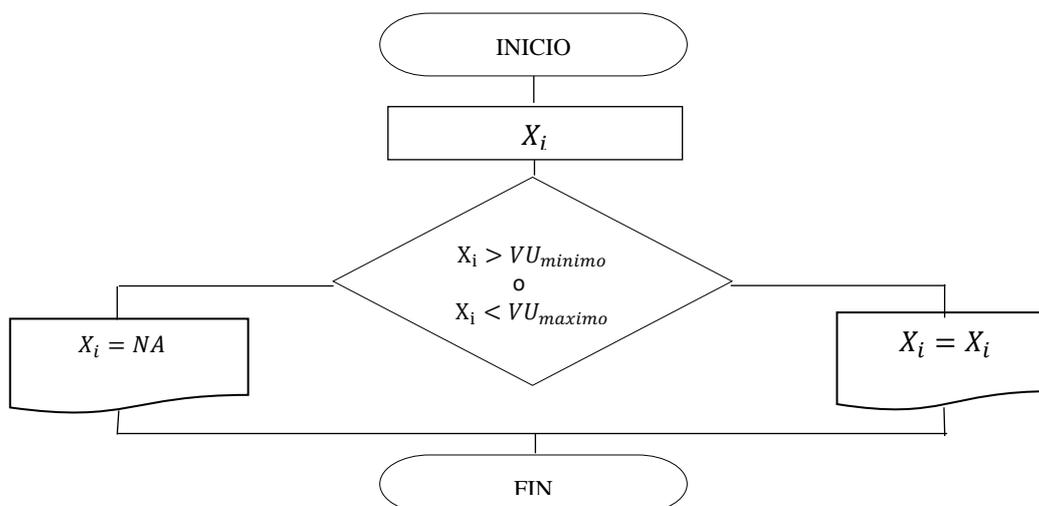


Gráfico 1-3: Validación de Información prueba de tolerancia

Realizado por: Mariño Gabriela,2021.

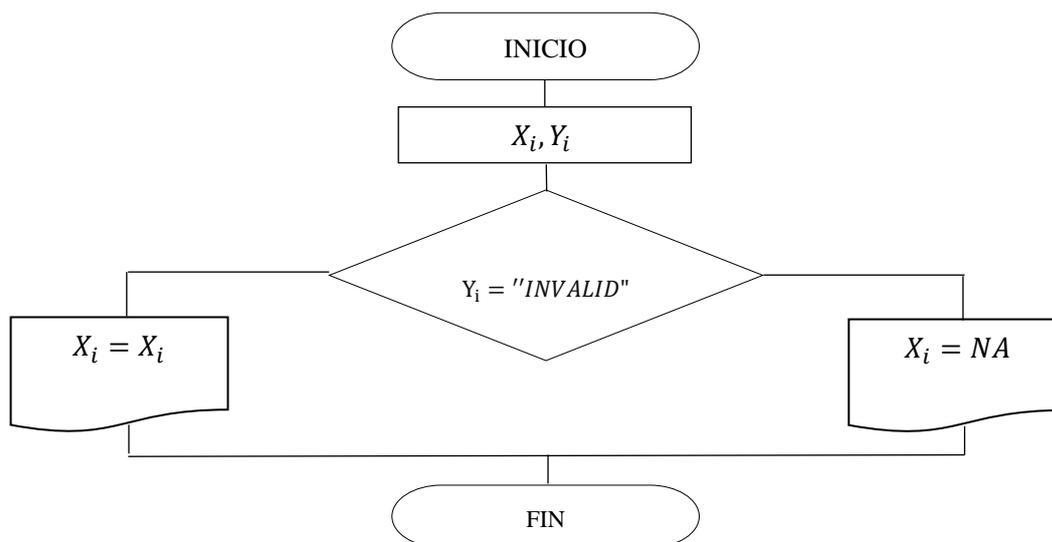


Gráfico 2-3: Validación de Información prueba de coherencia

Realizado por: Mariño Gabriela, 2021.

Se validaron los datos (Gráfico 1-3) utilizando valores umbrales, establecidos por la Organización Mundial Meteorológica, permitiendo así eliminar los valores erróneos de aquellos datos de temperatura que no estén dentro de los límites establecidos.

Existen valores que se registran en las bases de datos, los cuales están identificados con la variable Estado, con dos categorías: VALID e INVALID, que por error del instrumento de medición o el método de transmisión de los datos hacen que sean tomados de manera errónea. Para dar solución a este problema los valores que fueron registrados con el nombre INVALID fueron reemplazados por un NA, es decir fueron tomados en cuenta como dato faltante para luego ser imputado.

Tabla 2-3: Reemplazo de valores INVALID por NA

F. LOCAL	H. LOCAL	ESTADO	X1	F. LOCAL	H. LOCAL	ESTADO	X1
8/1/2014	16:00:07	VALID	-4,415	8/1/2014	16:00:07	VALID	-4,415
8/1/2014	17:00:07	INVALID	0	8/1/2014	17:00:07	INVALID	NA
8/1/2014	18:00:07	VALID	35,841	8/1/2014	18:00:07	VALID	35,841
8/1/2014	19:00:07	INVALID	0	8/1/2014	19:00:07	INVALID	NA
8/1/2014	20:00:08	INVALID	0	8/1/2014	20:00:08	INVALID	NA
8/1/2014	21:00:08	VALID	33,134	8/1/2014	21:00:08	VALID	33,134
8/1/2014	22:00:08	VALID	30,606	8/1/2014	22:00:08	VALID	30,606
8/1/2014	23:00:08	VALID	-53,4	8/1/2014	23:00:08	VALID	-53,4
9/1/2014	0:00:08	INVALID	0	9/1/2014	0:00:08	INVALID	NA
9/1/2014	1:00:08	INVALID	0	9/1/2014	1:00:08	INVALID	NA
9/1/2014	2:00:07	VALID	-57,43	9/1/2014	2:00:07	VALID	-57,43
9/1/2014	3:00:07	INVALID	0	9/1/2014	3:00:07	INVALID	NA
9/1/2014	4:00:07	VALID	-60,788	9/1/2014	4:00:07	VALID	-60,788
9/1/2014	5:00:07	INVALID	0	9/1/2014	5:00:07	INVALID	NA
9/1/2014	21:00:08	INVALID	0	9/1/2014	21:00:08	INVALID	NA
31/12/2014	19:00:00			31/12/2014	19:00:00		NA
31/12/2014	20:00:00			31/12/2014	20:00:00		NA
31/12/2014	21:00:00			31/12/2014	21:00:00		NA
31/12/2014	22:00:00			31/12/2014	22:00:00		NA

Fuente: GEAA, 2019.

Realizado por: Mariño Gabriela, 2021.

Para la validación de los datos, también es necesario la completitud de los mismos, es decir imputación de datos faltantes, y para ello se identificó la cantidad de datos faltantes por estación de la variable Temperatura Ambiental por año y por meses de las 11 estaciones meteorológicas. La variable Temperatura fue particionada por meses y clasificadas por sus respectivas horas, para utilizar el método de imputación de Random Forest, haciendo uso de la función rfImpute, permitiendo realizar imputaciones por una característica común, en este caso por horas. Se seleccionó 300 árboles aleatorios y una iteración de 10, que equivale a las veces que se va a repetir para encontrar el valor óptimo. Este valor es el recomendado según estudios anteriormente realizados.

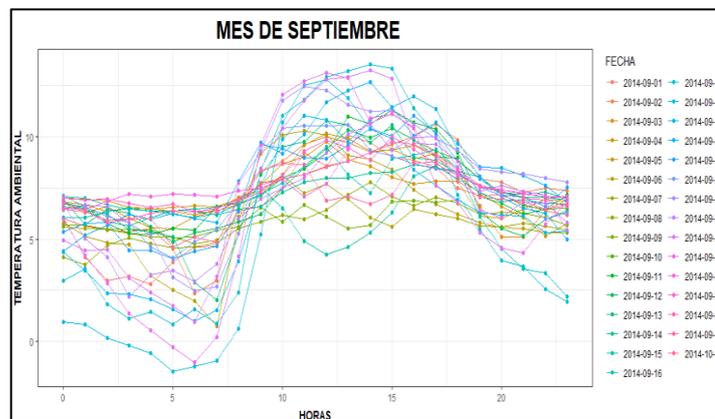


Gráfico 3-3: Faltantes de la Estación Atillo

Realizado por: Mariño Gabriela, 2021.

Se puede observar (Gráfico 3-3) los datos de la temperatura ambiental de la Estación Atillo, del mes de septiembre. La cantidad de faltantes presente es del 6.7 % para el año 2014, y para este mes es de 2.92%, la imputación se realizó hasta el 30 % de datos faltantes por año. Una vez particionada la base de datos de la variable de estudio por meses, se procede a realizar la imputación por medio de la función rfImpute con los parámetros mencionados anteriormente.

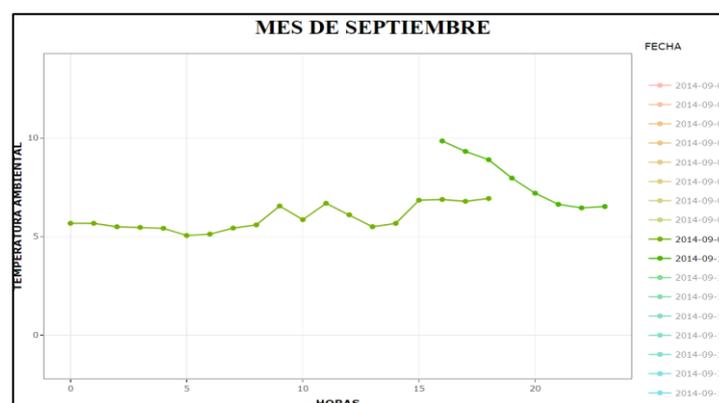


Gráfico 4-3: Faltantes de Estación Atillo según los días

Realizado por: Mariño Gabriela, 2021.

Se observa (Grafico 4-3) que en el día 9 de septiembre del 2014 existen datos faltantes desde las 00:00 hasta las 15:00 horas, registrándose valores de Temperatura únicamente desde las 16:00 a las 23:00 horas. De igual forma se observa que el 10 de septiembre del 2014 solo se presentan valores a partir de las 00:00 hasta las 18:00 horas, presentando faltantes desde las 19:00 hasta las 23:00 horas.

Se observa (Grafico 5-3) que al momento de realizar la imputacion, los datos siguen el mismo patrón que la gráfica con valores faltantes, además se visualiza que ya no presentan valores faltantes en los días 9 y 10 de septiembre. Este método de imputacion se utilizó para las 11 estaciones meteorológicas, para cada uno de los años y se almacenó en un R. Data, generando una sola base común llamada BASEESTACIONES, con el fin de facilitar la manipulación de los datos.

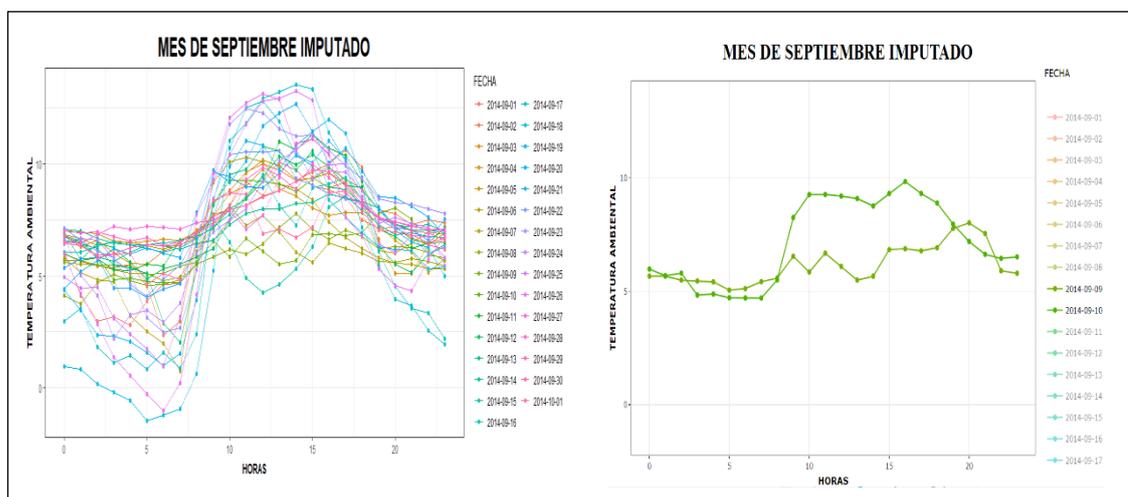


Gráfico 5-3: Datos Imputados de la Estación Atillo

Realizado por: Mariño Gabriela, 2021.

3.3. Base funcional óptima

Para seleccionar la base óptima, se utilizó la función `optim.basis`, de la librería `fda` de R, dando como resultado 23 bases. Se probó con bases impares de 5,7,11,13,15,17,19,21,23 de Fourier y B-Spline. Para escoger el número y tipo de base, se calculó el *MSE* con la función `plotfit.fd()`. Se observa (Tabla 3-3) los valores *MSE* utilizando bases de Fourier y B-splines, presentando un menor *MSE* al utilizar 15 bases de B-splines.

Tabla 3-3: MSE de suavizados con Fourier y B-splines

nbasis	Fourier	B-splines
5	0.66	1.07
7	0.54	0.58
9	0.46	0.45
11	0.40	0.37
13	0.36	0.29
15	0.33	0.25
17	0.36	0.29
19	0.30	0.39
21	0.32	0.40
23	0.35	0.48

Fuente: GEAA,2019

Realizado por: Mariño Gabriela,2021.

Se observa (Gráfico 6-3) que existe dispersión en los datos al utilizar una base de 23 B-splines, mientras que al utilizar 15 B-splines conserva los valores de los datos.

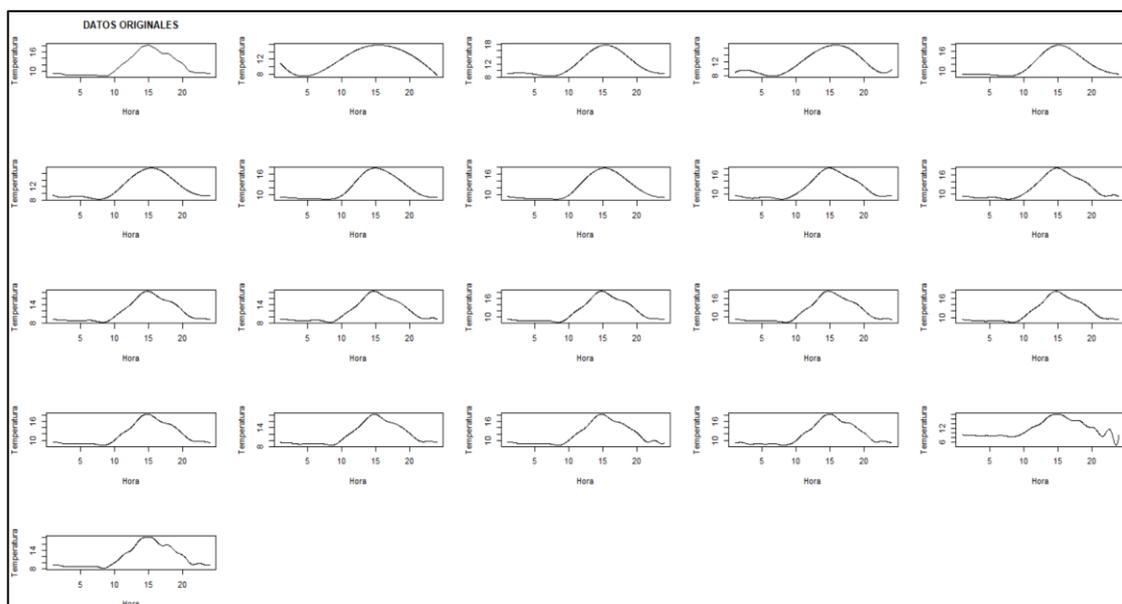


Gráfico 6-3: Suavizado con diferentes bases y número de bases

Realizado por: Mariño Gabriela,2021.

Con la ayuda de la función `plotfit.fd()` del paquete `fda`, se puede visualizar los valores de la temperatura ambiental por día, presentado el MSE (Error Cuadrático medio) de 12 días al utilizar una base de 15 B-splines.

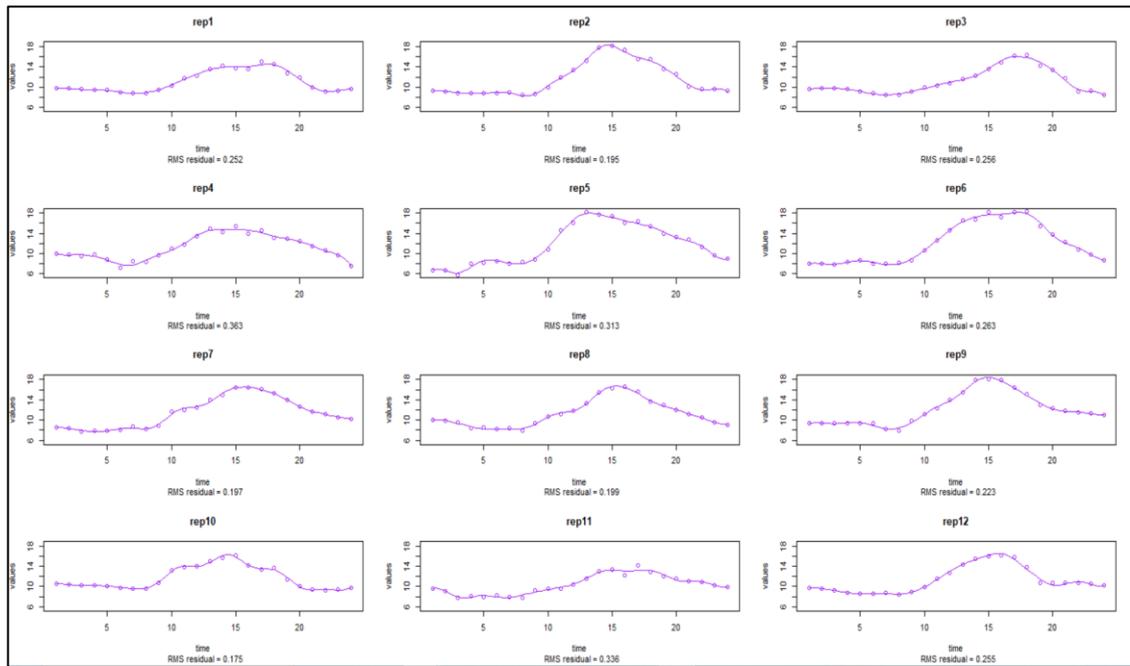


Gráfico 7-3: Errores con 15 B-splines

Realizado por: Mariño Gabriela,2021.

La base óptima utilizada en el estudio es una base de 15 B-Splines

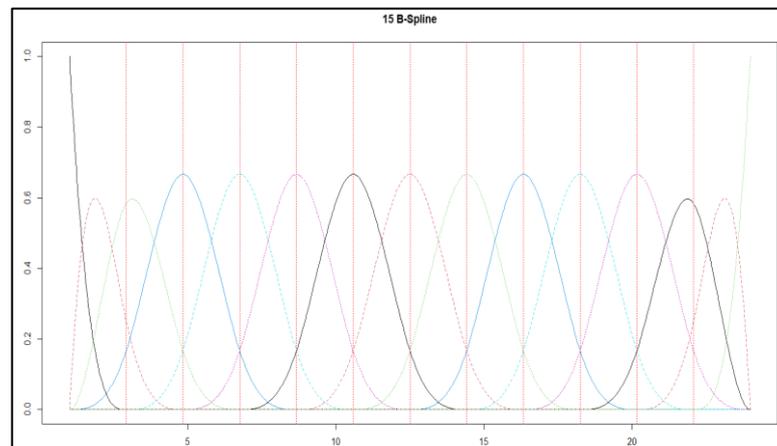


Gráfico 8-3: Base Optima 15 B-splines

Realizado por: Mariño Gabriela,2021.

3.4. Detección de datos Atípicos

Se ha utilizado la función `outliers.depth.trim` que está en el paquete `fda.usc` de R, para detectar datos atípicos mediante un Bootstrap basado en recorte. El proceso para detectar atípicos consiste en calcular la profundidad de la curva de la muestra, en este caso se utilizó profundidad modal, luego se obtiene b muestras Bootstrap, en este estudio se realizaron 200 réplicas con un valor de

corte de 0.01. El valor atípico se determinará por su baja profundidad, es decir todos los valores con profundidad menor a ese corte son los datos atípicos.

En los Gráficos (9-3) y (10-4) se puede observar la temperatura ambiental por día y está representada por cada una de las curvas, identificando curvas atípicas en cada una de las Estaciones.

En la estación de ALAO presentó un total de 54 curvas atípicas registradas a lo largo de los años del 2014 al 2019, en la estación ATILLO se identificó un total de 51 curvas atípicas a partir del 2014 al 2018, para la estación CUMANDÁ registro un total de 37 curvas atípicas en los años 2014,2015,2018 y 2019, en la estación ESPOCH un total de 39 curvas atípicas en los años 2015,2016,2017 y 2019, en la estación MATUS presentó un total de 34 curvas atípicas 2015,2016,2017 y 2018, en la estación MULTITUD se identificaron un total de 56 curvas atípicas 2014,2015,2016,2018 y 2019.

En la estación QUIMIAG un total de 66 curvas atípicas 2014,2015,2016,2017 y 2019, en SAN JUAN 69 curvas atípicas 2015 al 2019, en TIXÁN se identificó 48 curvas atípicas en los años 2014,2016,2017,2018, y 2019, en la estación TUNSHI 54 curvas atípicas del 2014 al 2017 y finalmente en la estación URBINA ,54 curvas atípicas del 2014, 2015,2016,2017 y 2019.

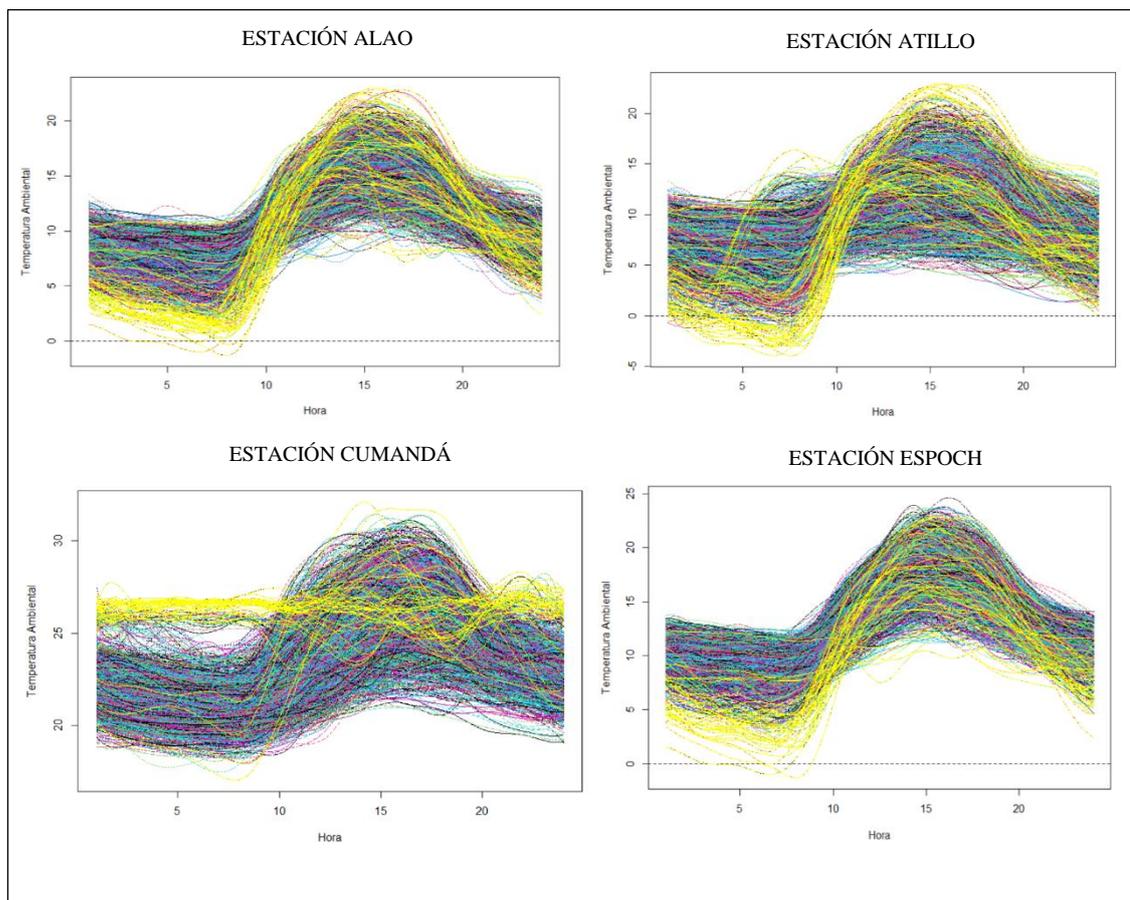


Gráfico 9-3: Curvas atípicas de las Estaciones suavizadas 1

Realizado por: Mariño Gabriela,2021.

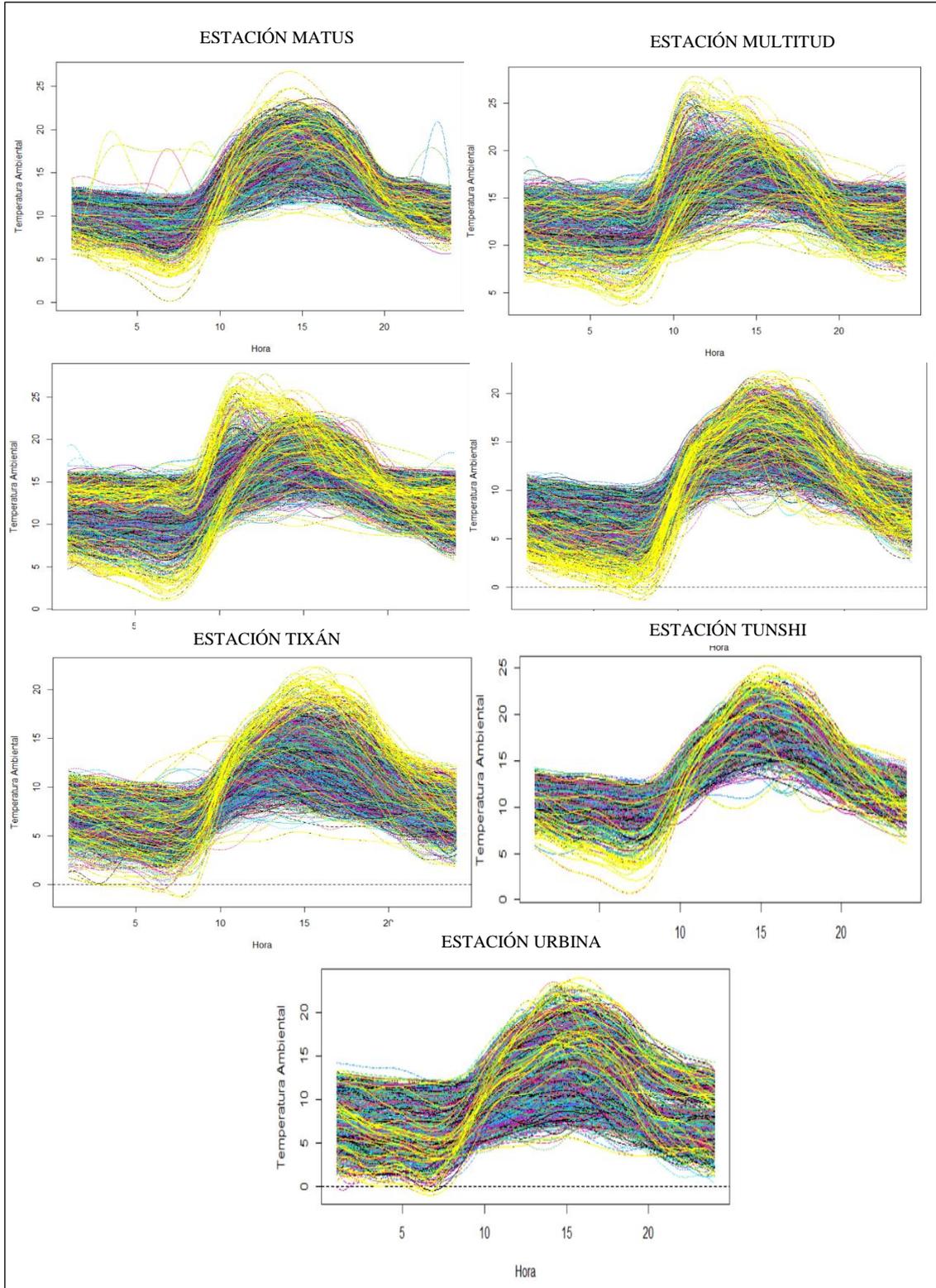


Gráfico 10-3: Curvas atípicas de las Estaciones suavizadas 2

Realizado por: Mariño Gabriela, 2021.

3.5. Análisis exploratorio de datos funcionales

3.5.1. Media y Desviación estándar Funcional

3.5.1.1. Estación Alao

Se observa (Gráfico 11-3) los valores medios de la temperatura ambiental de la Estación Alao, presentan un valor de temperatura promedio en el año 2014 de 15,68°C, en el 2015 de 15,68 °C, en el 2016 de 16,62°C, en el 2017 de 15,78 °C, en el 2018 de 15,82 °C y en el año 2019 de 15,77 °C, registrados a las 14:00 horas del día. Los valores de temperatura se incrementan en todos los años desde las 10:00 hasta las 16:00 horas. Presenta poca variabilidad, con un valor inferior a 2.41 sd registrado en el año 2016, es decir los valores de Temperatura ambiental están próximos a la media.

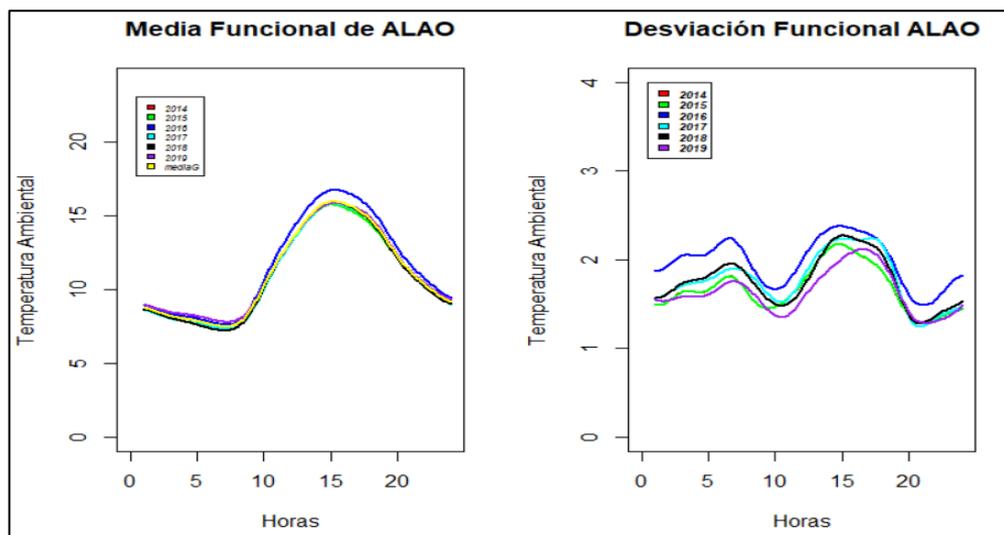


Gráfico 11-3: Media y Desviación Funcional Alao

Realizado por: Mariño Gabriela, 2021.

3.5.1.2. Estación Cumandá

En la gráfica se observa los valores medios de la temperatura ambiental de la Estación Cumandá, para el año 2014 la temperatura promedio llega a 26,32 °C, en el 2015 llega a una temperatura promedio de 26,65 °C, en el 2018 de 25,53 °C y en el año 2019 de 25,59 °C, registrados a las 15:00 horas. La temperatura incrementa a partir de las 10:00 hasta las 16:00, mientras que la desviación estándar funcional presenta poca variabilidad con valores que oscilan de 1.90 a 2,11 sd, en el año 2019 y 2018.

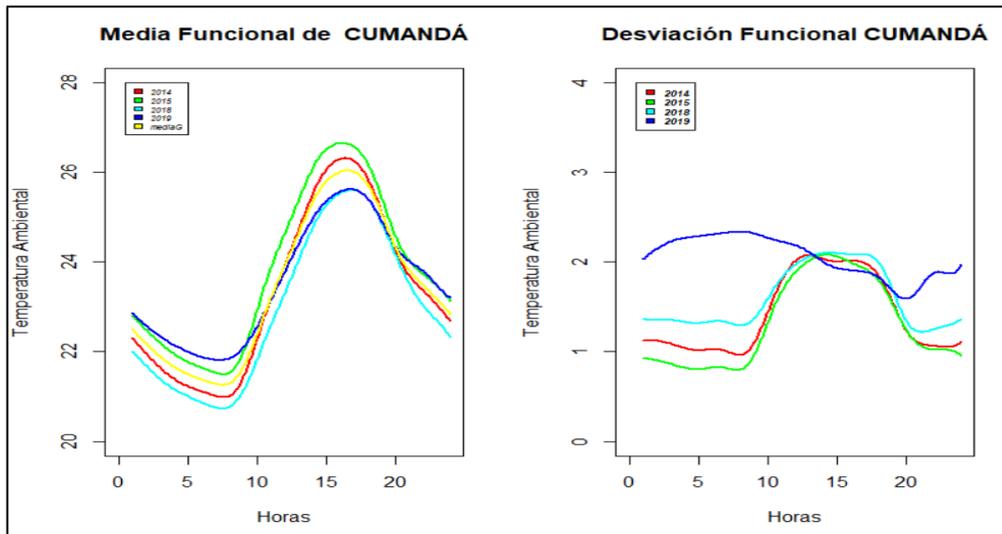


Gráfico 12-3: Media y Desviación Funcional Cumandá

Realizado por: Mariño Gabriela, 2021.

3.5.1.3. Estación Atillo

La temperatura ambiental en la Estación Atillo alcanzó una temperatura promedio en el año 2014 de 9,19 °C, en el 2015 de 10,62 °C, en el 2016 de 16,62 °C, en el 2017 de 11,03 °C, y en el año 2018 de 10,48 °C, registrados a las 14:00 horas. La temperatura incrementa a partir de las 9:00 am hasta las 17:00 pm. Presenta mayor variabilidad en el año 2014 en las horas comprendidas de 5:00 am a 10:00 am, registrando un sd de 3,24, es decir que los valores de Temperatura ambiental están alejados de la media para ese intervalo de tiempo.

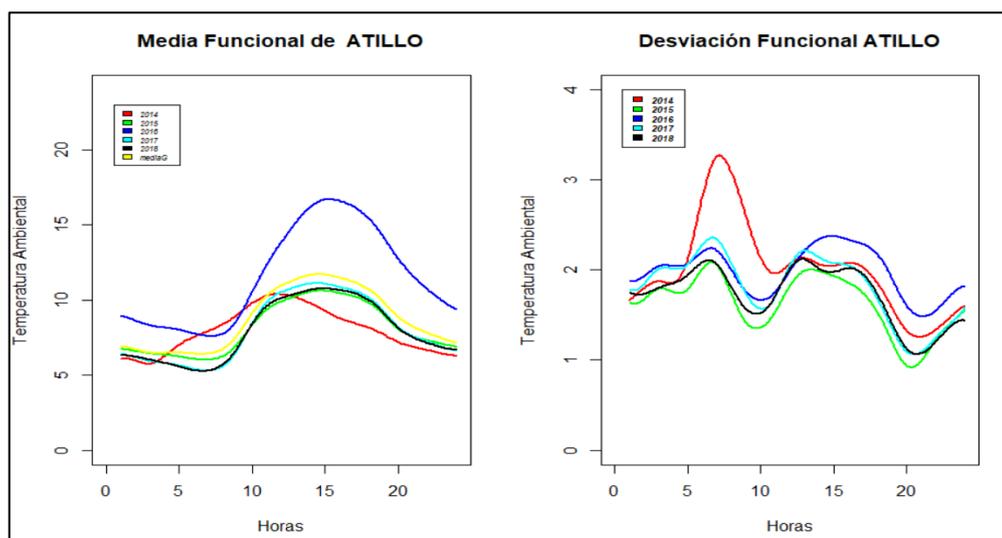


Gráfico 13-3: Media y Desviación Funcional Atillo

Realizado por: Mariño Gabriela, 2021.

3.5.1.4. Estación Multitud

En el año 2014 la temperatura ambiental promedio de la Estación Multitud fué de 17,29 °C, a las 14:00 pm, en el 2015 de 17,03 °C, a las 10:00 am, en el 2016 de 19, 11 °C, a las 10:00 am, en el 2018 de 17,89 °C a las 11 am y en el año 2019 de 17,05 °C, a las 12:00 am. La temperatura ambiental incrementa considerablemente a partir de las 10:00 am hasta las 17:00 pm. Presenta bastante variabilidad, específicamente en el año 2014.

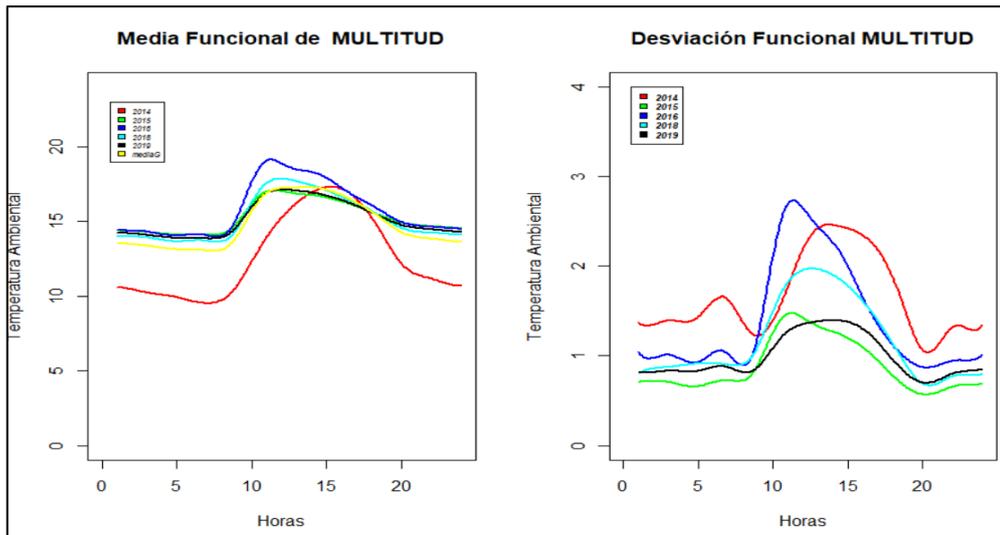


Gráfico 14-3: Media y Desviación Funcional Multitud

Realizado por: Mariño Gabriela, 2021.

3.5.1.5. Estación Quimiag

Los valores promedios de la temperatura ambiental de la Estación Quimiag en el año 2014 fué de 17,10 °C, en el 2015 de 17,30 °C, en el 2016 de 17,92 °C y 19,11 °C a las 10:00 am, en el 2017 de 18,03 °C, y en el 2019 de 17,35 °C, registrados a las 14:00 horas, mientras que la desviación estándar funcional presenta gran variabilidad en el año 2016, alcanzando un valor de sd de 2,84, a las 10:00 am, es decir los valores de temperatura ambiental en el año 2016 se encuentran alejados de la media.

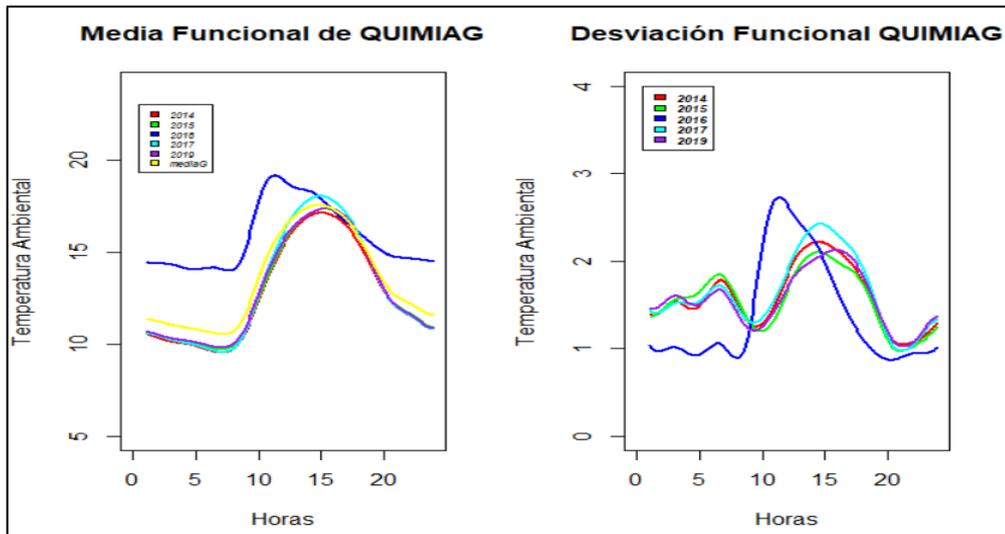


Gráfico 15-3: Media y Desviación Funcional Quimiag

Realizado por: Mariño Gabriela, 2021.

3.5.1.6. Estación San Juan

La temperatura ambiental de la Estación San Juan en el año 2014 alcanzó una temperatura promedio de 14,80 °C, en el 2015 de 15,13 °C, en el 2016 de 15,96 °C, en el 2017 de 15,10 °C, en el 2018 de 15,33 °C y en el 2019 de 15,25 °C, registrados a las 14:00 horas. La desviación estándar funcional, presenta poca variabilidad, es decir es decir que los valores de Temperatura ambiental están próximos a la media.

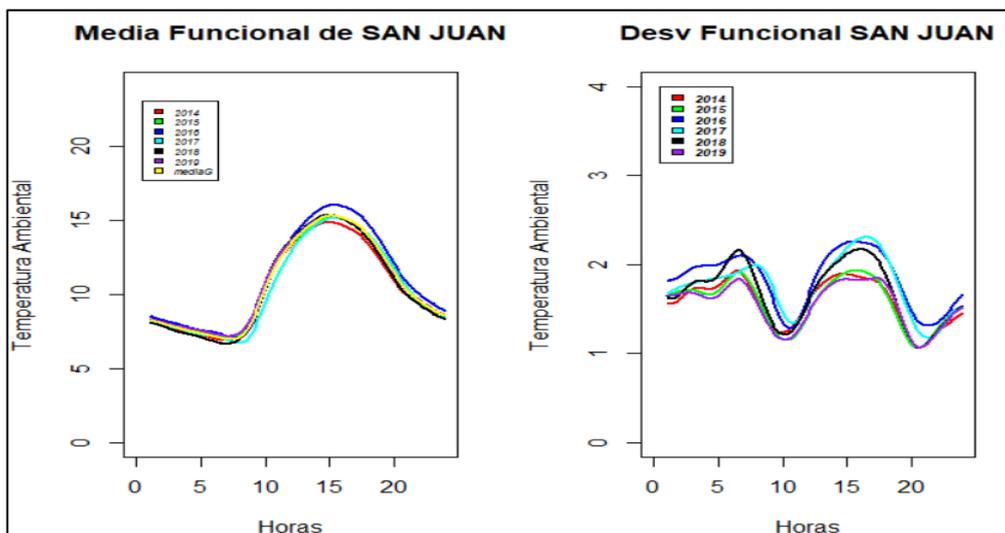


Gráfico 16-3: Media y Desviación Funcional San Juan

Realizado por: Mariño Gabriela, 2021.

3.5.1.7. Estación Tixán

Los valores promedios de la temperatura ambiental de la Estación Tixán, en el año 2014 fué de 11,95 °C, en el 2016 de 15,96 °C, en el 2017 de 11,79, en el 2018 de 11,89 °C y en el 2019 de 12,37, registrados a las 14:00 horas, la media general fue de 13 °C. Se observa que los valores de temperatura incrementan a partir de las 10:00 am hasta las 16:00 pm. La desviación estándar funcional presenta un valor que varía de 1,80 a 2,08 sd, presentes desde las 00:00 a 7:00 am, en el año 2016, es decir que los valores de temperatura ambiental están alejados de la media.

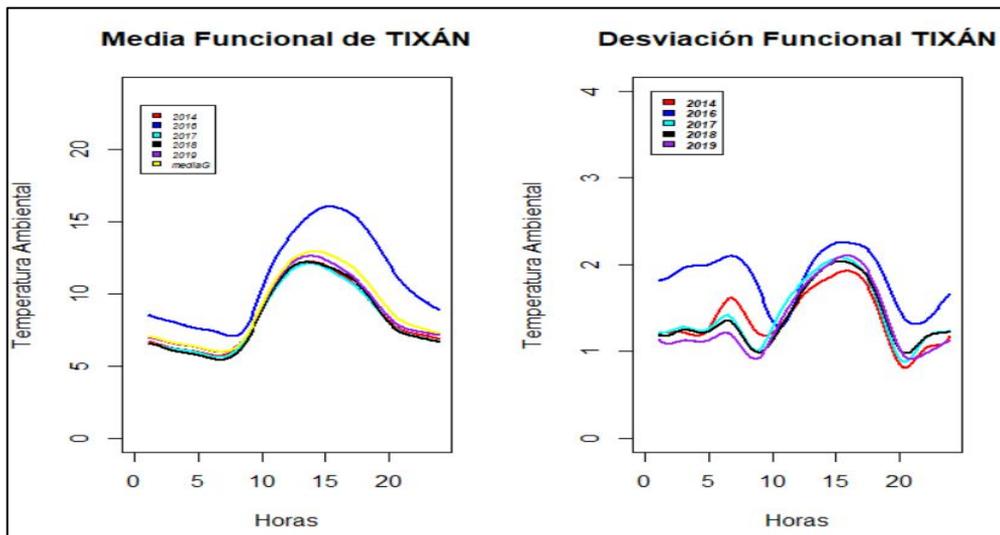


Gráfico 17-3: Media y Desviación Funcional Tixán

Realizado por: Mariño Gabriela, 2021.

3.5.1.8. Estación Tunshi

En la Estación Tunshi la temperatura ambiental en el año 2014 fué de 18,06 °C, en el 2015 de 18,06 °C, en el 2016 de 18,78 °C y en el 2017 de 18,06 °C, registrados a las 14:00 horas. Presentan un valor promedio de temperatura global de 18 °C a las 14:00 horas, los valores de temperatura incrementan en todos los años desde las 10:00 hasta las 16:00. Presentan poca variabilidad, es decir que la temperatura ambiental en esta estación registra valores similares año tras año.

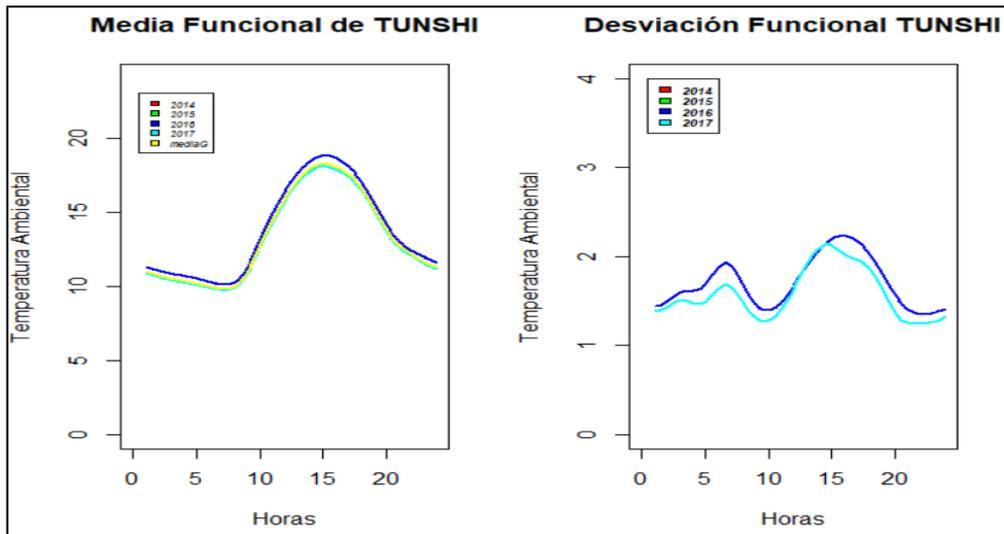


Gráfico 18-3: Media y Desviación Funcional Tunshi

Realizado por: Mariño Gabriela, 2021.

3.5.1.9. Estación Urbina

Para el año 2014 en la estación Urbina, la temperatura ambiental alcanzó un valor promedio de 18,06 °C, en el 2105 de 18,06 °C, en el 2016 de 11,35 °C, en el 2017 de 10,75 °C y en el 2019 de 10,72 °C, registrados a las 14:00 horas. La temperatura ambiental incrementa en todos los años, a partir de las 10:00 am hasta las 16:00 pm. La desviación estándar funcional osciló de 2,10 a 2,22 sd, a las 14:00 horas, es decir es decir que los valores de temperatura ambiental están próximos a la media.

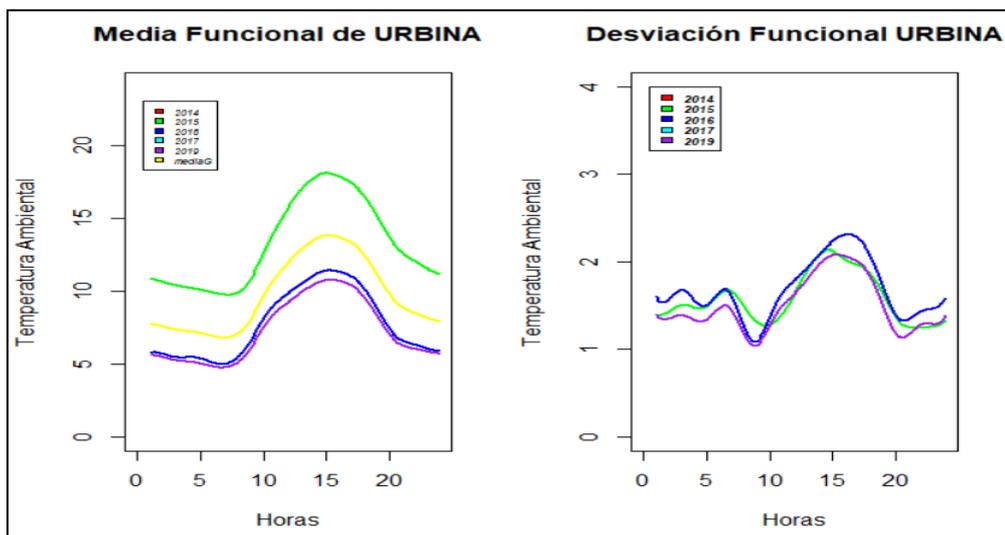


Gráfico 19-3: Media y Desviación Funcional Urbina

Realizado por: Mariño Gabriela, 2021.

3.5.1.10. Estación Epoch

La temperatura ambiental de la Estación Epoch, en el año 2015 fué de 18,53 ° C, en el 2016 de 16,62 ° C, en el 2017 de 18,53 ° C y en el 2019 de 18,55 ° C, registrados a las 14:00 horas. Los valores de temperatura incrementan en todos los años, desde las 10:00 am hasta las 16:00 pm, disminuyendo notablemente hasta llegar a las 23:00 horas. La desviación estándar funcional presenta una variabilidad significativa en el año 2016, con un sd de 2,42, es decir es decir que los valores de Temperatura ambiental están alejados de la media.

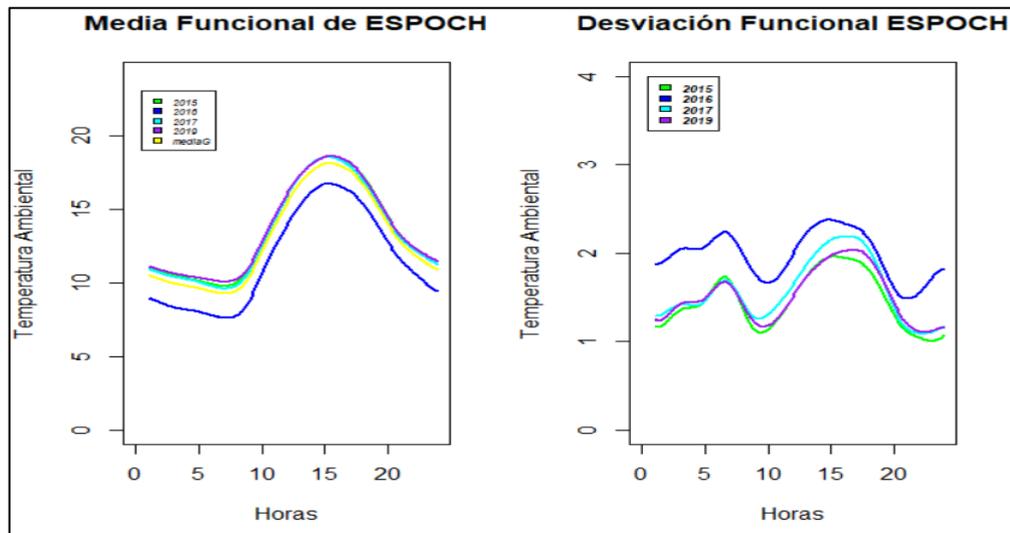


Gráfico 20-3: Media y Desviación Funcional Espoch

Realizado por: Mariño Gabriela, 2021.

3.5.1.11. Estación Matus

La temperatura ambiental de la Estación Matus en el 2015 fué de 17,06 °C, en el 2016 de 17,86 °C, en el 2017 de 17,29 °C y en el 2018 de 17,35 °C, registrados a las 14:00 horas. La temperatura ambiental aumenta a partir de las 10:00 am a 16:00 pm. La desviación estándar funcional presenta poca variabilidad, con un valor registrado de 2,65 en el 2016, es decir es decir que los valores de temperatura ambiental están próximos a la media.

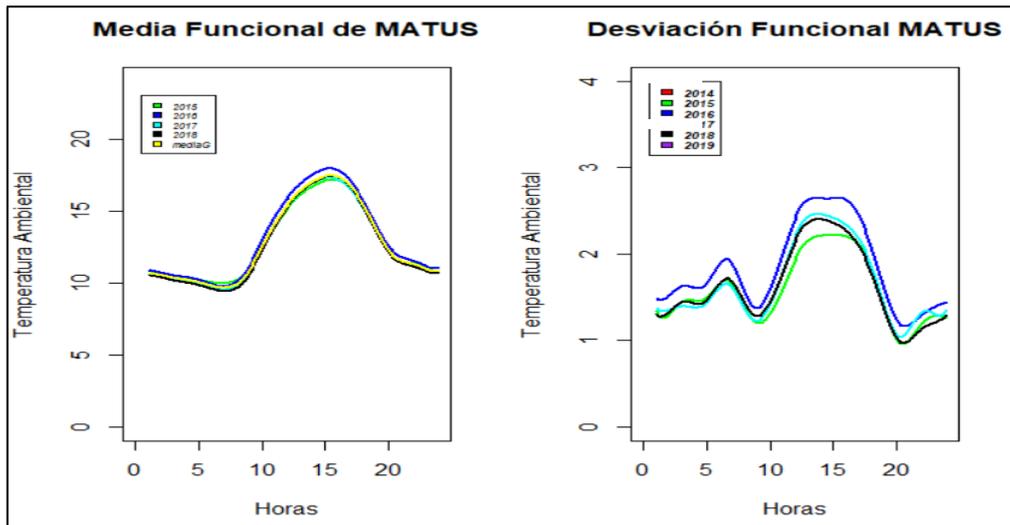


Gráfico 21-3: Media y Desviación Funcional Matus

Realizado por: Mariño Gabriela, 2021

3.5.2. Varianza Funcional

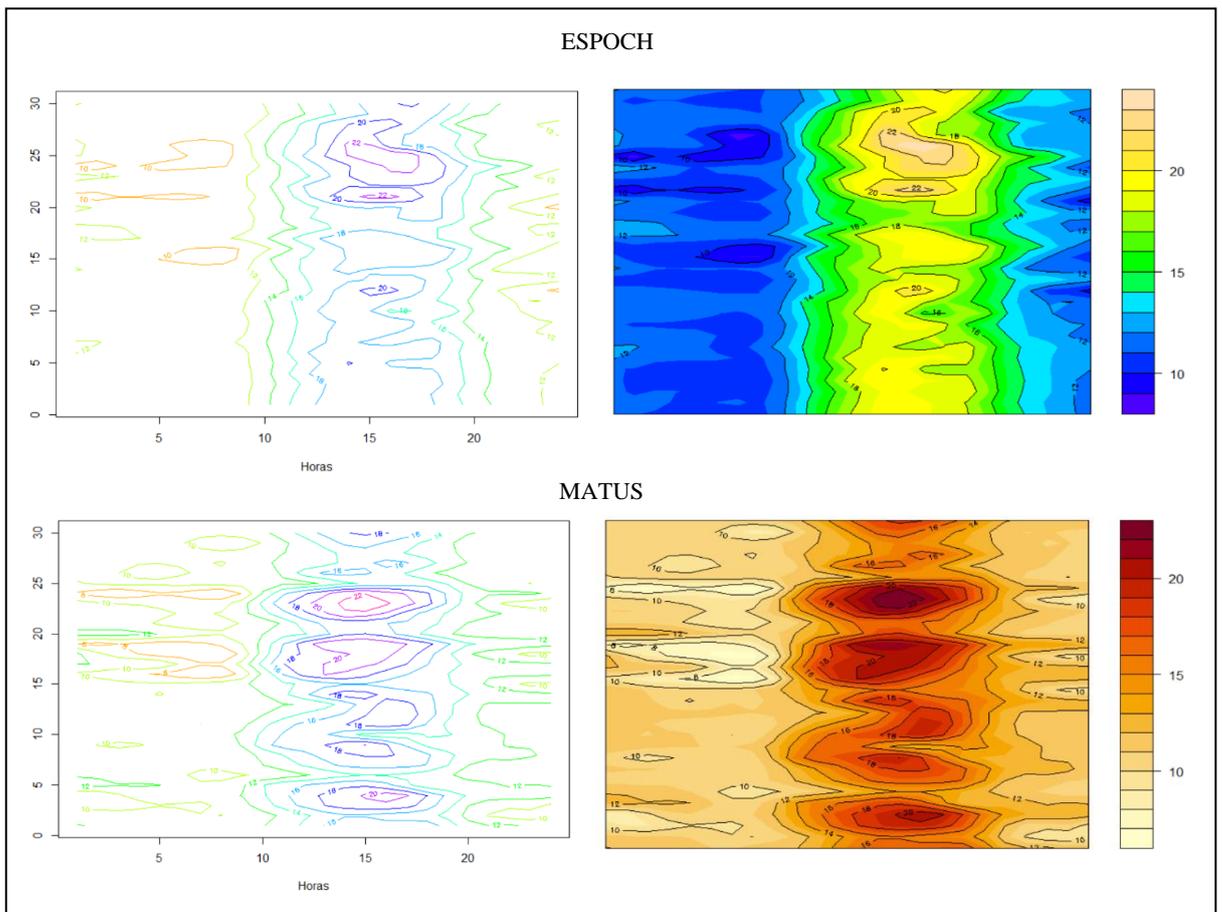


Gráfico 22-3: Varianza Funcional Espoch, Matus

Realizado por: Mariño Gabriela, 2021.

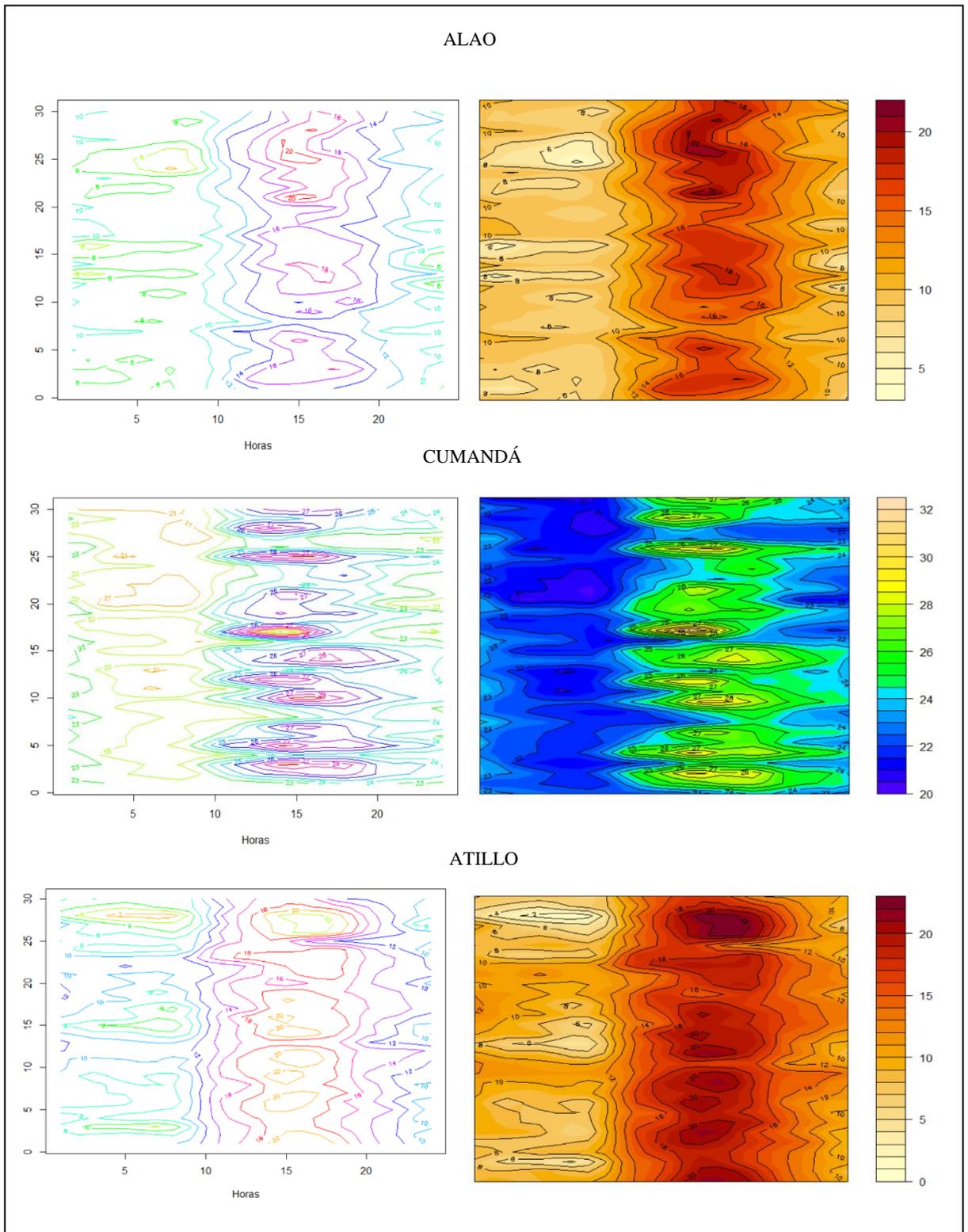


Gráfico 23-3: Varianza Funcional Alao, Cumandá, Atillo

Realizado por: Mariño Gabriela, 2021.

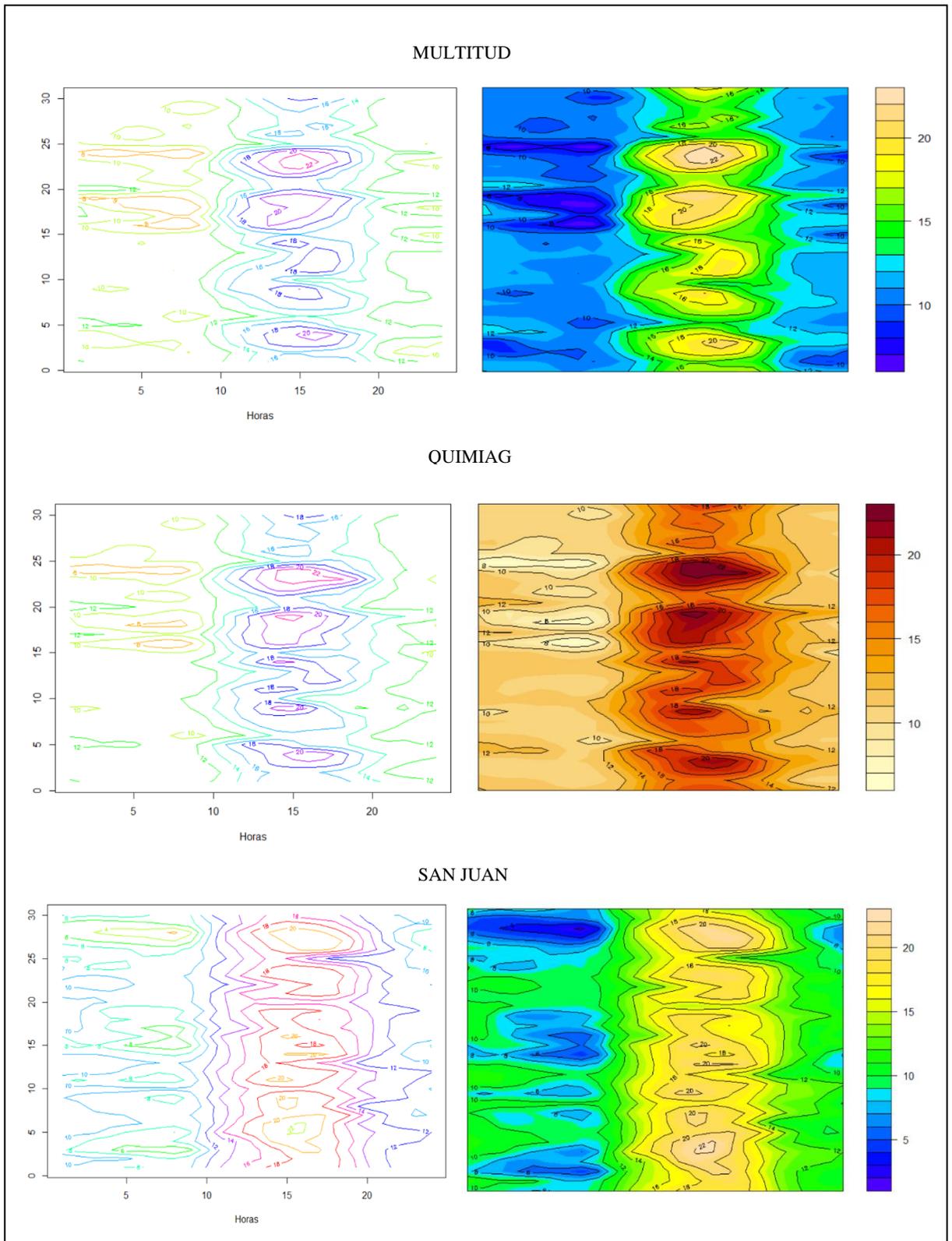


Gráfico 24-3: Varianza Funcional Multitud, Quimiag, San Juan

Realizado por: Mariño Gabriela, 2021.

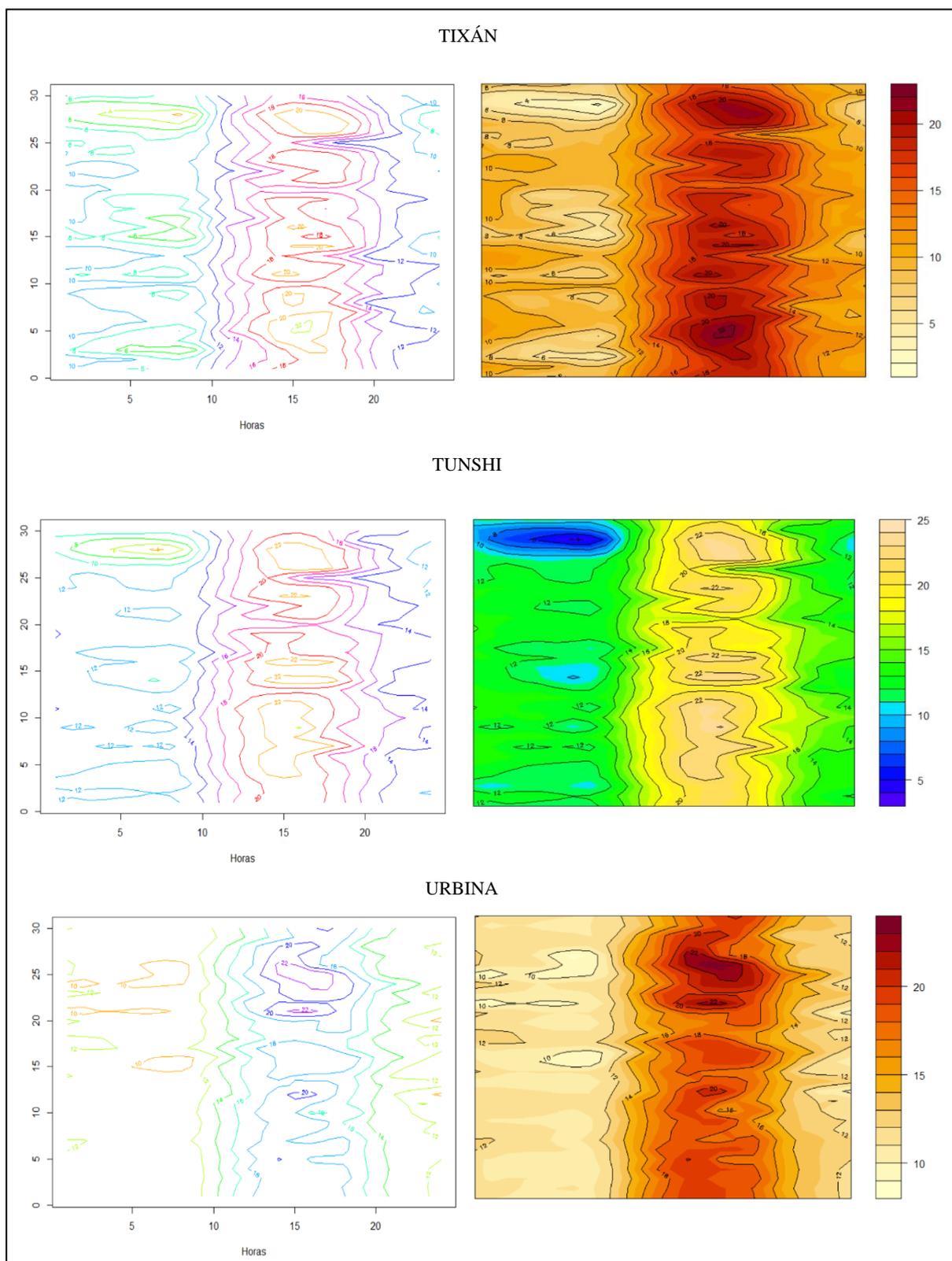


Gráfico 25-3: Varianza Funcional Tixán, Tunshi, Urbina

Realizado por: Mariño Gabriela, 2021.

En los gráficos se observa la variación de la Temperatura Ambiental de cada una de las estaciones, los colores en café y rosado claro representa las temperaturas altas. La temperatura promedio máxima en la Estación Alao fué de 22 °C y su temperatura mínima de 0.9 °C ,en la Estación de Atillo de 18 °C como máxima y -0.6 °C como mínima, en la Estación de Cumandá alcanzó una temperatura máxima de 32 °C y mínima de 17 °C, en la Estación Espoch una temperatura máxima de 23 °C y mínima de 2 °C ,en la Estación de Matus de 23 °C como máxima y 2 °C como temperatura mínima, en esta estación en el año 2016 presentó una temperatura máxima de 26,6 °C ,en la estación de Quimiag alcanzo un máximo de 23 °C y un mínimo de 1,8 °C , en el 2016 registró una temperatura promedio máxima de 27,5 °C , En la estación San Juan su temperatura máxima fue de 21 °C y mínima de 0.3 °C, en la estación Tixán se registró un máximo de 17 °C y un mínimo de 0.4 °C , solo en el 2016 presentó una temperatura máxima en promedio de 22,4 °C , en la Estación de Tunshi alcanzo una temperatura de 25 °C y 2.7 °C como mínima y en la Estación de Urbina su temperatura máxima fué de 20 °C y 1.40 °C de mínima, a partir del 2017 la temperatura ha ido disminuyendo, alcanzando un máximo en el 2019 de 16,06 °C.

3.6. Modelización de datos funcionales (Ajuste del Modelo)

La modelización de datos funcionales se la realiza mediante regresión del componente principal funcional, por una descomposición de los componentes principales funcionales de los datos de Temperatura Ambiental y sus puntuaciones asociadas. La función empleada es el ftsm del paquete ftsa, el cual contiene funciones para el análisis funcional de series de tiempo. Se utiliza los datos de cada estación meteorológica, particionándola en dos partes, 70 % para el modelado de los datos y el 30 % para la validación.

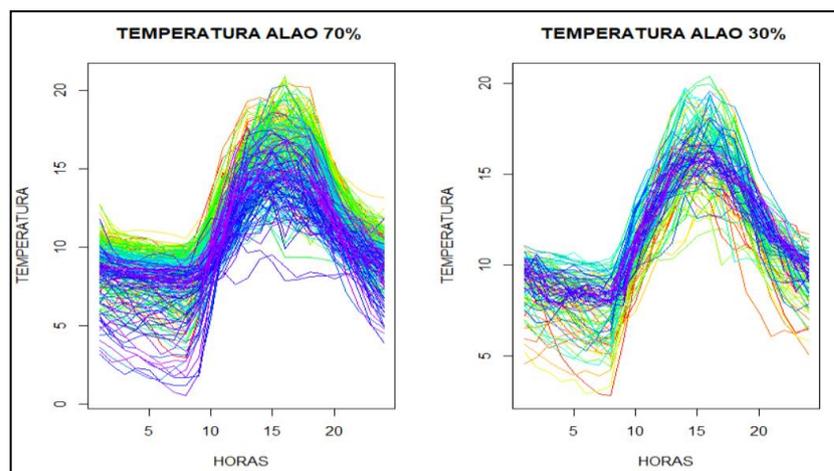


Gráfico 26-3: Datos particionados para modelización y validación

Realizado por: Mariño Gabriela, 2021.

3.6.1. Componentes Principales de Estación Alao

Tabla 4-3: Valores de componentes Principales de Alao

Horas	mean	phi1	phi2	phi3	phi4
0	8,894	0,084	0,277	0,036	-0,013
1	8,654	0,079	0,291	0,023	-0,039
2	8,428	0,079	0,306	0,025	-0,052
3	8,289	0,078	0,311	0,013	-0,064
4	8,162	0,076	0,320	0,001	-0,066
5	8,023	0,071	0,334	-0,019	-0,086
6	7,880	0,060	0,344	-0,026	-0,076
7	7,835	0,051	0,347	-0,013	-0,055
8	8,562	0,084	0,298	0,060	-0,001
9	10,141	0,146	0,134	0,210	0,084
10	11,515	0,177	0,033	0,323	0,096
11	12,993	0,195	-0,056	0,433	0,109
12	14,418	0,251	-0,094	0,434	0,092
13	15,274	0,293	-0,114	0,307	0,017
14	15,646	0,330	-0,122	0,122	-0,132
15	15,620	0,362	-0,116	-0,089	-0,187
16	15,367	0,358	-0,110	-0,236	-0,205
17	14,783	0,341	-0,088	-0,264	-0,174
18	13,685	0,281	-0,060	-0,255	-0,104
19	12,298	0,231	-0,017	-0,201	0,037
20	11,187	0,188	0,010	-0,168	0,241
21	10,378	0,154	0,037	-0,178	0,416
22	9,695	0,134	0,053	-0,176	0,523
23	9,216	0,134	0,075	-0,173	0,540

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

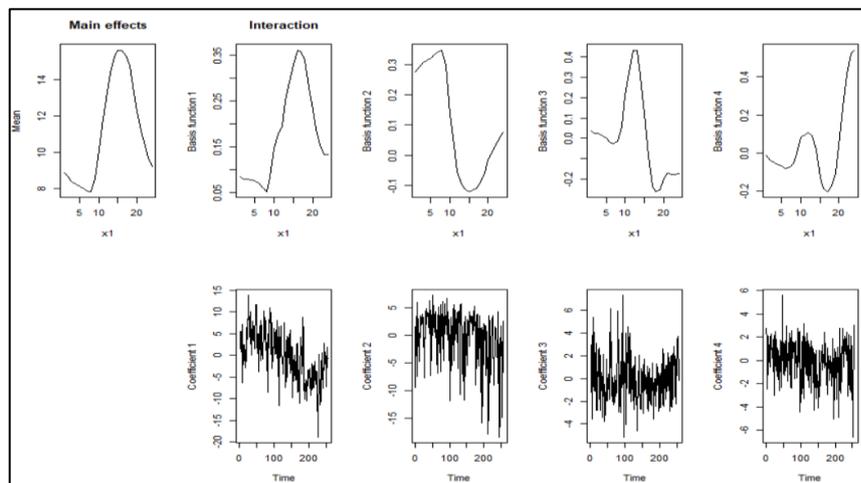


Gráfico 27-3: Componentes Principales Alao

Realizado por: Mariño Gabriela,2021.

Las cuatro componentes principales explican el 89% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.2. Componentes Principales de Estación Atillo

Tabla 5-3: Valores de componentes Principales de Atillo

Horas	mean	phi1	phi2	phi3	phi4
0	6,167	0,040	0,256	0,031	0,318
1	6,082	0,047	0,286	0,056	0,294
2	5,964	0,049	0,321	-0,038	0,045
3	5,718	0,040	0,345	-0,069	-0,076
4	5,467	0,026	0,379	-0,073	0,031
5	5,341	0,010	0,356	-0,165	-0,036
6	5,346	0,010	0,355	-0,178	0,042
7	5,524	0,021	0,346	-0,114	-0,186
8	6,727	0,086	0,223	0,057	-0,076
9	8,127	0,188	0,054	0,065	0,082
10	9,068	0,271	-0,044	-0,038	0,215
11	9,635	0,303	-0,073	-0,063	0,336
12	9,999	0,319	-0,087	-0,091	0,350
13	10,222	0,326	-0,074	-0,130	0,239
14	10,330	0,338	-0,065	-0,141	0,059
15	10,244	0,346	-0,044	-0,103	-0,260
16	10,090	0,331	-0,023	-0,071	-0,403
17	9,702	0,308	-0,010	-0,003	-0,334
18	8,943	0,254	-0,007	0,056	-0,228
19	7,993	0,181	0,024	0,206	-0,106
20	7,271	0,114	0,065	0,320	0,028
21	6,952	0,106	0,072	0,435	0,020
22	6,629	0,082	0,107	0,538	0,039
23	6,541	0,091	0,117	0,460	-0,064

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

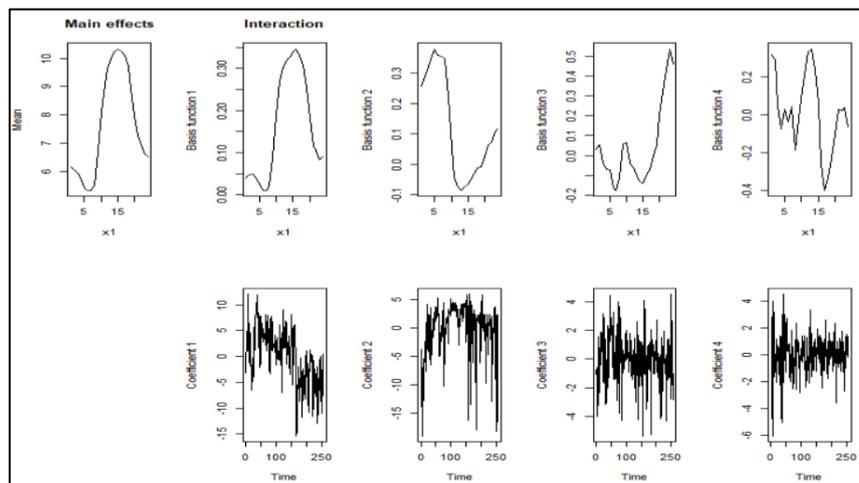


Gráfico 28-3: Componentes Principales Atillo

Realizado por: Mariño Gabriela,2021.

Para la estación Atillo las cuatro componentes principales explican el 87% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.3. Componentes Principales de Estación Cumandá

Tabla 6-3: Valores de componentes Principales de Cumandá

Horas	mean	phi1	phi2	phi3	phi4
0	23,467	0,218	-0,120	-0,001	0,018
1	23,224	0,235	-0,137	0,046	-0,028
2	23,009	0,245	-0,148	0,063	-0,066
3	22,802	0,252	-0,153	0,078	-0,138
4	22,624	0,257	-0,158	0,101	-0,179
5	22,512	0,259	-0,162	0,108	-0,191
6	22,447	0,262	-0,165	0,113	-0,206
7	22,423	0,265	-0,164	0,111	-0,201
8	22,724	0,264	-0,127	0,013	-0,109
9	23,258	0,249	-0,033	-0,253	0,116
10	23,774	0,226	0,057	-0,388	0,161
11	24,377	0,206	0,138	-0,440	0,125
12	25,026	0,185	0,218	-0,405	-0,071
13	25,469	0,156	0,293	-0,220	-0,142
14	25,790	0,122	0,337	-0,052	-0,207
15	26,050	0,102	0,365	0,100	-0,196
16	26,059	0,085	0,370	0,203	-0,170
17	25,843	0,074	0,360	0,251	-0,095
18	25,372	0,084	0,298	0,276	0,055
19	24,779	0,135	0,167	0,230	0,228
20	24,540	0,179	0,097	0,190	0,349
21	24,313	0,194	0,046	0,144	0,420
22	24,029	0,203	-0,008	0,118	0,412
23	23,764	0,211	-0,037	0,059	0,299

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

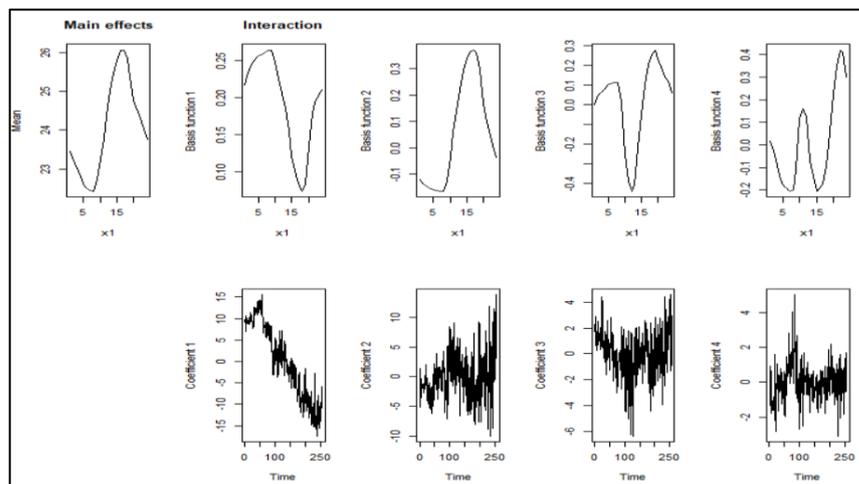


Gráfico 29-3: Componentes Principales Cumandá

Realizado por: Mariño Gabriela,2021.

En la estación Cumandá, se observa que las cuatro componentes principales explican el 96% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.4. Componentes Principales de Estación Epoch

Tabla 7-3: Valores de componentes Principales de Epoch

Horas	mean	phi1	phi2	phi3	phi4
0	11,038	0,127	0,226	0,017	-0,009
1	10,757	0,134	0,250	0,005	-0,065
2	10,541	0,144	0,273	0,016	-0,109
3	10,331	0,155	0,308	0,005	-0,127
4	10,211	0,151	0,320	-0,006	-0,145
5	10,108	0,156	0,332	-0,008	-0,091
6	10,002	0,152	0,339	-0,020	-0,052
7	10,026	0,149	0,337	-0,012	-0,018
8	11,014	0,131	0,186	0,076	0,113
9	12,535	0,137	0,081	0,178	0,197
10	14,229	0,158	0,008	0,300	0,234
11	15,794	0,184	-0,077	0,390	0,177
12	17,146	0,227	-0,164	0,412	0,076
13	18,089	0,251	-0,184	0,368	-0,124
14	18,614	0,284	-0,196	0,196	-0,341
15	18,649	0,308	-0,199	-0,044	-0,400
16	18,215	0,308	-0,187	-0,280	-0,260
17	17,388	0,305	-0,165	-0,366	-0,017
18	16,074	0,281	-0,122	-0,304	0,203
19	14,522	0,231	-0,081	-0,209	0,290
20	13,335	0,196	-0,046	-0,130	0,304
21	12,422	0,176	0,002	-0,074	0,296
22	11,856	0,168	0,032	-0,045	0,261
23	11,388	0,162	0,062	-0,012	0,243

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

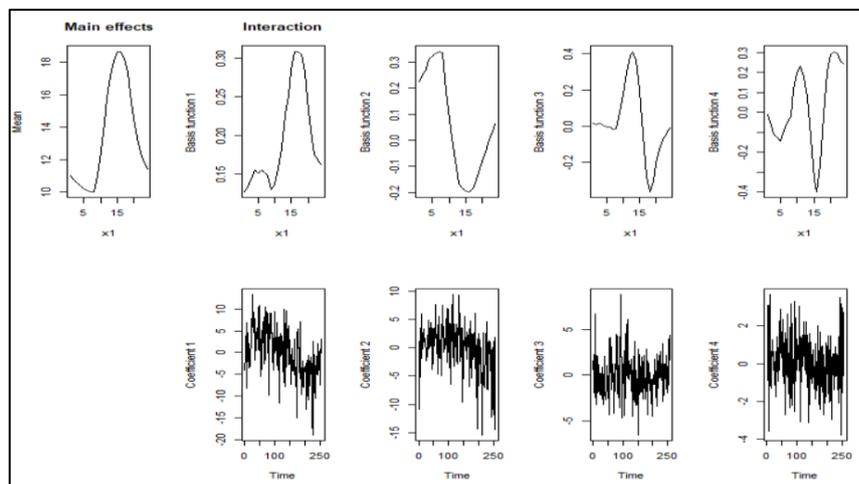


Gráfico 30-3: Componentes Principales Epoch

Realizado por: Mariño Gabriela,2021.

La Temperatura Ambiental en la Estación Epoch es representada por las cuatro componentes principales, explicando el 91% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.5. Componentes Principales de Estación Matus

Tabla 8-3: Valores de componentes Principales de Matus

Horas	mean	phi1	phi2	phi3	phi4
0	10,276	-0,046	0,233	-0,036	0,000
1	10,117	-0,065	0,267	-0,036	-0,013
2	9,930	-0,072	0,309	-0,024	-0,033
3	9,788	-0,078	0,330	-0,016	-0,040
4	9,630	-0,097	0,345	-0,023	-0,064
5	9,479	-0,113	0,356	-0,027	-0,056
6	9,319	-0,127	0,359	-0,036	-0,037
7	9,365	-0,126	0,352	-0,034	-0,041
8	10,321	-0,022	0,243	-0,050	-0,009
9	11,790	0,118	0,150	-0,122	-0,015
10	13,340	0,230	0,087	-0,195	-0,068
11	14,765	0,317	0,052	-0,302	-0,082
12	15,752	0,353	0,029	-0,337	0,035
13	16,498	0,369	0,023	-0,300	0,009
14	16,867	0,376	0,071	-0,136	-0,038
15	16,831	0,348	0,120	0,135	-0,081
16	16,312	0,320	0,091	0,454	-0,158
17	15,274	0,259	0,091	0,532	-0,167
18	13,307	0,173	0,079	0,279	0,012
19	11,973	0,124	0,071	0,140	0,138
20	11,236	0,092	0,079	0,075	0,307
21	10,842	0,078	0,081	0,065	0,457
22	10,597	0,068	0,098	0,069	0,535
23	10,422	0,054	0,094	0,065	0,552

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

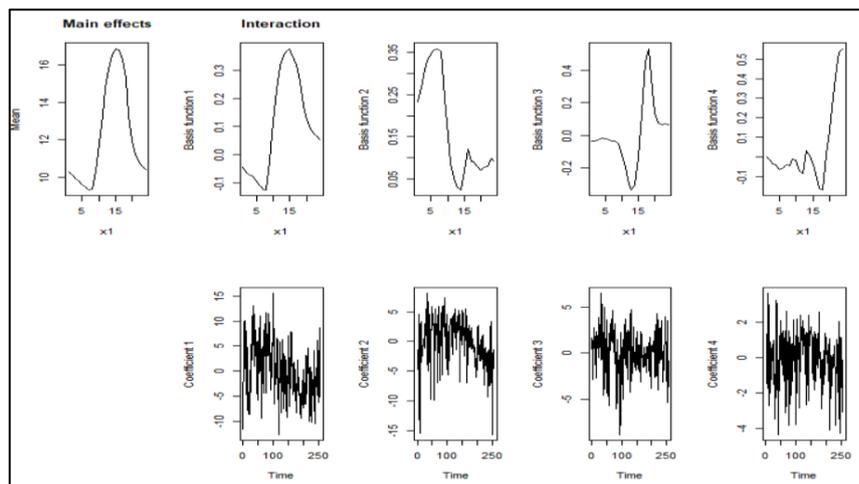


Gráfico 31-3: Componentes Principales Matus

Realizado por: Mariño Gabriela,2021.

En la Estación Matus las cuatro componentes principales explican el 88% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.6. Componentes Principales de Estación Multitud

Tabla 9-3: Valores de componentes Principales de Multitud

Horas	mean	phi1	phi2	phi3	phi4
0	14,360	0,048	0,233	0,062	-0,121
1	14,261	0,048	0,242	0,037	-0,196
2	14,167	0,064	0,263	0,035	-0,226
3	14,082	0,078	0,256	0,046	-0,248
4	13,992	0,087	0,272	0,035	-0,233
5	13,920	0,090	0,274	0,010	-0,210
6	13,923	0,097	0,273	0,002	-0,188
7	14,029	0,109	0,254	-0,023	-0,133
8	14,685	0,140	0,217	-0,117	-0,095
9	15,811	0,218	0,125	-0,244	0,051
10	16,861	0,337	0,047	-0,467	0,241
11	16,996	0,372	-0,037	-0,395	0,142
12	16,940	0,360	-0,138	-0,149	-0,019
13	16,838	0,344	-0,215	0,053	-0,168
14	16,767	0,344	-0,235	0,167	-0,192
15	16,503	0,321	-0,194	0,285	-0,131
16	16,120	0,266	-0,101	0,369	0,055
17	15,750	0,204	-0,008	0,345	0,131
18	15,321	0,144	0,080	0,244	0,179
19	14,877	0,093	0,172	0,174	0,227
20	14,673	0,079	0,207	0,147	0,257
21	14,561	0,074	0,225	0,136	0,320
22	14,480	0,053	0,234	0,101	0,343
23	14,416	0,052	0,212	0,111	0,347

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

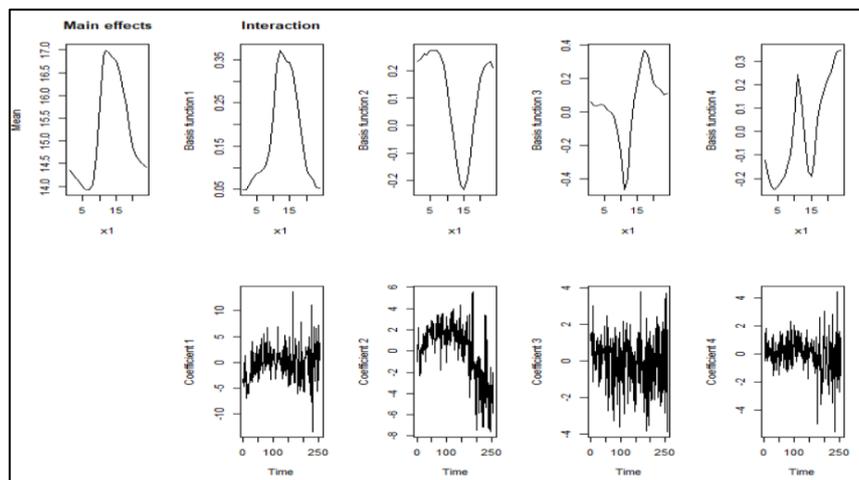


Gráfico 32-3: Componentes Principales Multitud

Realizado por: Mariño Gabriela,2021.

Para la Estación Multitud las cuatro componentes principales explican el 86% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.7. Componentes Principales de Estación Quimiag

Tabla 10-3: Valores de componentes Principales de Quimiag

Horas	mean	phi1	phi2	phi3	phi4
0	10,599	0,121	0,274	-0,037	-0,017
1	10,460	0,130	0,296	-0,041	-0,036
2	10,311	0,135	0,314	-0,028	-0,052
3	10,201	0,132	0,316	-0,003	-0,085
4	10,076	0,125	0,322	-0,008	-0,087
5	9,968	0,115	0,319	0,001	-0,107
6	9,898	0,114	0,322	0,015	-0,114
7	9,933	0,116	0,318	0,026	-0,076
8	10,924	0,129	0,187	-0,052	0,143
9	12,481	0,154	0,061	-0,165	0,382
10	14,099	0,187	-0,031	-0,248	0,452
11	15,423	0,243	-0,109	-0,368	0,274
12	16,339	0,277	-0,149	-0,380	-0,028
13	16,825	0,298	-0,154	-0,336	-0,253
14	17,163	0,320	-0,163	-0,174	-0,383
15	17,141	0,331	-0,182	0,115	-0,328
16	16,701	0,315	-0,177	0,367	-0,132
17	15,896	0,290	-0,128	0,402	0,143
18	14,385	0,238	-0,065	0,290	0,216
19	12,877	0,185	-0,035	0,185	0,159
20	12,063	0,158	-0,020	0,142	0,125
21	11,511	0,144	0,038	0,118	0,111
22	11,085	0,151	0,075	0,108	0,169
23	10,757	0,152	0,106	0,089	0,143

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

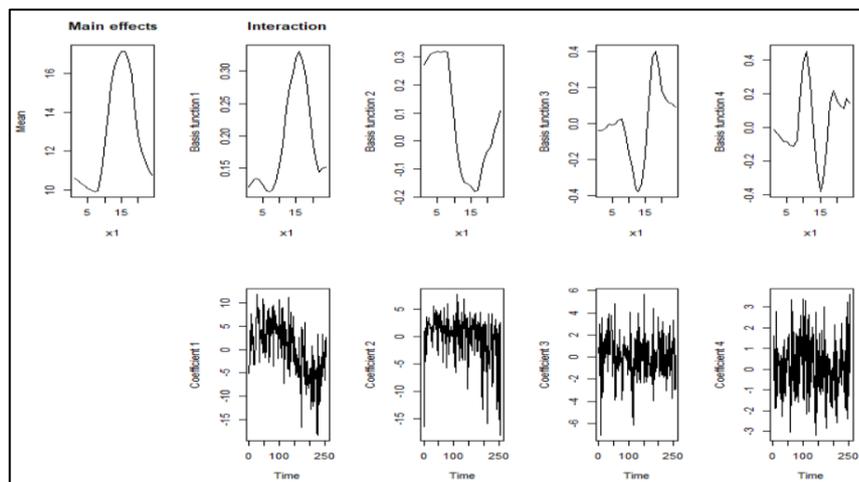


Gráfico 33-3: Componentes Principales Quimiag

Realizado por: Mariño Gabriela,2021.

En la Estación Quimiag las cuatro componentes principales explican el 90% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.8. Componentes Principales de Estación San Juan

Tabla 11-3: Valores de componentes Principales San Juan

Horas	mean	phi1	phi2	phi3	phi4
0	8,226	0,277	-0,077	0,068	-0,010
1	8,035	0,311	-0,073	0,063	-0,072
2	7,857	0,316	-0,069	0,049	-0,095
3	7,631	0,327	-0,078	0,035	-0,090
4	7,468	0,331	-0,079	0,032	-0,100
5	7,342	0,344	-0,078	0,033	-0,075
6	7,231	0,341	-0,078	0,058	-0,046
7	7,252	0,324	-0,089	0,062	-0,010
8	8,624	0,215	-0,013	0,145	0,018
9	10,734	0,092	0,100	0,212	0,079
10	12,350	0,031	0,174	0,311	0,078
11	13,537	0,002	0,239	0,350	0,090
12	14,405	0,004	0,310	0,355	0,098
13	15,024	0,005	0,336	0,326	0,020
14	15,264	0,018	0,355	0,194	-0,046
15	15,131	0,041	0,359	-0,108	-0,190
16	14,727	0,082	0,352	-0,299	-0,224
17	13,994	0,096	0,330	-0,316	-0,230
18	12,823	0,088	0,275	-0,239	-0,109
19	11,352	0,098	0,190	-0,173	0,039
20	10,277	0,107	0,139	-0,169	0,214
21	9,514	0,131	0,107	-0,213	0,395
22	8,940	0,154	0,091	-0,187	0,522
23	8,534	0,167	0,083	-0,161	0,547

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

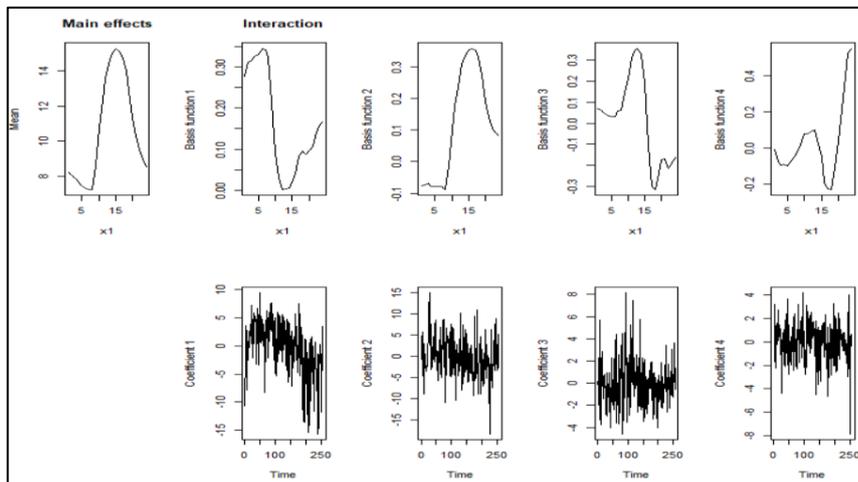


Gráfico 34-3: Componentes Principales San Juan

Realizado por: Mariño Gabriela,2021.

En la estación San Juan las cuatro componentes principales explican el 85% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.9. Componentes Principales de Estación Tixán

Tabla 12-3: Valores de componentes Principales de Tixán

Horas	mean	phi1	phi2	phi3	phi4
0	6,984	0,040	0,291	-0,011	-0,113
1	6,835	0,035	0,309	-0,022	-0,105
2	6,634	0,027	0,320	-0,005	-0,150
3	6,431	0,029	0,341	-0,002	-0,145
4	6,316	0,029	0,343	-0,023	-0,123
5	6,177	0,028	0,350	-0,029	-0,127
6	6,063	0,033	0,326	-0,012	-0,127
7	6,194	0,043	0,280	0,015	-0,065
8	7,482	0,077	0,186	0,185	-0,009
9	9,030	0,153	0,094	0,284	0,047
10	10,649	0,232	0,007	0,366	0,049
11	11,919	0,281	-0,064	0,430	0,027
12	12,572	0,314	-0,101	0,345	-0,083
13	12,787	0,357	-0,126	0,168	-0,201
14	12,625	0,373	-0,108	-0,081	-0,248
15	12,167	0,378	-0,073	-0,340	-0,168
16	11,582	0,340	-0,045	-0,408	-0,016
17	10,831	0,296	-0,007	-0,322	0,117
18	9,795	0,232	0,033	-0,155	0,207
19	8,506	0,151	0,105	-0,003	0,263
20	7,793	0,116	0,133	0,020	0,326
21	7,494	0,094	0,135	0,003	0,373
22	7,284	0,101	0,133	-0,018	0,437
23	7,162	0,092	0,137	-0,006	0,419

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

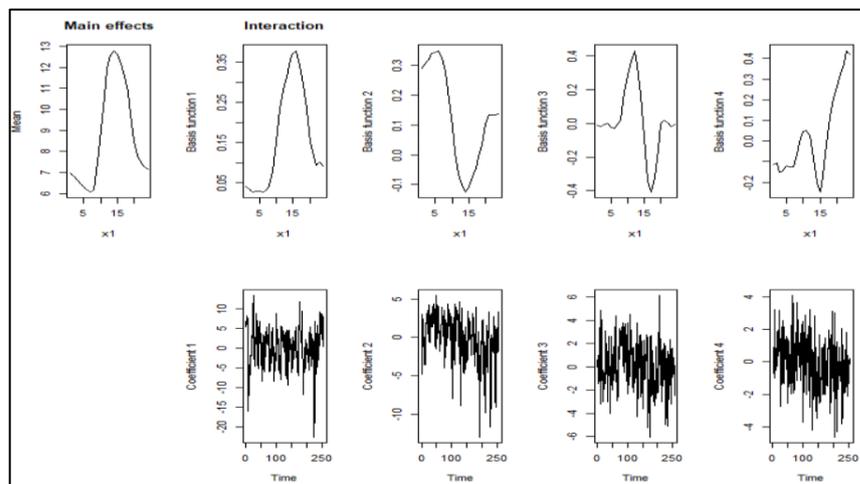


Gráfico 35-3: Componentes Principales Tixán

Realizado por: Mariño Gabriela,2021.

En la Estación Tixán las cuatro componentes principales explican el 85% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.10. Componentes Principales de Estación Tunshi

Tabla 13-3: Valores de componentes Principales de Tunshi

Horas	mean	phi1	phi2	phi3	phi4
0	10,854	0,151	0,245	-0,016	-0,012
1	10,637	0,150	0,273	-0,017	-0,041
2	10,458	0,149	0,280	-0,017	-0,096
3	10,310	0,146	0,285	-0,017	-0,121
4	10,166	0,146	0,292	-0,007	-0,138
5	10,044	0,142	0,298	-0,003	-0,133
6	9,946	0,141	0,316	0,005	-0,130
7	9,934	0,135	0,322	0,033	-0,125
8	10,789	0,135	0,232	-0,043	0,003
9	12,324	0,156	0,091	-0,145	0,131
10	13,938	0,179	-0,004	-0,234	0,193
11	15,391	0,214	-0,076	-0,296	0,167
12	16,555	0,248	-0,130	-0,321	0,111
13	17,388	0,273	-0,171	-0,360	0,026
14	17,735	0,306	-0,217	-0,304	-0,185
15	17,621	0,294	-0,237	-0,041	-0,354
16	17,211	0,282	-0,251	0,289	-0,323
17	16,459	0,272	-0,168	0,387	-0,145
18	15,204	0,239	-0,105	0,356	0,029
19	13,545	0,215	-0,032	0,251	0,216
20	12,549	0,197	-0,008	0,195	0,305
21	11,941	0,188	0,024	0,151	0,357
22	11,456	0,182	0,047	0,111	0,358
23	11,109	0,176	0,070	0,097	0,359

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

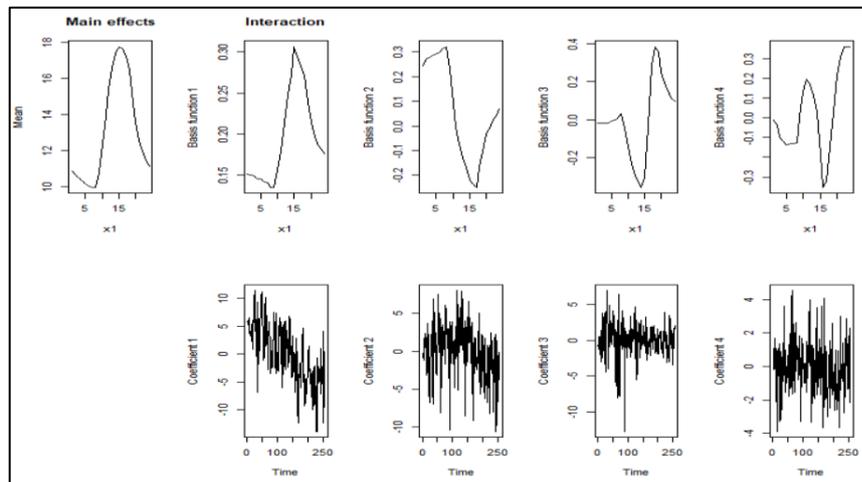


Gráfico 36-3: Componentes Principales Tunshi

Realizado por: Mariño Gabriela,2021

En la Estación Tunshi las cuatro componentes principales explican el 90% de variabilidad de las variables representadas por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.6.11. Componentes Principales de Estación Urbina

Tabla 14-3: Valores de componentes Principales de Urbina

Horas	mean	phi1	phi2	phi3	phi4
0	5,547	0,034	0,298	0,144	-0,104
1	5,446	0,026	0,301	0,125	-0,131
2	5,315	0,019	0,311	0,092	-0,149
3	5,182	0,014	0,307	0,075	-0,138
4	5,044	0,024	0,324	0,065	-0,148
5	4,950	0,027	0,313	0,045	-0,118
6	4,896	0,012	0,307	0,020	-0,121
7	4,901	-0,001	0,322	0,032	-0,111
8	5,998	0,085	0,149	0,103	-0,050
9	7,266	0,174	0,013	0,219	0,013
10	8,195	0,219	-0,021	0,316	0,059
11	8,977	0,254	-0,064	0,358	0,053
12	9,667	0,300	-0,079	0,310	0,042
13	10,057	0,336	-0,104	0,252	0,034
14	10,352	0,350	-0,121	0,145	-0,026
15	10,401	0,370	-0,099	-0,093	-0,184
16	10,076	0,348	-0,057	-0,348	-0,273
17	9,427	0,312	-0,014	-0,412	-0,214
18	8,345	0,241	0,057	-0,298	-0,071
19	6,937	0,185	0,128	-0,163	0,175
20	6,236	0,151	0,172	-0,142	0,336
21	5,966	0,143	0,185	-0,134	0,405
22	5,804	0,124	0,187	-0,111	0,461
23	5,679	0,122	0,196	-0,100	0,420

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

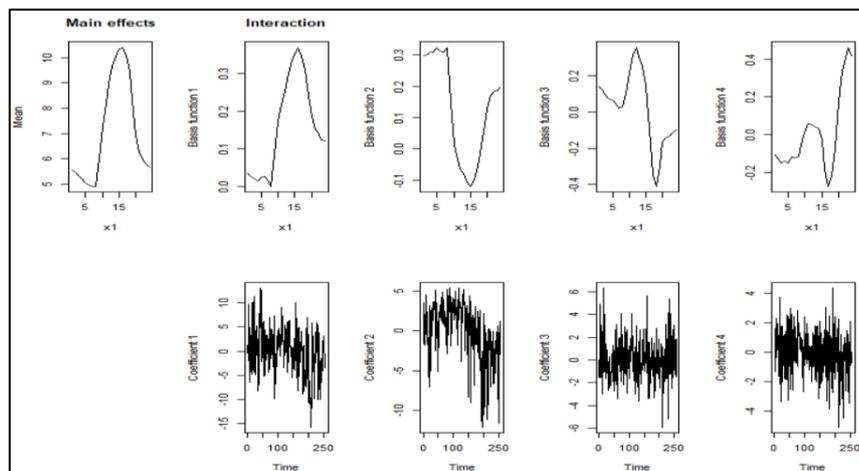


Gráfico 37-3: Componentes Principales Urbina

Realizado por: Mariño Gabriela,2021.

En la Estación Urbina las cuatro componentes principales explican el 86% de variación alrededor de la curva media de la temperatura ambiental por los días, existe un mayor ajuste con los datos históricos al utilizar las cuatro componentes principales.

3.7. Pronósticos de Temperatura Ambiental

Se utiliza las 4 componentes principales funcionales para el ajuste del modelo. Se proyecta los $\hat{\beta}_{t,k}$, usando de series de tiempo univariantes ARIMA y multivariantes VAR para pronosticar los coeficientes, y así obtener las predicciones a horizontes de $h = 1,2,3,4,5,6,7$ días, es decir de una semana.

3.7.1. Pronósticos de Estación Alao

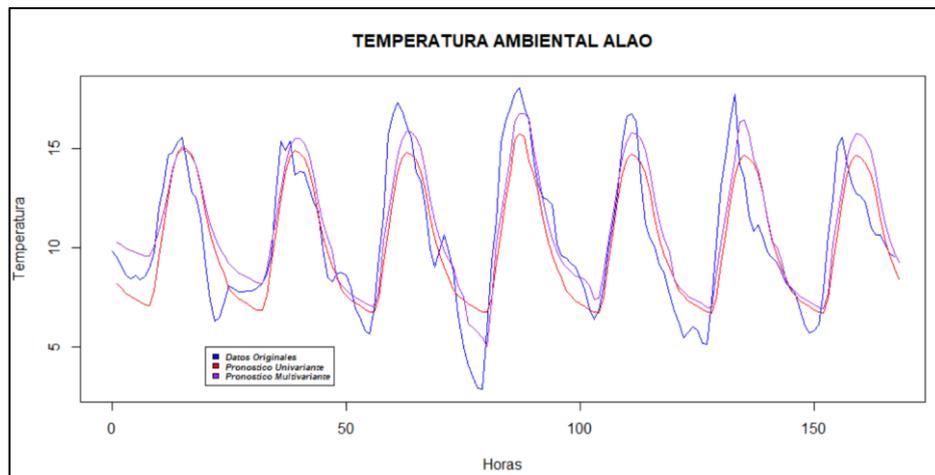


Gráfico 38-3: Pronósticos Alao

Realizado por: Mariño Gabriela,2021.

Para la estación de Alao se pronostican los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1,2,3,4$ por medio de modelos ARIMA, específicamente con un ARIMA (3,1,2), ARIMA (2,1,2), ARIMA (2,0,0) y ARIMA (1,1,2). De color morado se observan los pronósticos con coeficientes calculados por el método VAR y de color azul los datos originales.

3.7.2. Pronósticos de Estación Atillo

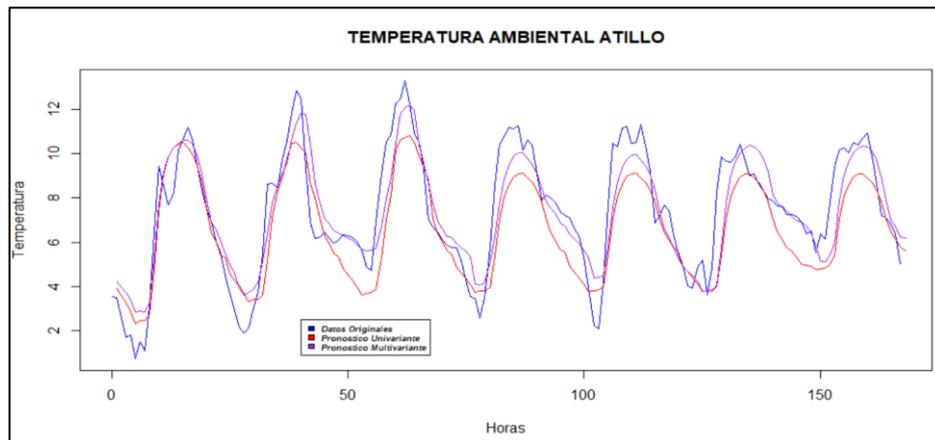


Gráfico 39-3: Pronósticos Atillo

Realizado por: Mariño Gabriela, 2021.

Para la estación de Atillo se pronostican los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1, 2, 3, 4$ por medio de modelos ARIMA, específicamente con un ARIMA (1,1,4), ARIMA (1,1,1), ARIMA (3,0,1) y ARIMA (2,0,1). De color morado se observan los pronósticos con coeficientes calculados por el método VAR.

3.7.3. Pronósticos de Estación Cumandá

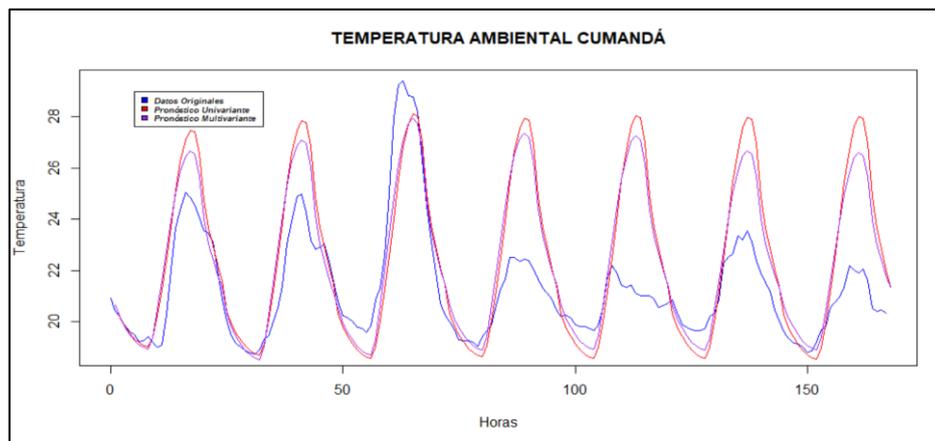


Gráfico 40-3: Pronósticos Cumandá

Realizado por: Mariño Gabriela, 2021.

Los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1, 2, 3, 4$ de la Estación Cumandá fueron pronosticados por medio de modelos ARIMA, específicamente con un ARIMA (0,1,2), ARIMA (1,1,3), ARIMA (0,1,1) y ARIMA (2,0,1). De color morado se observan los pronósticos con coeficientes calculados por el método VAR.

3.7.4. Pronósticos de Estación Epoch

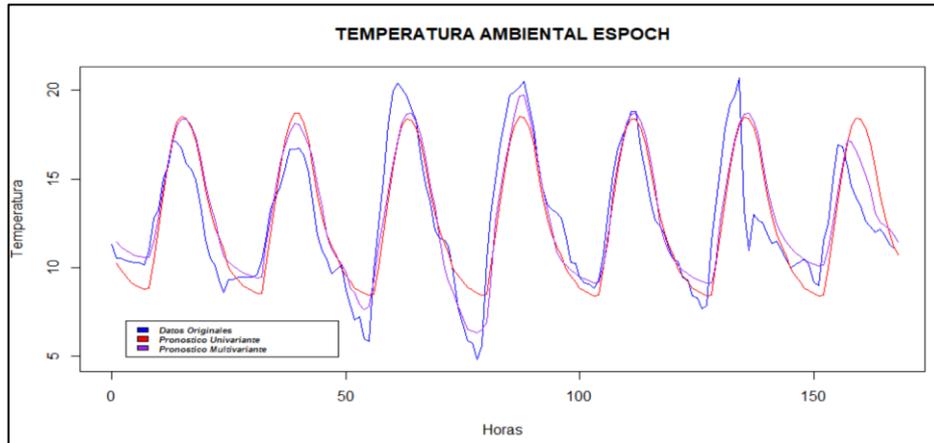


Gráfico 41-3: Pronósticos Epoch

Realizado por: Mariño Gabriela, 2021.

En la estación Epoch se pronostican los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1,2,3,4$ por medio de modelos ARIMA, específicamente con un ARIMA (1,1,4), ARIMA (1,1,1), ARIMA (3,0,1) y ARIMA (2,0,1). De color morado se observan los pronósticos con coeficientes calculados por el método VAR.

3.7.5. Pronósticos de Estación Matus

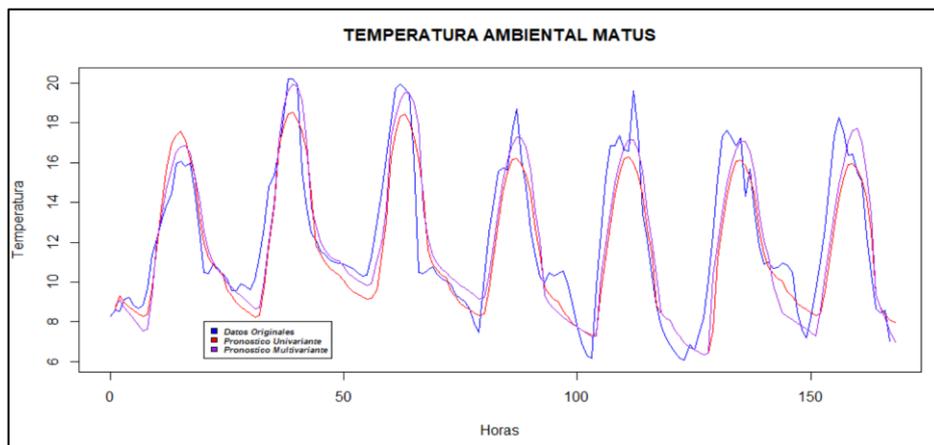


Gráfico 42-3: Pronósticos Matus

Realizado por: Mariño Gabriela, 2021.

En la estación Matus se pronostican los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1,2,3,4$ por medio de modelos ARIMA, específicamente con un ARIMA (2,1,2), ARIMA (1,1,1), ARIMA (1,1,1), ARIMA (1,1,2). De color morado se observan los pronósticos con coeficientes calculados por el método VAR.

3.7.6. Pronósticos de Estación Multitud

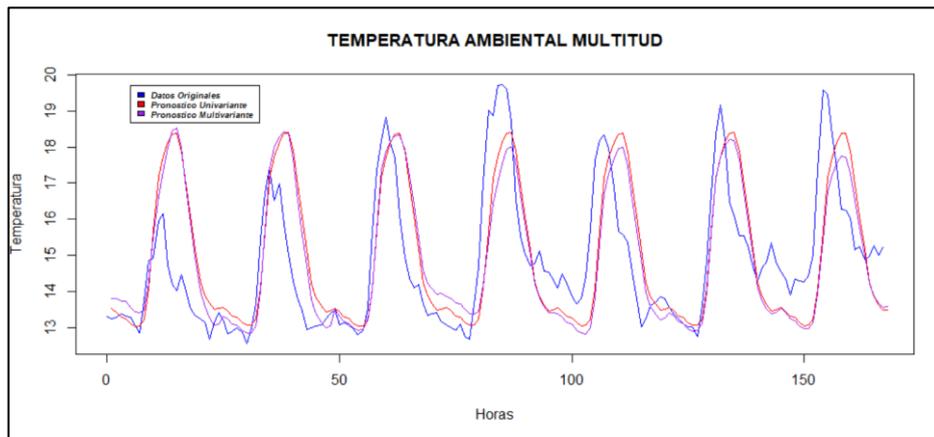


Gráfico 43-3: Pronósticos Multitud

Realizado por: Mariño Gabriela, 2021.

En la estación Multitud se pronostican los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1, 2, 3, 4$ por medio de modelos ARIMA, específicamente con un ARIMA (0,1,1), ARIMA (1,1,1), ARIMA (1,0,1) y ARIMA (1,1,2). De color morado se observan los pronósticos con coeficientes calculados por el método VAR.

3.7.7. Pronósticos de Estación Quimiag

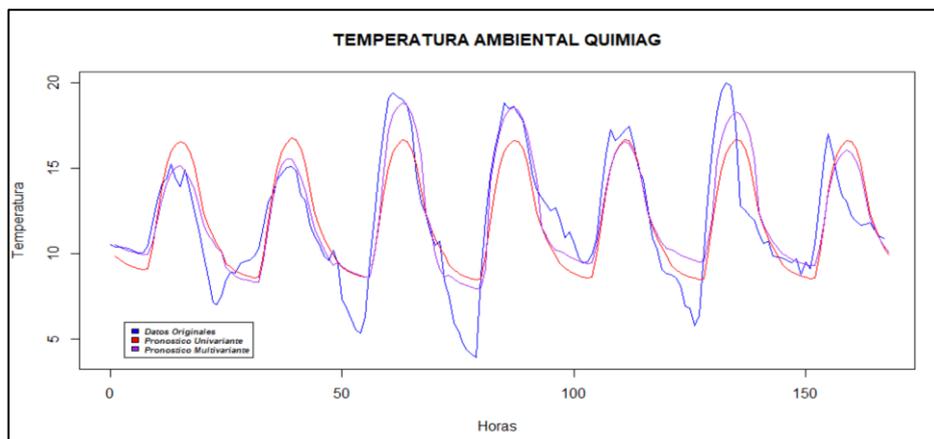


Gráfico 44-3: Pronósticos Quimiag

Realizado por: Mariño Gabriela, 2021.

En la estación Quimiag se pronostican los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1, 2, 3, 4$ por medio de modelos ARIMA, específicamente con un ARIMA (1,1,1), ARIMA (2,1,2), ARIMA (1,0,1) y ARIMA (2,0,2). De color morado se observan los pronósticos con coeficientes calculados por el método VAR.

3.7.8. Pronósticos de Estación San Juan

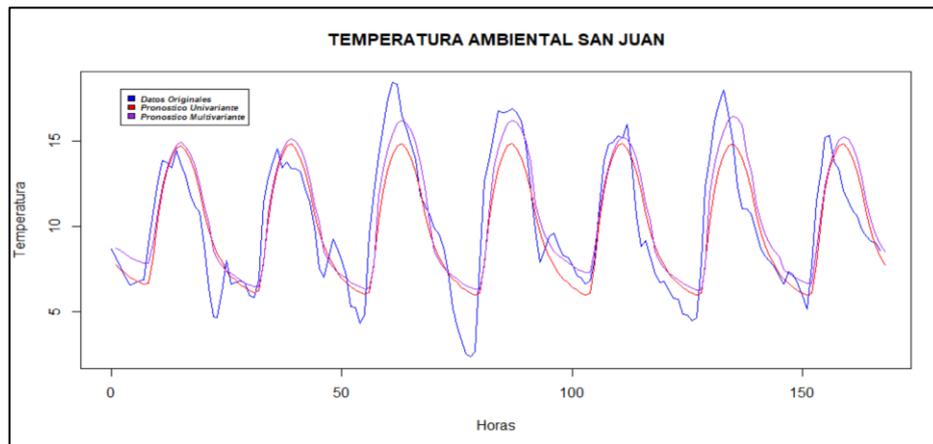


Gráfico 45-3: Pronósticos San Juan

Realizado por: Mariño Gabriela, 2021.

Los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1,2,3,4$ para la Estación San Juan fueron pronosticados por medio de modelos ARIMA, específicamente con un ARIMA (1,1,1), ARIMA (1,1,1), ARIMA (1,0,1) y ARIMA (1,0,1), mientras que los datos observados de color morado fueron pronosticados con coeficientes calculados por el método VAR.

3.7.9. Pronósticos de Estación Tixán

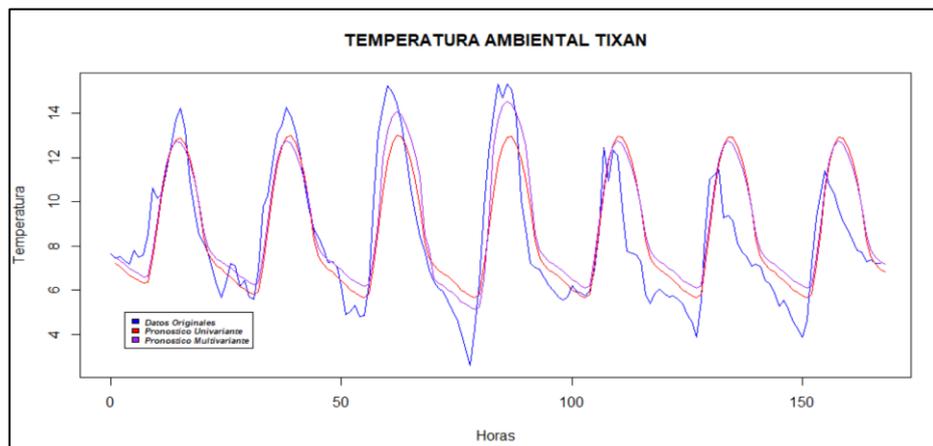


Gráfico 46-3: Pronósticos Tixán

Realizado por: Mariño Gabriela, 2021.

Los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1,2,3,4$ para la Estación San Juan fueron pronosticados por medio de modelos ARIMA, específicamente con un ARIMA (1,0,1), ARIMA (1,1,3), ARIMA (2,1,2) y ARIMA (2,1,3), mientras que los datos observados de color morado fueron pronosticados con coeficientes calculados por el método VAR.

3.7.10. Pronósticos de Estación Tunshi

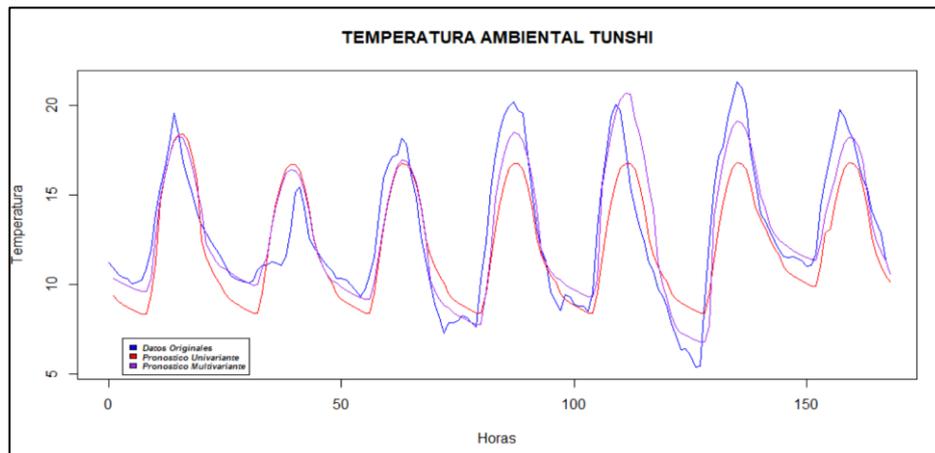


Gráfico 47-3: Pronósticos Tunshi

Realizado por: Mariño Gabriela, 2021.

Los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1,2,3,4$ para la Estación Tunshi fueron pronosticados por medio de modelos ARIMA, específicamente con un ARIMA (1,1,1), ARIMA (0,1,2), ARIMA (0,1,1) y ARIMA (1,1,1), mientras que los pronósticos de los coeficientes con series de tiempo multivariante fueron pronosticados con el método VAR.

3.7.11. Pronósticos de Estación Urbina

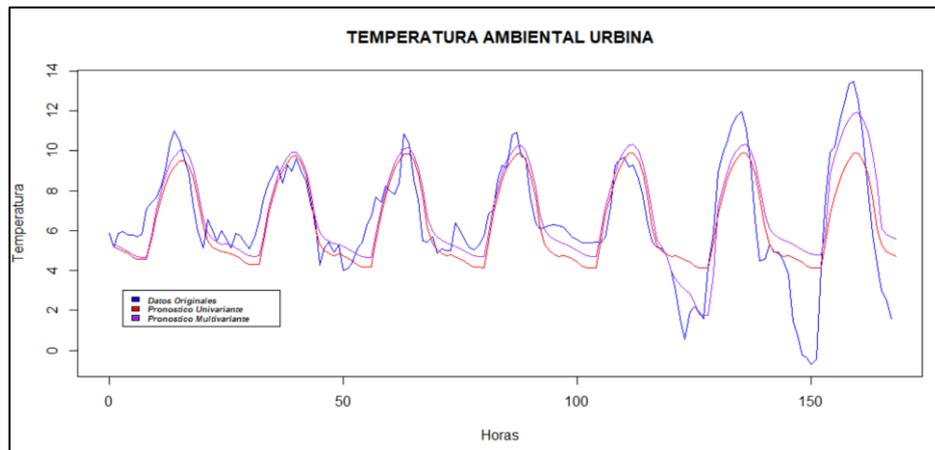


Gráfico 48-3: Pronósticos Urbina

Realizado por: Mariño Gabriela, 2021.

Los coeficientes de las 4 componentes $\hat{\beta}_{t,k}$ para $k = 1,2,3,4$ para la Estación Urbina fueron pronosticados por medio de modelos ARIMA (1,1,3), ARIMA (1,1,1), ARIMA (1,0,1) y ARIMA (4,1,2) para series de tiempo univariante), mientras que los pronósticos de los coeficientes con series de tiempo multivariante fueron pronosticados con el método VAR.

3.8. Criterios de Evaluación de los Modelos

Tabla 15-3: Criterios de evaluación del modelo

N°	ESTACIÓN	MODELO	MSE	RMSE	MAE	MDAE	MAPE	MDAPE	RMSPE	RMSDPE
1	ALAO	Univariante	1,7645	3,1133	1,4227	1,2469	0,1815	0,1123	0,1381	0,0950
		Multivariante	1,5507	2,4046	1,2625	1,0422	0,1421	0,1153	0,0163	0,0313
2	ATILLO	Univariante	1,6103	2,5931	1,3627	1,2945	0,2210	0,1754	0,0585	0,1269
		Multivariante	1,3568	1,8408	1,1026	1,0290	0,1709	0,1215	0,0550	0,0303
3	CUMANDÁ	Univariante	2,3811	5,6698	1,6898	1,0164	0,0775	0,0500	0,0511	0,0103
		Multivariante	2,0954	4,3906	1,5195	0,9745	0,0700	0,0445	0,0483	0,0078
4	ESPOCH	Univariante	1,9682	3,8737	1,4958	1,1586	0,1418	0,0894	0,1084	0,0630
		Multivariante	1,7112	2,9284	1,3343	1,0804	0,1150	0,0850	0,0237	0,0242
5	MATUS	Univariante	1,5971	2,5508	1,3147	1,1688	0,1142	0,0939	0,0259	0,0641
		Multivariante	1,5907	2,5303	1,2168	0,8925	0,1141	0,0735	0,0114	0,0537
6	MULTITUD	Univariante	1,2193	1,4867	0,9233	0,6759	0,0607	0,0502	0,0077	0,0779
		Multivariante	1,2093	1,4624	0,9064	0,6455	0,0599	0,0443	0,0050	0,0669
7	QUIMIAG	Univariante	1,8260	3,3343	1,4718	1,1090	0,1462	0,0977	0,0453	0,0152
		Multivariante	1,8056	3,2603	1,3345	0,9516	0,1414	0,0720	0,0446	0,0109
8	SAN JUAN	Univariante	1,7531	3,0734	1,4090	1,2087	0,1962	0,1084	0,1353	0,0611
		Multivariante	1,5375	2,3639	1,2330	1,0268	0,1510	0,1120	0,0316	0,0269
9	TIXÁN	Univariante	1,5081	2,2744	1,2201	1,0505	0,1704	0,1315	0,1143	0,0837
		Multivariante	1,4346	2,0580	1,1113	0,8774	0,1482	0,1155	0,0771	0,0429
10	TUNSHI	Univariante	2,0717	4,2918	1,7410	1,5999	0,1394	0,1403	0,0532	0,0985
		Multivariante	1,9275	3,7152	1,5915	1,3798	0,1326	0,1116	0,0123	0,0630
11	URBINA	Univariante	1,5504	2,4037	1,1034	0,7176	0,3116	0,1129	0,1269	0,0616
		Multivariante	1,4673	2,1530	1,0943	0,6332	0,2412	0,1029	0,1104	0,0577

Fuente: Software R

Realizado por: Mariño Gabriela, 2021.

Se calculan los errores generados al utilizar los modelos de regresión del componente principal funcional, con coeficientes pronosticados por series de tiempo univariante y multivariante, utilizados para pronosticar la Temperatura Ambiental, dando como resultado menores errores en las Estaciones de Alao, Atillo, Cumandá, Espoch, Matus, Multitud, Quimiag, San Juan, Tixán, Tunshi y Urbina con el modelo de coeficientes pronosticados con series de tiempo multivariante, utilizando el método VAR para el pronóstico de coeficientes.

3.9. Análisis de Precisión de los Modelos

3.9.1. Coeficiente U de Theil

Tabla 16-3: Coeficiente U de Theil

ESTACIÓN	COEFICIENTE U de THEIL	
	S.T Univariante	S.T. Multivariante
ALAO	0,071	0,068
ATILLO	0,110	0,105
CUMANDÁ	0,054	0,054
ESPOCH	0,065	0,063
MATUS	0,065	0,063
MULTITUD	0,040	0,040
QUIMIAG	0,073	0,073
SAN JUAN	0,072	0,071
TIXÁN	0,080	0,080
TUNSHI	0,080	0,079
URBINA	0,106	0,102

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

Los valores del coeficiente U de Theil (Tabla 16-3) están cercanos a cero, es decir que ambos modelos tienen buena precisión al pronosticar la Temperatura Ambiental.

3.9.2. Test de Diebold-Mariano

El test de Diebold Mariano permite ver la precisión del modelo, basado en algunas hipótesis, dependiendo de la instrucción (less, greater y two.sided) y objetivos del investigador, en este caso se utilizó la instrucción greater, cuyas hipótesis son las siguientes:

H0: Los dos métodos tienen la misma precisión de pronóstico

H1: el método 2 es más preciso que el método 1

Tabla 17-3: Test de Diebold-Mariano

ESTACIÓN	TEST DE DIEBOLD-MARIANO
	Univariante -Multivariante
ALAO	0,001
ATILLO	0,002
CUMANDÁ	0,890
ESPOCH	0,981
MATUS	0,954
MULTITUD	0,665
QUIMIAG	0,321
SAN JUAN	0,920
TIXÁN	0,989
TUNSHI	0,003
URBINA	0,001

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

Según el test de Diebold-Mariano para las Estaciones de Alao, Atillo, Tunshi y Urbina, el valor de p es menor al valor de alfa establecido a 0.05, por ende, el modelo de regresión del componente principal funcional, con coeficientes pronosticados por series de tiempo multivariante presenta mayor precisión de pronóstico, mientras que en las demás estaciones los dos modelos tienen la misma precisión de pronóstico.

3.10. Pronósticos de la Temperatura Ambiental

Los pronósticos de la Temperatura Ambiental vienen dados por el siguiente modelo matemático

$$y_{n+h|n}(x) = \bar{f}(x) + \sum_{k=1}^K \hat{\beta}_{n+h|n,k} \hat{\phi}_k(x) \quad (51)$$

En las siguientes ecuaciones se representan pronósticos para 7 días en la hora cero

$$\begin{aligned} y_1(x) &= \bar{f}_0(x) + (\hat{\beta}_{7|1,1} \hat{\phi}_1(x) + \hat{\beta}_{7|1,2} \hat{\phi}_2(x) + \hat{\beta}_{7|1,3} \hat{\phi}_3(x) + \hat{\beta}_{7|1,4} \hat{\phi}_4(x)) \\ y_2(x) &= \bar{f}_0(x) + (\hat{\beta}_{7|2,1} \hat{\phi}_1(x) + \hat{\beta}_{7|2,2} \hat{\phi}_2(x) + \hat{\beta}_{7|2,3} \hat{\phi}_3(x) + \hat{\beta}_{7|2,4} \hat{\phi}_4(x)) \\ y_3(x) &= \bar{f}_0(x) + (\hat{\beta}_{7|3,1} \hat{\phi}_1(x) + \hat{\beta}_{7|3,2} \hat{\phi}_2(x) + \hat{\beta}_{7|3,3} \hat{\phi}_3(x) + \hat{\beta}_{7|3,4} \hat{\phi}_4(x)) \\ y_4(x) &= \bar{f}_0(x) + (\hat{\beta}_{7|4,1} \hat{\phi}_1(x) + \hat{\beta}_{7|4,2} \hat{\phi}_2(x) + \hat{\beta}_{7|4,3} \hat{\phi}_3(x) + \hat{\beta}_{7|4,4} \hat{\phi}_4(x)) \\ y_5(x) &= \bar{f}_0(x) + (\hat{\beta}_{7|5,1} \hat{\phi}_1(x) + \hat{\beta}_{7|5,2} \hat{\phi}_2(x) + \hat{\beta}_{7|5,3} \hat{\phi}_3(x) + \hat{\beta}_{7|5,4} \hat{\phi}_4(x)) \\ y_6(x) &= \bar{f}_0(x) + (\hat{\beta}_{7|6,1} \hat{\phi}_1(x) + \hat{\beta}_{7|6,2} \hat{\phi}_2(x) + \hat{\beta}_{7|6,3} \hat{\phi}_3(x) + \hat{\beta}_{7|6,4} \hat{\phi}_4(x)) \\ y_7(x) &= \bar{f}_0(x) + (\hat{\beta}_{7|7,1} \hat{\phi}_1(x) + \hat{\beta}_{7|7,2} \hat{\phi}_2(x) + \hat{\beta}_{7|7,3} \hat{\phi}_3(x) + \hat{\beta}_{7|7,4} \hat{\phi}_4(x)) \end{aligned}$$

A continuación, se muestra los pronósticos de la temperatura ambiental, mediante el modelo de regresión del componente principal funcional, con coeficientes pronosticados por series de tiempo multivariante, de una semana, de todas las estaciones meteorológicas, ordenados según el número de horas.

3.10.1. Valores Pronosticados de Alao

Tabla 18-3: Valores Pronosticados de Alao

Horas	día1	día2	día3	día4	día5	día6	día7
0	10,2918	9,1908	8,0848	9,0440	9,0194	8,0067	7,9907
1	10,1178	8,9646	7,8445	8,1062	8,7806	7,7685	7,7525
2	9,9666	8,7527	7,6238	7,5053	8,5586	7,5464	7,5299
3	9,8474	8,6196	7,4846	6,1476	8,5005	7,4084	7,3920
4	9,7587	8,5028	7,3587	5,9476	8,3947	7,2828	7,2663
5	9,6846	8,3787	7,2181	5,7476	8,1060	7,1447	7,1283
6	9,5944	8,2479	7,0702	5,4476	7,4001	6,9999	6,9839
7	9,5743	8,2057	7,0213	5,0476	7,4617	6,9921	6,9366
8	10,0777	8,8794	7,7643	7,3199	8,6939	8,6806	7,6637
9	10,8773	10,2798	10,3463	10,2816	10,2649	10,2457	10,2285
10	11,7712	11,5347	11,7152	11,6405	11,6298	11,6078	11,5909
11	12,8324	12,9052	13,1815	13,0996	13,0946	13,0711	13,0551
12	14,0511	14,2935	14,6371	14,5443	14,5390	14,5103	14,4915
13	14,7426	15,1431	15,5062	16,4164	15,4099	16,3784	15,3581
14	14,9778	15,5195	15,8825	16,8043	15,7953	16,4623	15,7411
15	14,9008	15,5287	15,8654	16,7958	15,7843	15,7495	15,7272
16	14,6260	15,3008	15,5979	16,5398	15,5278	14,4943	15,4731
17	14,1574	14,7467	15,0070	14,9522	14,9397	13,9078	14,8871
18	13,2303	13,6797	13,8737	13,8283	13,8172	12,7908	13,7735
19	12,1135	12,3417	12,4740	12,4274	12,4167	11,3936	12,3778
20	11,2018	11,2768	11,3519	11,3015	11,2923	10,2717	11,2574
21	10,5682	10,5158	10,5325	10,4805	10,4722	10,0537	10,4406
22	9,9984	9,8626	9,8454	9,7918	9,7840	9,3667	9,7542
23	9,6335	8,4077	9,3758	9,3197	8,3105	8,2927	9,2794

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.2. Valores Pronosticados de Atillo

Tabla 19-3: Valores Pronosticados de Atillo

Horas	día1	día2	día3	día4	día5	día6	día7
0	4,2341	4,9912	6,2927	5,9371	6,3565	4,8931	7,0007
1	3,9692	4,5576	6,1946	5,7736	6,1460	4,6822	6,9063
2	3,7289	3,9046	6,1151	5,5988	5,9484	4,5032	6,7006
3	3,3891	3,6024	5,8875	5,3169	5,6529	4,2068	6,4071
4	2,8440	3,7045	5,6864	4,1013	5,3508	3,7749	5,8504
5	2,8796	3,8835	5,6144	4,0842	4,3790	3,7810	5,1352
6	2,8448	4,2010	5,6406	4,1376	4,4112	3,7636	5,1213
7	3,2452	5,1443	5,7216	5,1458	4,5085	4,0637	5,4010
8	5,2632	6,3859	6,7130	6,2677	5,8397	5,7238	5,9704
9	7,6963	7,7540	7,9685	7,7619	7,5575	7,7388	7,8451
10	9,1773	8,5922	8,8861	8,8462	8,7239	9,0096	9,0325
11	9,8475	9,0805	10,4512	9,4740	9,3492	9,6334	9,6414
12	10,2869	9,6034	11,8166	9,9031	9,7467	10,0443	10,0395
13	10,4747	10,6601	12,0771	10,0507	9,9267	10,2484	10,2463
14	10,6286	11,4091	12,2070	10,0651	9,9655	10,3812	10,3743
15	10,6110	11,8531	11,9634	9,7675	9,7212	10,3244	10,3174
16	10,4141	11,7712	10,4046	9,5178	9,4859	10,1418	10,1448
17	9,9221	10,3916	9,6079	9,1138	9,0669	9,7845	9,7078
18	9,1086	8,6801	8,6531	8,4215	8,3822	9,1961	8,9351
19	7,9345	7,7914	7,6968	7,8095	7,3939	8,1804	7,8868
20	6,9003	7,0927	6,9939	7,5095	6,6188	7,7660	7,0230
21	6,5799	6,8143	6,6050	7,2912	6,2100	7,6150	6,6970
22	6,0486	6,5067	6,2514	6,8808	5,7579	7,3880	6,2709
23	5,5017	6,4202	6,1965	6,7013	5,2657	7,1660	6,1722

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.3. Valores Pronosticados de Cumandá

Tabla 20-3: Valores Pronosticados de Cumandá

Horas	día1	día2	día3	día4	día5	día6	día7
0	20,6652	20,3392	20,5175	20,6541	20,6719	20,6293	20,6403
1	20,1735	19,8120	19,9902	20,1489	20,1701	20,1343	20,1475
2	19,8035	19,4192	19,5964	19,7696	19,7930	19,7622	19,7769
3	19,5044	19,1081	19,2885	19,4679	19,4922	19,4618	19,4771
4	19,2543	18,8480	19,0289	19,2144	19,2396	19,2108	19,2267
5	19,1086	18,6959	18,8729	19,0638	19,0899	19,0646	19,0811
6	18,9882	18,5675	18,7444	18,9403	18,9672	18,9434	18,9604
7	18,9469	18,5263	18,7094	18,9029	18,9293	18,9020	18,9186
8	19,4082	19,0386	19,2873	19,4250	19,4419	19,3663	19,3763
9	20,4866	20,2567	20,6505	20,6482	20,6412	20,4538	20,4473
10	21,6072	21,5169	22,0346	21,9002	21,8709	21,5841	21,5623
11	22,7590	22,7940	23,4211	23,1687	23,1196	22,7447	22,7092
12	23,9508	24,1101	24,8462	24,4766	24,4078	23,9451	23,8961
13	25,0106	25,2935	26,1123	25,6356	25,5491	25,0144	24,9532
14	25,8742	26,2432	27,0811	26,5440	26,4479	25,8856	25,8180
15	26,4639	26,8860	27,7378	27,1626	27,0605	26,4799	26,4082
16	26,6672	27,1076	27,9369	27,3587	27,2564	26,6856	26,6139
17	26,5340	26,9702	27,7621	27,2002	27,1011	26,5530	26,4836
18	25,6991	26,0427	26,7394	26,2699	26,1866	25,7120	25,6535
19	24,0317	24,1587	24,7087	24,4276	24,3754	24,0282	23,9911
20	23,0415	23,0408	23,5435	23,3591	23,3224	23,0273	23,0007
21	22,4453	22,3639	22,8018	22,6911	22,6669	22,4253	22,4071
22	21,8450	21,6828	22,0376	22,0077	21,9971	21,8197	21,8110
23	21,3692	21,1612	21,4790	21,4912	21,4876	21,3406	21,3368

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.4. Valores Pronosticados de Espoch

Tabla 21-3: Valores Pronosticados de Espoch

Horas	día1	día2	día3	día4	día5	día6	día7
0	11,4419	10,3347	10,2364	8,5978	10,1424	10,1194	11,1037
1	11,1715	10,0793	9,7743	7,8101	9,8715	9,8465	10,8295
2	10,9732	9,8888	8,7769	7,2072	9,6652	9,6382	10,6197
3	10,8100	9,7178	8,5926	6,5150	9,4684	9,4385	10,4181
4	10,6952	9,6066	7,9084	6,3990	9,3516	9,3212	10,3006
5	10,6391	9,5204	7,6041	6,3019	9,2530	9,2215	10,2002
6	10,5587	9,4200	7,7992	6,4960	9,1468	9,1152	10,0937
7	10,5947	9,4409	9,2997	6,9176	9,1690	9,1378	10,1165
8	11,4308	11,2809	11,1909	9,1397	10,1081	10,0871	11,0726
9	12,8576	12,7092	12,6536	12,6218	11,6000	12,5846	12,5737
10	14,4910	14,3479	14,3163	14,2963	14,2797	14,2670	14,2578
11	15,9337	15,8471	15,8474	15,8400	15,8289	15,8192	15,8120
12	17,1457	16,8045	17,1759	17,1784	17,1711	17,1633	17,1573
13	17,9746	17,5791	18,1237	18,5946	18,1177	18,1094	17,1032
14	18,3861	18,1138	18,6654	19,6631	18,6518	18,6421	16,6350
15	18,3888	18,0711	18,7172	19,7091	18,6952	18,6840	15,9560
16	18,0236	17,5032	18,2913	18,6686	18,2635	18,2519	15,2436
17	17,3398	16,8677	17,4727	17,4557	17,4392	17,4265	14,4174
18	16,1822	16,1829	16,1656	16,1450	16,1273	16,1137	13,1038
19	14,7026	14,6335	14,6066	14,5867	14,5706	14,5582	12,5492
20	13,5526	13,4512	13,4181	13,3977	12,4819	13,3700	12,3613
21	12,6921	11,7653	11,5203	12,1947	11,4765	12,4632	12,1537
22	11,4472	11,0182	9,9662	10,9366	10,9162	11,9016	11,8912
23	10,6099	10,5711	9,2114	10,3775	10,4548	11,4389	11,4277

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.5. Valores Pronosticados de Matus

Tabla 22-3: Valores Pronosticados de Matus

Horas	día1	día2	día3	día4	día5	día6	día7
0	8,7884	9,9629	11,0656	10,2118	8,2553	7,5353	8,2790
1	9,0863	9,7229	10,7898	10,0297	8,0843	7,2717	8,1150
2	8,7199	9,4896	10,4414	9,8364	7,8978	6,9376	7,9304
3	8,4629	9,3128	10,2418	9,6872	7,7535	6,7417	7,7882
4	8,1312	9,0927	10,0922	9,5074	7,5827	6,6041	7,6243
5	7,8334	8,8835	9,9508	9,3363	7,4190	6,4600	7,4667
6	7,5672	8,6765	9,8248	9,1591	7,2478	6,3331	7,3013
7	7,6399	8,7295	9,8958	9,2057	7,2934	6,4090	8,5463
8	9,3653	10,0747	11,1113	10,2831	10,3182	10,3292	10,3347
9	12,1057	12,0380	11,9272	11,9090	11,8791	11,8639	11,8502
10	13,2484	13,9874	13,6937	13,5848	13,5036	13,4675	13,4388
11	14,5881	17,1039	17,2647	15,0991	14,9807	14,9318	14,8928
12	15,6787	18,7910	18,3071	16,1193	15,9851	15,9322	15,8890
13	16,5540	19,5936	19,0961	16,8890	16,7475	16,6892	17,0429
14	16,8172	19,9577	19,5112	17,2876	17,1465	17,0787	17,6274
15	16,8227	19,8058	19,4824	17,2518	17,1245	17,0491	17,6955
16	16,5111	19,1379	19,0143	16,7435	16,6210	16,5335	17,0061
17	15,3540	17,0170	17,8841	15,6443	15,5453	15,4661	15,4157
18	14,2108	13,7609	13,6758	13,5389	13,4765	13,4298	13,3985
19	12,5483	12,8375	12,2036	12,1252	12,0836	12,0556	9,3513
20	11,5336	12,0279	11,3707	9,3354	10,3092	11,2939	8,6994
21	11,0078	11,6072	10,9330	8,9167	8,4607	9,8880	8,1762
22	10,6153	11,3077	10,6554	8,6555	8,2138	9,1372	7,5274
23	10,3574	11,1161	10,4540	8,4637	8,0715	8,4536	6,9960

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.6. Valores Pronosticados de Multitud

Tabla 23-3: Valores Pronosticados de Multitud

Horas	día1	día2	día3	día4	día5	día6	día7
0	13,8175	13,3107	13,5138	13,8515	13,3993	13,4299	13,5386
1	13,8182	13,2527	13,3268	13,8061	13,3063	13,3237	13,4195
2	13,7273	13,1037	13,1656	13,6949	13,1422	13,1689	13,2586
3	13,7299	13,0751	13,1251	13,6694	13,1001	13,1314	13,2034
4	13,5663	12,9208	12,9997	13,5150	12,9369	12,9849	13,0574
5	13,4428	12,8523	12,9324	13,4001	12,8479	12,9106	12,9768
6	13,4174	12,8609	12,9637	13,3804	12,8458	12,9224	12,9820
7	13,5022	13,0587	13,1910	13,4739	13,0112	13,1161	13,1524
8	14,2469	14,0085	14,0445	14,1772	13,8327	13,9913	13,9502
9	15,5268	15,7123	15,7302	15,4041	15,3257	15,6099	15,4163
10	16,6399	17,4003	17,3683	16,4470	16,6853	17,1674	16,7668
11	17,3333	17,9924	17,8623	17,0098	17,2441	17,7080	17,2264
12	17,9819	18,2744	18,1543	17,5502	17,6814	18,0283	17,5448
13	18,4607	18,4297	18,2957	17,9379	17,9687	18,2106	17,7318
14	18,5322	18,3518	18,3256	18,0023	17,9797	18,1901	17,7297
15	17,9961	17,6784	17,9293	17,5702	17,4716	17,6572	17,2919
16	16,8626	16,5231	17,2109	16,6729	16,5382	16,7153	16,5185
17	15,9146	15,5798	16,3909	15,8781	15,7009	15,8556	15,7835
18	14,9631	14,7117	15,4955	15,0439	14,8572	14,9998	15,0186
19	14,0013	13,7853	14,5978	14,1994	13,9758	14,1094	14,2227
20	13,5854	13,3947	14,2365	13,8308	13,5985	13,7367	13,8827
21	13,2826	13,1519	14,0798	13,5807	13,3733	13,5276	13,6960
22	13,0847	13,0151	13,9220	13,4153	13,2374	13,3864	13,5739
23	13,0965	13,0470	13,9469	13,4188	13,2697	13,4133	13,5909

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.7. Valores Pronosticados de Quimiag

Tabla 24-3: Valores Pronosticados de Quimiag

Horas	día1	día2	día3	día4	día5	día6	día7
0	10,5472	9,1723	9,5054	8,7370	10,1644	10,2530	10,0274
1	10,4076	8,9235	9,2809	8,5552	9,9925	10,0881	9,8451
2	10,2658	8,6844	9,0608	8,3685	9,8176	9,9214	9,6666
3	10,1587	8,5518	8,9367	8,2463	9,7027	9,8107	9,5555
4	10,0909	8,4736	8,8190	8,1385	9,5932	9,7030	9,4587
5	10,0241	8,4259	8,7395	8,0518	9,5052	9,6158	9,3849
6	9,9581	8,3407	8,6584	7,9738	9,4318	9,5452	9,3139
7	9,9665	8,3568	8,6942	8,0012	9,4613	9,5745	9,3394
8	10,6528	9,6634	10,0642	9,1375	10,5310	10,5857	10,3673
9	11,8621	11,4771	11,9684	14,1220	12,1447	12,1375	11,9199
10	13,1510	13,1829	14,7987	15,7959	13,7713	15,5183	13,4804
11	14,0711	14,4547	17,2560	17,1277	15,0571	16,7571	14,6676
12	14,7186	15,2382	18,1965	18,0008	15,9221	17,6042	15,4732
13	15,0511	15,5322	18,6164	18,4102	16,3480	18,0328	15,8659
14	15,1588	15,5408	18,8410	18,6042	16,5896	18,2907	16,0744
15	14,8620	15,1182	18,6962	18,3956	16,4570	18,1874	15,9215
16	14,3766	14,4630	18,1593	17,8400	15,9681	17,7336	15,4595
17	13,7943	13,6738	17,2644	17,0292	15,1778	16,9671	14,7160
18	12,7647	12,5163	15,7734	15,6576	13,7833	15,5837	13,4099
19	11,6760	11,4396	12,3770	14,3207	12,4131	12,4139	12,1255
20	11,0705	10,8364	11,6179	11,5896	11,6676	11,6698	11,4243
21	10,7329	10,2731	9,9480	11,0289	11,1180	11,1393	10,9054
22	10,3570	9,7317	9,1061	10,5567	10,6587	10,6916	10,4427
23	10,0987	9,3413	8,6920	10,2039	10,3131	10,3552	10,0999

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.8. Valores Pronosticados de San Juan

Tabla 25-3: Valores Pronosticados de San Juan

Horas	día1	día2	día3	día4	día5	día6	día7
0	8,7523	7,4285	7,3354	7,3021	8,2761	7,2604	8,2505
1	8,5904	7,2490	7,1483	7,1162	8,0893	7,0733	7,4630
2	8,4049	7,0673	6,9667	6,9366	7,9104	6,8948	7,2847
3	8,2019	6,8497	6,7449	6,7140	7,6870	6,6706	7,0601
4	8,0406	6,6870	6,5813	6,5508	7,5239	6,5075	6,8969
5	7,9394	6,5700	6,4612	6,4282	7,3999	6,3825	6,7715
6	7,8412	6,4654	6,3561	6,3200	7,2904	6,2725	6,6611
7	7,8595	6,4852	6,3775	6,3399	7,3102	6,2921	6,6806
8	9,0207	7,7801	7,7138	7,6831	8,6613	7,6496	8,6421
9	10,8547	10,7874	11,7765	10,7563	10,7450	11,9417	10,7394
10	12,3310	12,3538	13,3706	13,3539	12,3469	13,5481	12,3486
11	13,4170	13,5036	14,5414	14,5283	13,5247	14,7292	13,5317
12	14,2049	14,3399	15,3943	15,3856	14,3848	15,5919	14,3961
13	14,7649	14,9356	15,9974	15,9968	14,9997	16,2092	15,0147
14	14,9298	15,1408	16,2116	16,2247	15,2336	16,4458	15,2530
15	14,6664	14,9449	16,0280	16,0695	15,0908	16,3077	15,1177
16	14,2416	14,5240	15,6066	15,6602	14,6860	15,9036	14,7141
17	13,5471	13,8052	14,8798	14,9322	13,9568	14,1731	13,9829
18	12,4890	12,6800	13,7394	13,7762	12,7937	13,2001	12,8127
19	11,1987	11,2812	12,3139	11,3308	11,3384	11,3435	11,3467
20	10,2375	10,2481	10,2672	10,2700	10,2712	10,2714	10,2718
21	8,5698	8,5176	8,5264	9,5185	8,5147	9,5107	9,5086
22	8,0873	7,9786	7,9771	8,9575	7,9480	8,9403	8,9361
23	7,7254	7,5901	7,5821	8,5575	7,5454	8,5362	8,5311

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.9. Valores Pronosticados de Tixán

Tabla 26-3: Valores Pronosticados de Tixán

Horas	día1	día2	día3	día4	día5	día6	día7
0	7,4775	7,1608	7,0761	6,0374	7,0180	7,0069	7,0002
1	7,3555	7,0220	6,9320	5,8913	6,8709	6,8592	6,8522
2	7,1752	6,8257	6,7342	5,6925	6,6717	6,6598	6,6526
3	7,0029	6,6350	6,5379	5,4935	6,4714	6,4587	6,4510
4	6,8912	6,5223	6,4233	5,3786	6,3564	6,3436	6,3359
5	6,7659	6,3876	6,2862	5,2406	6,2179	6,2049	6,1970
6	6,6123	6,2598	6,1658	5,1230	6,1017	6,0894	6,0820
7	6,6485	6,3610	6,2820	5,2453	6,2266	6,2157	6,2091
8	7,7302	7,5822	7,5408	6,5165	7,5029	7,4945	7,4894
9	9,1178	9,0789	9,0647	9,0493	9,0391	9,0325	9,0284
10	10,6078	10,6541	11,9632	12,4553	10,6481	10,6430	10,6397
11	11,7735	11,8861	13,2149	13,7137	11,9095	11,9059	11,9035
12	12,4464	12,5358	13,8627	14,3623	12,5592	12,5566	12,5548
13	12,7380	12,7651	14,0781	14,5751	12,7719	12,7696	12,7681
14	12,7132	12,6433	13,9283	14,4185	12,6134	12,6104	12,6086
15	12,3793	12,2307	13,4854	13,9681	12,1603	12,1562	12,1539
16	11,7995	11,6568	12,9045	13,3862	11,5778	11,5735	11,5710
17	11,0185	10,9073	12,1582	12,5399	10,8307	10,8257	10,8228
18	9,9302	9,8628	11,1243	10,6075	9,7981	9,7927	9,7894
19	8,6483	8,5844	8,5471	8,5283	8,5172	8,5105	8,5064
20	7,9366	7,8769	7,8384	7,8188	7,8072	7,8001	7,7957
21	7,6236	7,5757	6,5380	7,5196	7,5085	7,5017	7,4975
22	7,4011	7,3652	6,3268	7,3085	7,2973	7,2905	7,2863
23	7,2838	7,2441	6,2060	7,1876	7,1764	7,1696	7,1654

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.10. Valores Pronosticados de Tunshi

Tabla 27-3: Valores Pronosticados de Tunshi

Horas	día1	día2	día3	día4	día5	día6	día7
0	10,3365	10,8876	10,0654	8,6640	10,2165	8,2006	12,2896
1	10,1652	10,6508	9,8325	8,4366	9,9869	7,5629	12,0631
2	10,0417	10,4672	9,6631	8,2675	9,8154	7,2913	11,8908
3	9,9338	10,3316	9,5295	8,1334	9,6789	7,1537	11,7520
4	9,7993	10,1809	9,3830	7,9870	9,5326	7,0078	11,6064
5	9,6893	10,0715	9,2661	7,8704	9,4147	6,8892	11,4874
6	9,5956	9,9633	9,1536	7,7609	9,3051	6,7808	11,3802
7	9,5815	9,9725	9,1526	7,7571	9,3011	6,7755	11,3742
8	10,3507	10,9179	10,0715	9,6657	10,2104	7,6778	12,2714
9	12,6989	11,4814	11,6439	11,7244	15,3779	13,7813	13,8318
10	14,7916	13,0760	13,2625	14,8371	16,9974	15,4006	15,0503
11	16,1275	14,4173	14,6682	16,2412	18,4120	16,8210	15,9339
12	17,1874	15,4638	15,7821	17,3525	19,5342	17,9493	17,0658
13	17,9995	16,2097	16,5915	18,1609	20,3478	18,7670	17,8850
14	18,3049	16,4206	16,9130	18,4698	20,6685	19,0932	18,2133
15	18,1374	16,3666	16,8883	18,4065	20,6149	19,0320	18,1464
16	17,4449	16,0210	16,5105	17,9889	19,2183	18,6296	17,7434
17	16,5356	15,2488	15,6606	17,1497	18,3863	17,8030	16,9248
18	15,3332	14,1040	14,4154	15,9182	17,1439	16,5567	15,0778
19	14,1554	12,5124	12,7308	14,2593	15,4717	14,8844	13,4069
20	12,2228	11,5843	10,2525	11,7900	14,1937	14,2435	12,4245
21	11,6667	10,9977	9,6359	11,1840	11,1818	13,2918	11,8133
22	11,2566	10,5221	9,1510	10,7071	9,9989	12,8091	11,3303
23	10,9617	10,1875	8,8050	10,3662	9,1539	12,4639	10,5850

Fuente: Software R

Realizado por: Mariño Gabriela,2021.

3.10.11. Valores Pronosticados de Urbina

Tabla 28-3: Valores Pronosticados de Urbina

Horas	día1	día2	día3	día4	día5	día6	día7
0	5,2884	5,3354	5,2977	5,3210	5,3553	3,5856	5,4123
1	5,1989	5,2452	5,2047	5,2256	5,2583	3,2877	5,3140
2	5,0705	5,1175	5,0748	5,0939	5,1258	3,0551	5,1816
3	4,9445	4,9923	4,9490	4,9661	4,9968	2,8255	5,0515
4	4,7829	4,8307	4,7905	4,8126	4,8462	2,3766	4,9043
5	4,6900	4,7373	4,7012	4,7239	4,7570	1,7869	4,8139
6	4,6588	4,7070	4,6667	4,6835	4,7133	1,7417	4,7674
7	4,6662	4,7212	4,6719	4,6835	4,7131	1,7426	4,7692
8	5,7859	5,7967	5,8068	5,8460	5,8738	3,8926	5,9090
9	7,0425	7,0175	7,0799	7,1489	7,1797	7,1923	8,8021
10	7,9354	7,9012	7,9828	8,0677	8,1042	8,1179	9,7277
11	8,7100	8,6610	8,7616	8,8588	8,8970	8,9085	10,5163
12	9,3574	9,2934	9,4210	9,5376	9,5805	9,5918	11,1998
13	9,7251	9,6466	9,7981	9,9294	9,9743	9,9838	11,5908
14	10,0243	9,9312	10,0989	10,2378	10,2808	10,2867	11,8916
15	10,0578	9,9437	10,1338	10,2868	10,3291	10,3312	11,9360
16	9,7397	9,6230	9,8148	9,9648	10,0028	10,0028	11,6079
17	9,0920	8,9946	9,1683	9,3051	9,3416	9,3447	10,9521
18	8,0194	7,9668	8,0900	8,1963	8,2325	8,2440	9,8564
19	6,5855	6,5845	6,6630	6,7440	6,8840	6,8073	8,4265
20	5,8645	5,8946	5,9501	6,0165	5,5583	6,0886	6,1117
21	5,5830	5,6235	5,6734	5,7362	5,1791	5,8120	5,8365
22	5,4342	5,4848	5,5242	5,5781	4,7195	5,6533	5,6779
23	5,3122	5,3616	5,3977	5,4513	3,9927	5,5265	5,5514

Fuente: Software R

Realizado por: Mariño Gabriela, 2021.

CONCLUSIONES

- Los datos de las estaciones meteorológicas proporcionados por el GEAA, de la variable Temperatura Ambiental, del año 2014 al 2019, fueron validados, según como lo establece la Organización Mundial Meteorológica, utilizando valores umbrales, eliminando los valores erróneos de aquellos datos de temperatura, que no estaban dentro de los límites establecidos, además, se completaron datos faltantes de 53 bases de las 66 proporcionadas, con un máximo de imputación del 30 %.
- Por la cantidad de base de datos almacenadas, se consolidó la información en una sola base, formando una matriz de 19355 filas por 27 columnas, donde las filas representaban los días y las columnas las horas, identificadas por el año y la estación a la que pertenecen los valores de temperaturas registradas, siendo esto necesario para la optimización de tiempo y procesamiento de la información.
- Mediante la metodología de datos funcionales, se utilizó una base de 15 B-splines para el suavizado de las curvas, se identificó las curvas medias anuales de las 11 estaciones meteorológicas, observando que la variabilidad de la temperatura ambiental por horas, aumenta considerablemente en el lapso de 10:00 am a 16:00 pm. Por medio de curvas de contornos y niveles se observó que las estaciones de Cumandá y Tunshi registran valores de temperaturas promedios más altas, bordeando los 32 °C y 25 °C respectivamente.
- El modelado de los datos fue, mediante regresión del componente principal funcional, utilizando 4 componentes principales funcionales, explicando más del 80 % de la variabilidad total. Los 4 componentes en la estación de Alao explicaron el 89%, en Atillo el 87%, en Cumandá el 96%, en Espoch el 91%, en Matus 88%, en Multitud 86%, en Quimiag 90%, en San Juan 85%, en Tixán 85%, en Tunshi 90% y en Urbina 86% de variabilidad total. Para los pronósticos de coeficientes $\hat{\beta}_{t,k}$ se utilizaron series de tiempo univariante y multivariante.
- El modelo que presentó menores errores fue el de regresión del componente principal funcional, con coeficientes pronosticados por series de tiempo multivariante. Se calculó el coeficiente U de Theil, cuyos valores oscilaron de 0,040 a 0,106, siendo estos valores próximos a cero, concluyendo que, ambos modelos tienen una buena precisión de pronóstico, además el test Diebold-Mariano indicó que los modelos utilizados presentan la misma precisión de pronóstico en las estaciones de Cumandá, Espoch, Matus, Multitud, Quimiag, San Juan y Tixán, mientras que en las estaciones de Alao, Atillo, Tunshi y Urbina, el modelo de regresión del componente principal funcional, con coeficientes pronosticados por series de tiempo multivariante tiene mejor precisión.
- Se utilizó el modelo de regresión del componente principal funcional, con coeficientes pronosticados por series de tiempo multivariante, para pronosticar los valores de temperatura ambiental promedio para una semana, con un total de 168 horas, para las 11 estaciones.

RECOMENDACIONES

- Analizar y dar mantenimiento a las estaciones de forma constante, para evitar la pérdida de información de las variables meteorológicas, disminuyendo así la cantidad de datos faltantes, obteniendo un análisis más preciso y veraz a la hora del análisis.
- Realizar análisis descriptivos funcionales de forma continua a la variable temperatura ambiental u otras variables meteorológicas, para observar el comportamiento de dicha variable, permitiendo conocer sus valores mínimos y máximos, registrados en diferentes periodos de tiempo.
- Aplicar los métodos de pronósticos analizados en esta investigación con nuevas variables meteorológicas.
- Motivar a los estudiantes a utilizar técnicas de análisis de datos funcionales para resolver diferentes problemas, en donde existan gran cantidad de datos, optimizando así el tratamiento y análisis de los datos.

GLOSARIO

- 1. ACPF:** Análisis de Componentes Principales Funcionales
- 2. ADF:** Análisis de datos funcionales
- 3. ARIMA:** Autoregressive Integrated Moving Average (Modelo Autoregresivo Integrado de media móvil)
- 4. GEAA:** Grupo de Energías Alternativas y Ambiente
- 5. OMM:** Organización Mundial de Meteorología
- 6. RSM:** Error Cuadrático Medio
- 7. VAR:** Vector Autoregressive Models for Multivariate Time Series (Modelos de vectores autorregresivos para series de tiempo multivariante)

BIBLIOGRAFÍA

ACURIO, W. & PILCO, V. Técnicas Estadísticas Para La Modelación Y Predicción. (Trabajo de titulación) (Grado). [en línea] Escuela Superior Politecnica de Chimborazo, Riobamba, Ecuador.2019.pp.45-55 [Consulta:2019-11-19]. Disponible en: <http://dspace.espoch.edu.ec/handle/123456789/10955>

AGUILERA, M. Estimación Penalizada con datos Funcionales (Trabajo de titulación) (Maestría). [en línea] Universidad de Granada, Granada, España.2009.pp.12-23 [Consulta:2019-10-15]. Disponible en: [https://masteres.ugr.es/moea/pages/tfm0809/estimacin-penalizada-con-datos-funcionales/!](https://masteres.ugr.es/moea/pages/tfm0809/estimacin-penalizada-con-datos-funcionales/)

ANDRADES, M. & MUÑEZ, C. Fundamentos de Climatología [en línea]. Universidad de la Rioja, Logroño.2012 ,ISBN 978-84-695-2799-3,Disponible en: <http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf>.

BÁRCENAS, R. Estudio de Perfiles de Olas vía Análisis de Datos Funcionales (Trabajo de titulación) (Maestría). [en línea] Centro de Investigación en Matemáticas, Guanajuato, Mexico.2014.pp.20-30 [Consulta:2019-10-16]. Disponible en: <http://ciimat.repositorioinstitucional.mx/jspui/handle/1008/335>

CAMPETELLA, C. & CERNE, B. *Entornos invisibles de la ciencia y la tecnología, Estación Meteorológica.* [en línea]. Instituto Nacional de Educación Tecnológica. Buenos Aires, 2011. [Consulta: 19 Octubre 2019. Disponible en: http://www.inet.edu.ar/wp-content/uploads/2012/11/C9_Estacion_meteorologicaR.pdf

CAMPOS, V. Modelado mediante Random Forest de las emisiones de autobuses urbanos en funcion de los ciclos cinemáticos (Trabajo de titulación) (Maestría). [en línea] Universidad Politécnica de Madrid, Madrid, España.2017.pp.30-40 [Consulta:2019-10-19]. Disponible en: http://oa.upm.es/45914/1/TFG_VICTOR_PITA_GONZALEZ_CAMPOS.pdf

CARREÑO, S. Aplicaciones en meteorología de modelos funcionales de predicción en componentes principales (Trabajo de titulación) (Maestría).[en línea] Universidad de Granada, Granada, España.2013.pp.42-48[Consulta:2019-10-19]. Disponible en: [http://masteres.ugr.es/moea/pages/tfm-1213/tfm_carrenosheila/!](http://masteres.ugr.es/moea/pages/tfm-1213/tfm_carrenosheila/)

CARRILLO, A. & GARATEJO, O Analisis de datos Funcionales Aplicado en

Electroencefalogramas (Trabajo de titulación) (Maestría) [en línea] Fundación Universitaria Los Libertadores, Bogotá, Colombia. 2016. pp. 31-48 [Consulta: 2019-10-19]. Disponible en: https://pdfs.semanticscholar.org/3092/81dc3c91225aab6257e5dd3c6014e812631f.pdf?_ga=2.33596193.1261649383.1618891657-68700725.1618891657

CASTILLO, S., et al. "Análisis de Datos Funcionales Aplicado al Riesgo Financiero: Un Caso de Estudio en Cooperativas de Ahorro y Crédito Ecuatorianas". *Latin American Journal of Computing*, vol. 15, n°. 1 (2015), (Ecuador) pp. 19-25.

CHÁVEZ, C., et al. "Análisis de componentes principales funcionales en series de tiempo económicas". *GECONTEC: Revista Internacional de Gestión del Conocimiento y la Tecnología* [en línea], 2015, (España) 3(2), pp. 13-25. [Consulta: 20 octubre 2019]. Disponible en: https://rei.iteso.mx/bitstream/handle/11117/2757/Análisis_de_componentes_principales_funcionales_en_series_de_tiempo_económicas_Chávez_Sánchez_DelaCerde-2015%28002%29.pdf?sequence=2.

CALCETERO, S. Análisis estadístico de datos funcionales y su aplicación en el estudio del PM10 en Bogotá, Colombia (Trabajo de titulación) (Maestría). [en línea] Universidad de los Andes, Bogotá, Colombia. 2017. pp. 51-55 [Consulta: 2019-10-20]. Disponible en: <https://repositorio.uniandes.edu.co/flexpaper/handle/1992/34058/u806803.pdf?sequence=1&isAllowed=y#page=1>

CUITIÑO, F., et al. Modelos univariados de series de tiempo para predecir la inflación de corto plazo Modelos univariados de series de tiempo para predecir la inflación de corto plazo [en línea] Universidad de los Andes, Montevideo, Uruguay. 2010. [Consulta: 2019-10-20]. Disponible en: <https://www.bcu.gub.uy/Comunicaciones/Jornadas%20de%20Economía/iees03j3101010.pdf>

ESCABIAS, M. & AGUILERA, A. *Functional data analysis for Statfda* [en línea], United State of America .2008. pp. 1-19. [Consulta: 2019-10-20]. Disponible en: <http://www.statfda.com/media/files/theory.pdf>

ESCUADERO, I., 2016. Modelos funcionales para el tratamiento de datos de Radiación Solar Global [en línea] (Trabajo de titulación) (Maestría). [en línea] Universidad de Granada, Granada, España. 2016. pp. 13-16 [Consulta: 2019-10-21]. Disponible en: http://masteres.ugr.es/moea/pages/curso201516/tfm1516/TFM_Escudero_Villa

ESTRADA, A. Nociones de Profundidad en Análisis de Datos Funcionales [en línea]

Bogotá, Colombia. 2014. pp.4-6 [Consulta: 2019-10-22]. Disponible en: <https://repositorio.uniandes.edu.co/bitstream/handle/1992/16927/u703226.pdf?sequence=1>

NEIRA, L. Predicción de tasas de mortalidad por cáncer colorrectal y por cáncer en España (Trabajo de titulación). (Maestría). [en línea] España. 2016. pp.12-16 [Consulta: 2019-10-22]. Disponible en: http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_614.pdf

GONZÁLEZ, O. "Modelización de curvas de rentabilidad del IBEX 35 basada en Análisis en componentes principales funcional" (Trabajo de titulación) (Maestría). [en línea] Universidad de Granada, Granada, España. 2016. pp.13-16 [Consulta: 2019-10-22]. Disponible en: [https://masteres.ugr.es/moea/pages/curso201516/tfm1516/gonzalezfrutos_tfm/!](https://masteres.ugr.es/moea/pages/curso201516/tfm1516/gonzalezfrutos_tfm/)

LEVITIN, D., et al. "Introduction to functional data analysis". Canadian Psychology [en línea], 2007, (United State of America) 48(3), pp. 135-155. [Consulta: 2019-10-25]. ISSN 07085591. DOI 10.1037/cp2007014. Disponible en: https://www.researchgate.net/publication/232480916_Introduction_to_Functional_Data_Analysis

MILLÁN, L. Outliers de datos funcionales para la detección de caudales anómalos en el sector hidráulico (Trabajo de titulación) (Maestría). [en línea] Universidad Jaume, Castellón de la Plana, España. 2017. pp.19-31 [Consulta: 2019-10-28]. Disponible en: http://repositori.uji.es/xmlui/bitstream/handle/10234/174477/TFM_2017_Millan%20Roures_Laura.pdf?sequence=1&isAllowed=y

NAVARRO, V. Análisis De Datos Funcionales. Implementación Y Aplicaciones (Trabajo de titulación) (Grado). [en línea] Universidad Politécnica de Cataluña, Barcelona, España. 2004. pp.11-27 [Consulta: 2019-11-5]. Disponible en: <https://upcommons.upc.edu/handle/2099.1/13315>

Organización Meteorológica Mundial. [en línea]. OMM, Suiza. 2016 [Consulta: 6 Noviembre 2019]. Disponible en: https://library.wmo.int/doc_num.php?explnum_id=3327_2016

Organización meteorológica mundial. Guía de prácticas climatológicas. Suiza. 2018 [Consulta: 6 Noviembre 2019]. Disponible en: https://library.wmo.int/doc_num.php?explnum_id=10027

PÉREZ, A. Análisis de datos funcionales: una aplicación en componentes principales. (Trabajo de titulación) (Grado). [en línea] Universidad de Cádiz, Puerto Real, España. 2016 pp.21-33 [Consulta: 2019-11-6]. Disponible en:

<https://rodin.uca.es/xmlui/bitstream/handle/10498/18510/TFGPrincipal.pdf?sequence=1&isAllowed=y>

RODRÍGUEZ, R., et al. Meteorología Y Climatología. *Fundación Española para la Ciencia y la Tecnología* [en línea], 2004, (España), pp. 12-17. [Consulta: 7 noviembre 2019]. Disponible en: https://www.academia.edu/38729866/Meteorolog%C3%ADa_y_Climatolog%C3%ADa_Semana_de_la_Ciencia_y_la_Tecnolog%C3%ADa_2004_FUNDACION_ESPA%C3%91OLA_PARA_LA_CIENCIA_Y_LA_TECNOLOG%C3%8DA

SANCHEZ, C. “ MEDICIÓN EN LA PRECISIÓN DE LOS COSTOS DE LA EMPRESA ” (Trabajo de titulación) (Grado). [en línea] Universidad Andres Bello, Concepción, Chile. 2017. pp.19-25 [Consulta: 2019-11-7]. Disponible en: http://repositorio.unab.cl/xmlui/bitstream/handle/ria/3167/a117570_Sanchez_C_Medicion_en_la_precision_2017_Tesis.pdf?sequence=1

SANCHEZ, D. & RODRIGUEZ, M. Estimación Funcional Robusta de la media de la humedad relativa (Trabajo de titulación) (Grado). [en línea] Universidad del Valle, Santiago de Cali, Colombia. 2018. pp.7-12 [Consulta: 2019-11-7]. Disponible en: https://bibliotecadigital.univalle.edu.co/bitstream/handle/10893/17630/CB_0577800.pdf?sequence=1&isAllowed=y

SANTANA, J. & FARFÁN, E. *El Arte de programar en R: un lenguaje para la estadística* [en línea]. Morelos, Mexico: Instituto Mexicano de Tecnología del Agua, 2014. [Consulta: 8 de noviembre 2019]. Disponible en: https://cran.r-project.org/doc/contrib/Santana_El_arte_de_programar_en_R.pdf

SHANG, H.L. *ftsa: An R package for analyzing functional time series*. R Journal [en línea], 2013, (United State of America) 5(1), pp. 64-72. [Consulta: 2019-11-8]. ISSN 20734859. DOI 10.32614/rj-2013-006. Disponible en: <https://journal.r-project.org/archive/2013-1/shang.pdf>

LEVITIN, D., et al. "Introduction to functional data analysis". Canadian Psychology [en línea], 2007, (United State of America) 48(3), pp. 135-155. [Consulta: 2019-11-16]. ISSN 07085591. DOI 10.1037/cp2007014. Disponible en: https://www.researchgate.net/publication/232480916_Introduction_to_Functional_Data_Analysis

SHANG, H. *Functional time series forecasting with dynamic updating: An application of intraday*

particulate matter concentration. *Econometrics and Statistics* [en línea], 2017, (Australia) 1(2), pp. 184-200. ISSN 24523062 . [Consulta:2019-11-16]. DOI 10.1016/j.ecosta.2016.08.004
Disponible en: <https://arxiv.org/pdf/1608.07029.pdf>

SERRANO, V., et al. Análisis estadístico de datos meteorológicos mensuales y diarios para la determinación de variabilidad climática y cambio climático en el distrito metropolitano de Quito (Trabajo de titulación) (Grado). [en línea] Universidad Politécnica Salesiana, Ecuador. 2012. pp.20-21 [Consulta:2019-11-7]. Disponible en: <https://dspace.ups.edu.ec/handle/123456789/8828>

TAGUACUNDO, O. “ Componentes Principales Funcionales De La Radiación Solar Global De La Provincia De Chimborazo, 2014-2017”. (Trabajo de titulación) (Grado). [en línea] Escuela Superior Politécnica De Chimborazo, Riobamba, Ecuador. 2019. pp.20-21 [Consulta:2019-11-17]. Disponible en: <http://dspace.espech.edu.ec/handle/123456789/13600>

ZAHUMENSKÝ, I. "*Guidelines on quality control procedures for data from automatic weather stations*". World Meteorological Organization [en línea], 2004, (Slovakia) (955), pp. 2-6. [Consulta: 18 noviembre 2019]. Disponible en: [http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82TECO_2005/Papers/3\(14\)_Slovakia_2_Zahumensky.pdf](http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82TECO_2005/Papers/3(14)_Slovakia_2_Zahumensky.pdf)

ANEXOS

ANEXO A: Imputación de datos Faltantes

Las librerías a utilizadas

```
library(readxl) # lectura de datos en excel
library(openxlsx) # exportar datos a excel
library(data.table)#manipulacion de data.frame
library(tidyverse) #manipulacion de datos
library(lubridate)#manipulacion de fechas
library(randomForest) #imputacion de datos
library(fda) #analisis de datos funcional
library(fda.usc)# análisis de datos funcional
library(ftsa) #Pronóstico Funcional de series de tiempo
library(rainbow) #Graficas funcionales
library(tidyverse) # manipulacion de dataframe
library(readxl) # lectura de datos en excel
library(openxlsx) # exportar datos a excel
```

Lectura de datos y cantidad de faltantes

```
getwd()
dir()
Alao2014impu <- read_excel ("ATILLO2014.xls")
names (Alao2014impu) <- c ("FECHA LOCAL", "HORA
LOCAL", "VALIDAvg", "X1Avg", "VALIDMax", "X1Max", "VALIDMin", "X1Min")
Alao2014impuu <- Alao2014impu [-c (1:2,5:8)]
Fecha <- Alao2014impu[1]
total<-sum (is.na (Alao2014impuu$X1Avg))
porcentaje <- total/length(n)*100
```

Se revisa las bases de datos y se las valida, identificando cada valor invalido y reemplazando por NA, para que forme parte de datos faltantes, se procede a calcular nuevamente la cantidad de faltantes.

```
for (i in n) {
if (is.na (Alao2014impuu[i,1]) | Alao2014impuu [i,1] == "INVALID")
Alao2014impuu[i,2] <- NA
}
total <- sum (is.na (Alao2014impuu$X1Avg))
porcentaje <- total/length (n)*100
```

```

for (i in n) {
if (is.na (Alao2014impuu[i,2]) | Alao2014impuu [i,2] <= -80 | Alao2014impuu [i,2] <= -60)
Alao2014impuu[i,2] <- NA
}

```

```

total <- sum (is.na (Alao2014impuu$X1Avg))
porcentaje <- total/length (n)*100

```

Con la función rfImpute del paquete randomForest se procede a imputar los datos faltantes hasta el 30% de faltantes, con un total de 300 árboles aleatorios y una iteración de 10

```

relleno <- function(base) {
impuda <- rfImpute(HORA~ ., base,ntrees=300,iter=10)
return(impuda)
}
imputada<-relleno(base=BaseVaciameses)

```

Anexo B: Lectura de datos

Se almacena los datos imputados de todas las estaciones en una sola base, guardándolo en un R.data con el nombre BASEESTACIONES

```

{
Alao2014<-read_xlsx(path = "DATOS/ALAO2014LLENA.xlsx")
Alao2015<-read_xlsx(path = "DATOS/ALAO2015LLENA.xlsx")
Alao2016<-read_xlsx(path = "DATOS/ALAO2016LLENA.xlsx")
Alao2017<-read_xlsx(path = "DATOS/ALAO2017LLENA.xlsx")
Alao2018<-read_xlsx(path = "DATOS/ALAO2018LLENA.xlsx")
Alao2019<-read_xlsx(path = "DATOS/ALAO2019LLENA.xlsx")
ALAO<- rbind(Alao2014,Alao2015,Alao2016,Alao2017,Alao2018,Alao2019)
ALAO <- ALAO %>% mutate(ESTACION = "ALAO")
ESTACIONES <- rbind(ALAO,ATILLO,
CUMANDA,ESPOCH,MATUS,MULTITUD,QUIMIAG,SANJUAN,TIXAN,TUNSHI,URBI)
options(digits=3)
ESTACIONES <- ESTACIONES %>% mutate(AÑO =year(ESTACIONES$FECHA.LOCAL))
BASEESTACIONES<- spread (data = ESTACIONES,key = HORA.LOCAL,value = X1)
dim(BASEESTACIONES)
save(BASEESTACIONES,file = "DATOS/Dato.RData")
}

```

Anexo C: Base Funcional y Suavizado

Se filtra por estación y se prueba con varias bases de Fourier y B-splines. Se escoge la que presente errores mínimos, en este caso una base de 15 B-splines y se realiza el suavizado mediante mínimos cuadrados con la función `smooth.basis`

```
h<-t(BASEESTACIONES[,4:27])
A<-BASEESTACIONES%>%filter(ESTACION == "ALAO")
TALAO<-as.matrix(t(A[,4:27]))
argm<-1:nrow(TALAO)
baseALAO=create.bspline.basis(rangeval=c(1,24), nbasis=15)
ALAO<- smooth.basis(argvals=argm, y=TALAO,fdParobj=baseALAO)$fd
```

Anexo D: Cálculo de datos atípicos

Luego se calcula los datos atípicos con la función `outliers.depth.thim` de todas las Estación , en este caso de Alao

```
AALAO <- outliers.depth.trim(fdataobj = ALAO, nb = 200,smo = 0.05, quan = 0.5, trim = 0.01,
dfunc=depth.mode, ns = 0.01)
oALAO <- AALAO$outliers
sALAO <- strsplit(oALAO, split = "rep" )
UALAO <- as.integer(unlist(sALAO))
ALAOO<- na.omit(UALAO)
ALAOO
```

Anexo E: Análisis Descriptivo Funcional

Mediante la función `mean.fd` de la librería `fda.usc` se calcula las medias funcionales de cada una de las estaciones , de igual forma con la función `sd.fd` se calcula la desviación funcional.

```
plot(c(0, 24), c(0,24), xlab="Horas", ylab="Temperatura Ambiental",
type='n',main="Media Funcional de ALAO")
lines(mean.fd(ALAO2014),col="red",lwd=2);
lines(mean.fd(ALAO2015),col="green",lwd=2);
lines(mean.fd(ALAO2016),col="blue",lwd=2);
lines(mean.fd(ALAO2017),col="cyan",lwd=2);
lines(mean.fd(ALAO2018),col="black",lwd=2);
lines(mean.fd(ALAO2019),col="purple",lwd=2);
lines(mean.fd(ALAO),col="yellow",lwd=2);
```

```

legend(x = 0.5,y = 23,legend = c ( "2014","2015", "2016",
"2017","2018","2019", "mediaG"),
fill = c("red", "green", "blue", "cyan", "black","purple","yellow"),
cex =0.5, text.font = 3, bg = 'white')
plot(c(0, 24), c(0, 4), xlab="Horas", ylab="Temperatura Ambiental",
type='n',main="Desviación Funcional ALAO",ccex.main=1)
lines(sd.fd(ALAO2014),col="red",lwd=2);
lines(sd.fd(ALAO2015),col="green",lwd=2);
lines(sd.fd(ALAO2016),col="blue",lwd=2);
lines(sd.fd(ALAO2017),col="cyan",lwd=2);
lines(sd.fd(ALAO2018),col="black",lwd=2);
lines(sd.fd(ALAO2019),col="purple",lwd=2);
legend(x = 0.5,y = 4,legend = c ( "2014","2015", "2016",
"2017","2018","2019"),
fill = c("red", "green", "blue", "cyan", "black","purple"),
cex =0.55, text.font = 4.5, bg = 'white')

```

Para calcular la variabilidad de la temperatura ambiental, se realiza un gráfico de curvas de contornos y de niveles, mediante los comandos `contour` y `filled.contour` para que resalte tipos con mayor énfasis las temperaturas bajas y altas.

```

ALAO2014m<-ALAO2014m[1:24,1:30]
contour(x = 1:24, y = 1:30, z =ALAO2014m ,xlab = "Horas",ylab = "",
main = "Curvas de Contornos, ALAO 2014 ", col = rainbow(9))
filled.contour(x=1:24, y=1:30, z=ALAO2014m,
main="Curvas de Contornos y Niveles, ALAO 2014",
plot.axes=contour(1:24, 1:30, ALAO2014m, add=TRUE))
ALAO2015m<-ALAO2015m[1:24,1:30]
contour(x = 1:24, y = 1:30, z =ALAO2015m ,xlab = "Horas",ylab = "",
main = "Curvas de Contornos, ALAO 2015 ", col = rainbow(9))
filled.contour(x=1:24, y=1:30, z=ALAO2015m,
main="Curvas de Contornos y Niveles, ALAO 2015",
plot.axes=contour(1:24, 1:30, ALAO2015m, add=TRUE))

```

Anexo F: Modelado y Pronóstico

Transformación de las bases de datos a una serie de tiempo funcional

```

x1 <- 1:24
y1 <-(data.frame(Entrenamiento))

```

```

ftsE <- rainbow::fts(x1,y1)
colnames(ftsE$y) <- seq_along(colnames(ftsE$y))
x2 <- 1:24
y2 <-(data.frame(Validacion))
ftsV <- rainbow::fts(x2,y2)
colnames(ftsV$y) <- seq_along(colnames(ftsV$y))
par(mfrow = c(1,2))
plot(ftsE,main="TEMPERATURA ALAO 70% ",xlab="HORAS",ylab="TEMPERATURA")
plot(ftsV,main="TEMPERATURA ALAO 30% ",xlab="HORAS",ylab="TEMPERATURA")

```

Ajuste del modelo por regresión de componentes principales funcionales, por medio de la función ftsm de la librería ftsa

```

modelo1<-ftsm(ftsE,orde=2,method = "classical") # modelo con 2 componentes principales
modelo2<-ftsm(ftsE,orde=3,method = "classical") # modelo con 3 componentes principales
modelo3<-ftsm(ftsE,orde=4,method = "classical") # modelo con 4 componentes principales

```

Variabilidad explicada por cada una de los compontes

```

varexplcompm1<-sum(modelo1$varprop)
varexplcompm2<-sum(modelo2$varprop)
varexplcompm3<-sum(modelo3$varprop)

```

Varia<-as.data.frame(c(varexplcompm1,varexplcompm2,varexplcompm3))

Se ajusta el modelo con 4 componentes principales, siendo la cantidad de componentes utilizadas en el pronóstico.

```

promodelo1<-forecast.ftsm(modelo3,h = 7 ,method = "arima")
promodelo2<-farforecast(ftsE,h = 7,Dmax_value = 4,Pmax_value =6,var_type = "const",PI = TRUE,level = 90)

```

Gráficas de Comparación de los modelos de Predicción

```

{
lunes<-mutate(as.data.frame(PRONOSTICOM1[,1]),horas=as.matrix(0:23))
martes<-mutate(as.data.frame(PRONOSTICOM1[,2]),horas=as.matrix(24:47))
miercoles<-mutate(as.data.frame(PRONOSTICOM1[,3]),horas=as.matrix(48:71))
jueves<-mutate(as.data.frame(PRONOSTICOM1[,4]),horas=as.matrix(72:95))
viernes<-mutate(as.data.frame(PRONOSTICOM1[,5]),horas=as.matrix(96:119))
sabado<-mutate(as.data.frame(PRONOSTICOM1[,6]),horas=as.matrix(120:143))
domingo<-mutate(as.data.frame(PRONOSTICOM1[,7]),horas=as.matrix(144:167))
names(lunes)<-c("x","hora")
names(martes)<-c("x","hora")
names(miercoles)<-c("x","hora")
names(jueves)<-c("x","hora")

```

```

names(viernes)<-c("x","hora")
names(sabado)<-c("x","hora")
names(domingo)<-c("x","hora")
Pro1<-rbind(lunes,martes,miercoles,jueves,viernes,sabado,domingo)
lunes<-mutate(as.data.frame(PRONOSTICOM2[,1]),horas=as.matrix(0:23))
martes<-mutate(as.data.frame(PRONOSTICOM2[,2]),horas=as.matrix(24:47))
miercoles<-mutate(as.data.frame(PRONOSTICOM2[,3]),horas=as.matrix(48:71))
jueves<-mutate(as.data.frame(PRONOSTICOM2[,4]),horas=as.matrix(72:95))
viernes<-mutate(as.data.frame(PRONOSTICOM2[,5]),horas=as.matrix(96:119))
sabado<-mutate(as.data.frame(PRONOSTICOM2[,6]),horas=as.matrix(120:143))
domingo<-mutate(as.data.frame(PRONOSTICOM2[,7]),horas=as.matrix(144:167))
names(lunes)<-c("x","hora")
names(martes)<-c("x","hora")
names(miercoles)<-c("x","hora")
names(jueves)<-c("x","hora")
names(viernes)<-c("x","hora")
names(sabado)<-c("x","hora")
names(domingo)<-c("x","hora")
Pro2<-rbind(lunes,martes,miercoles,jueves,viernes,sabado,domingo)
}
plot(x = DValidacion$hora,y=DValidacion$x,
type="l",xlab="Horas",ylab="Temperatura",main="TEMPERATURA AMBIENTAL
ALAO",col="blue")
lines(Pro1$x,type="l",col="red")
lines(Pro2$x,type="l",col="purple")
legend(x = 10,y = 5,legend = c ("Pronóstico Univariante","Pronóstico Multivariante"),fill =
c("red","purple"),cex =0.55, text.font = 4.5, bg = 'white')

```

Anexo G: Errores de Pronóstico

Medidas dependientes de la Escala

$$MSE1 = (\text{mean}(((DValidacion\$x)-(Pro1\$x))^2))$$

$$MSE2 = (\text{mean}(((DValidacion\$x)-(Pro2\$x))^2))$$

$$RMSE1 = \text{sqrt}(\text{mean}(((DValidacion\$x)-(Pro1\$x))^2))$$

$$RMSE2 = \text{sqrt}(\text{mean}(((DValidacion\$x)-(Pro2\$x))^2))$$

$$MAE1 = (\text{mean}(\text{abs}((DValidacion\$x)-(Pro1\$x))))$$

$$MAE2 = (\text{mean}(\text{abs}((DValidacion\$x)-(Pro2\$x))))$$

MDAE1 = (median(abs((DValidacion\$x)-(Pro1\$x))))

MDAE2 = (median(abs((DValidacion\$x)-(Pro2\$x))))

Medidas basadas en porcentaje

MAPE1 = mean(abs(((DValidacion\$x)-(Pro1\$x))/(DValidacion\$x)))

MAPE2 = mean(abs(((DValidacion\$x)-(Pro2\$x))/(DValidacion\$x)))

MDAPE1 = median(abs(((DValidacion\$x)-(Pro1\$x))/(DValidacion\$x)))

MDAPE2 = median(abs(((DValidacion\$x)-(Pro2\$x))/(DValidacion\$x)))

RMSPE1 = sqrt(mean(((DValidacion\$x)-(Pro1\$x))/(DValidacion\$x))^2)

RMSPE2 = sqrt(mean(((DValidacion\$x)-(Pro2\$x))/(DValidacion\$x))^2)

RMDSPE1 = sqrt(median(((DValidacion\$x)-(Pro1\$x))/(DValidacion\$x))^2)

RMDSPE2 = sqrt(median(((DValidacion\$x)-(Pro2\$x))/(DValidacion\$x))^2)

Coefficiente U de Theil(Medida de Presición del modelo)

U1<-(sqrt(mean(((DValidacion\$x)-(Pro1\$x))^2)))/((sqrt(mean((Pro1\$x)^2)))+(sqrt(mean((DValidacion\$x)^2))))

U2<-(sqrt(mean(((DValidacion\$x)-(Pro1\$x))^2)))/((sqrt(mean((Pro2\$x)^2)))+(sqrt(mean((DValidacion\$x)^2))))

Diebold-Mariano test

Mt1<-dm.test((DValidacion\$x)-(Pro1\$x),(DValidacion\$x)-(Pro2\$x),alternative ="greater")

hipo1<-Mt1\$p.value

e1<-rbind(RMSE1 ,RMSE2)

e2<-rbind(MSE1 ,MSE2)

e3<-rbind(MAE1 ,MAE2)

e4<-rbind(MDAE1 ,MDAE2)

e5<-rbind(MAPE1 ,MAPE2)

e6<-rbind(MDAPE1,MDAPE2)

e7<-rbind(RMSPE1,RMSPE2)

e8<-rbind(RMDSPE1 ,RMDSPE2)

e9<-rbind(U1 ,U2)

Exportar los datos de los pronósticos

PRONOSTICOM1<-promodelo1\$mean

PRONOSTICOM1<-PRONOSTICOM1\$y

PRONOSTICOM2<-promodelo2\$point_fore

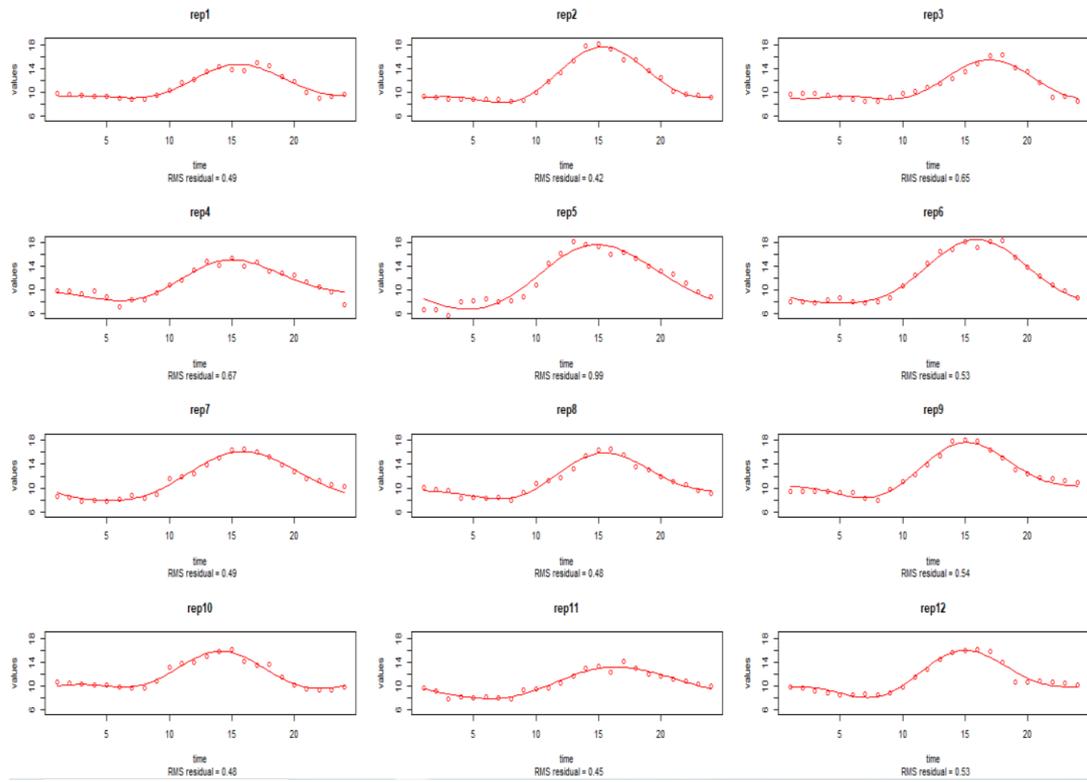
PRONOSTICOM2<-PRONOSTICOM2\$y

write.xlsx(PRONOSTICOM1, file="PRONOSALAOUnivariado.xlsx")

write.xlsx(PRONOSTICOM2, file="PRONOSALAOMultivariado.xlsx")

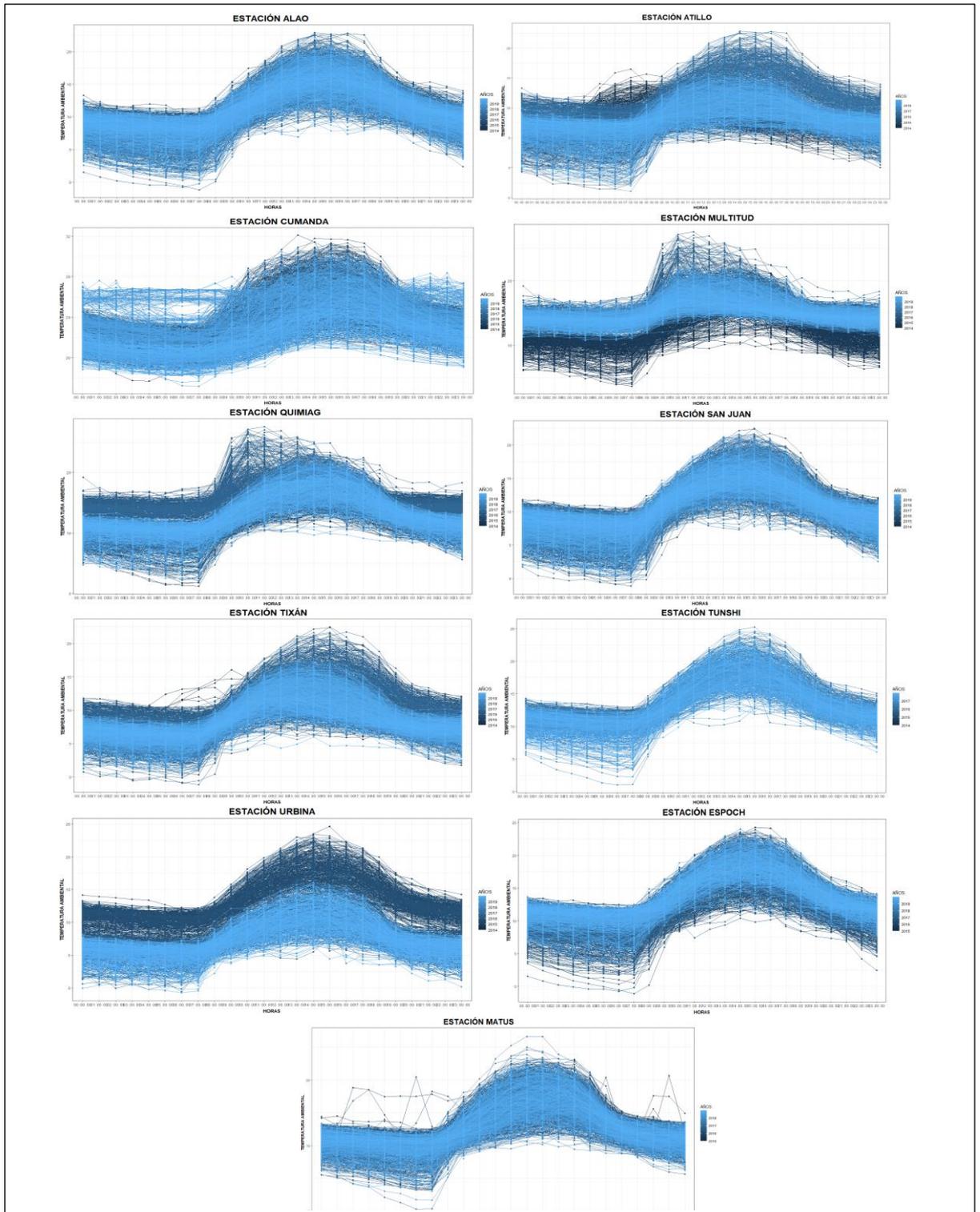
Anexo H: Errores al escoger la base

17 Bases B-splines



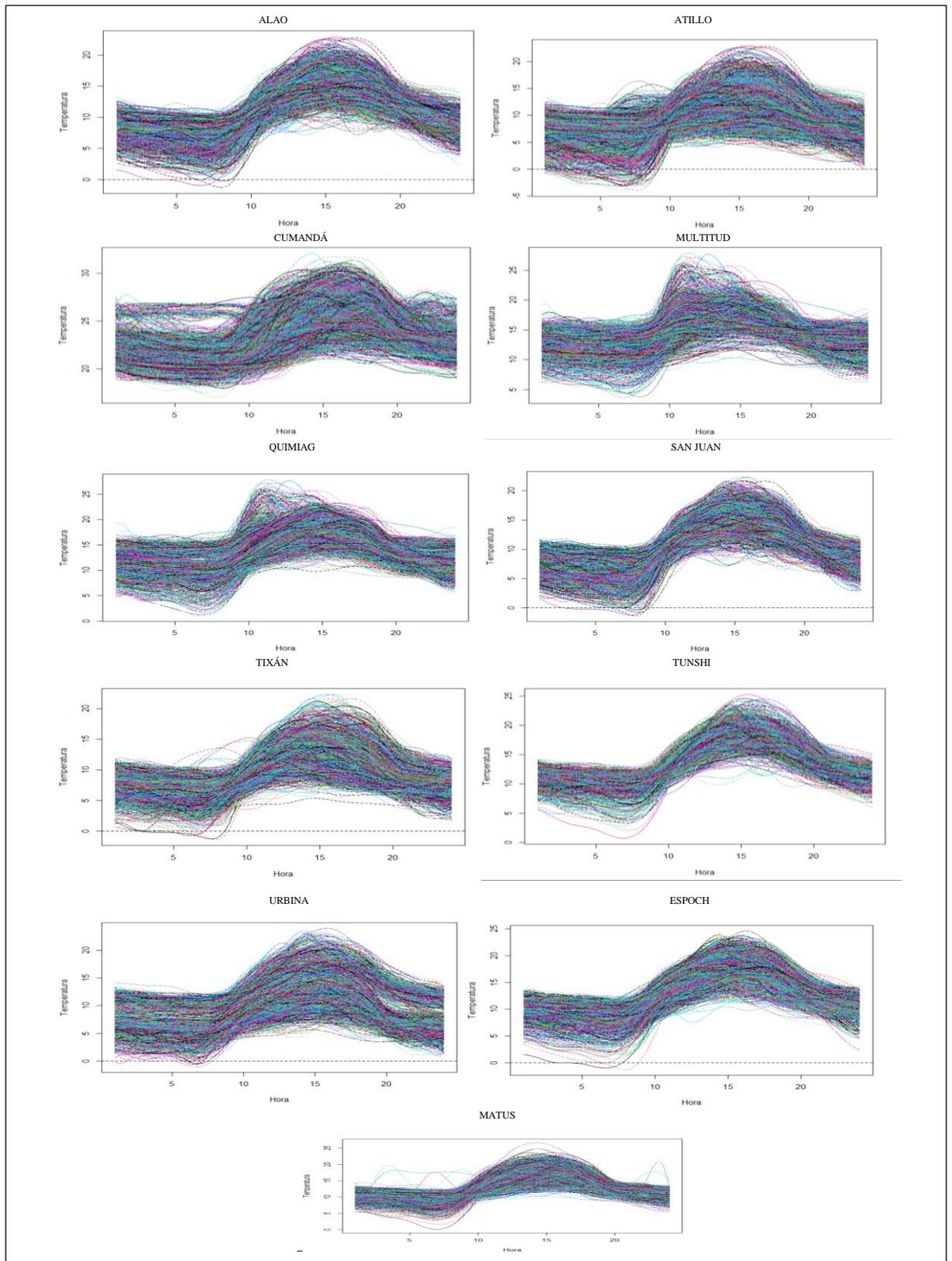
Realizado por: Mariño Gabriela, 2021.

Anexo I: Estaciones graficadas por hora y años



Realizado por: Mariño Gabriela, 2021.

Anexo J: Suavizado de los datos de todas las Estaciones



Realizado por: Mariño Gabriela, 2021.



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

**DIRECCIÓN DE BIBLIOTECAS Y RECURSOS DEL APRENDIZAJE
UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y DOCUMENTAL**

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 29 / 07 / 2021

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: <i>Roxana Gabriela Mariño Cajo</i>
INFORMACIÓN INSTITUCIONAL
Facultad: <i>Ciencias</i>
Carrera: <i>Estadística</i>
Título a optar: <i>Ingeniera en Estadística Informática</i>
f. Analista de Biblioteca responsable: <i>Ing. Leonardo Medina Ñuste MSc.</i>

LEONARDO
FABIO MEDINA
NUSTE

Firmado digitalmente por
LEONARDO FABIO MEDINA
NUSTE
Fecha: 2021.07.29 11:30:50
-05'00'



1466-DBRA-UTP-2021