



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE CIENCIAS

CARRERA DE INGENIERÍA EN ESTADÍSTICA INFORMÁTICA

**“COMPONENTES PRINCIPALES FUNCIONALES DE LA
RADIACIÓN SOLAR GLOBAL DE LA PROVINCIA DE
CHIMBORAZO, 2014-2017”**

Trabajo de Titulación

Tipo: Proyecto de investigación

Presentado para obtener el grado académico de:

INGENIERO EN ESTADÍSTICA INFORMÁTICA

AUTOR: OSCAR EMANUEL TAGUACUNDO ORTA

DIRECTORA: ING. AMALIA ISABEL ESCUDERO VILLA

Riobamba – Ecuador

2019

© 2019, Oscar Emanuel Taguacundo Orta

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

Yo, **Oscar Emanuel Taguacundo Orta**, declaro que el presente trabajo de titulación 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 titulación; El patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo

Riobamba, 15 de noviembre del 2019.

Oscar Emanuel Taguacundo Orta.

CI: 060474228-8

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

El Tribunal del Trabajo de Titulación, certifica que: El trabajo de investigación: "**Componentes Principales Funcionales de la radiación solar global de la provincia de Chimborazo, 2014-2017**", de responsabilidad del señor Oscar Emanuel Taguacundo Orta, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, quedando autorizada su presentación.

	FIRMA	FECHA
Ing. Amalia Isabel Escudero Villa DIRECTORA DEL TRABAJO DE TITULACION		2019-11-15
Mat. Luis Marcelo Cortéz Borilla MIEMBRO DE TRIBUNAL		2019-11-15
Dra. Silvia Mariana Haro Rivera MIEMBRO DE TRIBUNAL		2019-11-15

DEDICATORIA

Dedico este trabajo a cada una de las personas quienes estuvieron en los buenos y malos momentos de mi vida en especial es un logro alcanzado junto a mi Esposa, Padres y Suegros, quienes son los pilares de mi vida he impartieron en todo momento sus consejos y palabras de aliento, esto fue un impulso y apoyo incondicional, siendo mi principal motivación mi amada hija, quien es la razón de mi superación, día a día.

AGRADECIMIENTO

Mi profundo agradecimiento:

Primeramente, a Dios por brindar el privilegio de vivir y conocer a personas maravillosas en el ámbito familiar y curricular.

A mis padres que supieron guiarme por el camino del bien.

A mis suegros y esposa quienes supieron confiarme y brindar apoyo en todo momento.

A mis maestros, en especial a la Ing. Isabel Escudero y Mat. Marcelo Cortéz quienes guiaron mis pasos en este trabajo, además a la Dra. Silvia Haro, Ing. Alexandra Marcatoma, Ing. Héctor Mullo, Mat. Alberto Vilañez, Dr. Jorge Congacha, Ing. Milton López, Ing. Carlos Rosero, Ing. Lourdes Paredes, Ing. Pablo Flores, Ing. Carola Flores y al Dr. Luis Vera, quienes compartieron día a día su ejemplo, sabiduría y amistad.

A mis Amigos y compañeros de clase con quienes compartimos momentos inéditos y únicos durante los diferentes periodos académicos.

Al Grupo de Energías Alternativas y Ambiente (GEAA) por motivar la investigación y facilitar toda la información meteorológica utilizada en este estudio.

ÍNDICE GENERAL

RESUMEN	xv
THESIS ABSTRACT	xvi
INTRODUCCIÓN	1
CAPÍTULO I.....	3
1 MARCO REFERENCIAL	3
1.1 Antecedentes	3
1.2 Planteamiento del problema	6
1.2.1 <i>Formulación del problema</i>	6
1.3 Justificación	7
1.3.1 <i>Justificación teórica</i>	7
1.3.2 <i>Justificación Aplicativa</i>	7
1.4 Objetivos	8
1.4.1 <i>Objetivos Generales</i>	8
1.4.2 <i>Objetivos Específicos</i>	8
CAPÍTULO II.....	9
2 MARCO TEÓRICO REFERENCIAL	9
2.1 Meteorología	9
2.2 Efectos de la atmósfera sobre la radiación.	9
2.3 Radiación solar	9
2.4 Tipos de Radiación Solar	10
2.4.1 <i>Radiación solar directa</i>	10
2.4.2 <i>Radiación solar difusa</i>	10
2.4.3 <i>Radiación solar reflejada</i>	11
2.4.4 <i>Radiación solar global</i>	11
2.4.5 <i>Radiación solar extraterrestre</i>	11
2.4.6 <i>Medida de la radiación global y difusa</i>	12
2.4.7 <i>Unidades de medida de la radiación solar global</i>	12

2.5	Aproximación mediante funciones	13
2.5.1	<i>Bases Fourier</i>	13
2.5.2	<i>Bases B-spline.....</i>	14
2.5.3	<i>Bases Wavelets.....</i>	15
2.5.4	<i>Base exponencial y de potencia.....</i>	15
2.5.5	<i>Bases constantes</i>	16
2.5.6	<i>Bases polinomiales</i>	16
2.6	Correlación cruzada.....	17
2.6.1	<i>Suavizado de Kernel</i>	17
2.6.2	<i>Criterio de Validación</i>	17
2.7	Análisis de datos funcionales	18
2.7.1	<i>Espacios de Hilbert.....</i>	18
2.7.2	<i>Espacio con Producto escalar</i>	18
2.8	Dato funcional.....	19
2.9	Datos Funcionales atípicos	19
2.10	Profundidad funcional	19
2.10.1	<i>Profundidad de Fraiman y Muniz (FM).....</i>	20
2.10.2	<i>Profundidad Modal (MD).....</i>	20
2.10.3	<i>Profundidad de proyección aleatoria (RPD).....</i>	20
2.11	Análisis Descriptivo Funcional	21
2.11.1	<i>Media funcional.....</i>	21
2.11.2	<i>Varianza funcional.....</i>	21
2.11.3	<i>Covarianza funcional</i>	21
2.11.4	<i>Correlación funcional</i>	22
2.12	Análisis de Componentes Principales Funcionales	22
2.13	Análisis Clúster Funcional.....	24
2.13.1	<i>Algoritmo de alineación de k-medias</i>	24
2.14	Software Estadístico R	25
CAPÍTULO III.....		26

3	METODOLOGÍA	26
3.1	Tipo y Diseño de Investigación	26
3.1.1	Área de estudio	26
3.1.2	Población en estudio.....	26
3.1.3	Recolección de información	27
3.1.4	Operacionalización de variables.....	28
3.1.5	Análisis de Datos	28
3.1.6	Alcance de la investigación	28
	CAPÍTULO IV	29
4	RESULTADOS Y DISCUSIÓN	29
4.1	Construcción de la base de datos.....	29
4.2	Suavizados B-spline y Fourier	29
4.3	Suavizado de curvas y detección de curvas atípicas.....	34
4.4	Análisis exploratorio de datos.....	39
4.4.1	Análisis estadístico descriptivo funcional de la radiación solar global	39
4.4.2	Análisis de Varianza funcional	47
4.4.3	Análisis de Componentes principales funcionales.....	49
4.4.4	Análisis clúster por estación y Año de estudio	66
4.4.5	Análisis de Componentes principales por estación	77
4.4.6	Análisis Clúster Funcional por Estación	78
4.4.7	Análisis de componentes principales y Clúster funcional general por Estación.....	79
	CONCLUSIONES.....	80
	RECOMENDACIONES.....	82
	GLOSARIO	
	BIBLIOGRAFÍA	
	ANEXOS	

ÍNDICE DE TABLAS

Tabla 1-3: Operacionalización de variable.....	28
Tabla 2-4: Radiación Solar Global por estación y año de estudio	29
Tabla 3-4: Valores propios de la estación Alao.....	49
Tabla 4-4: Funciones propias de la estación Alao 2014, 2015, 2016, 2017.....	49
Tabla 5-4: Variabilidad explicada por cada componente Alao por año.....	50
Tabla 6-4: Valores propios de la estación Cumanda	51
Tabla 7-4: Funciones propias de la estación Cumanda 2014, 2015	51
Tabla 8-4: Variabilidad explicada por componente Cumanda por año.....	51
Tabla 9-4: Valores propios de la estación Atillo	52
Tabla 10-4: Funciones propias de la estación Atillo 2015, 2016, 2017	52
Tabla 11-4: Variabilidad explicada por cada función propia, Atillo	53
Tabla 12-4: Valores propios de la estación San Juan	54
Tabla 13-4: Funciones propias de la estación San Juan 2014, 2016, 2017	54
Tabla 14-4: Variabilidad explicada por Componentes, San Juan	54
Tabla 15-4: Valores propios de la estación ESPOCH	55
Tabla 16-4: Funciones propias de la estación ESPOCH 2015, 2016, 2017	55
Tabla 17-4: Variabilidad explicada por Componentes, ESPOCH	56
Tabla 18-4: Valores propios de la estación Tixan	57
Tabla 19-4: Funciones propias de la estación Tixan 2014, 2015, 2016, 2017	57
Tabla 20-4: Variabilidad explicada por Componentes, Tixan	57
Tabla 21-4: Valores propios de la estación Matus	58
Tabla 22-4: Funciones propias de la estación Matus 2015,2017	58
Tabla 23-4: Variabilidad explicada por Componentes, Matus	59
Tabla 24-4: Valores propios de la estación Tunshi	60
Tabla 25-4: Funciones propias de la estación Tunshi 2014, 2015, 2016	60
Tabla 26-4: Variabilidad explicada por Componentes, Tunshi	60
Tabla 27-4: Valores propios de la estación Multitud.....	61
Tabla 28-4: Funciones propias de la estación multitud 2014, 2015, 2016.....	61
Tabla 29-4: Variabilidad explicada por Componentes, Multitud	62
Tabla 30-4: Valores propios de la estación Urbina	63
Tabla 31-4: Funciones propias de la estación Urbina 2014, 2016, 2017	63
Tabla 32-4: Variabilidad explicada por Componentes, Urbina	63
Tabla 33-4: Valores propios de la estación Quimiag.....	64
Tabla 34-4: Funciones propias de la estación Quimiag 2014, 2015, 2016, 2017.....	65
Tabla 35-4: Variabilidad explicada por Componentes, Quimiag.....	65

ÍNDICE DE FIGURAS

Figura 1-2: Ingreso de Radiación solar a la superficie terrestre	11
Figura 2-2: Piranómetro.....	12
Figura 3-3: Provincia de Chimborazo	26

ÍNDICE DE GRÁFICO

Gráfico 1-3: Ubicación de las estaciones meteorológicas	27
Gráfico 3-4: Residuos por número de bases, B-spline y Fourier	32
Gráfico 4-4: Residuos, B-spline y Fourier, con bases de 13 al 23.	33
Gráfico 5-4: Residuos, B-spline y Fourier, con bases del 5 al 23	33
Gráfico 6-4: Residuo promedio por número de bases.	34
Gráfico 7-4: Numero de bases B-spline óptimos.....	35
Gráfico 8-4: Curvas suavizadas de 11 estaciones por año con sus respectivas curvas atípicas..	37
Gráfico 9-4: Curvas suavizadas de 11 estaciones por año con sus respectivas curvas atípicas..	38
Gráfico 10-4: Media y Desviación Funcional por año, Alao	39
Gráfico 11-4: Media y desviación funcional, Cumanda	40
Gráfico 12-4: Media y desviación funcional por año, Atillo	40
Gráfico 13-4: Media y desviación funcional por año, San Juan	41
Gráfico 14-4: Media y desviación funcional por año, ESPOCH	42
Gráfico 15-4: Media y desviación funcional por año, Tixan	42
Gráfico 16-4: Media y desviación funcional por año, Matus	43
Gráfico 17-4: Media y desviación funcional por año, Tunshi	44
Gráfico 18-4: Media y desviación funcional por año, Multitud	44
Gráfico 19-4: Media y desviación funcional por año, Urbina	45
Gráfico 20- 4: Media y desviación funcional por año, Quimiag.....	46
Gráfico 21-4: Variación de la rsg según estación.....	47
Gráfico 22-4: Raster de superficie por estación	48
Gráfico 23-4: Primera y segunda componente principal funcional de rsg, Alao	50
Gráfico 24-4: Primera y segunda componente principal funcional de rsg, Cumanda.....	52
Gráfico 25-4: Primera y segunda componente principal funcional de rsg, Atillo.....	53
Gráfico 26-4: Primera y segunda componente principal funcional de rsg, San Juan.....	55
Gráfico 27-4: Primera y segunda componente principal funcional de rsg, ESPOCH.....	56
Gráfico 28-4: Primera y segunda componente principal funcional de rsg, Tixan.....	58
Gráfico 29-4: Primera y segunda componente principal funcional de rsg, Matus	59
Gráfico 30-4: Primera y segunda componente principal funcional de rsg, Tunshi	61
Gráfico 31-4: Primera y segunda componente principal funcional de rsg, Multitud	62
Gráfico 32-4: Primera y segunda componente principal funcional de rsg, Urbina	64
Gráfico 33-4: Primera y segunda componente principal funcional de rsg, Quimiag	65
Gráfico 34-4: Análisis clúster funcional de la estación Alao	66
Gráfico 35-4: Análisis clúster funcional de la estación Cumanda	67

Gráfico 36-4: Análisis clúster funcional de la estación Atillo	68
Gráfico 37-4: Análisis clúster funcional de la estación San Juan	69
Gráfico 38-4: Análisis clúster funcional de la estación ESPOCH	70
Gráfico 39-4: Análisis clúster funcional de la estación Tixan	71
Gráfico 40-4: Análisis clúster funcional de la estación Matus	72
Gráfico 41-4: Análisis clúster funcional de la estación Tunshi	73
Gráfico 42-4: Análisis clúster funcional de la estación Multitud	74
Gráfico 43-4: Análisis clúster funcional de la estación Urbina	75
Gráfico 44-4: Análisis clúster funcional de la estación Quimiag	76
Gráfico 45-4: Componentes principales funcionales por estación y año de estudio	77
Gráfico 46-4: Análisis clúster funcional por estación y año de estudio	78
Gráfico 47-4: a) Gráfico de la primera y segunda componentes principal funcional; b) Análisis clúster funcional	79

ÍNDICE DE ANEXOS

Anexo A: Lista de Librerías de R utilizadas

Anexo B: Importación datos a Rstudio

Anexo C: Limitación de las horas de estudio

Anexo D: Suavizado de curvas y optimización de número de bases B-spline y Fourier

Anexo E: Detección de curvas atípicas

Anexo F: Análisis de estadísticas descriptivas funcionales

Anexo G: Análisis de componentes principales funcionales

Anexo H: Análisis clúster

RESUMEN

Esta investigación tiene la finalidad de identificar patrones de comportamiento de la radiación solar global (rsg) en la provincia de Chimborazo, del 2014 al 2017 a través del análisis estadístico de datos funcionales, recolectados por el Grupo de Energías Alternativas y Ambiente de la Escuela Superior Politécnica de Chimborazo en las estaciones de: Alao, Atillo, Cumanda, ESPOCH, Matus, Multitud, Quimiag, San Juan, Tixan, Tunshi y Urbina. El estudio de variables meteorológicas toma relevancia al buscar posibles fuentes de energía alternativa y renovable, por tanto, surge la necesidad de conocer las características de cada una de estas, entre las cuales está la rsg, considerando como dato funcional la intensidad de radiación recibida durante las 12 horas (luz) del día, trabajando así con 365 datos funcionales por año. Mediante correlación cruzada de la función *min.basis* de R se identificó como óptimo el suavizado B-spline cúbico con 15 bases y un error cuadrático medio de $4.273653e-6$. Se identificó atípicos: concentrados en el 2015 para Alao y Atillo, en 2016 Tunshi y Multitud; con comportamientos extremos en Tixan 2014 y 2017, y ninguno en San Juan 2015, mismas que fueron retiradas para el análisis. Se encontró curvas medias de rsg que fluctúan desde 400 W/m^2 (Multitud y Cumanda) hasta 800 W/m^2 (ESPOCH). Se graficó las curvas de nivel y las imágenes superficiales raster en las cuales se observó mayor variabilidad aproximadamente desde las 10h00 hasta las 15h00. Las dos primeras componentes principales funcionales en general explican una variabilidad total mayor que el 80% para el 2014, 2015 y 2016 mientras que para el 2017 del 74%. Y el análisis clúster clasificó comportamientos de rsg similares en tres grupos: Alao, Cumanda, Atillo, San Juan, Urbina y Quimiag; ESPOCH, Tixan, Matus y Tunshi, y por último Multitud.

Palabras clave: <ESTADÍSTICA>, <ANÁLISIS DE DATOS FUNCIONALES>, <COMPONENTES PRINCIPALES FUNCIONALES>, <CLÚSTER FUNCIONAL>, <METEOROLOGÍA>, <RADIACIÓN SOLAR GLOBAL>.

ESPOCH - DBRAI
PROCESOS TÉCNICOS Y ANÁLISIS
BIBLIOGRÁFICO Y DOCUMENTAL



15 OCT 2013

REVISIÓN DE RESPONDER Y BIBLIOGRAFÍA

Por: le Hora: 09:49

THESIS ABSTRACT

This research aims to identify patterns of behaviour of global solar radiation (GSR) in the province of Chimborazo, from 2014 to 2017 through the statistical analysis of functional data, collected by the Alternative Energies and Environment Group of the "Escuela Superior Politécnica of Chimborazo" in the stations of: Alao, Atillo, Cumandá, ESPOCH, Matus, Multitud, Quimiag, San Juan, Tixan, Tunshi and Urbina. The study of meteorological variables becomes relevant when looking for possible sources of alternative and renewable energy, therefore, there is a need to know the characteristics of each of these, among which is the GSR, considering as functional data the intensity of radiation received during the 12 hours' daylight thus working with 365 functional data per year. Through cross-correlation of the *min.basis* function of R, the smoothing cubic B-spline was identified as optimal with 15 bases and an average square error of $4.273653e-6$. Atypical ones were identified: concentrated in 2015 for Alao and Atillo, in 2016 Tunshi and Multitud; with extreme behaviours in Tixan 2014 and 2017, and none in San Juan 2015, which were withdrawn for analysis. We found average GSR curves that range from 400 W / m² (Multitud and Cumandá) to 800 W / m² (ESPOCH). The level curves and raster surface images were plotted in which greater variability was observed from approximately 10 am to 3 pm. The first two main functional components, in general, explain a total variability greater than 80% for 2014, 2015 and 2016 while for 2017 of 74%. And the cluster analysis classified similar GSR behaviours into three groups: Alao, Cumandá, Atillo, San Juan, Urbina and Quimiag; ESPOCH, Tixan, Matus and Tunshi, and finally Multitud.

Key words: <STATISTICS>, <FUNCTIONAL DATA ANALYSIS>, <MAIN FUNCTIONAL COMPONENTS>, <FUNCTIONAL CLUSTER>, <METEOROLOGY>, <GLOBAL SOLAR RADIATION>.



INTRODUCCIÓN

Desde la aparición del universo y el sistema solar, la tierra ofrece diversas fuentes naturales, que pueden ser utilizadas como alternativas que coadyuven a la disminución de la contaminación ambiental. El sol provee energía lumínica y calórica a cada uno de los planetas que giran alrededor de ella, cuyos rayos solares emitidos son considerados como una fuente natural que es necesario aprovecharla en beneficio del hombre y el medio en el que vivimos, a pesar de contar con estas bendiciones el hombre prefirió el avance eléctrico e industrial, el cual ha deteriorado la capa de ozono y contaminado el medio ambiente, dando lugar a que aumente el flujo de radiación en toda la superficie terrestre afectando a todos los seres existentes en la tierra. Por tal motivo se intensifican los estudios de la evolución y el comportamiento de la radiación solar.

La estadística como ciencia proporciona variedad de técnicas y métodos que conjuntamente con los avances informáticos contribuyen de gran manera a la recolección y almacenamiento de información, dando lugar al almacenamiento y proceso de una gran cantidad de datos, Data Meaning (DM) profundiza el análisis de información con el objetivo de encontrar patrones que permitan simplificar la información y optimicen la toma de decisiones en diferentes áreas como: física, química, agronomía, medicina, meteorología, etc., con el menor error posible.

El Análisis Estadístico con Datos Funcionales (AEDF) nace por la necesidad de buscar posibles soluciones que no se han podido resolver por medio de métodos convencionales debido a la extensa cantidad de datos que son almacenados en la actualidad. Los datos funcionales es un conjunto de curvas $\{x_1(t), x_2(t), \dots, x_n(t)\}$ con $t \in T$, cada una de ellas son observadas en distintos instantes o puntos de argumento t de la forma $x_i = x_i(t_{i0}), x_i(t_{i1}), \dots, x_i(t_{im_i})'$, para el conjunto de instantes t_{i0}, \dots, t_{im_i} , con $i = 1, \dots, n$, para cada curva, no necesariamente igualmente espaciados.

La provincia de Chimborazo cuenta con estaciones meteorológicas que recolectan información en: Alao, Atillo, Cumanda, Espoch, Matus, Multitud, Quimiag, San Juan, Tixan, Tunshi, Urbina, a cargo de la Escuela Superior Politécnica de Chimborazo mediante el Grupo de Energías Alternativas y Ambiente quienes incentivan estudios con variables como: velocidad de viento, temperatura, precipitación, humedad, radiación solar global, entre otras. El objetivo de este trabajo de investigación fue identificar características esenciales de la radiación solar global (rsg) mediante el Análisis Estadístico con Datos Funcionales, considerando las mediciones de rsg registradas de durante las 12 horas luz de cada día $x_i(t_j)$ de 07h00 a 19h00 en unidades de

potencia por área (W/m^2), desde el 01 de enero del 2014 al 31 de diciembre del 2017, formando así matrices aproximadamente de 161096 datos, equivalentes a 12392 filas y 13 columnas igualmente espaciados de cada estación meteorológica. Se realizó suavizado de datos mediante B-splines cúbicos y Fourier, con el fin de buscar el ajuste óptimo con el menor error cuadrático medio, la identificación de funciones atípicas, estadísticas descriptivas funcionales y análisis de componentes principales funcionales se realizó con el fin de caracterizar comportamientos similares entre estaciones meteorológicas; entre los meses de los diferentes años en cada estación. Para la validación de grupos mediante componentes principales, se aplicó el análisis clúster funcional con agrupación de curvas de acuerdo a una curva centroide.

Esta investigación está formada por cuatro capítulos: El capítulo I describe los antecedentes, planteamientos y justificación del problema y los objetivos. Seguidamente en el capítulo II muestra la fundamentación teórica de la meteorología, radiación solar global y las técnicas y métodos estadísticos con datos funcionales. El capítulo III detalla la metodología aplicada a los datos de rsg y algoritmos de Rstudio. El capítulo IV muestra los resultados obtenidos. Y finalmente las conclusiones, recomendaciones, referencias bibliográficas y Anexos.

CAPÍTULO I

1 MARCO REFERENCIAL

1.1 Antecedentes

El deterioro de la capa de ozono (gas que recubre la tierra en la atmosfera) (Kriner, et al., 2003, p.2) durante las últimas décadas, permite el ingreso de energía emitida por el sol, la misma que se ha deteriorado según el avance de la industria, esto afecta a todo ser vivo en el planeta provocado por el efecto invernadero; por tanto la Federación Ecuatoriana de la Salud (FEPSO), Instituto de Física Rosario de Argentina (IFIR) y la Organización Mundial de la Salud (OMS) recomiendan protegerse ante este fenómeno que evoluciona diariamente, ya que esto afecta a la salud de la piel, sin embargo, a pesar de estas complicaciones aprovechar esta energía natural ha generado importancia entre los meteorólogos, por tanto es necesario (Urdiales y Espinoza, 2015, p.8-25) contar con datos que coadyuven al estudio de la radiación solar global (rsg) para caracterizar su comportamiento y estudiar sus ventajas lumínicas, además recalcan que Ecuador tiene una gran cantidad de recursos renovables, por ello el Consejo Nacional de Electricidad (CONELEC) pone en marcha la Ley de Régimen del sector Eléctrico (art. 13: (a)) que ayuda e incentiva a buscar nuevas fuentes de energía renovables. En la Universidad de Cuenca realizaron un estudio que ayuda a la generación de energía fotovoltaica que ya está conectada a la red eléctrica pública; para identificar el funcionamiento del proyecto crearon un pequeño panel que ayude a verificar su factibilidad (Lojano y Ríos, 2013, p.32).

La estadística es una técnica que ha recabado gran estadía en el mundo moderno, la cual se encarga de buscar información histórica a lo largo del tiempo o recolectar mediante encuestas, planillas, entrevistas, etc. Dependiendo de las características y de las variables del fenómeno en estudio, pues un buen análisis de los datos no asegura la solución inmediata del problema, más ayudara a la toma de decisiones con el fin de buscar las posibles soluciones a diferentes problemas de investigación, ya sea en el área de la medicina, economía, agricultura, medio ambiente, meteorología y entre otros, donde exista información almacenada (Wackerly et al., 2010, p.1-2).

La matemática ha evolucionado mediante investigaciones continuas realizadas desde principios de siglos y en la actualidad presenta al Análisis Funcional (AF) encargada de analizar funciones $y = f(x)$, donde una de sus magnitudes varia dentro de un espacio de dimensión infinita, con el fin de resolver problemas utilizando esta metodología buscando un ajuste optimo al fenómeno en

estudio, Dieudonné lo define como “*El estudio de los espacios vectoriales topológicos y de las aplicaciones definidas entre subconjuntos de los mismos, sujetas a distintas condiciones algebraicas y topológicas (Bombal, 1994, p.1)*”, desencadenando una nueva rama de la matemática que ha surgido como alternativa en algunos casos de estudio.

Quienes inducen el Análisis de Datos funcionales (ADF), fueron Ramsay y Silverman (1997), publicando su primer trabajo titulado “*Functional Data Analysis*”, seguidamente del libro “*Applied Functional Data Analysis: Methods and Case Studies*”, Ramsay y Silverman (2002), desarrollan distintos casos en el ámbito de la medicina donde analizan mediante el AEDF, la cual detalla de mejor manera los fenómenos estudiados. En el mundo actual existe una gran cantidad de datos, las cuales se almacenan en distintos formatos y el ADF nace por la necesidad de tratar grandes masas de información mediante diferentes métodos de suavizado, estas utilizan trozos de funciones conocidas, Al igual que la estadística convencional el AEDF cuenta con diferentes metodologías para el tratamiento de datos, en este caso el análisis de Componentes Principales Funcional (ACPF) la cual nace referida del Análisis de Componentes Principales (ACP) donde el ACPF reemplaza los vectores por funciones (Ramsay y Silverman 2005, p.1-18).

La meteorología estudia fenómenos aleatorios que ocurren en la atmosfera con el fin de observar y pronosticar la evolución diaria de cada uno, en todo el mundo se establecen centros meteorológicos con el fin de recolectar información, comprender el comportamiento y evaluar características, Carreño (2013, p.12) detalla que los modelos numéricos están muy lejos de representar el comportamiento de estas variables, por esta razón propone unir el “Método Operativo de Senlle (MOS) y el análisis de Componentes Principales Funcionales”, con el objetivo de realizar predicciones de la variable temperatura, de esta manera concluye que el ACPF en el campo de la meteorología y los modelo de predicción creada por el método funcional es mucho mejor para realizar predicciones, ya que el ADF tiene el objetivo de analizar y estudiar “*los espacios vectoriales topológicos, espacios normados, espacios de Banach, el espacio topológico dual, calculo operacional, la noción de espectro del operador y el espacio de Hilbert*”, esto ayuda a encontrar posibles soluciones a problemas que se plantean en la topología algebraica donde se estudia diferentes situaciones en otros campos de la ciencia (Fernández, 2017, p.16-19).

El ACPF puede ser aplicado en Series de tiempo económicas, en el estudio “*Análisis de componentes principales funcionales en series de tiempo económicas*” Chávez et al. (2015, p.16) proponen utilizar este método donde analiza datos de venta anuales de 30 empresas líderes en México durante el periodo 1978-2012, además realizan el mismo estudio con el Análisis de componentes principales clásico, de esta manera realiza una comparación de los resultados obtenidos utilizando las dos metodologías planteadas, donde destaca al ACPF con resultados más

próximos a la serie temporal. También Ramírez et al. (2016, p.79-80) han estudiado la contaminación del aire, de esta manera plantea el “*Análisis de componentes principales funcionales para el estudio de material particulado (pm10) en la ciudad de Bogotá*”, este estudio se realizó en Bogotá-Colombia, donde indica que la primera componente explica el 44.3% de la variabilidad y que la segunda componente explica el 38,3%, con esto concluyen que el ACPF caracteriza de mejor manera la evolución de la contaminación en esta ciudad.

“*El Grupo de Energías Alternativas y Medio Ambiente (GEAA) de la Escuela Superior Politécnica de Chimborazo (ESPOCH) viene trabajando en el ámbito de la energía y el medio ambiente desde 1995. En julio de 2014 mediante resolución N°. 036.CP.2014, tanto el GEAA-ESPOCH como el GEAA-FIE se unieron para formar el Centro de Investigación de Energías Alternativas y Ambiente (CEAA)*” y trabajar en diversas líneas de investigación con el fin de generar energías limpias sin descuidar el medio ambiente, permitiendo que la estadística sea una herramienta primordial en su trabajo con el fin de cumplir su objetivo de “*Desarrollar antecedentes científicos, tecnológicos e innovadores y los relacionados con el conocimiento ancestral, para la aplicación de Energías Alternativas y Protección Ambiental*” (CEAA, 2018), Por ello cuenta con estaciones meteorológicas que se encuentran en las comunidades, Alao, Atillo, Cumanda, EsPOCH, Matus, Multitud, Quimiag, San Juan, Tixan, Tunshi, Urbina, pertenecientes a la provincia de Chimborazo.

Escudero (2012, p.9), analiza de forma funcional el comportamiento de la radiación solar global registrados en la estación, meteorológica de la ESPOCH durante el año 2011, donde considera que el análisis exploratorio funcional, donde la cual permite detallar de manera más concreta y detallada el comportamiento del rsg en relación a los meses diciembre, enero y febrero donde el periodo de verano no está fuertemente definidas en nuestro país. Además, recomienda empezar por la detección de datos atípicos para un estudio más significativo. De esta forma marca un valor metodológico con respecto al uso de estas técnicas que poco a poco ayudan a mejorar el análisis de grades masas de datos.

Escudero et al. (2016, p.2), realizan un estudio de la radiación solar global, donde aplican el Spline cúbico como métodos para brindar una base funcional a los datos de rsg, con el objetivo de analizar y caracterizar el comportamiento de este fenómeno, lo cual conlleva a concluir que el análisis de datos funcionales ayudo a detallar de mejor manera la evolución a través del tiempo, el análisis fue realizado en el software Rstudio que cuenta con las librerías *fda.usc()* y *fda()* que facilita la suavización y el análisis respectivo al estudio, las mismas que fueron creados gracias al impulso del Ramsay y Silverman (1997) quienes promueven el estudio de datos a través de funciones.

Para poder analizar y describir las principales características de los habitantes de Ecuador Naula (2016, p.61) utiliza la base de datos del Censo de Población y Vivienda de año 2010, mediante la ayuda del software R y la librería *fda.usc* con la ayuda de la función *min.basis()*, esta se basa en la Validación Cruzada Generalizada, la cual ayuda a encontrar el número de bases óptimas para poder representar el comportamiento de la población, y el suavizado o representación a funciones se lo realiza mediante B-Spline, de esta realiza un análisis descriptivo funcional y seguidamente aplica el ACPF, la cual ayuda a reducir la direccionalidad donde las dos primeras componentes explican el 96.5% de la variabilidad total de la población ecuatoriana.

1.2 Planteamiento del problema

El avance industrial ha deteriorado paulatinamente la capa de ozono en el planeta Tierra, dando lugar a un libre ingreso de rayos solares, sumado a esto la posición geográfica del Ecuador permite un ingreso de dicha radiación en forma perpendicular a la superficie terrestre, generando una gran fuente de energía, que lamentablemente no es aprovechada al máximo, debido a que es un fenómeno poco estudiado; por ello surge la necesidad de conocer su comportamiento. Este estudio pretende caracterizar la rsg de las zonas aledañas a las estaciones meteorológicas instaladas en la provincia de Chimborazo, mediante métodos estadísticos como el Análisis de Componentes Principales Funcionales.

1.2.1 Formulación del problema

¿Cuál es el número de bases b-splines cúbicos o Fourier óptimo para la base funcional de la rsg?

¿Cómo el ACPF facilita la caracterización de la rsg?

¿Cuál es el comportamiento de la rsg en las zonas aledañas a las estaciones meteorológicas en Chimborazo?

1.3 Justificación

1.3.1 Justificación teórica

La aparición del Big Data (BD) ha generado grandes cambios en el análisis estadístico de datos, además ha generado ventajas en el estudio de un problema, el ACPF ha incursionado en el estudio de gran masa de datos, eliminando el vector de datos de la estadística clásica mediante funciones, lo cual ayuda a verificar la evolución de un fenómeno a través del tiempo.

EL presente estudio aprovechará los datos de rsg de las 11 estaciones meteorológicas instaladas en la provincia de Chimborazo almacenados por el GEAA-ESPOCH, con el fin de caracterizar el comportamiento de dicha variable y establecer directrices del uso del ACPF que permita ser replicado con otras variables, dando lugar a planteamiento de nuevos proyectos de investigación.

1.3.2 Justificación Aplicativa

El GEAA tiene como objetivo principal de aprovechar las energías limpias y renovables existentes en el planeta, por tanto, recolecta información de 14 variables meteorológicas, acumulando una gran cantidad de datos, lo cual dificulta su análisis con estadísticas convencionales.

Este estudio fue de interés del GEAA, ya que, plantean como un proyecto futuro la *"Aplicación de la energía solar para el diseño y adaptación tecnológica de dispositivos solares para la eficiencia energética en edificaciones en la zona central Andina"* que consiste en definir criterios básicos de diseño e implementación, certificación y control de dispositivos solares, buscando la eficiencia de edificaciones en la zona central Andina con la intención de aprovechar la gran cantidad de energía solar presente hoy en día.

1.4 Objetivos

1.4.1 Objetivos Generales

Determinar el comportamiento de la rsg en la provincia de Chimborazo en el período 2014-2017, a través del Análisis de Componentes principales funcional.

1.4.2 Objetivos Específicos

- Estructurar una base de datos común de la rsg registrada en las estaciones meteorológicas ubicadas en la provincia de Chimborazo
- Identificar datos faltantes, e imputarlos de ser necesario.
- Identificar el suavizado óptimo para la base funcional de los datos de rsg.
- Realizar un análisis estadístico descriptivo de los datos funcionales.
- Realizar el Análisis de Componentes Principales Funcional
- Identificar las estaciones meteorológicas con características de rsg similares.

CAPÍTULO II

2 MARCO TEÓRICO REFERENCIAL

2.1 Meteorología

La Meteorología es la ciencia encargada del estudio de la atmósfera, de sus propiedades y de los fenómenos que en ella tienen lugar. El estudio de la atmósfera se basa en el conocimiento de una serie de magnitudes, o variables meteorológicas, como la temperatura, radiación solar global, la presión atmosférica o la humedad, las cuales varían tanto en el espacio como en el tiempo (Rodríguez et al., 2004, p.6).

2.2 Efectos de la atmósfera sobre la radiación.

Las ondas electromagnéticas emitidas por el sol atraviesan la atmósfera antes de llegar a la superficie terrestre, ésta se ve alterada por vapor de agua, aire, suciedad, aerosoles en suspensión, y otros elementos presentes, dichas alteraciones son de diferente tipo las cuales Arenas y Zapata (2011, p.27) detallan a continuación:

Absorción. – la radiación que ingresa en la atmósfera es transformada en energía térmica lo cual provoca que aumente su temperatura y hace irradiar calor hacia la tierra y espacio interplanetario.

Reflexión. – Producida cuando la radiación solar al incidir sobre un cuerpo presente en la atmósfera es desviada o devuelta al espacio, las nubes son la principal causante de reflexión.

Dispersión. - Fenómeno similar a la reflexión, pero la radiación modifica sus caracteres al ser devuelta o desviada por polvo, aerosoles, moléculas de aire y gotas de agua.

2.3 Radiación solar

La energía transferida por el Sol a la Tierra es lo que se conoce como energía radiante o radiación. Ésta viaja a través del espacio en forma de ondas que llevan asociada una determinada cantidad de energía. Según lo energéticas que sean estas ondas se clasifican en lo que se conoce como el espectro electromagnético. Las ondas más energéticas son las correspondientes al rango del ultravioleta, seguidas por la luz visible, infrarroja y así hasta las menos energéticas que

corresponden a las ondas de radio. Donde la longitud de onda y la frecuencia de las ondas electromagnéticas, son importantes para determinar su energía, su visibilidad y su poder de penetración. Todas las ondas electromagnéticas se desplazan en el vacío a una velocidad de 299.792 Km/s. Estas ondas electromagnéticas pueden tener diferentes longitudes de onda. El conjunto de todas las longitudes de onda se denomina espectro electromagnético. El conjunto de las longitudes de onda emitidas por el Sol se denomina espectro solar (Arenas y Zapata, 2011, p.22).

Casi la mitad de la radiación solar que recibe nuestro planeta son frecuencias comprendidas entre 0.4 μ m y 0.7 μ m, pueden ser detectadas por el ojo humano, constituyendo lo que conocemos como luz visible. La otra mitad, mayormente está situada en la parte infrarroja del espectro y una pequeña parte en la ultravioleta. La radiación solar se mide normalmente con un instrumento denominado piranómetro (Arenas y Zapata, 2011, p.25).

2.4 Tipos de Radiación Solar

Los diferentes tipos de radiación solar se clasifican en función de cómo reciben la energía lumínica cada uno de los objetos situados en la superficie terrestre y de algunos de los cuerpos que están en la atmósfera, de la cuales Arenas y Zapata (2011, p.21-23) distinguen los siguientes:

2.4.1 Radiación solar directa

Es la radiación que llega directamente del Sol sin haber sufrido cambio alguno en su dirección. Este tipo de radiación se caracteriza por proyectar una sombra definida de los objetos opacos que la interceptan.

2.4.2 Radiación solar difusa

Una parte de la radiación que atraviesa la atmósfera es reflejada por las nubes o absorbida por éstas. Esta radiación, llamada difusa, va en todas direcciones, efecto producido por las reflexiones y absorciones, no sólo de las nubes sino de las partículas de polvo atmosférico, montañas, árboles, edificios, el propio suelo, etc. Este tipo de radiación se caracteriza por no producir sombra alguna respecto a los objetos opacos interpuestos. Las superficies horizontales son las que más radiación difusa reciben, ya que ven toda la bóveda celeste, mientras que las verticales reciben menos porque sólo ven la mitad.

2.4.3 Radiación solar reflejada

Este tipo de radiación solar es la que refleja la superficie terrestre. La cantidad de radiación depende del coeficiente de reflexión de la superficie, también llamado albedo. Las superficies horizontales no reciben ninguna radiación reflejada, porque no ven ninguna superficie terrestre y las superficies verticales son las que más radiación reflejada reciben.

2.4.4 Radiación solar global

La radiación solar global es la radiación total. Esta constituye la suma de las tres radiaciones anteriormente nombradas. En un día despejado, la radiación directa es preponderante sobre la radiación difusa. Por el contrario, en un día nublado no hay radiación directa y la totalidad de la radiación que incide es difusa.

2.4.5 Radiación solar extraterrestre

La radiación solar extraterrestre es la energía que recibe sobre una superficie horizontal situada en el límite superior de la atmósfera, para calcular la radiación solar extraterrestre se debe corregir la constante solar considerando que la distancia Sol-Tierra varía a lo largo del año, y pasando también de una superficie perpendicular a los rayos solares a una horizontal a la Tierra.



Figura 1-2: Ingreso de Radiación solar a la superficie terrestre

Fuente: Mi sistema Solar

2.4.6 Medida de la radiación global y difusa

El descubrimiento de la radiación solar (r_s) ayudo a que el hombre buscara una forma de cuantificar la incidencia de este fenómeno en la superficie terrestre, la necesidad de obtener información y comprender las características y comportamiento de la r_s es el objetivo de muchos meteorólogos quienes buscan predecir y analizar las ventajas y desventajas de la misma. El instrumento necesario para cuantificar la radiación global es el piranómetro. Este se utiliza a veces para medir la radiación incidente sobre superficies inclinadas y se dispone en posición invertida para medir la radiación global reflejada (Arenas y Zapata, 2011, p.21).

Los distintos tipos de colectores solares aprovechan de forma distinta la radiación solar. Los colectores solares planos, por ejemplo, captan la radiación total (directa + difusa), sin embargo, los colectores de concentración sólo captan la radiación directa. Por esta razón, los colectores de concentración suelen situarse en zonas de muy poca nubosidad y con pocas brumas, en el interior, alejadas de las costas. Los colectores solares planos pueden colocarse en cualquier lugar, siempre que la insolación sea suficiente. Para medir solamente la componente difusa de la radiación solar, la componente directa se cubre por medio de un sistema de pantalla o sombreado.



Figura 2-2: Piranómetro

Fuente: Empresa (kipp y Zonen)

2.4.7 Unidades de medida de la radiación solar global

El impacto de la radiación solar global en la superficie terrestre es expresada en términos de exposición radiante o irradiación, esta se recaba como medida del flujo de energía ingresada por unidad de área $\left(\frac{\text{Energía}}{\text{Área}}\right)$, su unidad de medida se expresa como: $\frac{\text{watt}}{\text{m}^2}$ (Arenas y Zapata, 2011, p.21), (Escudero, 2012, p.18).

2.5 Aproximación mediante funciones

La información conseguida y la creación de bases de datos de diversas variables, ha cobrado una gran relevancia ya sea en el ámbito social, económico, cultural, ambiental, entre otras; estudiar y analizar el comportamiento de cada fenómeno es el objetivo principal de la estadística, el salto de lo finito a lo infinito es posible gracias al análisis de datos funcionales el cual busca una serie de funciones conocidas que se adapte de manera óptima al fenómeno en estudio (Ramsay y Silverman, 200, p.37-40).

Existen diversos métodos de suavizado según la naturaleza de la información, mencionados por: Torrecilla (2010, p.13), Benko (2006, p.29), Hooker (2017, p.20), Ramsay y Silverman (2005, p.45-56), quienes hacen referencia a algunos de los métodos de suavizado más comunes, cabe recalcar que el uso de funciones es necesario aplicar a los datos discretos siendo estos representados como funciones de tiempo continuo.

2.5.1 Bases Fourier

Las bases Fourier son antiguas y poco utilizadas, son funciones usadas para aproximar datos periódicos, su función base es la del seno y coseno; la serie de Fourier como función periódica está definido por:

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

dónde: T define el periodo de oscilación

por su estructura aditiva se puede usar como una expansión de bases, definido como:

$$\phi_0(t) = \frac{1}{\sqrt{T}}$$
$$\phi_{2r-1}(t) = \frac{1}{\sqrt{T/2}} \text{sen}(rwt)$$
$$\phi_{2r}(t) = \frac{1}{\sqrt{T/2}} \text{cos}(rwt)$$

para $r = 1, \dots, L/2$, donde L es un entero par. La frecuencia determina el período y la duración del intervalo $|J| = 2$ (aquí $J = [0, 1]$ es asumido por lo tanto $|J| = 1$). La base de Fourier

definida anteriormente es una base ortonormal. La popularidad de esta base se basa parcialmente en la posibilidad de un cálculo rápido del coeficiente por el algoritmo de transformación de Fourier rápida (Fast Fourier Transformation FFT).

Otra característica importante de la serie de Fourier es la existencia de continuos:

$$D^m \phi_0(t) = 0, m \geq 1$$

$$D^m \phi_{2r-1}(t) = (rw)^m \text{sen} \left(rwt + \frac{m\pi}{2w} \right), m \geq 0$$

$$D^m \phi_{2r}(t) = (rw)^m \text{cos} \left(rwt + \frac{m\pi}{2w} \right), m \geq 0$$

para $r = 1, \dots, L/2$.

2.5.2 Bases B-spline

Las Bases spline son las más utilizadas para aproximar datos no periódicos, por su gran facilidad y adaptabilidad a la mayoría de datos, dichas curvas son de grado p , unidas entre sí. Las cuales cumplen con las siguientes características.

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

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

con $j = -1, 0, 1, 2, \dots, m + 4$

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

Con $p = 1, 2, \dots$ y $j = -1, 0, 1, 2, \dots, m - p + 4$, si $p = 4$ estas bases son B-spline 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 $j = -1, 0, 1, 2, \dots, m + 1$.

2.5.3 Bases Wavelets

Las funciones son utilizadas para filtrar datos de tipo serie temporal, es utilizada en el análisis de la variación espectral, una de las ventajas es que representa tiempo-frecuencia con más precisión. Se obtienen por la dilación y translación de una wavelet ψ .

$$\int \psi(t) dt = 0$$

Y la función de escala o wavelet padre:

$$\int \varphi(t) dt = 0$$

por tanto, queda definido como

$$\Psi_{j,k}(t) = \frac{1}{2} \Psi\left(\frac{t - 2^j k}{2^j}\right), \quad \Phi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \varphi\left(\frac{t - 2^j k}{2^j}\right)$$

Con $j = 1, 2, \dots, J$ son índices de escala (dilatación) y $k = 1, 2, \dots, K$ (traducción), de esta forma las escalas S_j se descomponen en coeficiente de detalle d_j, k y coeficiente de aproximación $a_{j,k}$

2.5.4 Base exponencial y de potencia

Los sistemas de base exponencial consisten en una serie de funciones exponenciales,

$$e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_j t} \text{ con } j = 1, 2, \dots$$

Donde los parámetros λ_j son todos distintos, y a menudo $\lambda_1 = 0$, las ecuaciones diferenciales lineales con coeficientes constantes tienen como soluciones, expansiones en términos de bases exponenciales.

Las bases de potencias

$$t^{\lambda_1}, t^{\lambda_2}, \dots, t^{\lambda_j} \text{ con } j = 1, 2, \dots$$

Su ventaja es utilizar cuando t es estrictamente positivo, para que las potencias negativas sean posibles.

2.5.5 Bases constantes

En algunos casos se presentan trayectorias de procesos puntuales y de recuento las cuales son constantes en intervalos aleatorios definidos, en los que se producen nuevas llegadas. Para reconstruir las funciones muestrales de este tipo de procesos es la base ortogonal de funciones constantes sobre los intervalos de una partición fijada.

Si fijamos un intervalo de observación T definida por los nodos $0 = a_1 < a_2 < \dots < a_p$ una base ortonormal del subespacio de las funciones constantes sobre cada uno de los intervalos $(a_{j-1}, a_j]$, $j = 1, \dots, p$, definida como:

$$\delta_j(t) = (a_j - a_{j-1})^{\frac{1}{2}} I_j(t)$$

Donde $I_j(t)$ es la función indicadora en el intervalo $(t_{j-1}, t_j]$ que toma el valor uno en dicho intervalo y el valor cero fuera de él.

2.5.6 Bases polinomiales

Se define la base polinomial $\phi_j(t_{ik}) = (t_{ik} - w)^j$, con $j = 0, \dots, p$ e $i = 1, \dots, n$.

dónde

p es el número de funciones base,

n el número de observaciones funcionales

w es un parámetro que normalmente se elige en el centro del intervalo de aproximación.

Tal parámetro se puede tomar cuidadosamente para evitar el error de redondeo en las computaciones, ya que los valores están más altamente correlacionados cuando el grado incrementa. Sin embargo, si los t_{ik} están igualmente espaciados o se pueden elegir para mostrar algunos patrones estándar, las extensiones polinomiales ortogonales se pueden obtener implementando $O((m_i + g)p)$ operaciones, llamando g al grado del polinomio.

2.6 Correlación cruzada

Según Febrero et al. (2013, p.131), presentan la librería *fda.usc* usado el software R para el análisis de datos funcionales, mediante la función *min.basis* con el objetivo de analizar el número de bases óptimos para el suavizado de curvas en diferentes estudio, para ello emplea el suavizado de Kernel con dos criterios de validación, la cual detalla Febrero y Oviedo (2012, p.6-8).

2.6.1 Suavizado de Kernel

La metodología es usada para representar datos funcionales, esto es dado por la suavización de la matriz $S = (s_{ij})$, donde el ejemplo del estimador de Nadaraya-Watson (función *S.NW()*):

$$s_j(t_i) = K\left(\frac{t_i - t_j}{k}\right) / \sum_{k=1}^m K\left(\frac{t_i - t_j}{k}\right), \text{ donde } K(\cdot) \text{ es la función Kernel.}$$

2.6.2 Criterio de Validación

Elegir un parámetro que represente de manera óptima al conjunto de datos es primordial, para ello se tiene dos criterios para poder elegir el parámetro v las cuales se detallan a continuación según Febrero y Oviedo (2012, p. 6-8): Validación Cruzada (Cross-validation CV) y Validación cruzada generalizada (Generalized cross-validation GCV).

$$\text{Validación Cruzada: } CV(v) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{-i}^v)^2 w_i$$

dónde \hat{y}_{-i}^v , indica el estimador basado en omitir el par (t_i, y_i) y w_i es el peso en el punto t_i . Este criterio es implementado por la función *CV.S ()*.

$$\text{Validación cruzada generalizada: } GCV(v) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i^v)^2 w_i \mathcal{E}(v)$$

dónde: $\mathcal{E}(v)$ denota el tipo de función penalizadora.

Además, se implementa en la función *GCV.S ()* con los siguientes tipos de funciones penalizadas:

- Validación cruzada generalizada (GCV): $\mathcal{E}(v) = (1 - \text{tr}(S_v)n^{-1})^{-2}$
- Criterio de información de Akaike (AIC): $\mathcal{E}(v) = e^{\text{tr}(S_v)n^{-1}}$
- Error de predicción finita (FPE): $\mathcal{E}(v) = (1 + \text{tr}(S_v)n^{-1}) / (1 - \text{tr}(S_v)n^{-1})$

- Selector de modelo Shibata (Shibata): $E(v) = 1 + 2tr(S_v)n^{-1}$
- Rice's bandwidth selector (Rice): $E(v) = (1 - 2tr(S_v)n^{-1})^{-1}$

2.7 Análisis de datos funcionales

El análisis de datos funcionales (ADF) surge en el siglo XIX, como la evolución del análisis de datos clásico (Estadística clásica) con el objetivo de procesar grandes cantidades de datos, donde sus mediciones discretas son remplazadas por muestras de funciones aleatorias, además las estadísticas descriptivas e Inferencial son análogas para el estudio de datos funcionales. En este tipo de estudio es primordial trabajar en un espacio de cuadrado integrable L^2 , la cual es conocida comúnmente con espacio de Hilbert (Navarro, 2004, p.7).

2.7.1 Espacios de Hilbert

Se conoce como espacios de Hilbert al espacio de funciones de cuadrado integrable L^2 , por tanto es común trabajar con vectores conformados por infinitas componentes, es decir para cada $t \in R$, su producto escalar estará compuesto por la suma del producto de sus infinitas componentes (Marrero, 2011, p.9-10), (González, 2016, p.4-5).

$$L^2 = \left\{ f: R \rightarrow R: \int_R f^2(t)dt < \infty \right\}$$

2.7.2 Espacio con Producto escalar

El espacio con producto escalar, se identifica con diversos nombres y de diversos autores, sin embargo, el objetivo de la teoría no se altera, los nombres comúnmente estudiados son: espacio prehilbertiano, espacio con producto interior, espacio pre-Hilbert, sin embargo, en el caso de ADF se define elementos x e y como funciones de R en R (Bruzual y Domínguez 2005, p.1).

Definición 2.1.- Sea L^2 un espacio de funciones de cuadrado integrable donde:

$L^2 = \{f: \}$ vectorial sobre K . Se define como producto escalar en E una función $\langle \cdot, \cdot \rangle: E \times E \rightarrow K$

$$(x, y) \rightarrow \int_R x(t)y(t)dt$$

que satisface las siguientes propiedades:

1. Simetría $\langle x, y \rangle = \int_R x(t)y(t)dt = \int_R y(t)x(t)dt = \langle y, x \rangle$, para todo x e y .

2. Positividad $\langle x, x \rangle = \int_R x(t)x(t)dt \geq 0$ para todo x ,
con $\langle x, x \rangle = \int_R x(t)x(t)dt = 0 \leftrightarrow x = 0$
3. Bilinealidad Para todo $a, b \in \mathbb{R}$ $\langle ax + by, z \rangle = \int_R [a \cdot x(t) + b \cdot y(t)]z(t)dt =$
 $\int_R [a \cdot x(t) \cdot z(t) + b \cdot y(t) \cdot z(t)]dt = a\langle x, z \rangle + b\langle y, z \rangle$

El producto escalar se puede interpretar como medida de cantidad de asociación entre dos elementos del espacio vectorial en el que trabajamos.

2.8 Dato funcional

Los datos funcionales es un conjunto de curvas $\{x_1(t), x_2(t), \dots, x_n(t)\}$ con $t \in T$, cada una de ellas son observadas en distintos instantes o puntos de argumento t de la forma $x_i = (x_i(t_{i0}), x_i(t_{i1}), \dots, x_i(t_{im_i}))'$, para el conjunto de instantes t_{i0}, \dots, t_{im_i} , con $i = 1, \dots, n$, para cada curva, no es necesario tener espacios iguales (Escudero, 2016, p.11).

2.9 Datos Funcionales atípicos

Al igual que en el análisis descriptivo clásico (Observaciones puntuales), el ADF necesita realizar la detección de curvas atípicas, Millán (2017, p.50-60) indica las diferentes causas de ser llamadas atípicas, además presenta métodos computacionales en R que pueden ser aplicados para detectar anomalías dentro de los datos funcionales.

La función *outlier.depth.trim()* y *outlier.depth.prod()* son métodos que se basan en cálculos de profundidad de las curvas recortadas en un $\alpha\%$ especificada en función de B muestras bootstrap de tamaño n denotadas por x_i^b , con $i = 1, \dots, n$ y $b = 1, \dots, B$, se consideran atípicos si las curvas son aislados, persistentes, por la forma y amplitud (Millán, 2017, p.50-56).

2.10 Profundidad funcional

Millán (2017, p. 58-59) realiza un estudio de los diferentes métodos existentes para explorar datos atípicos mediante el cálculo de distancias, las más conocidas son las profundidades de: FM, MD y RPD, que se detallan a continuación.

2.10.1 Profundidad de Fraiman y Muniz (FM)

Fraiman y Muniz (2001, p.422-423) son los primeros en integrar la noción de cálculo de profundidad para datos funcionales.

Si $F_{n,t}(x_i(t))$ la distribución empírica acumulada de las curvas $x_1(t), \dots, x_n(t)$, en un tiempo $t \in [a, b]$, la profundidad FM de una curva x_i , respecto al conjunto x_1, \dots, x_n es

$$F_{n,t}(x_i(t)) = \int_a^b D_n(x_i(t)) dt$$

donde $D_n(x_i(t)) = 1 - \left| \frac{1}{2} - F_{n,t}(x_i(t)) \right|$ es la profundidad univariada de la curva para cada $1 \leq i \leq n$.

2.10.2 Profundidad Modal (MD)

Sea x_1, \dots, x_n un conjunto de curvas, la profundidad MD de una curva x_i esta dada por

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

siendo $\| \cdot \|$ la norma del espacio funcional, $K: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ la función Kernel y h denota el ancho de banda, por tanto curva que alcanza el máximo valor de $MD_n(x_i, h)$ será la moda funcional.

2.10.3 Profundidad de proyección aleatoria (RPD)

Se basa en cuantificar la profundidad de las curvas por medio de proyecciones de los datos funcionales a su primera derivada sobre una dirección aleatoria. La profundidad de datos en \mathbb{R}^2 se define por el punto medio según el número de proyecciones.

Sea x_1, \dots, x_n un conjunto de curvas, una dirección v , entonces $T_{i,v} = \langle v, x_i \rangle$ es la proyección de la curva x_i sobre v , denotada por

$$T_{i,v} = \langle v, x_i \rangle = \int_a^b v(t) x_i(t) dt$$

y la primera derivada de la proyección está dado por $T'_{i,v} = \langle v, x'_i \rangle$, con la dirección v , entonces el punto que se denota en \mathbb{R}^2 está dado por el par $(T_{i,v}, T'_{i,v})$ y si v_1, \dots, v_p son las direcciones aleatorias e independientes la profundidad RPD de una curva viene dado por

$$RPD_n x_i = \frac{1}{P} \sum_{p=1}^P D_n \langle v, x_i \rangle, \langle v, x'_i \rangle$$

y D_n es una profundidad cualquiera definida en \mathbb{R}^2 por el par $\langle v, x_i \rangle, \langle v, x'_i \rangle$.

2.11 Análisis Descriptivo Funcional

Las medidas de análisis descriptivo pueden ser expresadas con notación algebraica. Ramsay y Silverman (2005, p.22-26) las definen así para que sean fácilmente adaptables a la naturaleza de los datos: variables, vectores o funciones del fenómeno en estudio.

2.11.1 Media funcional

Definimos a la media de la función $x(t)$ o también valor medio de la función $x(t)$ como:

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

dónde la función $\mathbf{1}(t) = 1 \forall t \in \mathbb{R}$

2.11.2 Varianza funcional

Sea $x_1(t), x_2(t), \dots, x_n(t)$ una muestra de funciones de una función aleatoria (t) , definida en $[0, T]$. Se define la función varianza muestral de (t) como:

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

dónde su desviación estándar es la raíz de la varianza.

2.11.3 Covarianza funcional

Sea $x_1(t), x_2(t), \dots, x_n(t)$ una muestra de funciones de una función aleatoria $x(t)$, definida en $[0, T]$. Se define la función covarianza muestral de (t) entre t_1 y t_2 como:

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

2.11.4 Correlación funcional

Sea $x_1(t), x_2(t), \dots, x_n(t)$ una muestra de funciones de una función aleatoria $x(t)$, definida en $[0, T]$. Se define la función correlación muestral de (t) entre t_1 y t_2 como:

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

2.12 Análisis de Componentes Principales Funcionales

Uno de los problemas más comunes en el tratamiento estadístico de datos es la presencia de una gran cantidad de variables o datos funcionales, esto es motivo para evaluar la correlación de cada una de ellas, para ello el análisis de componentes principales funcionales (Functional Principal Components Analysis FPCA), nace del paso de lo finito a lo infinito del Análisis de Componentes Principales Convencional donde trabaja con observaciones discretas con el fin de reducir la dimensionalidad de p variables en estudio, en cambio el ACPF cuenta con observaciones como trayectorias de un proceso estocástico de segundo orden, continuo (Escudero, 2016: p.43-44) en media cuadrática con trayectoria de cuadrado integrable en el espacio $L^2(T)$. Denotadas con $\{\xi_i\}$ definidas como combinaciones lineales generalizadas de variables funcionales, incorreladas de varianza máxima dadas por (Carreño, 2013, p.21-25).

$$\xi_i = \int_T x(t)f_i(t)dt$$

donde la primera componente principal funcional debe explicar la variabilidad máxima posible del conjunto de información.

Cada componente principal está asociada a un autovalor $\lambda_i = \text{Var}[\xi_i]$, la selección es de acuerdo a la proporción de variabilidad explicada lo más cercana a 1.

$$V_T = \sum_{i=1}^{\infty} \lambda_j$$

Generalmente seleccionan las q -primeras componentes principales cuya proporción de varianza explicada acumulada sean tan próxima a uno.

$$V^q = \frac{\sum_{i=1}^q \lambda_i}{V_T}$$

Si $\{x_1(t), x_2(t), \dots, x_n(t)\}$ una muestra de curvas del proceso estocástico $\{x(t): t \in T\}$, las w componentes principales seleccionadas son vectores de dimensión n y están dadas por

$$\xi_{wi} = \int_T x_w(t) f_i(t) dt$$

con $w = 1, \dots, n$ y f_i función peso asociado a la i -ésima función propia del operador de covarianza muestral (Escudero, 2016, p.43-44)

$$\int_T \hat{C}(s, t) f(s) ds = \lambda f(t)$$

donde f_i son las funciones asociadas a los autovalores λ_i , y $\hat{C}(s, t)$ la covarianza muestral

$$\hat{C}(s, t) = \frac{1}{n-1} \sum_{i=1}^n (x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t))$$

de modo que se pueden representar y aproximar las curvas muestrales en términos de las primeras q componentes principales

$$x_w^q(t) = \sum_{i=1}^{n-1} \xi_{wi} f_i(t)$$

Para el cálculo de las CPF a partir de las representaciones básicas de las curvas muestrales

$$x_i(t) = \sum_{p=1}^P a_{ip} \phi_p(t)$$

con $i = 1, \dots, n$, está demostrado que, el problema de ACPF se reduce a calcular el ACP multivariante de la matriz $A\Psi^{1/2}$, donde $A = (a_{ij})_{n \times P}$ es la matriz que contiene por filas los coeficientes básicos de las trayectorias muestrales y $\Psi = (\Psi_{jk})_{P \times P}$ es la matriz de elementos los productos escalares entre las funciones básicas dados por

$$\Psi_{jk} = \int_T \phi_j \phi_k dt$$

con $j, k = 1, \dots, P$.

Suponiendo que $\Gamma = (\xi_{ij})_{n \times p}$ es la matriz cuyas columnas son las componentes principales de $A\Psi^{1/2}$ y G es la matriz cuyas columnas son los vectores propios asociados, tenemos que $\Gamma = (A\Psi^{1/2})G$ (Carreño, 2013, p.21-25), (Escudero, 2016, p.43-44).

Las funciones propias del operador de covarianza muestral admiten entonces la siguiente expansión básica:

$$f_i(t) = \sum_{l=1}^P f_{li} \phi_l(t)$$

con $i = 1, \dots, P$, donde los coeficientes son los elementos de la matriz $F = (f_{li})_{P \times P}$ dada por $F = (\Psi^{-1/2})G$ (Carreño, 2013, p.21-25) (Escudero, 2016: p.43-44).

2.13 Análisis Clúster Funcional

El análisis clúster puede ser usado para construir subgrupos de un conjunto de p variables, al igual que ACPF es un método de reducción de dimensionalidad, las curvas dentro de los grupos deben ser lo más homogéneos posibles mientras que entre los grupos deben ser lo más heterogéneos.

Uno de los métodos de clasificación no supervisada es la de K-mean denotado por *kmean.fd()*, su característica principal es asignar curvas basadas en similitud o medidas de distancia entre k número de clúster. Jacques y Preda (2014, p.14) analizaron de forma detalla las diferentes funciones que se aplican para clasificar datos funcionales, así como también el algoritmo que determina el número de clústeres optimo, aunque y Strandberg (2013, p.15-16), detalla el algoritmo que determina el número de clúster óptimos aunque la clasificación deber ser mínimo con $k = 2$.

2.13.1 Algoritmo de alineación de k -medias

- Elegir k clúster representativos $\varphi_1(t), \dots, \varphi_k(t)$, dado un conjunto de curvas $x_1(t), \dots, x_n(t)$.
- Asignar y alinear las curva al clúster con similitud, entonces la curva $x_i(t)$ es asignado y alineado a la j -ésimo clúster si

$$\max_{l=1, \dots, k} \rho(x_i(h_{il}^*), \varphi_l) = \rho(x_i(h_{ij}^*), \varphi_j)$$

donde $h_{ij}^*(t)$ es la deformación funcional óptima que maximiza la medida $\rho(x_i(h_{ij}^*), \varphi_j)$.

- Necesitamos realizar un paso de la normalización. Para $j = 1, \dots, k$, que permita a N_j denotar el número de curvas y permita a Γ_j ser el conjunto de todas las curvas $i = 1, \dots, n$ perteneciente al clúster j . Para $j = 1, \dots, k$ todas las N_j curvas que se asignaron y alinearon al clúster j en el paso 2, ahora debe formarse a lo largo de la función $\tilde{h}_j^{-1}(t)$, dónde

$$\tilde{h}_j^{-1}(t) = \frac{1}{N_j} \sum_{i \in \Gamma_j} h_{ij}^*(t)$$

- Para todas las curvas calcular las medidas de similaridad entre la alineación de la curva con su clúster más representativo usando $\rho(x_i(\tilde{h}_j^{-1}), \varphi_j)$. Si los incrementos de la medida de similaridad son muy bajos detener el procedimiento.
- Calcular nuevo clúster representativo usando la asignación y alineación de la curva por las medias y regresar al paso 2.

2.14 Software Estadístico R

El proyecto R fue desarrollado en Nueva Zelanda, Universidad de Auckland por Robert Gentleman y Ross Ihaka y recibe su nombre que conmemora las iniciales de los creadores, es un software libre en condiciones de licencia GNU-GPL y gracias a la contribución de investigadores cuenta con diversas librerías, con el fin de analizar los datos o cualquier fenómeno en estudio, ya sea descriptivo, series temporales, diseño de experimento, análisis multivariante, análisis de datos funcionales y entre otros (Arriaza et al., 2008, p.2-5).

El proyecto R se encuentra en su página oficial de: <http://www.r-project.org>, donde se puede descargar en diferentes Sistemas Operativos, las indicaciones de la descarga e instalación vienen incluidas para todas aquellas personas la obtengan y haga uso de ella. Las librerías están subidas al repositorio CRAN de donde se puede descargar según la necesidad y situación ante la matriz de datos, Además el código fuente puede ser manipulado directamente por los usuarios (Arriaza et al., 2008, p.2-5).

CAPÍTULO III

3 METODOLOGÍA

3.1 Tipo y Diseño de Investigación

3.1.1 Área de estudio

La presente investigación abarca como área de estudio toda la provincia de Chimborazo.



Figura 3-3: Provincia de Chimborazo

Fuente: Google Maps

3.1.2 Población en estudio

Para este estudio se tomó los datos de la radiación solar global durante los años 2014 hasta el 2017 según la disposición de datos, de 11 estaciones: Alao, Atillo, Cumanda, EsPOCH, Matus,

Multitud, Quimiag, San Juan, Tixan, Tunshi y Urbina, ubicadas en lugares estratégicos en la provincia de Chimborazo.

3.1.3 *Recolección de información*

La radiación solar global en las diferentes estaciones es tomada mediante piranómetros, la información es almacenada en el repositorio del GEAA, las cuales son utilizadas en diferentes ámbitos de estudio.



Gráfico 1-3. Ubicación de las estaciones meteorológicas

Elaborado por: Taguacundo Oscar, 2019

3.1.4 Operacionalización de variables

La siguiente tabla muestra la variable meteorológica en estudio

Tabla 1-3: Operacionalización de variable

Variable	Definición Conceptual	Definición operacional	Unidad de medida	Escala
Radiación solar global	Flujo de energía recibida desde el Sol en forma de ondas electromagnéticas de distintas frecuencias	Mide la intensidad de ondas electromagnéticas emitidas por el sol por metro cuadrado	$\frac{watt}{m^2}$	Razón

Fuente: GEAA

Elaborado por: Taguacundo Oscar, 2019

3.1.5 Análisis de Datos

De la información proporcionado por el GEAA, se estudió la variable meteorológica radiación solar global, el análisis primordial fue explorar la información faltante, se realizó un estudio intensivo en el suavizado de datos funcionales con base B-spline y Fourier con el fin de identificar el mejor suavizado, posteriormente se prosiguió con el análisis descriptivo funcional para ello se utilizó las librerías *fda* y *fda.usc* del software R, las cuales ayudaron con los cálculos necesarios, donde cada una de las curvas representan la rsg de cada día en W/m^2 durante los años 2014, 2015, 2016, 2017. De la misma forma para caracterizar su comportamiento se analizó por meses, por años y estaciones. Con la ayuda de la reducción de dimensionalidad y la gráfica de las dos primeras componentes donde se observó las diferentes agrupaciones de comportamiento de la rsg en la provincia de Chimborazo.

3.1.6 Alcance de la investigación

El siguiente trabajo consideró el análisis de la variable radiación solar global de las 11 estaciones meteorológicas sondeadas por el GEAA de la Facultad de Ciencias, además de ser un impulso más para que el AEDF sea empleada en las demás variables¹ que recolecta el GEAA, con el fin de conocer el comportamiento de dichas variables y optimizar la toma de decisiones.

¹ Temperatura de aire (°C), Humedad relativa (%), Presión barométrica (hPa), Radiación solar difusa (w/m2), Radiación solar global (w/m2), Temperatura del suelo (5M) (°C), Temperatura del suelo (0CM) (°C), Temperaturas del suelo (-5CM) (°C), Temperaturas del suelo (-10CM) (°C), Temperaturas del suelo (-20CM) (°C), Temperaturas del suelo (-50CM) (°C), Temperaturas del suelo (-100CM) (°C), Velocidad de viento, Dirección de viento, Sensación térmica

CAPÍTULO IV

4 RESULTADOS Y DISCUSIÓN

4.1 Construcción de la base de datos

Con la información proporcionada por el GEAA de cada una de las estaciones, se identificó que fueron imputadas todas aquellas estaciones con información faltante menor del 20%, mediante regresión lineal múltiple en Rstudio teniendo como resultado una base de datos con la información según muestra la tabla 2-4.

Tabla 2-4: Radiación Solar Global por estación y año de estudio

Estación	Año	Estación	Año
Alao	2014-2017	Matus	2015, 2017
Cumanda	2014-2015	Tunshi	2014-2016
Atillo	2015-2017	Multitud	2014-2016
San Juan	2014, 2016, 2017	Urbina	2014, 2016, 2017
Epoch	2015-2017	Quimiag	2014-2017
Tixan	2014-2017		

Elaborado por: Taguacundo Oscar, 2019.

4.2 Suavizados B-spline y Fourier

La literatura recomienda suavizar los datos funcionales con el fin de reducir el sesgo en los resultados, por ello se consideró las bases más comunes empleadas para estudio de diversos fenómenos que son B-spline y Fourier.

Existen dos formas de encontrar el número de bases optimas que pueden ser utilizadas para el suavizado de datos:

1. Sin fijar el número de bases (numbasis)

En este caso la búsqueda de bases optimas no se limita, lo que significa que serán buscadas en el intervalo $[4, \infty]$, sin olvidar que generalmente para las bases B-spline y Fourier un número par de

bases no aporta significativamente como un número impar de bases, por ello se iniciara las simulaciones de suavizado partir de `numbasis = 5`.

```
nbopt <- min.basis(rsgestdia, type.CV = GCV.S, verbose = T, type
.basis = "bspline")
The minimum GCV (GCV.OPT=27905.31) is achieved with
the number of basis (numbasis.opt=23)
and lambda value (lambda.opt=0)
```

2. Fijando el número de bases (`numbasis`)

En este caso la búsqueda de bases óptimas se limita, es decir, serán buscadas en el intervalo $[a, b]$, mismas que serán fijadas por el investigador experimentado.

```
nb <- seq(5,23, by = 2)
nbopt <- min.basis(rsgestdia, type.CV = GCV.S, verbose = T, type
.basis = "bspline", numbasis = nb)
The minimum GCV (GCV.OPT=27962.54) is achieved with
the number of basis (numbasis.opt=7)
and lambda value (lambda.opt=0)
```

La búsqueda del número de bases óptimo se realizó mediante la función `min.basis()` que trabaja con suavizado B-spline en el que se obtuvo 7 y 23 bases, sin embargo, con el fin de garantizar el número de bases óptimas para los datos de rsg se probó el suavizado con: 7, 9, 11, 13, 15, 17, 19, 21 y 23, para Fourier y B-spline cúbico y se identificó el error cuadrático medio(RMS) de cada uno con la función `plotfit.fd()`, de la librería `fda` de Rstudio, el gráfico 2-4 muestra los resultados del RMS con las diferentes bases propuestas, el Anexo D contiene más resultados acerca de la optimización del número de bases.

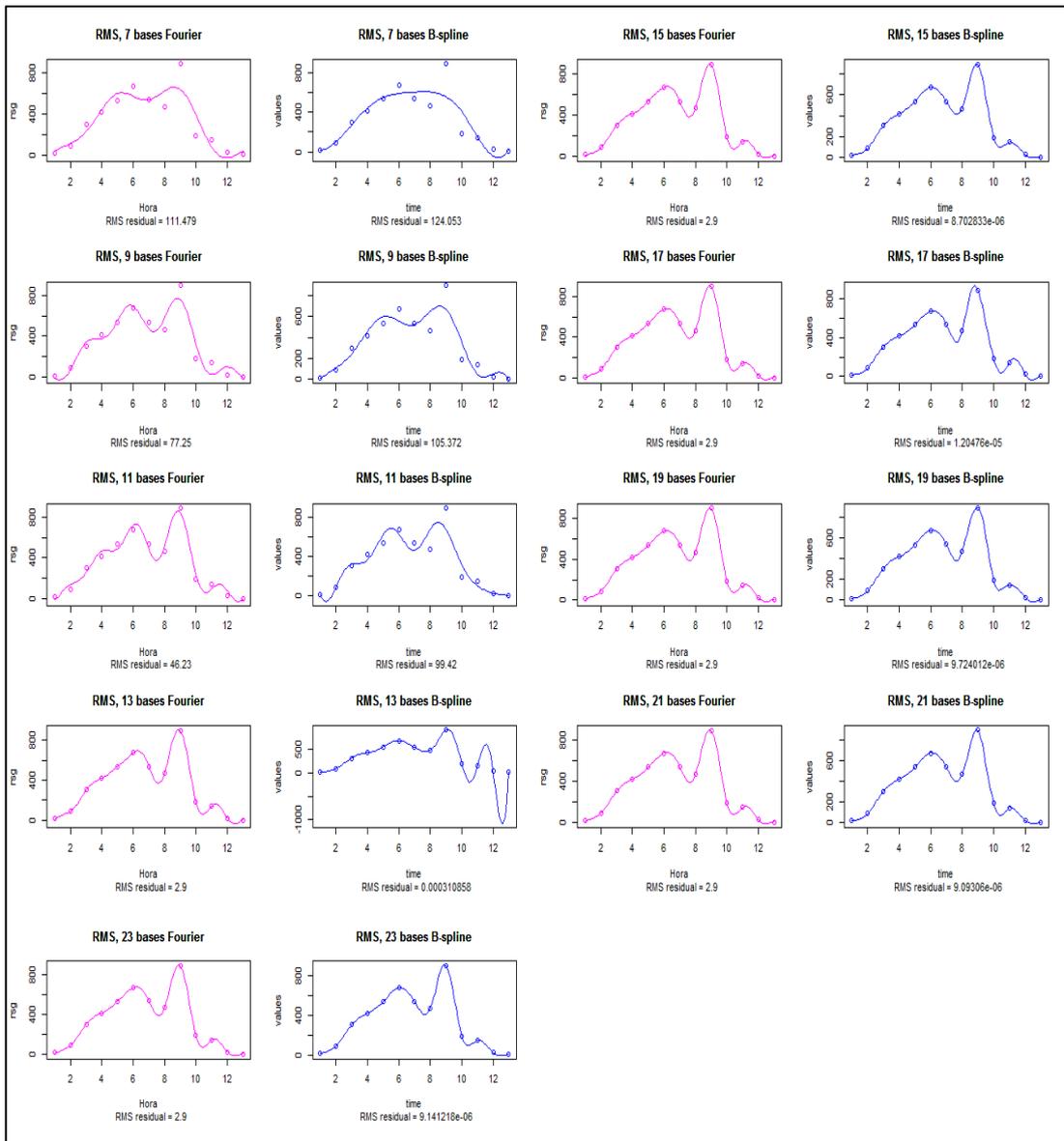


Gráfico 2-4: Error cuadrático medio de suavizado Fourier y B-Spline cúbico, según el número de bases

Elaborado por: Taguacundo Oscar, 2019

Se realizó una comparación de errores entre Fourier y B-spline cúbico, ordenados de forma ascendente (Gráfico 3-4).

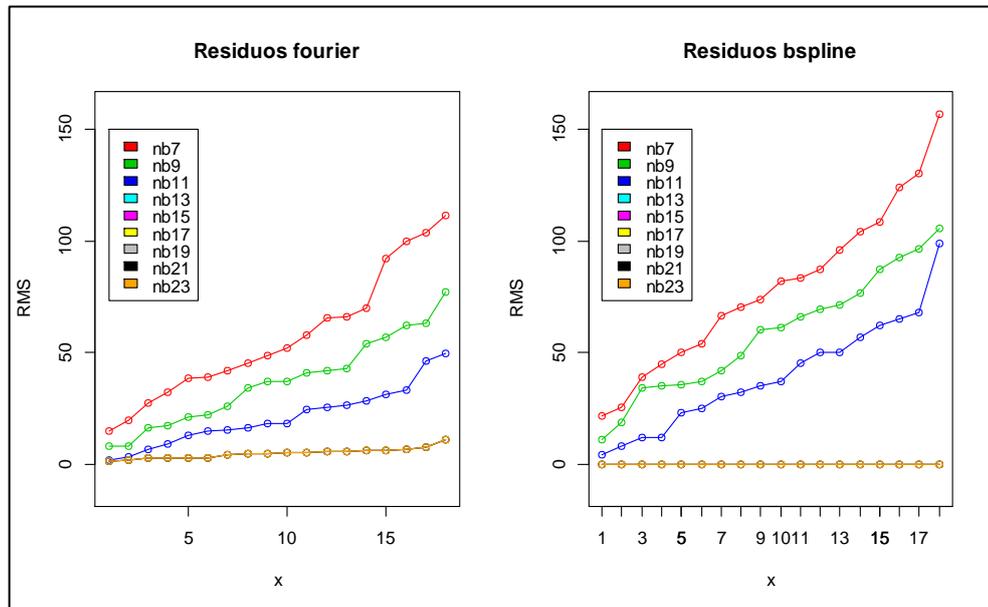


Gráfico 3-4: Residuos por número de bases, B-spline y Fourier

Elaborado por: Taguacundo Oscar, 2019

El gráfico de residuos muestra el error cuadrático medio suavizado mediante Fourier y B-spline, donde se observa un error alto para 7, 9 y 11 bases de suavizado, además Fourier no altera su error desde 15 hasta 23 bases Fourier, mientras que para los B-spline la base 13 al 23 son a simple vista similares, sin embargo, los errores son inferiores a 1. De esta forma se descarta el suavizado para las bases 7, 9 y 11 para los dos métodos de suavizado y se intensifica el análisis con 13 hasta 23 bases (Gráfico 4-4).

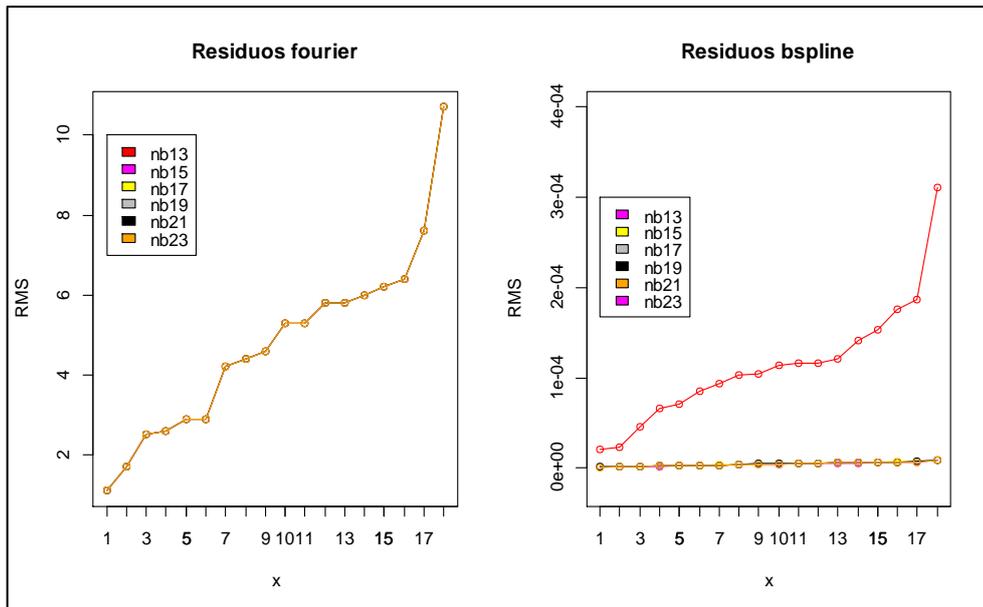


Gráfico 4-4: Residuos, B-spline y Fourier, con bases de 13 al 23.

Elaborado por: Taguacundo Oscar, 2019

El gráfico muestra que para las bases Fourier desde 13 hasta 23 el error se mantiene constante, mientras que para los B-spline la base 13 tiene un error mayor que las bases 15, 17, 19, 21 y 23. Con el fin de identificar diferencias entre el suavizado B-spline con número de bases del 15 al 23, se analiza dichos residuos.

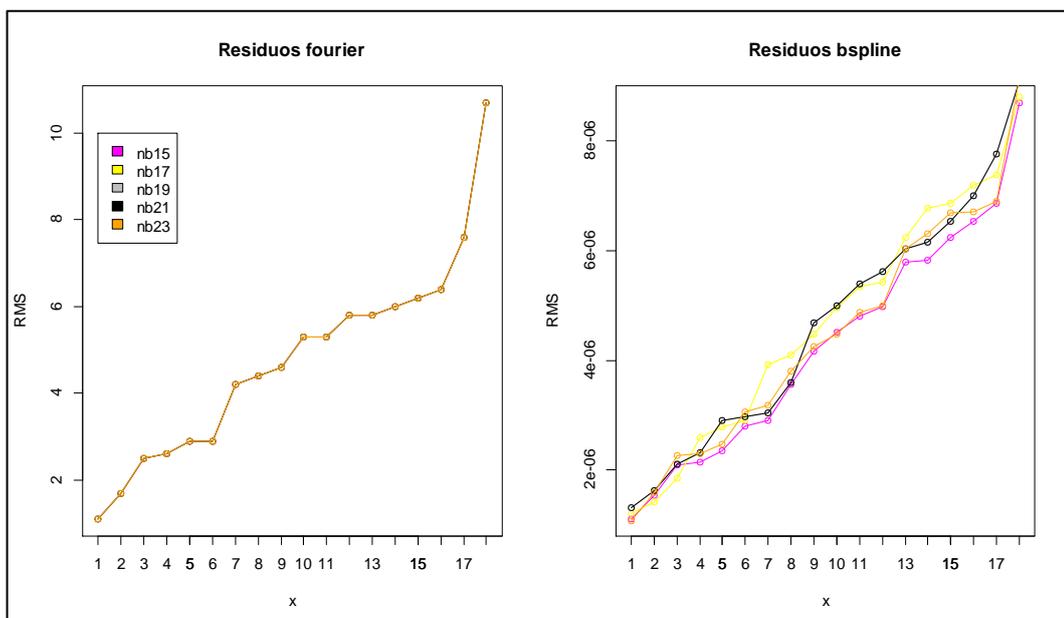


Gráfico 5-4: Residuos, B-spline y Fourier, con bases del 5 al 23

Elaborado por: Taguacundo Oscar, 2019

El gráfico muestra que los errores para los B-spline cúbicos con base 15, 17, 19, 21 y 23, son menores que los de Fourier. De esta manera se elige como método de suavizado al B-spline cubico, sin embargo, a simple vista no se identifica el número de bases óptimo, por lo que, se calcula el promedio del error cuadrático medio.

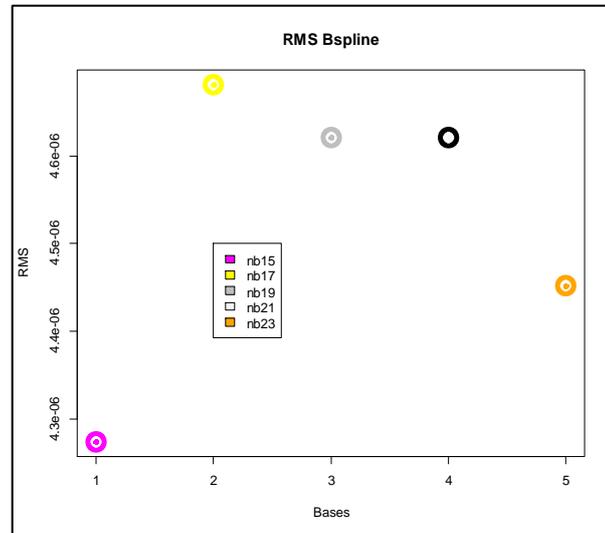


Gráfico 6-4: Residuo promedio por número de bases.

Elaborado por: Taguacundo Oscar, 2019

En el gráfico se observa que el error cuadrático medio con un ajuste B-spline de 15 bases es el óptimo con un RMS= 4.273653e-06 o RMS= 0,000004273653.

4.3 Suavizado de curvas y detección de curvas atípicas

Se realizó el cambio de los datos discretos a los datos funcionales con la ayuda de las librerías *fda* y *fda.usc*, mediante la función *min.basis()* con el criterio de validación cruzada generalizada, con un error cuadrático medio de 4.273653e-06 se elige el método de suavizado B-spline con 15 bases funcionales, la cual será utilizada en toda la investigación.

Se creó la base B-spline cúbica con la función *create.bsplie.basis()*.

```
argm <- 1:nrow(alao2014mt)
```

```
bsrsg <- create.bspline.basis(rangeval=c(min(argm), max(argm)),
nbasis= opt$numbasis.opt)
```

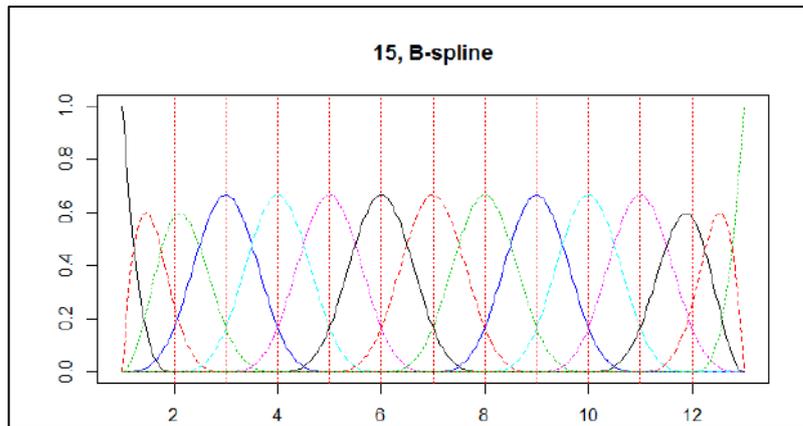


Gráfico 7-4: Numero de bases B-spline óptimos

Elaborado por: Taguacundo Oscar, 2019

Se suavizó los datos de radiación solar global a cada una de las estaciones en los diferentes años de estudio y con la función *Data2fd()* se obtuvo los datos funcionales suavizados.

```
fdalao2014 <- Data2fd(alao2014mt, argvals = argm, bsrsg)
fdalao2015 <- Data2fd(alao2015mt, argvals = argm, bsrsg)
fdalao2016 <- Data2fd(alao2016mt, argvals = argm, bsrsg)
fdalao2017 <- Data2fd(alao2017mt, argvals = argm, bsrsg)
```

Utilizando la función *outliers.depth.trim()* se analizó las curvas atípicas mediante remuestras bootstrap, considerando la profundidad al cuantil 0.5 (mediana), con un $\alpha = 0.01$, implementando el siguiente código en Rstudio para la estación de Alao 2014, se obtuvo las curvas atípicas, las cuales hacen referencia a los días del año en secuencia.

```
outalao2014 <- outliers.depth.trim(fdataobj = fdalao2014, nb = 200,
smo = 0.05, trim = 0.01, quan = 0.5, dfunc=depth.mode, ns = 0.01)

oalao2014 <- outalao2014$outliers

sepalao2014 <- strsplit(oalao2014, split = "rep" )
colalao2014 <- as.integer(unlist(sepalao2014))
opalao2014 <- na.omit(colalao2014)

opalao2014

[1] 61 243 268 300 301 312 337 47 87 313 347 14 91 303 348 88 323
[18] 317
```

Se encontró un total de 18 atípicos para la estación de Alao en el año 2014: día 61 muestra al 02 de marzo del 2014, 14 hace referencia al 14 de enero del 2014, 47 indico 16 de febrero del 2014 y así sucesivamente, concluyendo que se tienen un dato atípico en enero uno en febrero, 2 marzo, 2 de agosto, 2 de octubre, 5 de noviembre y 3 datos de diciembre. Este procedimiento se realizó a cada una de las estaciones meteorológicas con el fin de identificar comportamiento atípico y reducir el sesgo en los resultados.

La siguiente lista de gráficos contiene las curvas diarias suavizadas de cada estación meteorológica con sus respectivas curvas atípicas.

```
plot(fdalao2014, main = " RSG, Alao 2014", las= 1, col =  
"yellowgreen", xlab = "Hora", ylab = TeX(" Radiación Solar Global,  
$\w$/\m^2$" ), cex.lab = 0.8)
```

```
plot(fdalao2015, main = " RSG, Alao 2015", las= 1, col =  
"yellowgreen", xlab = "Hora", ylab = TeX(" Radiación Solar Global,  
$\w$/\m^2$" ), cex.lab = 0.8)
```

```
plot(fdalao2016, main = " RSG, Alao 2016", las= 1, col =  
"yellowgreen", xlab = "Hora", ylab = TeX(" Radiación Solar Global,  
$\w$/\m^2$" ), cex.lab = 0.8)
```

```
plot(fdalao2017, main = " RSG, Alao 2017", las= 1, col =  
"yellowgreen", xlab = "Hora", ylab = TeX(" Radiación Solar Global,  
$\w$/\m^2$" ), cex.lab = 0.8)
```

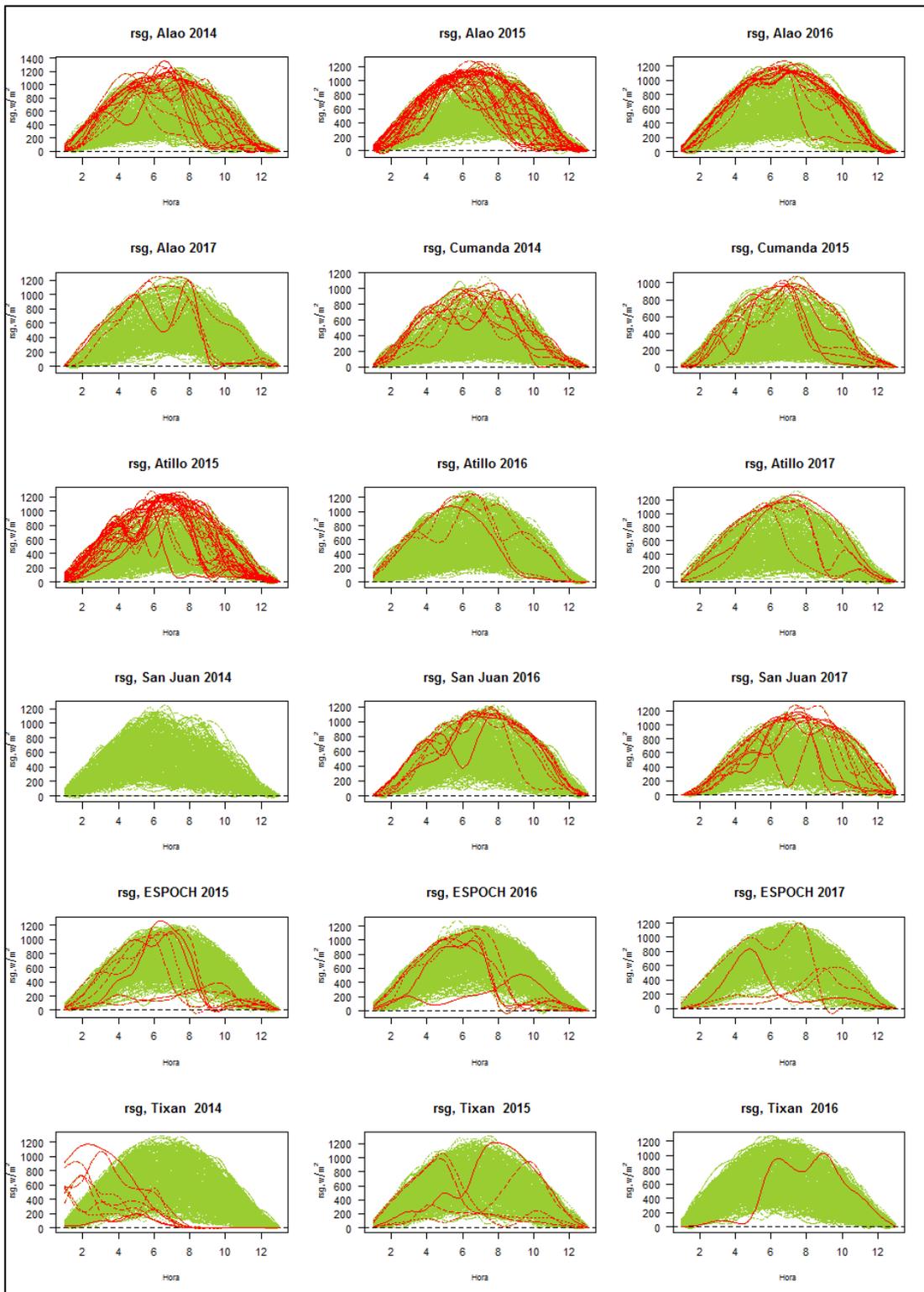


Gráfico 8-4: Curvas suavizadas de 11 estaciones por año con sus respectivas curvas atípicas

Elaborado por: Taguacundo Oscar, 2019.

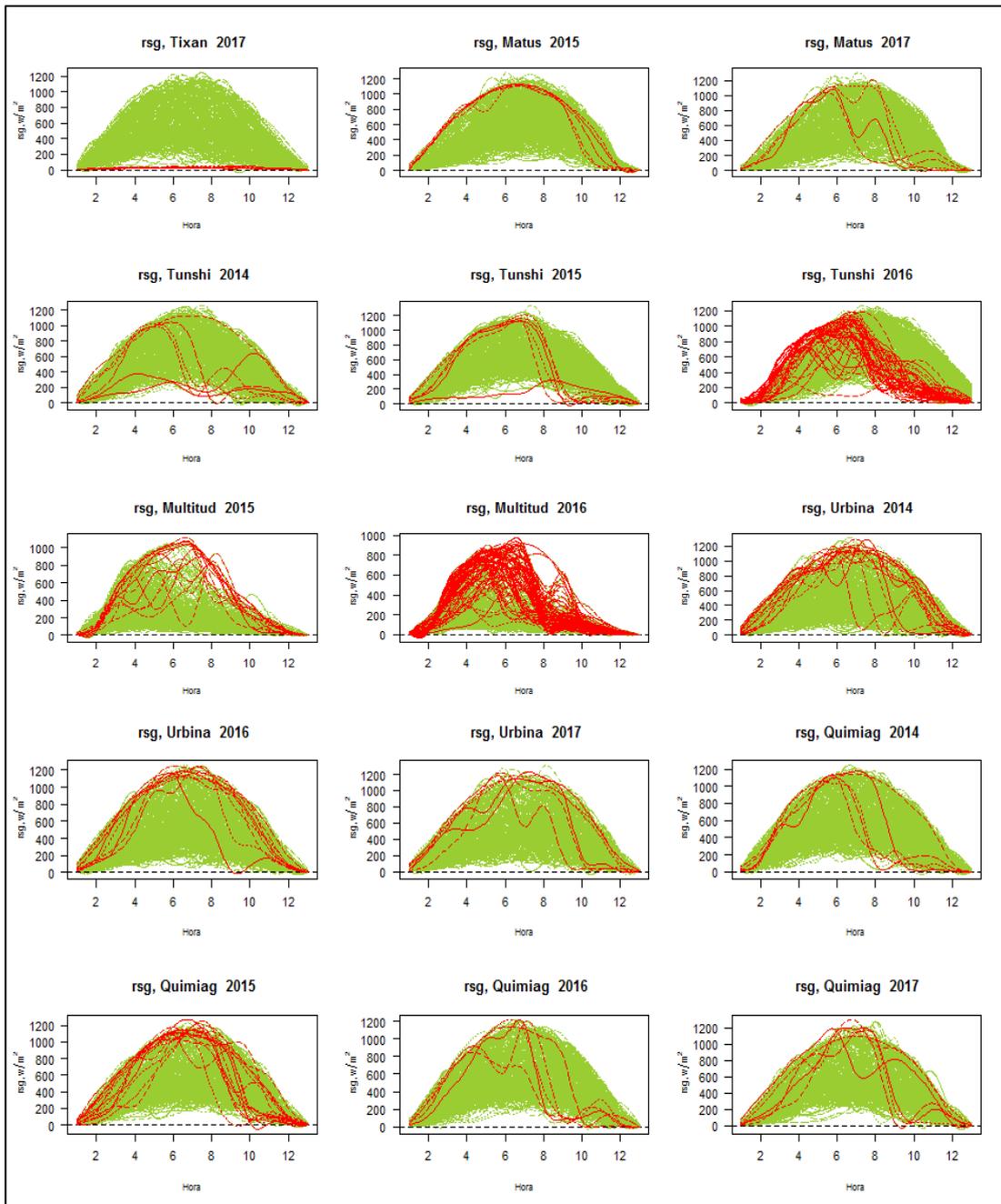


Gráfico 9-4: Curvas suavizadas de 11 estaciones por año con sus respectivas curvas atípicas

Elaborado por: Taguacundo Oscar, 2019.

Cada uno de los gráficos representa curvas diarias que reflejan la evolución de la radiación solar global durante el día, en general la rsg alcanza alrededor de 1200 W/m^2 , sin embargo en Multitud 2016 y Cumanda 2014, 2015 están alrededor de los 1000 W/m^2 , además se observó que la estación Alao 2014 registro rsg que sobrepasa los 1300 W/m^2 .

Las curvas atípicas identificadas se muestran en color rojo, en el 2015 Alao y Atillo, y en el 2016 Tunshi y Multitud tienen una mayor concentración de atípicos, San Juan 2014 no tiene atípicos y

en Tixan 2014, 2017 se observa comportamiento anormal (improbables) en las curvas atípicas. Por tal motivo se tomó la decisión de retirar dichas curvas del análisis de datos.

4.4 Análisis exploratorio de datos

4.4.1 Análisis estadístico descriptivo funcional de la radiación solar global

4.4.1.1 Estación meteorológica Alao

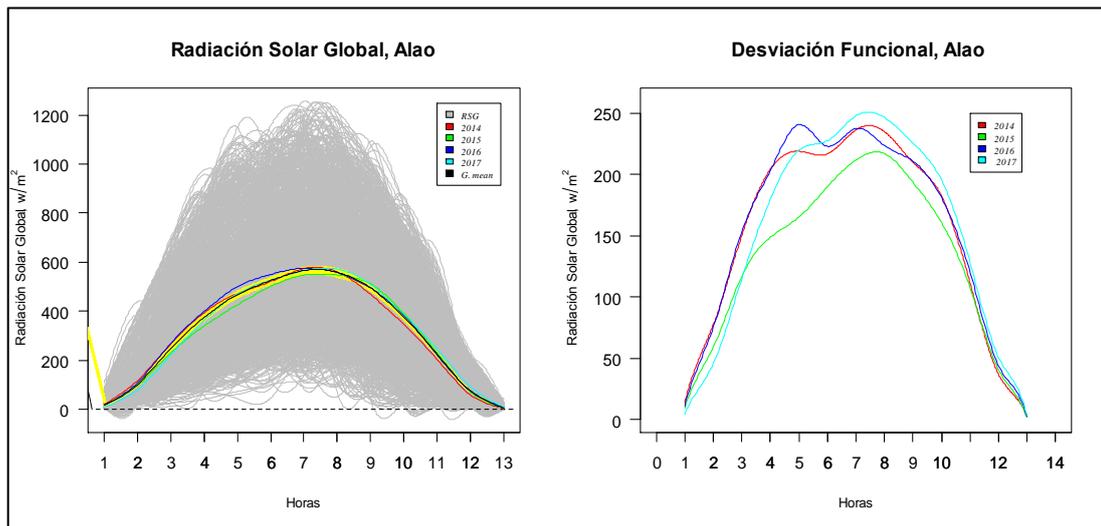


Gráfico 10-4: Media y Desviación Funcional por año, Alao

Elaborado por: Taguacundo Oscar, 2019.

El gráfico izquierdo representa el suavizado de datos con sus respectivas curvas medias e indica la radiación promedio durante el día, el comportamiento de la rsg es muy notoria, al inicio del día no existe una alta concentración, mientras que al medio día se observa un alcance de $1200 W/m^2$, y antes de finalizar el día la radiación disminuye, la desviación estándar funcional en dicho sector muestra poca variabilidad con respecto a su media funcional en el cual su curva se asemeja a una campana, sin embargo, durante el año 2014 y 2016 existe inestabilidad de 10h00 y 15h00.

4.4.1.2 Estación meteorológica Cumanda

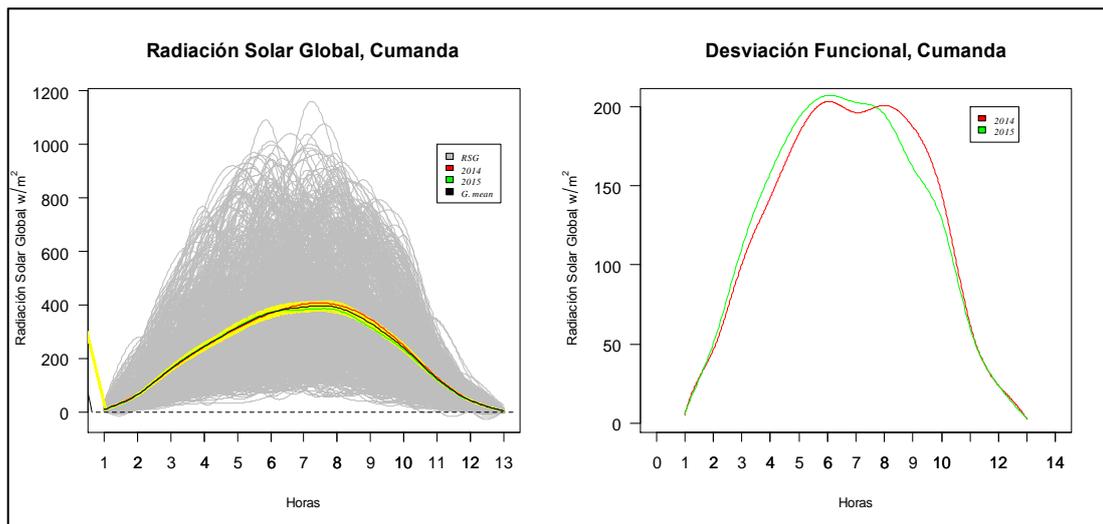


Gráfico 11-4: Media y desviación funcional, Cumanda

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda indica el avance de la radiación solar global en la estación Cumanda, se observa que al inicio del día no existe una alta concentración de calor, mientras que al medio día llega hasta los $1200 W/m^2$ se observa una alta radiación, y antes de finalizar el día la radiación disminuye significativamente, la desviación estándar funcional se indica en la ilustración derecha y se notó poca variabilidad con respecto a su media funcional, su curva se asemeja a una campana, sin embargo, durante el año 2014 existe una variabilidad significativa durante las 11h00 y 14h00.

4.4.1.3 Estación meteorológica Atillo

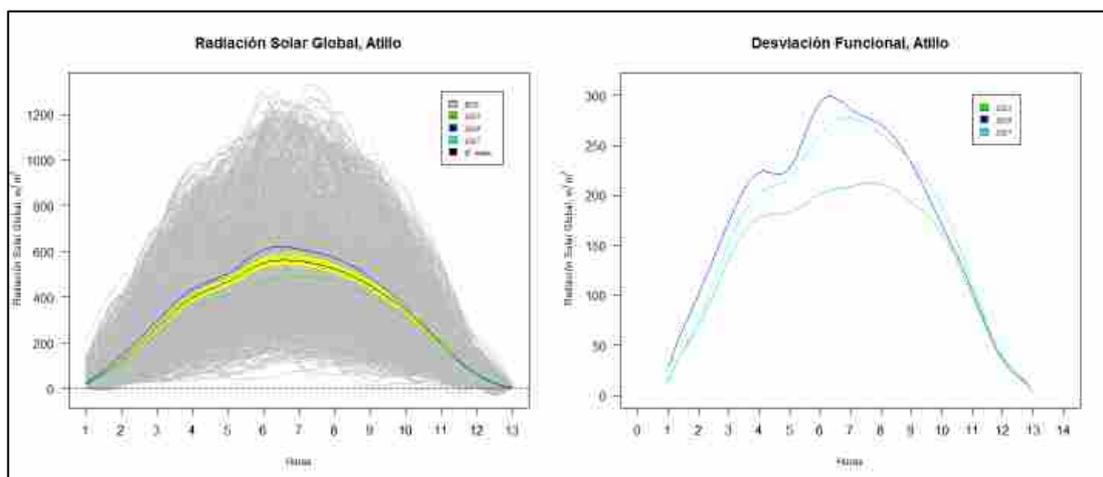


Gráfico 12-4: Media y desviación funcional por año, Atillo

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda representa el avance de la radiación solar global en la estación Atillo, el comportamiento de la rsg es muy notoria y se observó que al inicio del día no existe una alta concentración para los años 2015, 2016, 2017, y antes de finalizar el día la radiación disminuye significativamente. A la gráfica de la derecha muestra la desviación estándar funcional, se notó poca variabilidad con respecto a su media funcional, mientras que, durante los años 2015, 2016, 2017, tiene una pequeña variación entre 11h00 y 14h00 horas.

4.4.1.4 Estación meteorológica San Juan

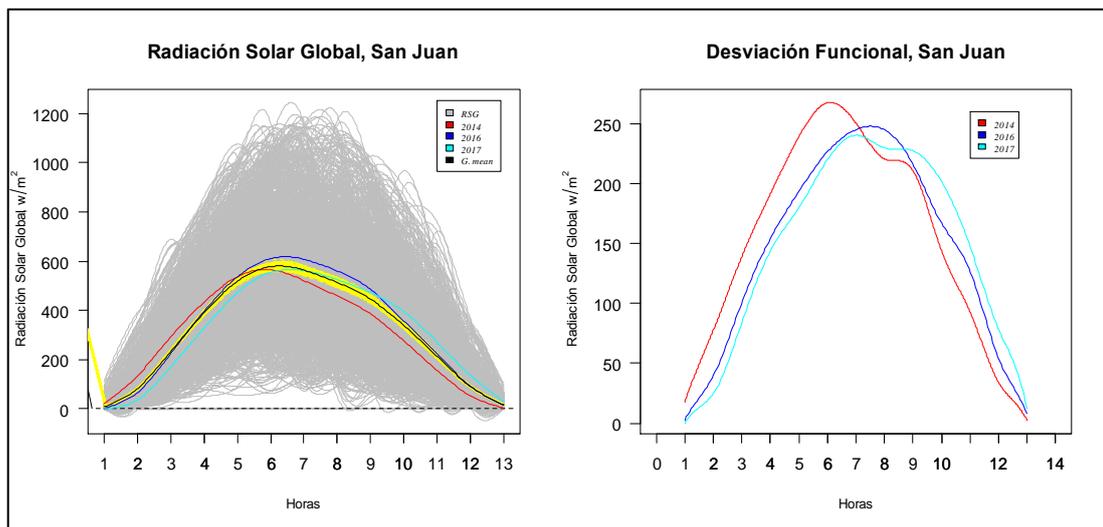


Gráfico 13-4: Media y desviación funcional por año, San Juan

Elaborado por: Taguacundo Oscar, 2019

El gráfico del lado izquierdo, indica la radiación solar global en la estación San Juan, el comportamiento de la rsg es muy notoria y se observa que al inicio del día no existe una alta concentración, mientras que al medio día la rsg alcanza los $1200 W/m^2$, y antes de finalizar el día esta disminuye significativamente. La gráfica de la desviación estándar funcional indica poca variabilidad con respecto a su media funcional, sin embargo, durante el año 2014 existe una variabilidad significativa durante las 11h00 y 14h00.

4.4.1.5 Estación meteorológica ESPOCH

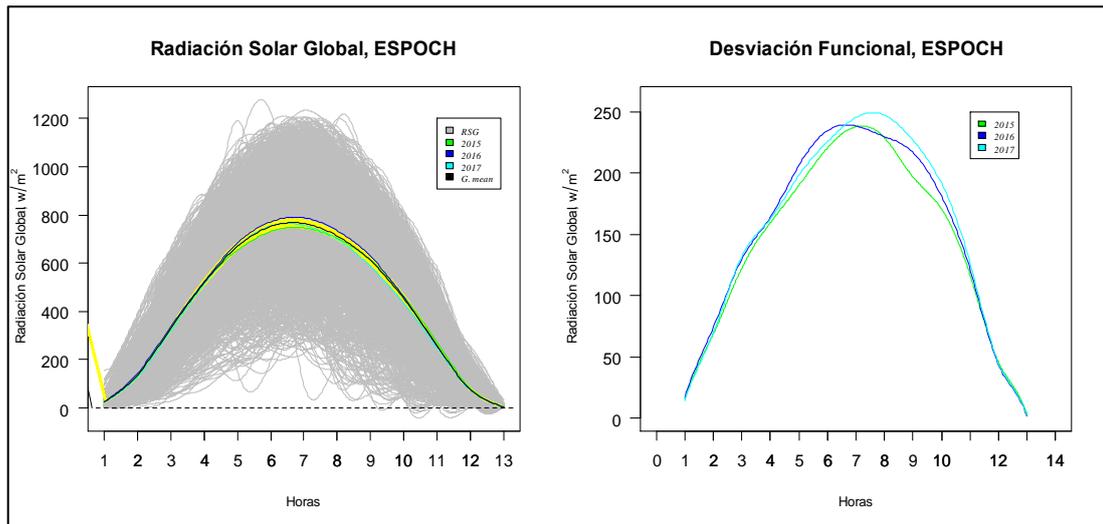


Gráfico 14-4: Media y desviación funcional por año, ESPOCH

Elaborado por: Taguacundo Oscar, 2019

La ilustración de la izquierda indica el avance de la radiación solar global en la estación ESPOCH, se observa que al inicio del día no existe una alta concentración, mientras que al medio día se observa una alta radiación, y antes de finalizar el día la radiación disminuye significativamente, sin embargo, La desviación estándar funcional indica poca variabilidad con respecto a su media funcional, ésta se asemeja a una campana, la cual describe una variabilidad significativa durante las 11h00 y 14h00.

4.4.1.6 Estación meteorológica Tixan

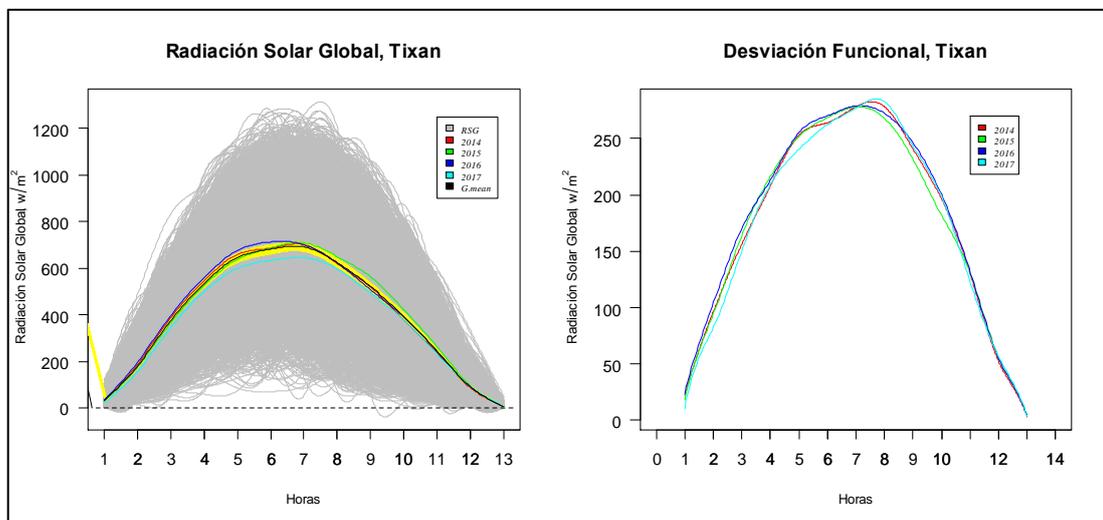


Gráfico 15-4: Media y desviación funcional por año, Tixan

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda indica el avance de la radiación solar global en la estación Tixan, se observa que al inicio del día la rsg es baja y se observó una alta concentración de radiación durante el mediodía, al finalizar el día la radiación disminuye significativamente. La desviación estándar funcional mostró poca variabilidad durante las 10h00 hasta 13h00 con respecto a su media funcional.

4.4.1.7 Estación meteorológica Matus

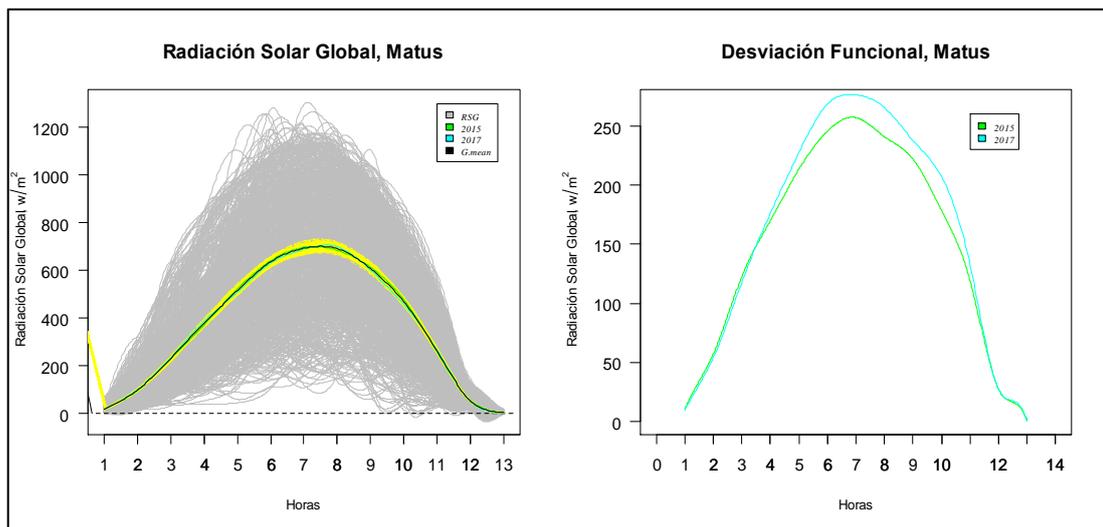


Gráfico 16-4: Media y desviación funcional por año, Matus

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda indica el avance de la radiación solar global en la estación Matus, se observa que al inicio del día no existe una alta concentración, mientras que al medio día se observa una alta radiación, y antes de finalizar el día esta disminuye significativamente. La desviación estándar funcional mostró poca variabilidad con respecto a su media funcional, su curva se asemeja a una campana, sin embargo, durante el año 2017 existe una variabilidad significativa durante las 11h00 y 14h00.

4.4.1.8 Estación meteorológica Tunshi

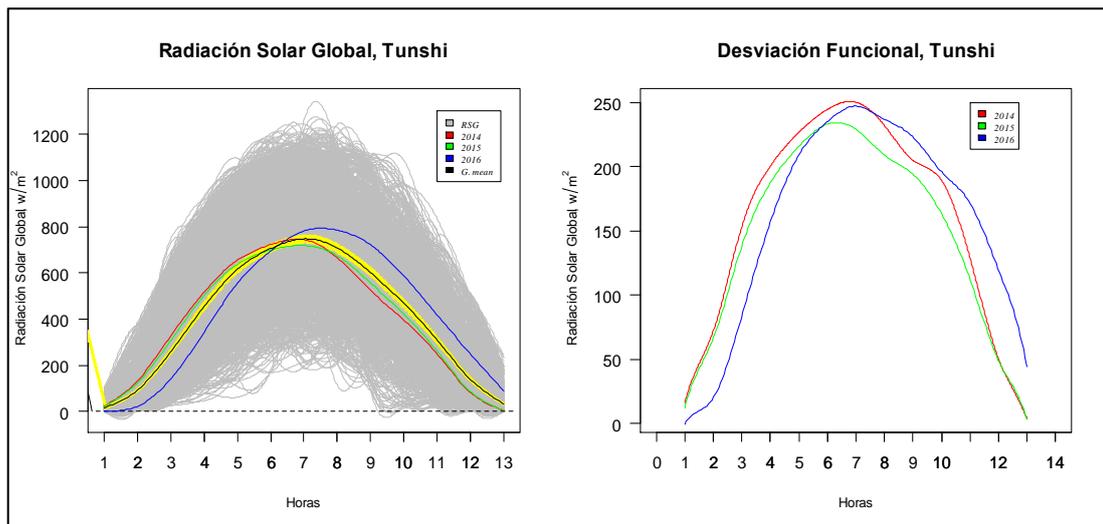


Gráfico 17-4: Media y desviación funcional por año, Tunshi

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda muestra el avance de la radiación solar global en la estación Tunshi, se observa que al inicio del día no existe una alta concentración, mientras que al medio día se observa una alta radiación sin embargo hay pocos días donde la intensidad solar es muy baja, y antes de finalizar el día la radiación disminuye significativamente. La desviación estándar funcional en Tunshi, indica poca variabilidad con respecto a su media funcional, su curva se asemeja a una campana y existe una variabilidad significativa durante las 11h00 y 14h00.

4.4.1.9 Estación meteorológica Multitud

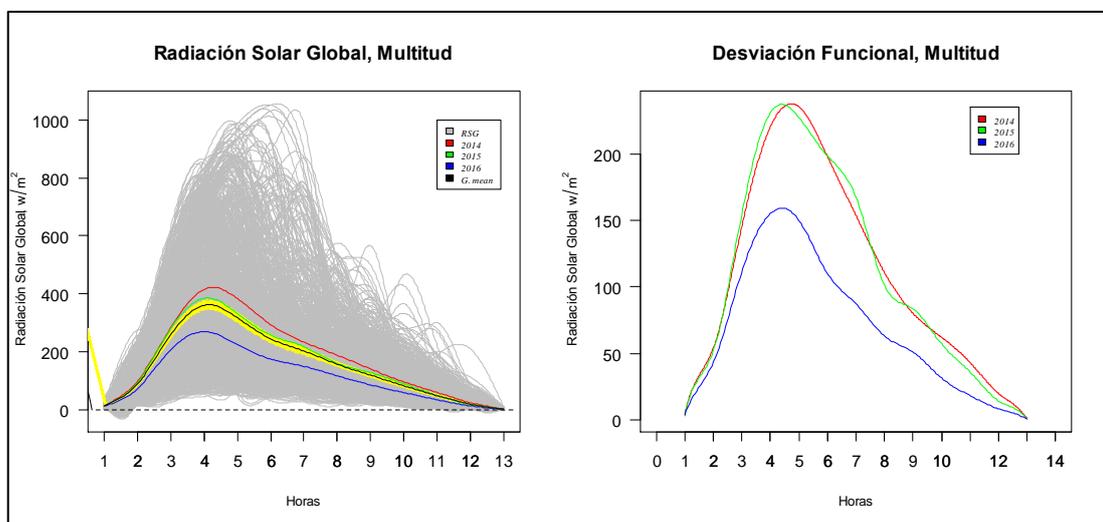


Gráfico 18-4: Media y desviación funcional por año, Multitud

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda muestra el avance de la radiación solar global en la estación Multitud, se observa que al inicio del día no existe una alta concentración de calor, mientras que desde las 09h00 hasta 15h00 existe una gran radiación solar en esta zona. La desviación estándar funcional mostró poca variabilidad con respecto a su media funcional, con una intensidad de rsg baja en el 2016, además la variabilidad durante los 3 años es similar en forma.

4.4.1.10 Estación meteorológica Urbina

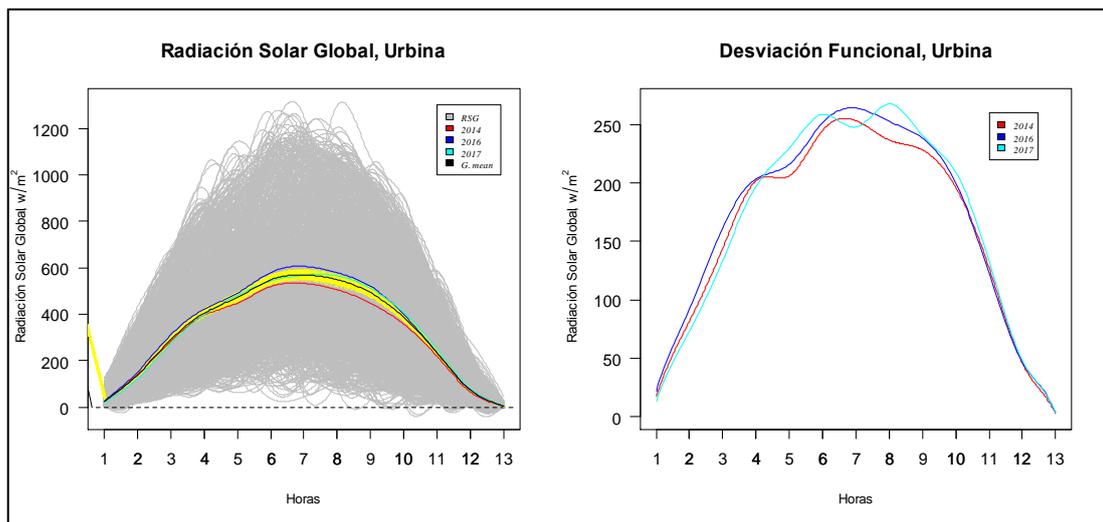


Gráfico 19-4: Media y desviación funcional por año, Urbina

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda indica el avance de la radiación solar global en la estación Urbina, se observa que al inicio del día existe poca intensidad, mientras que al medio día se observa una alta radiación, y antes de finalizar el día esta disminuye significativamente, sin embargo, existen unas cuantas curvas que están por debajo de cero esto es causado por el ajuste de los B-spline. La desviación estándar funcional mostró poca variabilidad con respecto a su media funcional, su curva se asemeja a una campana, sin embargo, durante el año 2017 existe una variabilidad significativa durante las 11h00 y 14h00.

4.4.1.11 Estación meteorológica Quimiag

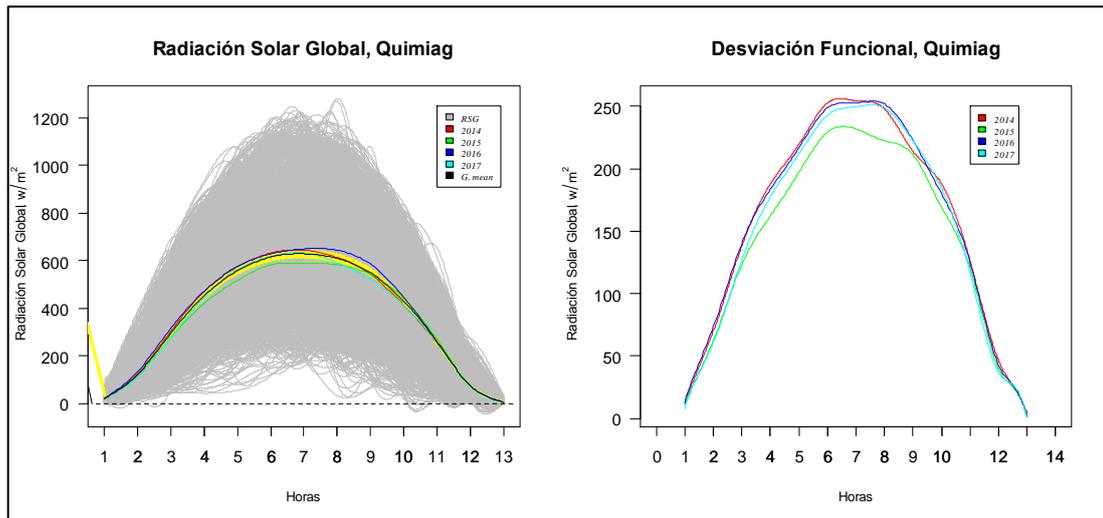


Gráfico 20-4: Media y desviación funcional por año, Quimiag

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda muestra el avance de la radiación solar global en la estación Quimiag, donde se observó que al inicio del día no existe una alta concentración, mientras que al medio día se observa una alta radiación, y antes de finalizar el día esta disminuye significativamente, sin embargo. La desviación estándar funcional mostró poca variabilidad con respecto a su media funcional, además la variabilidad es mínima durante las 10h00 hasta 15h00 en cada uno de los años estudiados.

4.4.2 Análisis de Varianza funcional

4.4.2.1 Curvas de nivel

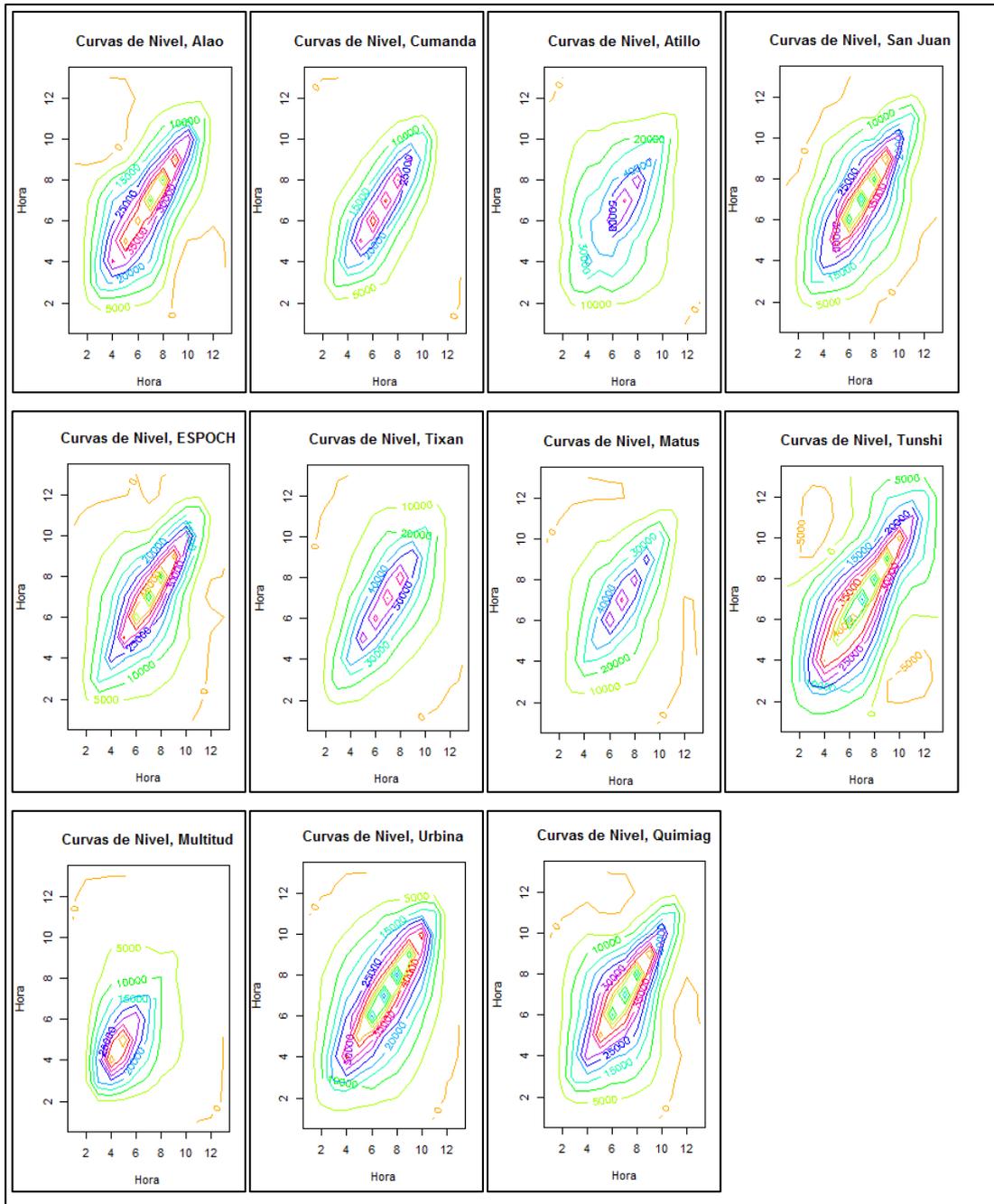


Gráfico 21-4: Variación de la rsg según estación

Elaborado por: Taguacundo Oscar, 2019

Las curvas de nivel indica la variación de la rsg según su coloración, las curvas naranjas denotan una mínima radiación en la superficie y el color rojo denota una magnitud significativa de luz solar, se observó el comportamiento de la varianza en cada una de las estaciones analizadas donde

sus pequeños círculos en el centro de las curvas representan la variación, cada una de ellas tienen un comportamiento según la posición geográfica donde está ubicada la estación meteorológica, esto se notó en Multitud que tiene una alta concentración de radiación en la mañana la cual se reduce en la tarde (Gráfico 21-4).

4.4.2.2 Análisis Raster

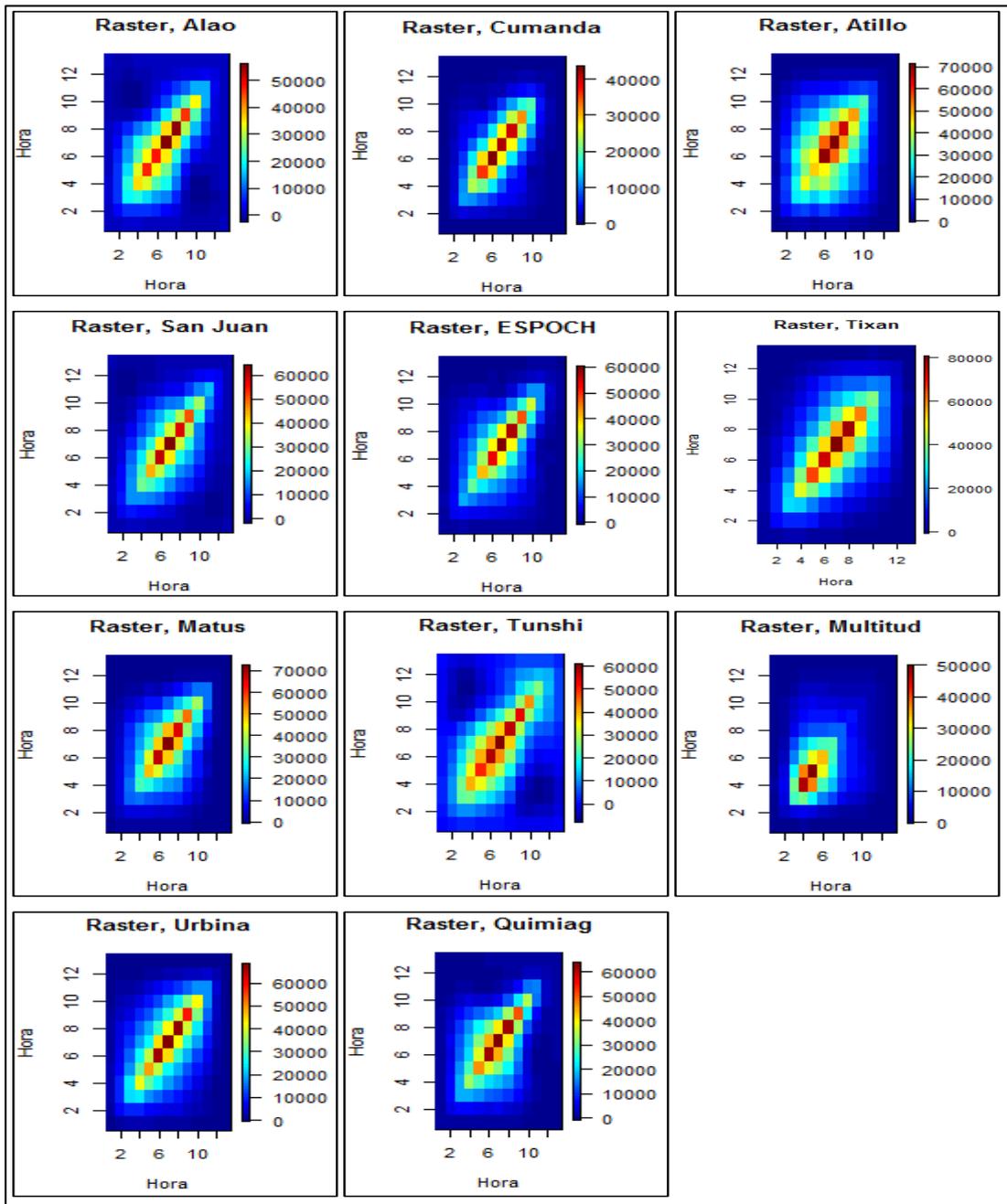


Gráfico 22-4: Raster de superficie por estación

Elaborado por: Taguacundo Oscar, 2019

El gráfico raster mejoró lo apreciado en las curvas de nivel, los detalles de la escala de colores muestran en cada una de sus gráficas, el azul claro indica la variación mínima y la parte roja la máxima variación existente por cada estación, en este caso la superficie de mayor variación es pequeña, lo cual concuerda con el comportamiento de la rsg durante el día, esta se da de 10h00 a 15h00 que concuerdan como máximas.

4.4.3 Análisis de Componentes principales funcionales

En esta sección se presenta el análisis de componentes principales, con el fin de reducir la dimensionalidad de la rsg en Chimborazo por mes y estación. Para dicho análisis se consideró la utilización de graficas de la primera y segunda componente, en donde se identificó en color azul los meses de la época lluviosa y de color naranja los meses considerados de época seca.

En la tabla de valores y funciones propias muestran las magnitudes (valores y funciones propias, y variabilidad explicada por cada componente) para los primeros 5 B-spline, cada uno de estos valores están asociados al número de bases óptimos (15 B-spline).

4.4.3.1 Estación meteorológica, Alao

Tabla 3-4: Valores propios de la estación Alao

	2014	2015	2016	2017
bspl4.1	42129,37	39604,38	73943,07	11955,39
bspl4.2	6740,14	5660,96	8114,22	3964,27
bspl4.3	1066,65	1718,85	1277,63	1034,14
bspl4.4	866,65	982,34	990,64	923,54
bspl4.5	535,06	539,85	734,91	574,88

Elaborado por: Taguacundo Oscar, 2019

Tabla 4-4: Funciones propias de la estación Alao 2014, 2015, 2016, 2017

	2014		2015		2016		2017	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	0,05	-0,03	0,02	-0,09	0,03	-0,04	0,02	0,01
bspl4.2	0,10	-0,04	0,07	-0,13	0,08	-0,10	0,07	-0,01
bspl4.3	0,21	-0,05	0,16	-0,22	0,19	-0,23	0,17	-0,05
bspl4.4	0,31	0,09	0,24	-0,35	0,37	-0,13	0,34	-0,12
bspl4.5	0,45	-0,12	0,35	-0,25	0,41	-0,23	0,16	-0,15

Elaborado por: Taguacundo Oscar, 2019

La primera componentes explica el 81%, 81%, 86% y 68% para los años 2014, 2015, 2016, 2017 respectivamente, de la misma forma la segunda componente explica el 13%, 12% 9% y 21 % de la variabilidad total para cada año (Tabla 5-4).

Tabla 5-4: Variabilidad explicada por cada componente Alao por año

	2014	2015	2016	2017
λ_1	0,81	0,81	0,86	0,68
λ_2	0,13	0,12	0,09	0,21
λ_3	0,02	0,04	0,01	0,04
λ_4	0,02	0,02	0,01	0,03

Elaborado por: Taguacundo Oscar, 2019

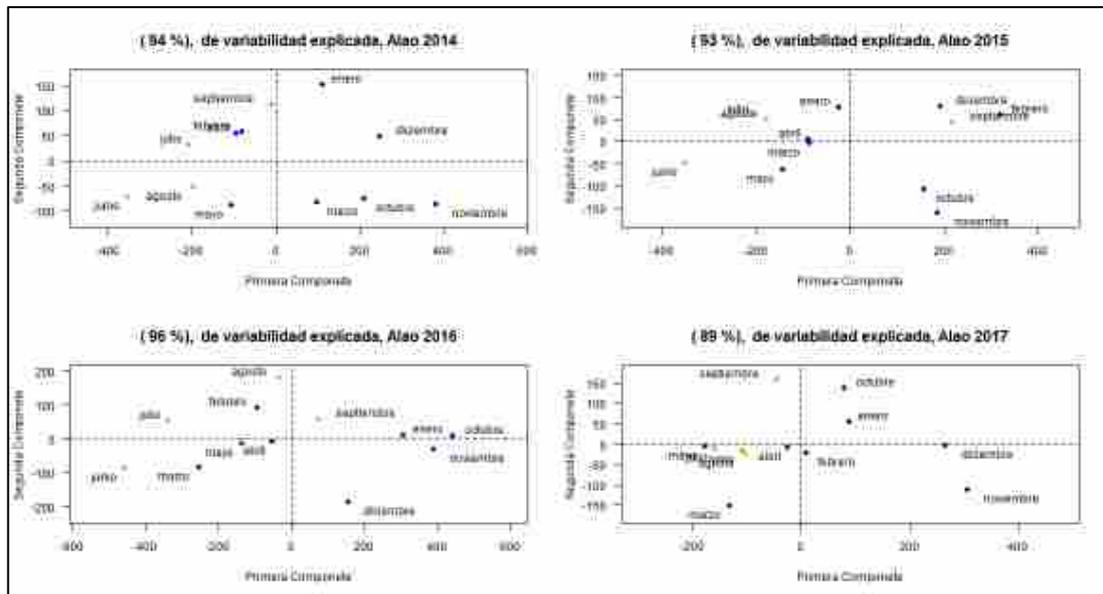


Gráfico 23-4: Primera y segunda componente principal funcional de rsg, Alao

Elaborado por: Taguacundo Oscar, 2019

El gráfico superior izquierdo contiene la agrupación de la estación Alao del año 2014, donde se observó que forma parte de la primera componente los meses enero, marzo, octubre, noviembre y diciembre, y para la segunda componte febrero, abril, mayo, junio, julio, agosto y septiembre, cada componente contiene meses con características similares. De la misma manera se detallan las agrupaciones para 2015, 2016, 2017.

4.4.3.2 Estación meteorológica, Cumanda

Tabla 6-4: Valores propios de la estación Cumanda

	2014	2015
bspl4.1	18520,94	13304,29
bspl4.2	4756,98	6685,24
bspl4.3	545,61	1064,61
bspl4.4	323,50	543,40
bspl4.5	138,48	216,32

Elaborado por: Taguacundo Oscar, 2019

Tabla 7-4: Funciones propias de la estación Cumanda 2014, 2015

	2014		2015	
	PC1	PC2	PC1	PC2
bspl4.1	0,00	0,01	0,00	0,01
bspl4.2	0,02	-0,01	0,02	0,05
bspl4.3	0,05	-0,04	0,06	0,15
bspl4.4	0,08	-0,04	0,26	0,19
bspl4.5	0,46	-0,11	0,48	-0,11

Elaborado por: Taguacundo Oscar, 2019

La primera componente del año 2014 explica el 76% de la variabilidad total y la segunda explica el 19%, las dos primeras componentes explican aproximadamente el 95% de la variabilidad total, la primera componentes del año 2015 explica el 60% y la segunda componentes explica el 30% de la variabilidad total de todas las variables, las dos primeras componentes juntas explican el 90% de la variabilidad total.

Tabla 8-4: Variabilidad explicada por componente Cumanda por año

	2014	2015
λ_1	0,76	0,60
λ_2	0,19	0,30
λ_3	0,02	0,05
λ_4	0,01	0,02

Elaborado por: Taguacundo Oscar, 2019

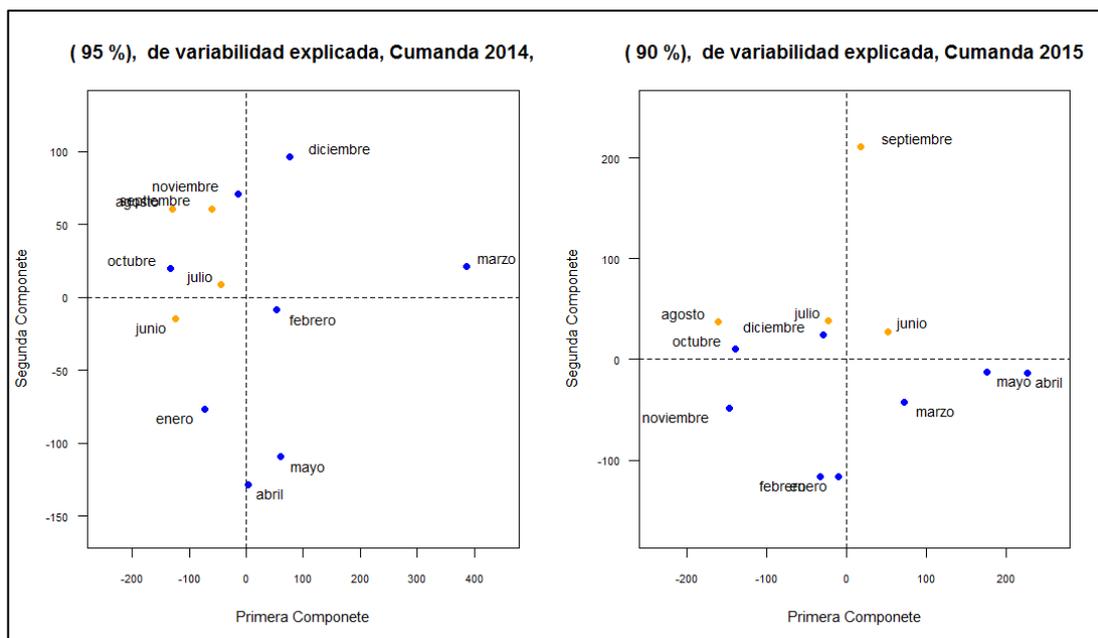


Gráfico 24-4: Primera y segunda componente principal funcional de rsg, Cumanda

Elaborado por: Taguacundo Oscar, 2019

El gráfico de la izquierda muestra la primera y segunda componente principal del año 2014, donde se identificó al mes de marzo quien se comporta diferente a los demás meses, por tanto, enero, febrero, abril, mayo, junio, julio, agosto, septiembre, octubre, noviembre y diciembre, tienen un comportamiento similar y en 2015 marzo, abril, mayo y junio.

4.4.3.3 Estación meteorológica, Atillo

Tabla 9-4: Valores propios de la estación Atillo

	2015	2016	2017
bspl4.1	30860,35	91443,58	12116,29
bspl4.2	7064,90	3887,83	1840,11
bspl4.3	1226,79	2056,78	936,79
bspl4.4	959,72	1007,83	650,82
bspl4.5	421,56	721,89	304,16

Elaborado por: Taguacundo Oscar, 2019

Tabla 10-4: Funciones propias de la estación Atillo 2015, 2016, 2017

	2015		2016		2017	
	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	0,04	-0,11	0,05	-0,03	-0,01	0,00

bsp14.2	0,05	-0,17	0,09	-0,04	-0,01	0,03
bsp14.3	0,09	-0,29	0,17	-0,05	0,00	0,11
bsp14.4	0,18	-0,48	0,30	-0,08	0,08	0,47
bsp14.5	0,42	-0,38	0,40	-0,07	0,41	0,44

Elaborado por: Taguacundo Oscar, 2019

La tabla muestra la variabilidad explicada por cada componente, el 75% de variabilidad explicada para el 2015, el 92% de variabilidad en 2016, el 75% de variabilidad para el 2017, las dos primeras componentes explican el 92%, 96%, 86% de la variabilidad total de la información, para los años 2015, 2016, 2017 respectivamente.

Tabla 11-4: Variabilidad explicada por cada función propia, Atillo

	2015	2016	2017
λ_1	0,75	0,92	0,75
λ_2	0,17	0,04	0,11
λ_3	0,03	0,02	0,06
λ_4	0,02	0,01	0,04

Elaborado por: Taguacundo Oscar, 2019

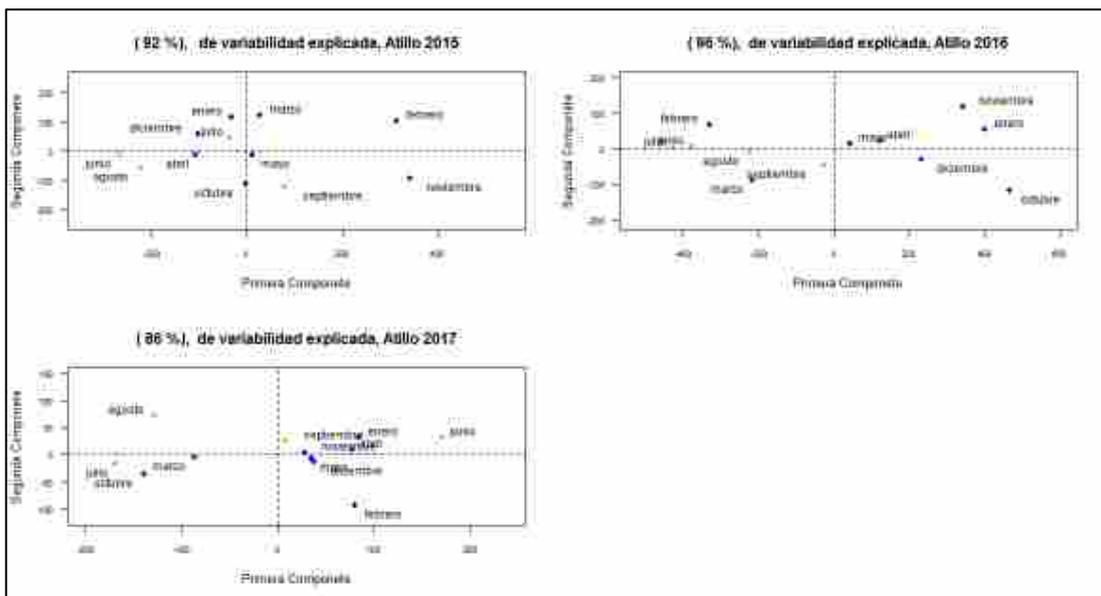


Gráfico 25-4: Primera y segunda componente principal funcional de rsg, Atillo

Elaborado por: Taguacundo Oscar, 2019

Los grupos con características similares de la estación Atillo, para 2015 fueron los meses de febrero, septiembre y noviembre tienen un comportamiento similar, para 2016 enero, abril, mayo,

octubre, noviembre y diciembre presentan comportamientos homogéneos, para 2017 marzo, julio, agosto y octubre tienen características similares.

4.4.3.4 Estación meteorológica, San Juan

Tabla 12-4: Valores propios de la estación San Juan

	2014	2016	2017
bspl4.1	21572,93	25761,07	11031,39
bspl4.2	2215,86	8344,76	2911,80
bspl4.3	1711,07	4062,00	1887,03
bspl4.4	1310,31	1607,24	1010,18
bspl4.5	514,76	464,37	792,28

Elaborado por: Taguacundo Oscar, 2019

Tabla 13-4: Funciones propias de la estación San Juan 2014, 2016, 2017

	2014		2016		2017	
	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	0,06	0,02	0,01	0,01	0,00	0,00
bspl4.2	0,11	0,05	0,03	0,05	0,00	0,03
bspl4.3	0,22	0,11	0,08	0,13	-0,01	0,08
bspl4.4	0,30	0,62	0,16	0,30	0,11	0,51
bspl4.5	0,43	0,21	0,31	0,32	0,17	0,47

Elaborado por: Taguacundo Oscar, 2019

La proporción de variabilidad explicada por cada componente, es del 77%, 63%, 61% explicada por la primera componente para el año 2014, 2016, 2017, el 85%, 84% y 75% son explicados por las dos primeras componentes respectivamente.

Tabla 14-4: Variabilidad explicada por Componentes, San Juan

	2014	2016	2017
λ_1	0,77	0,63	0,48
λ_2	0,08	0,21	0,27
λ_3	0,06	0,10	0,13
λ_4	0,05	0,04	0,04

Elaborado por: Taguacundo Oscar, 2019

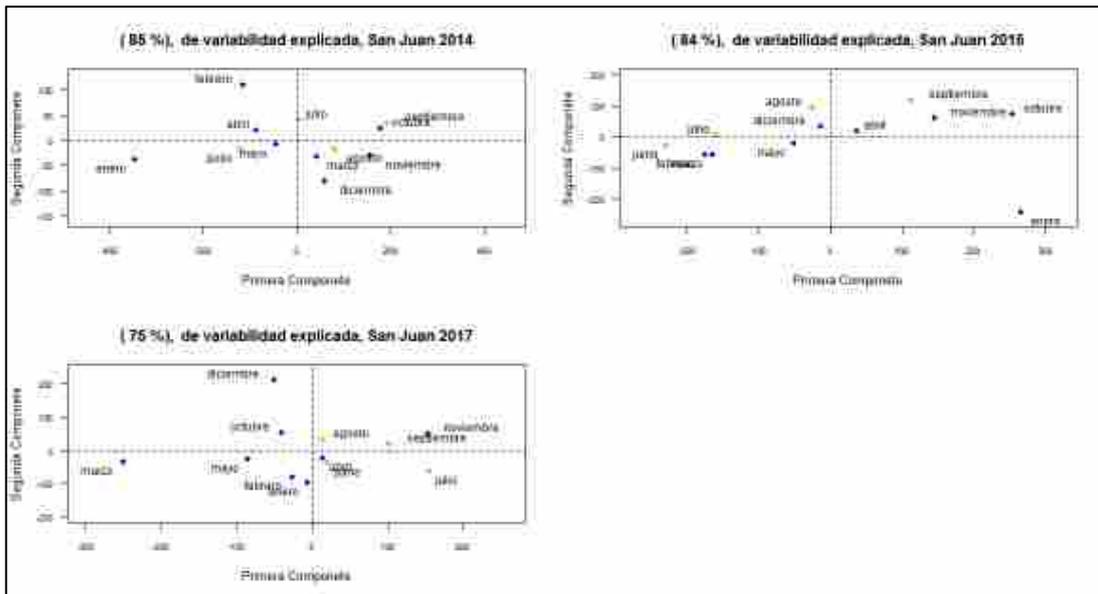


Gráfico 26-4: Primera y segunda componente principal funcional de rsg, San Juan

Elaborado por: Taguacundo Oscar, 2019

Se observó que la primera componente principal está formada por los meses de: marzo, abril, mayo, julio, agosto, septiembre, octubre, noviembre y diciembre tienen comportamiento similar, para el 2016 febrero, marzo, junio, julio, agosto y septiembre, y 2017 enero, febrero, abril, junio, julio, agosto, septiembre y octubre forman el primer grupo, de la misma forma los demás meses forman el segundo grupo respectivamente para cada año en estudio.

4.4.3.5 Estación meteorológica, Espoch

Tabla 15-4: Valores propios de la estación ESPOCH

	2015	2016	2017
bspl4.1	12450,63	26111,36	14698,35
bspl4.2	7211,51	5378,62	5268,67
bspl4.3	2520,92	3292,52	3831,19
bspl4.4	1243,81	947,64	1576,11
bspl4.5	639,50	765,40	482,20

Elaborado por: Taguacundo Oscar, 2019

Tabla 16-4: Funciones propias de la estación ESPOCH 2015, 2016, 2017

	2015		2016		2017	
	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	0,03	0,09	0,06	-0,08	0,05	0,03

bspl4.2	0,07	0,16	0,10	-0,14	0,08	-0,01
bspl4.3	0,16	0,30	0,20	-0,25	0,15	-0,10
bspl4.4	0,19	0,49	0,27	-0,34	0,39	0,00
bspl4.5	0,39	0,49	0,24	-0,38	0,38	-0,29

Elaborado por: Taguacundo Oscar, 2019

La tabla contiene valores de variabilidad explicada por cada componente, el 51%, 71%, 55%, de variabilidad es explicada por la primera componente, y 29%, 14%, 25%, de variabilidad es explicada por la segunda componente, para los años 2015, 2016, 2017 respectivamente.

Tabla 17-4: Variabilidad explicada por Componentes, ESPOCH

	2015	2016	2017
λ_1	0,51	0,71	0,55
λ_2	0,29	0,14	0,25
λ_3	0,10	0,09	0,09
λ_4	0,05	0,03	0,04

Elaborado por: Taguacundo Oscar, 2019

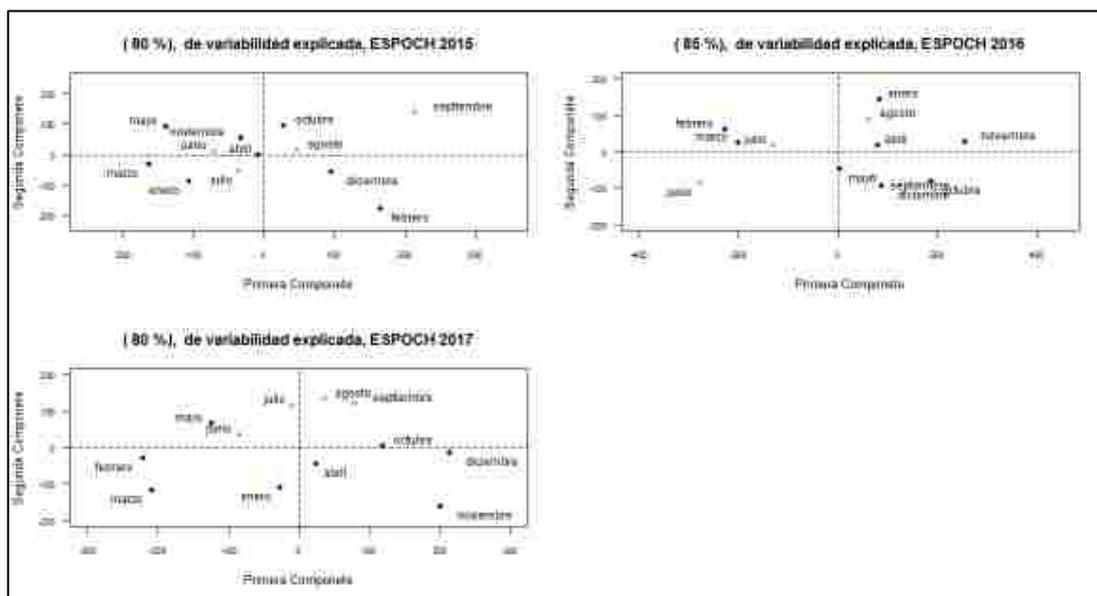


Gráfico 27-4: Primera y segunda componente principal funcional de rsg, ESPOCH

Elaborado por: Taguacundo Oscar, 2019

Se observó que la primera componente principal está formada por los meses de: marzo, abril, mayo, julio, agosto, septiembre, octubre, noviembre y diciembre son de comportamiento similar, para el 2016 febrero, marzo, junio, julio, agosto y septiembre, y 2017 enero, febrero, abril, junio,

julio, agosto, septiembre y octubre forman el primer grupo, los demás meses forman el segundo grupo respectivamente para cada año en estudio.

4.4.3.6 Estación meteorológica, Tixan

Tabla 17-4: Valores propios de la estación Tixan

	2014	2015	2016	2017
bspl4.1	35622,85	45212,45	36432,36	64391,32
bspl4.2	5535,70	12681,77	16711,37	3100,61
bspl4.3	3270,28	2413,78	1825,54	1879,61
bspl4.4	1307,12	1111,81	924,90	1065,44
bspl4.5	493,32	611,60	698,56	512,21

Elaborado por: Taguacundo Oscar, 2019

Tabla 19-4: Funciones propias de la estación Tixan 2014, 2015, 2016, 2017

	2014		2015		2016		2017	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	-0,10	0,14	-0,02	0,12	0,01	0,14	0,00	0,08
bspl4.2	-0,12	0,24	-0,02	0,22	0,03	0,25	0,01	0,16
bspl4.3	-0,16	0,42	-0,03	0,42	0,06	0,49	0,02	0,33
bspl4.4	-0,02	0,76	-0,01	0,38	0,14	0,63	0,10	0,49
bspl4.5	0,07	0,33	0,15	0,73	0,07	0,37	0,10	0,71

Elaborado por: Taguacundo Oscar, 2019

La variabilidad explicada por cada componente, es del 76%, 72%, 64%, 86% de variabilidad es explicada por la primera componente, y 12%, 20%, 29%, 8% de variabilidad es explicada por la segunda componente, para los años 2014, 2015, 2016, 2017 respectivamente.

Tabla 20-4: Variabilidad explicada por Componentes, Tixan

	2014	2015	2016	2017
λ_1	0,76	0,72	0,64	0,86
λ_2	0,12	0,20	0,29	0,08
λ_3	0,07	0,04	0,03	0,03
λ_4	0,03	0,02	0,02	0,02

Elaborado por: Taguacundo Oscar, 2019

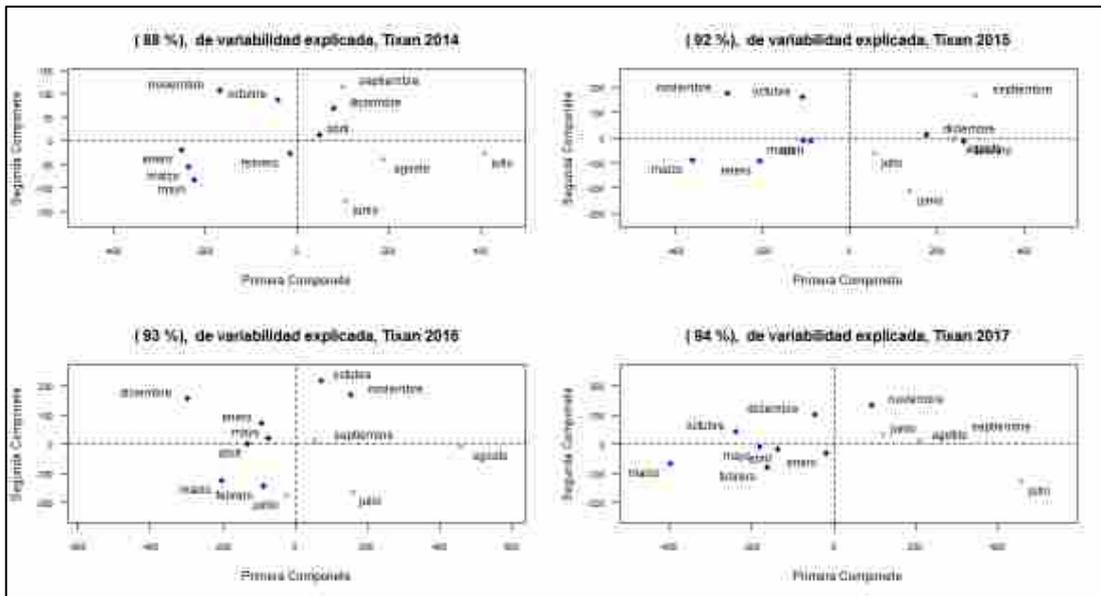


Gráfico 28-4: Primera y segunda componente principal funcional de rsg, Tixan

Elaborado por: Taguacundo Oscar, 2019

Se observó que la primera componente principal está formada por los meses de: enero, febrero, marzo, abril, mayo, octubre y noviembre; enero, marzo, abril, mayo, octubre y noviembre; enero, febrero, marzo, abril, mayo, junio, septiembre, octubre y diciembre; enero, febrero, marzo, abril y mayo, para los años 2014, 2015, 2016, 2017 respectivamente.

4.4.3.7 Estación meteorológica, Matus

Tabla 21-4: Valores propios de la estación Matus

	2015	2017
bspl4.1	26188,34	16543,29
bspl4.2	7129,36	2674,21
bspl4.3	2481,55	1519,12
bspl4.4	502,14	553,62
bspl4.5	394,18	469,23

Elaborado por: Taguacundo Oscar, 2019

Tabla 22-4: Funciones propias de la estación Matus 2015,2017

	2015		2017	
	PC1	PC2	PC1	PC2
bspl4.1	0,02	-0,08	0,02	-0,05
bspl4.2	0,03	-0,15	0,05	-0,08

bspl4.3	0,06	-0,30	0,12	-0,16
bspl4.4	0,22	-0,35	0,19	0,04
bspl4.5	0,28	-0,19	0,36	-0,12

Elaborado por: Taguacundo Oscar, 2019

La tabla contiene valores de variabilidad explicada por cada componente, el 70% y 76% de variabilidad es explicada por la primera componente, 19% y 8% de variabilidad es explicada por la segunda componente, para los años 2015 y 2017 respectivamente.

Tabla 23-4: Variabilidad explicada por Componentes, Matus

	2015	2017
λ_1	0,70	0,76
λ_2	0,19	0,08
λ_3	0,07	0,07
λ_4	0,01	0,05

Elaborado por: Taguacundo Oscar, 2019

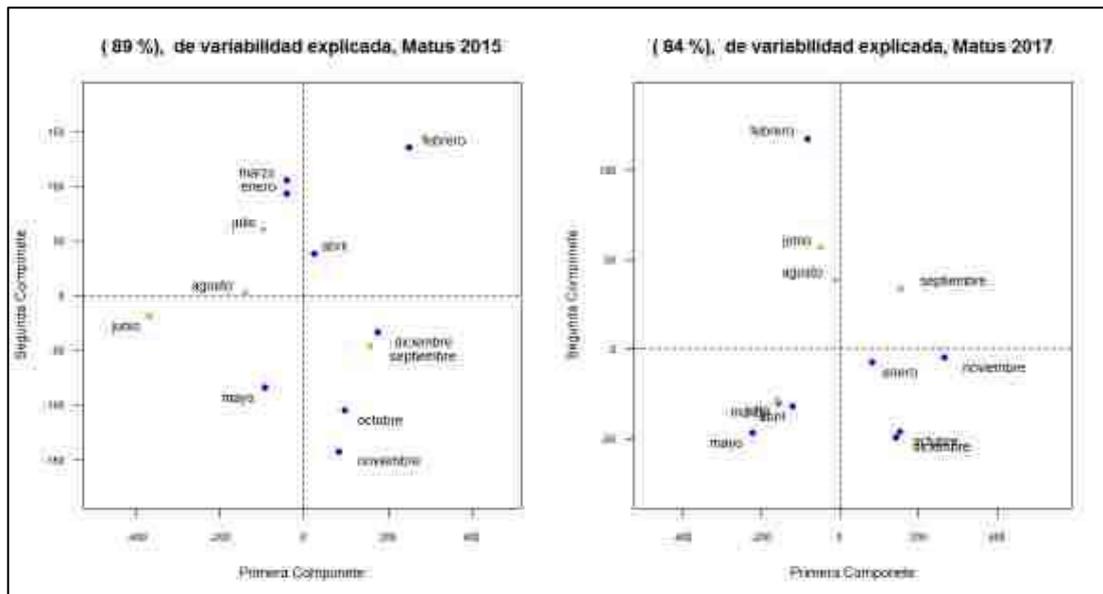


Gráfico 29-4: Primera y segunda componente principal funcional de rsg, Matus

Elaborado por: Taguacundo Oscar, 2019

Se observó que la primera componente principal está formada por los meses de: enero, febrero, marzo, abril, septiembre, octubre, noviembre y diciembre, y febrero, marzo, abril, mayo, junio y julio forman los primeros grupos para los años 2015 y 2017 respectivamente.

4.4.3.8 Estación meteorológica, Tunshi

Tabla 24-4: Valores propios de la estación Tunshi

	2014	2015	2016
bspl4.1	27878,78	21326,78	46003,35
bspl4.2	3319,02	10615,56	5234,19
bspl4.3	1716,24	1349,45	2975,42
bspl4.4	904,32	740,96	1085,38
bspl4.5	666,40	493,64	898,33

Elaborado por: Taguacundo Oscar, 2019

Tabla 25-4: Funciones propias de la estación Tunshi 2014, 2015, 2016

	2014		2015		2016	
	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	0,07	0,06	0,03	-0,08	0,00	0,00
bspl4.2	0,11	0,08	0,05	-0,14	0,01	-0,01
bspl4.3	0,19	0,14	0,11	-0,24	0,03	-0,03
bspl4.4	0,33	0,54	0,30	-0,32	0,18	-0,01
bspl4.5	0,40	0,51	0,44	-0,43	0,42	0,07

Elaborado por: Taguacundo Oscar, 2019

La tabla contiene valores de variabilidad explicada por cada componente, el 80%, 61%, 81%, de variabilidad es explicada por la primera componente, y 9%, 30%, 9%, de variabilidad es explicada por la segunda componente, para los años 2014, 2015, 2016 respectivamente.

Tabla 26-4: Variabilidad explicada por Componentes, Tunshi

	2014	2015	2016
λ_1	0,80	0,61	0,81
λ_2	0,09	0,30	0,09
λ_3	0,05	0,04	0,05

λ_4	0,03	0,02	0,02
-------------	------	------	------

Elaborado por: Taguacundo Oscar, 2019

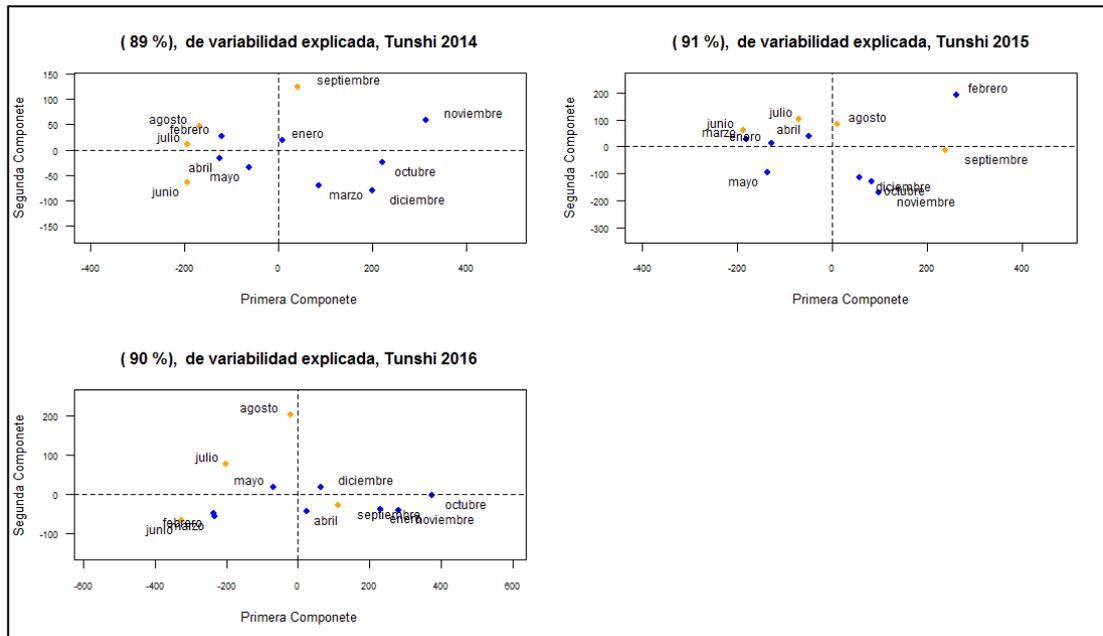


Gráfico 30-4: Primera y segunda componente principal funcional de rsg, Tunshi

Elaborado por: Taguacundo Oscar, 2019

El gráfico muestra que la primera componente principal está formada por los meses de: enero, febrero, abril, mayo, junio, julio, agosto y septiembre; enero, marzo, abril, mayo, junio, julio, agosto, octubre, noviembre y diciembre; febrero, marzo, abril, mayo, junio, julio, agosto y diciembre forman los primeros grupos de las componentes principales funcionales para los años 2014, 2015 y 2016 respectivamente.

4.4.3.9 Estación meteorológica, Multitud

Tabla 27-4: Valores propios de la estación Multitud

	2014	2015	2016
bspl4.1	87790,60	73764,91	78198,44
bspl4.2	3017,80	8383,44	4960,81
bspl4.3	1147,18	658,04	1801,55
bspl4.4	891,92	384,34	670,37
bspl4.5	399,65	184,41	330,38

Elaborado por: Taguacundo Oscar, 2019

Tabla 28-4: Funciones propias de la estación multitud 2014, 2015, 2016

	2014		2015		2016	
	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	0,01	0,03	0,01	-0,04	0,01	-0,04
bspl4.2	0,01	0,10	0,00	-0,08	-0,01	0,03
bspl4.3	0,01	0,25	-0,02	-0,17	-0,06	0,16
bspl4.4	0,25	0,93	0,37	-0,53	0,30	-0,29
bspl4.5	0,43	-0,03	0,49	-0,50	0,38	-0,45

Elaborado por: Taguacundo Oscar, 2019

La variabilidad explicada por la primera componente principal es del 94%, 88% y 91%, la segunda explica el 3%, 10% y 6% de la variabilidad total, para 2014, 2015 y 2016 respectivamente.

Tabla 29-4: Variabilidad explicada por Componentes, Multitud

	2014	2015	2016
λ_1	0,94	0,88	0,91
λ_2	0,03	0,10	0,06
λ_3	0,01	0,01	0,02
λ_4	0,01	0,00	0,01

Elaborado por: Taguacundo Oscar, 2019

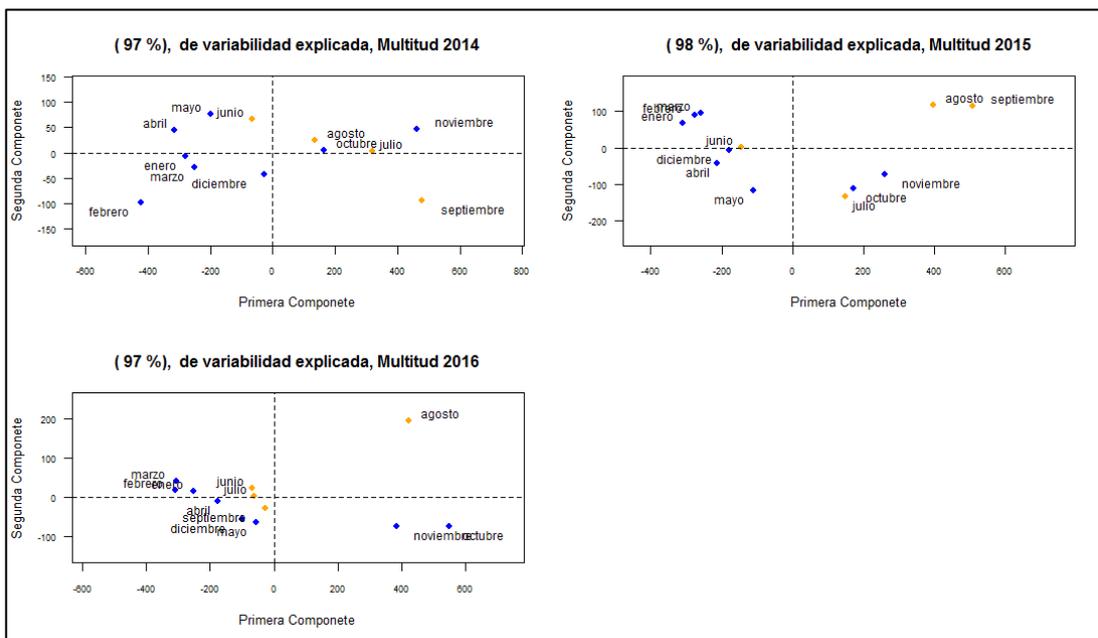


Gráfico 31-4: Primera y segunda componente principal funcional de rsg, Multitud

Elaborado por: Taguacundo Oscar, 2019

El gráfico muestra que la primera componente principal está formada por los meses de: enero, febrero, marzo, abril, mayo, junio y diciembre; enero, febrero, marzo, abril, mayo, junio y diciembre; enero, febrero, marzo, abril, mayo, junio, julio, septiembre y diciembre forman los primeros grupos de las componentes principales funcionales para los años 2014, 2015 y 2016 respectivamente.

4.4.3.10 Estación meteorológica, Urbina

Tabla 30-4: Valores propios de la estación Urbina

	2014	2016	2017
bspl4.1	29504,17	54015,44	17408,26
bspl4.2	7690,95	5679,50	2873,51
bspl4.3	1672,57	1757,75	1551,90
bspl4.4	1320,24	1550,40	719,91
bspl4.5	604,89	912,97	580,75

Elaborado por: Taguacundo Oscar, 2019

Tabla 31-4: Funciones propias de la estación Urbina 2014, 2016, 2017

	2014		2016		2017	
	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	0,04	-0,09	0,06	-0,05	0,04	0,01
bspl4.2	0,09	-0,15	0,11	-0,11	0,07	0,02
bspl4.3	0,18	-0,25	0,21	-0,23	0,13	0,06
bspl4.4	0,25	-0,26	0,33	-0,24	0,11	0,04
bspl4.5	0,44	-0,38	0,45	-0,41	0,23	0,53

Elaborado por: Taguacundo Oscar, 2019

La tabla contiene las proporciones de variabilidad explicada por cada componente principal funcional; la primera componente explica el 71%, 84%, 68% de la variabilidad total y la segunda explica el 18%, 8%, 20% de la variabilidad total para los años 2014, 2016 y 2017 respectivamente.

Tabla 32-4: Variabilidad explicada por Componentes, Urbina

	2014	2016	2017
λ_1	0,71	0,84	0,68
λ_2	0,18	0,08	0,20

λ_3	0,04	0,03	0,06
λ_4	0,03	0,02	0,03

Elaborado por: Taguacundo Oscar, 2019

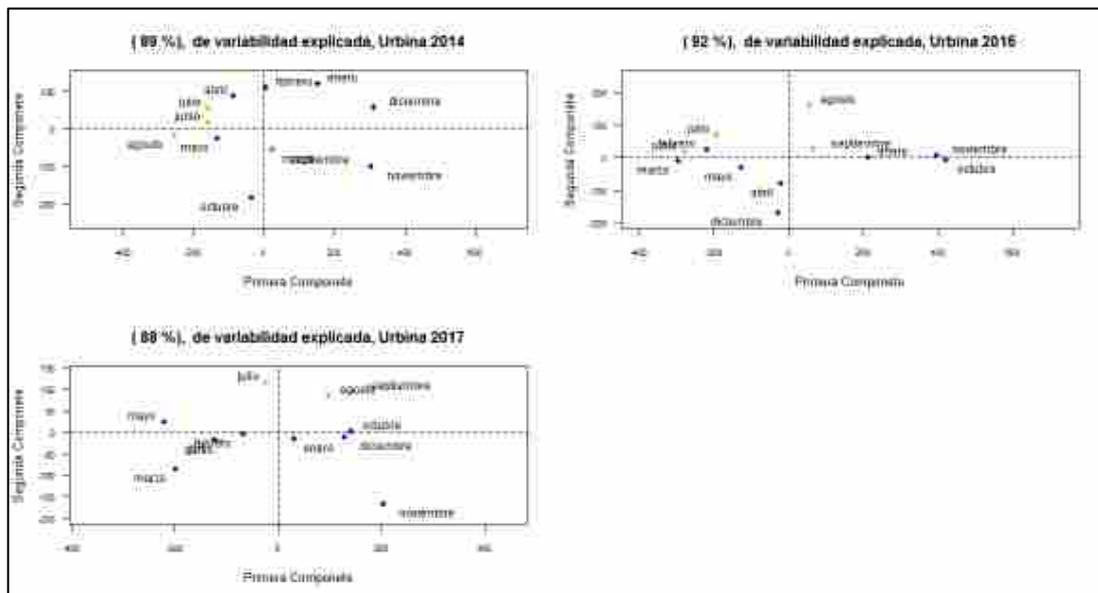


Gráfico 32-4: Primera y segunda componente principal funcional de rsg, Urbina

Elaborado por: Taguacundo Oscar, 2019

El gráfico muestra que la primera componente principal está formada por los meses de: febrero, marzo, abril, mayo, junio, julio, agosto, septiembre y octubre; febrero, marzo, abril, mayo, junio, julio, agosto, septiembre y diciembre; febrero, marzo, abril, mayo y junio forman los primeros grupos de las componentes principales funcionales para los años 2014, 2016 y 2017 respectivamente.

4.4.3.11 Estación meteorológica, Quimiag

Tabla 33-4: Valores propios de la estación Quimiag

	2014	2015	2016	2017
bsp14.1	30138,41	20690,77	52879,18	7577,41
bsp14.2	5231,13	6286,61	5391,85	7114,44
bsp14.3	3892,84	2616,86	2275,17	2318,14
bsp14.4	1314,35	1309,35	1808,79	901,23
bsp14.5	919,99	864,61	577,62	814,34

Elaborado por: Taguacundo Oscar, 2019

Tabla 34-4: Funciones propias de la estación Quimiag 2014, 2015, 2016, 2017

	2014		2015		2016		2017	
	PC1	PC2	PC1	PC2	PC1	PC2	PC1	PC2
bspl4.1	0,04	-0,05	0,01	0,09	0,03	-0,05	0,00	0,04
bspl4.2	0,07	-0,14	0,01	0,16	0,06	-0,09	0,02	0,10
bspl4.3	0,12	-0,32	0,03	0,29	0,12	-0,16	0,06	0,21
bspl4.4	0,28	-0,42	0,14	0,64	0,31	0,05	0,09	0,34
bspl4.5	0,26	-0,35	0,16	0,45	0,32	0,20	-0,08	0,65

Elaborado por: Taguacundo Oscar, 2019

La dos primeras componetes principales funcionales explicadan el 84%, 84%, 92% y 80% de variabilidad total para los años 2014, 2015, 2016 y 2017 respectivamente.

Tabla 35-4: Variabilidad explicada por Componentes, Quimiag

	2014	2015	2016	2017
λ_1	0,72	0,65	0,83	0,56
λ_2	0,12	0,19	0,09	0,24
λ_3	0,09	0,08	0,04	0,09
λ_4	0,03	0,04	0,03	0,06

Elaborado por: Taguacundo Oscar, 2019

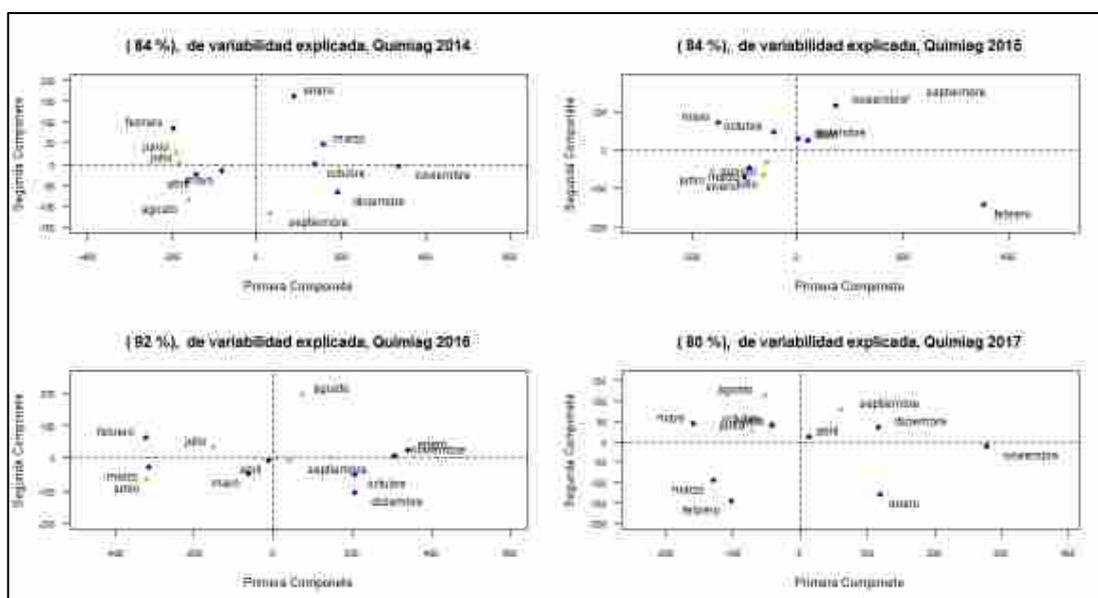


Gráfico 33-4: Primera y segunda componente principal funcional de rsg, Quimiag

Elaborado por: Taguacundo Oscar, 2019

El gráfico muestra que la primera componente principal está formada por los meses de: enero, febrero, abril, mayo, junio, julio, agosto y septiembre; enero, marzo, abril, mayo, junio, julio, agosto, octubre, noviembre y diciembre; enero, agosto, septiembre, octubre, noviembre y diciembre; febrero, marzo, mayo, junio, julio, agosto y octubre forman los primeros grupos de las componentes principales funcionales para los años 2014, 2015 2016 y 2017 respectivamente.

4.4.4 Análisis clúster por estación y Año de estudio

El objetivo principal del análisis clúster funcional es clasificar a cada una de las variables con respecto a una curva centroide homogénea, esto lo realiza calculando la profundidad de cada dato funcional a k diferentes centroide.

En el estudio de la radiación solar global se estable $k = 2$, con el fin de indagar posibles agrupaciones según su comportamiento diario.

4.4.4.1 Análisis clúster, Alao

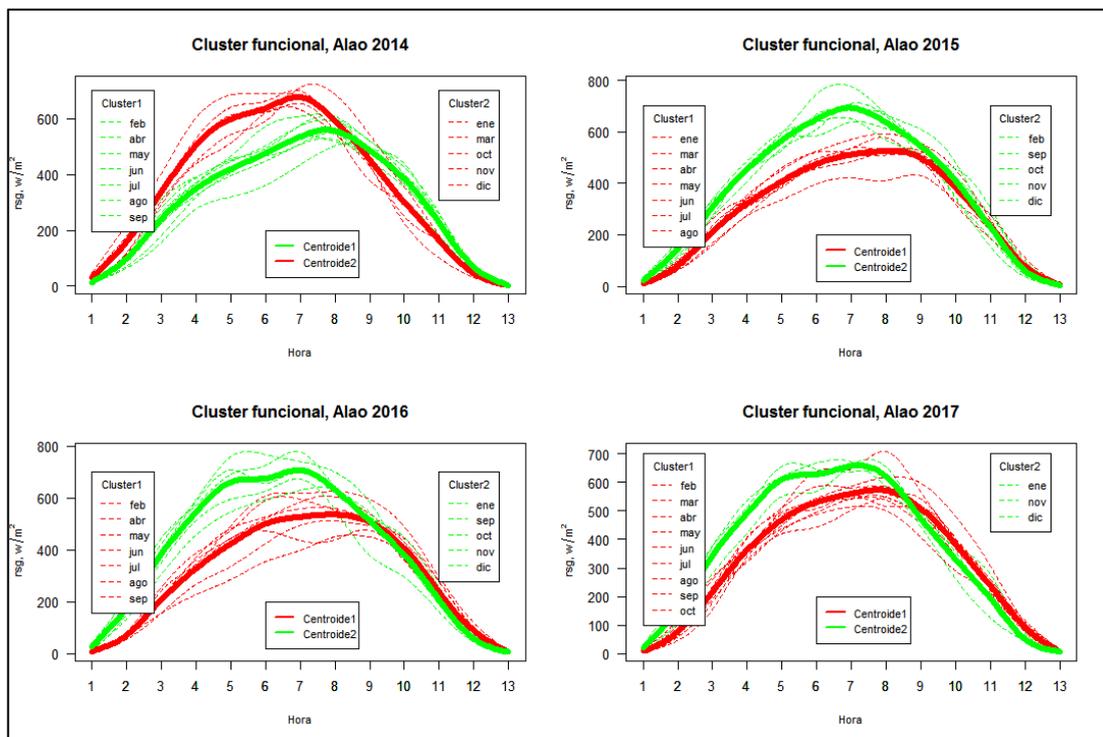


Gráfico 34-4: Análisis clúster funcional de la estación Alao

Elaborado por: Taguacundo Oscar, 2019

El gráfico 34-4 indica el análisis clúster funcional de la estación Alao, la rsg durante el 2014, 2015 y 2016 muestra alta intensidad desde tempranas horas en la mañana y dos comportamientos

heterogéneos. En 2014 y 2017 alcanza hasta los 700 W/m^2 , mientras que 2015 y 2016 hasta los 800 W/m^2 . Los meses de octubre, noviembre y diciembre muestran alta intensidad de rsg, mientras que mayo, junio, julio y agosto con baja rsg.

4.4.4.2 Análisis clúster, Cumanda

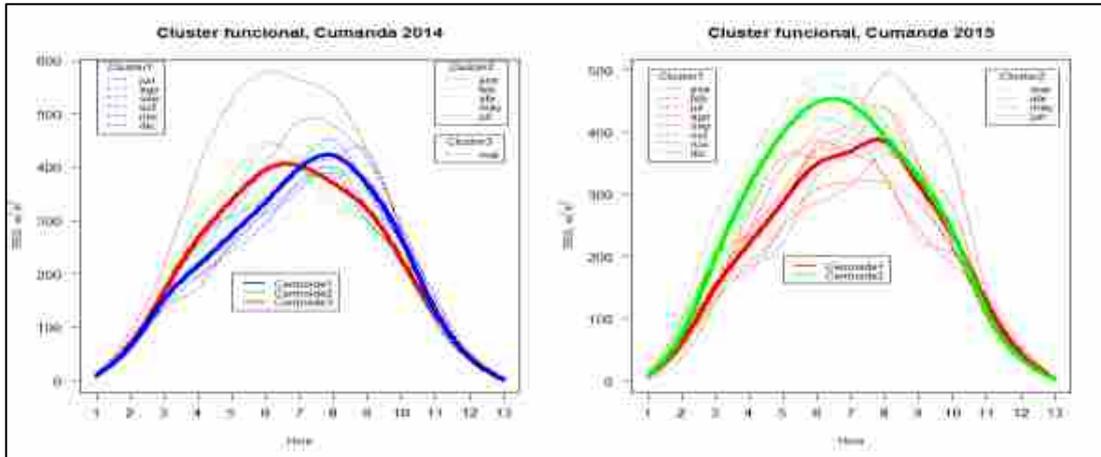


Gráfico 35-4: Análisis clúster funcional de la estación Cumanda

Elaborado por: Taguacundo Oscar, 2019

El gráfico 35-4 indica el análisis clúster funcional de la estación Cumanda, alcanza hasta 500 W/m^2 , sin embargo marzo es el único mes que presenta rsg mayor que 500 W/m^2 en el 2014, por lo que se considera como un tercer grupo. En el 2014 la intensidad de rsg es homogénea en todos los meses el grupo 1 (junio, agosto, septiembre, octubre, noviembre y diciembre) presenta la más alta rsg a las 13h30, mientras que el grupo 2 (enero, febrero, abril, mayo y julio) a las 12h00. En el 2015 se verifica dos comportamientos de rsg durante el año, el grupo 1 (enero, febrero, julio, agosto, septiembre, octubre, noviembre y diciembre) presenta la más alta rsg a las 12h00, mientras que el grupo 2 (marzo, abril, mayo y junio) a las 13h30 y con baja rsg.

4.4.4.3 Análisis clúster, Atillo

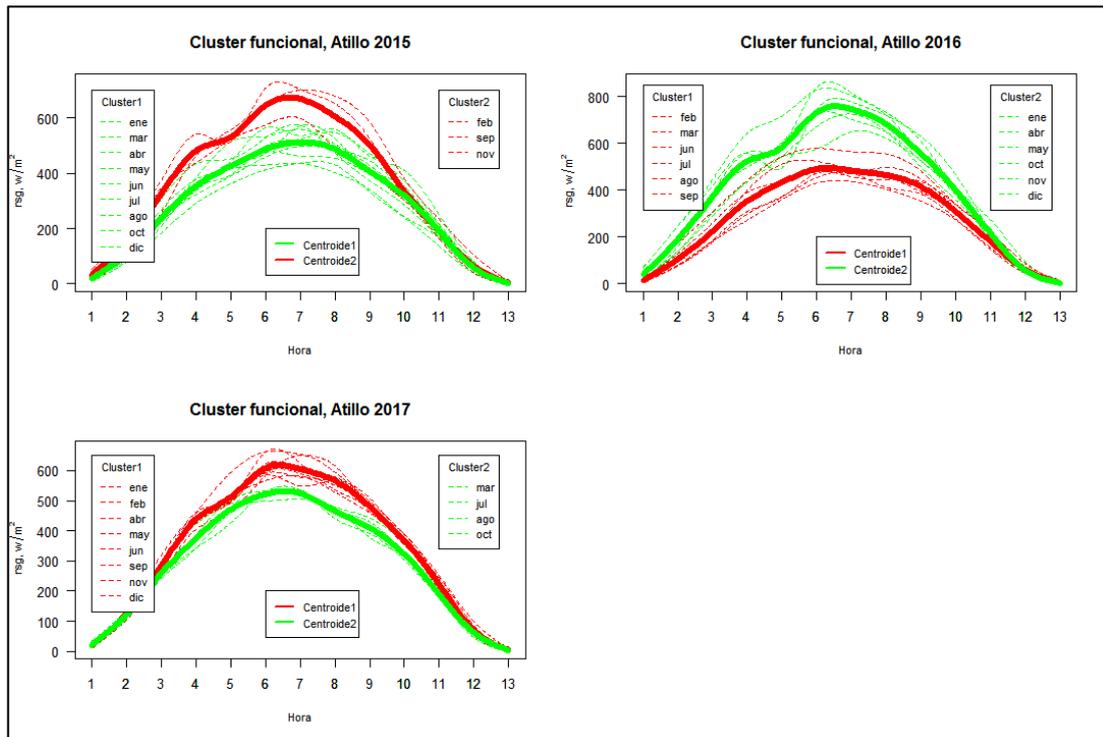


Gráfico 36-4: Análisis clúster funcional de la estación Atillo

Elaborado por: Taguacundo Oscar, 2019

El gráfico 36-4 indica el análisis clúster funcional con el fin de conocer el comportamiento de rsg homogéneos de la estación Atillo, en el 2015 se identifica 2 grupos, el primero sobrepasa los 600 W/m^2 y en el segundo son inferiores a 600 W/m^2 , en el 2016 el primero alcanza los 600 W/m^2 y en el segundo alcanza hasta 800 W/m^2 , mientras que en el año 2017 el grupo 1 rodea los 600 W/m^2 y el grupo 2 está alrededor de los 500 W/m^2 , mostrando que hubo una reducción del rsg. Los meses de marzo, julio y agosto durante los 3 años estudiados mantuvieron rsg baja y estable, mientras que los meses restantes mostraron cambios bruscos de rsg.

4.4.4.4 Análisis clúster, San Juan

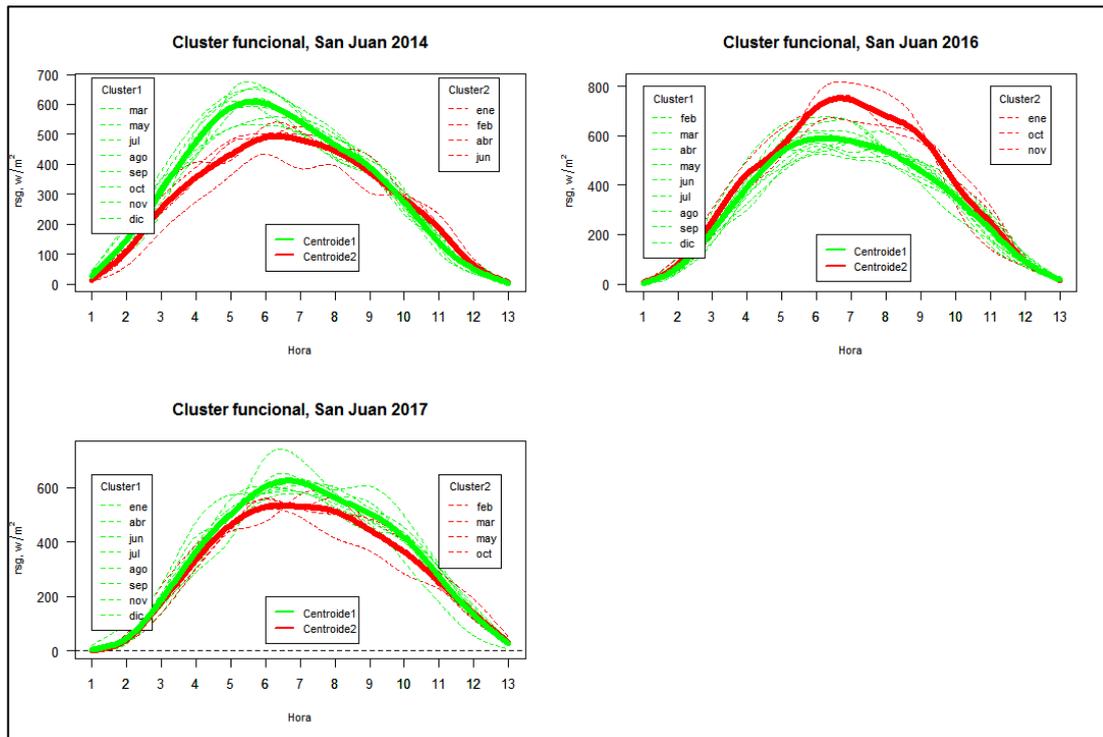


Gráfico 37-4: Análisis clúster funcional de la estación San Juan

Elaborado por: Taguacundo Oscar, 2019

El gráfico 37-4 indica el análisis clúster funcional de la estación San Juan. Se encontró similitud en la intensidad de rsg en los primeros grupos de los 3 años de estudio que fluctúa en $600 W/m^2$, mientras que el grupo 2 en 2014 y 2016 tiene intensidades alrededor de $500 W/m^2$, además 2016 sobrepasa los $600 W/m^2$ en enero, octubre y noviembre, mostrando incremento de la rsg en dichos meses. En el 2017 el grupo 1 fluctúa los $600 W/m^2$ y el grupo 2 alcanza dicho valor.

4.4.4.5 Análisis clúster, ESPOCH

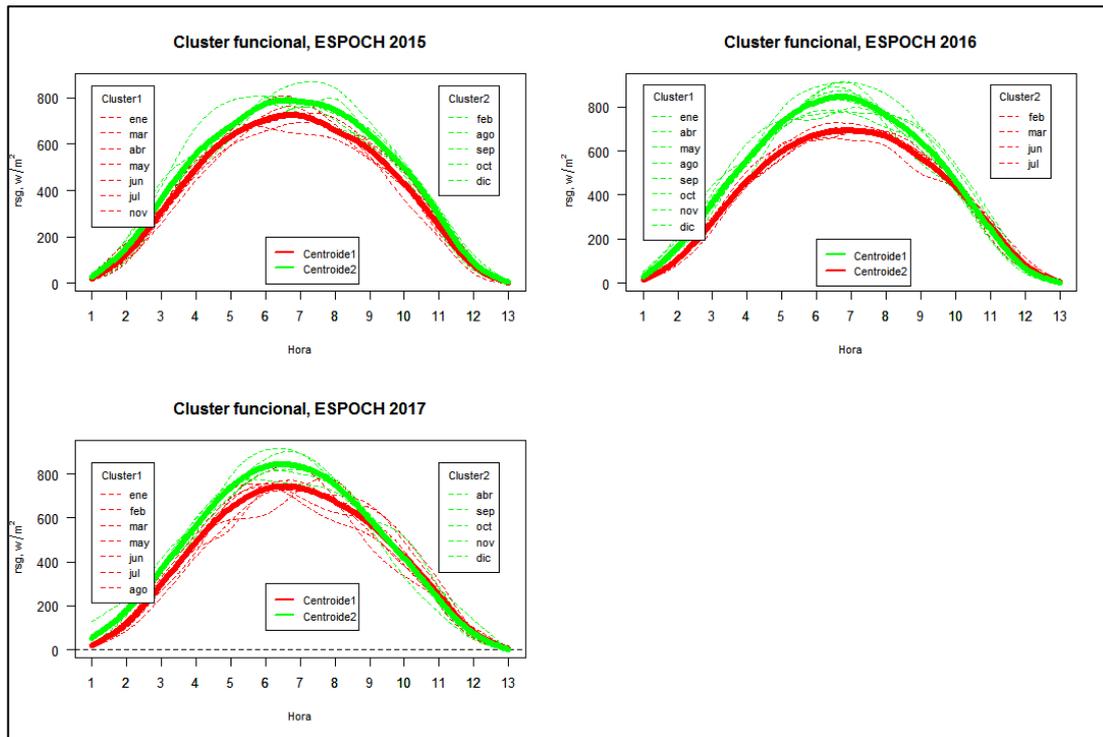


Gráfico 38-4: Análisis clúster funcional de la estación ESPOCH

Elaborado por: Taguacundo Oscar, 2019

El gráfico 38-4 indica el análisis clúster funcional de la estación ESPOCH, la rsg en esta estación llega hasta $800 W/m^2$. Agosto, septiembre y octubre presentan incidencia en la rsg elevada mientras que marzo y junio se mantienen estables en los tres años de estudio con rsg baja.

4.4.4.6 Análisis clúster, Tixan

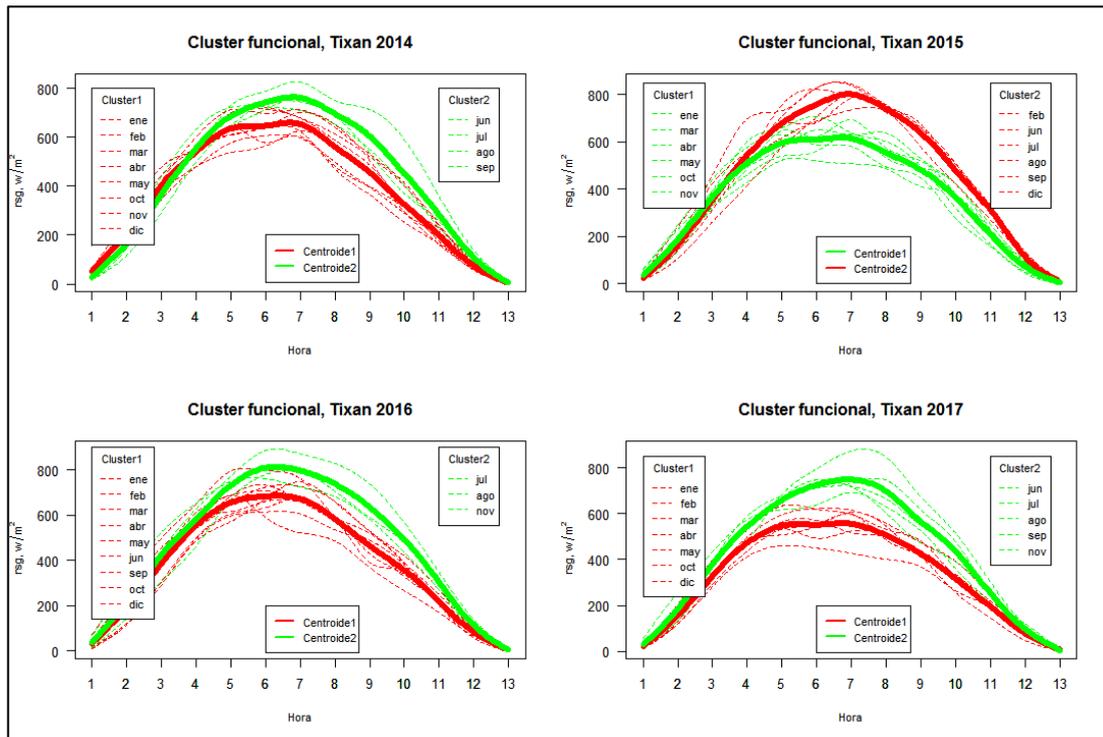


Gráfico 39-4: Análisis clúster funcional de la estación Tixan

Elaborado por: Taguacundo Oscar, 2019

El gráfico 39-4 indica el análisis clúster funcional de la estación Tixan, En esta estación la intensidad de rsg en general llega hasta los de $800 W/m^2$, los meses de julio agosto permanecen estables en los 4 años de estudio mientras que los demás que los demás fluctúan alrededor de los de $600 W/m^2$.

4.4.4.7 Análisis clúster, Matus

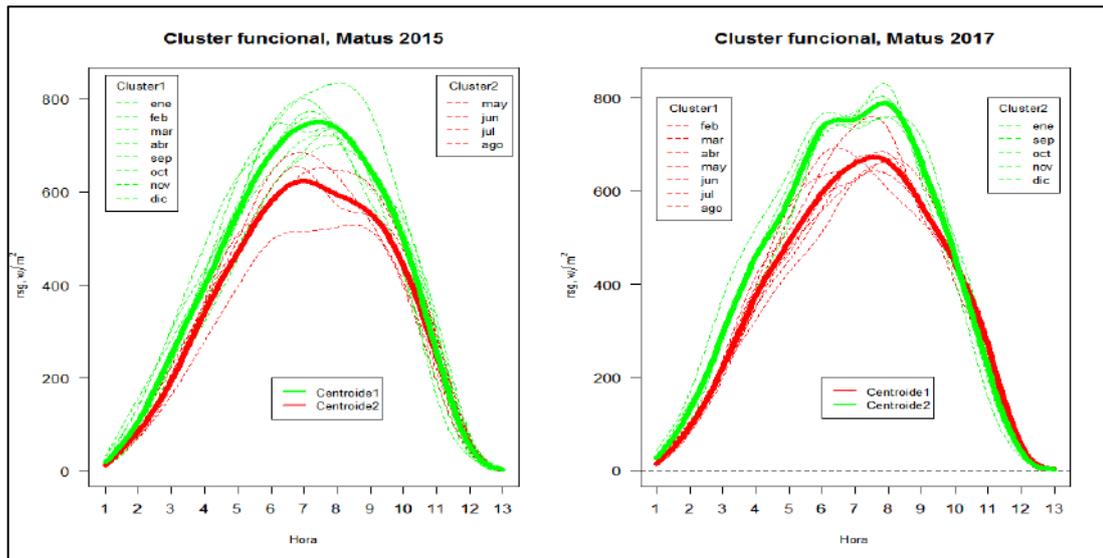


Gráfico 40-4: Análisis clúster funcional de la estación Matus

Elaborado por: Taguacundo Oscar, 2019

El gráfico 40-4 indica el análisis clúster funcional de la estación Matus, en los 2 años de estudio se identifica rsg que llega de 800 W/m^2 , los meses de enero, septiembre y octubre presentan rsg más elevada, mientras que mayo, junio y julio son bajas las cuales sobrepasan los 600 W/m^2 .

4.4.4.8 Análisis clúster, Tunshi

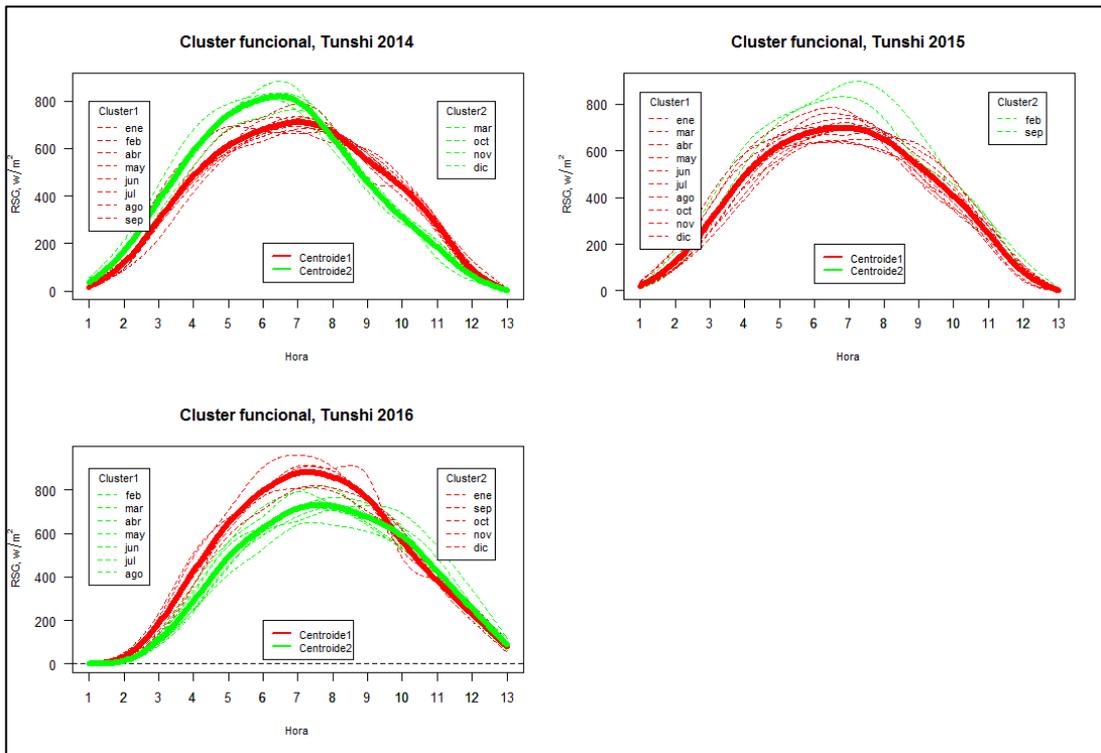


Gráfico 41-4: Análisis clúster funcional de la estación Tunshi

Elaborado por: Taguacundo Oscar, 2019

El gráfico 41-4 indica el análisis clúster funcional de la estación Tunshi sobrepasa los $800 W/m^2$, sobretodo en el año 2016. Los meses de mayo, junio, julio y agosto presentan baja rsg.

4.4.4.9 Análisis clúster, Multitud

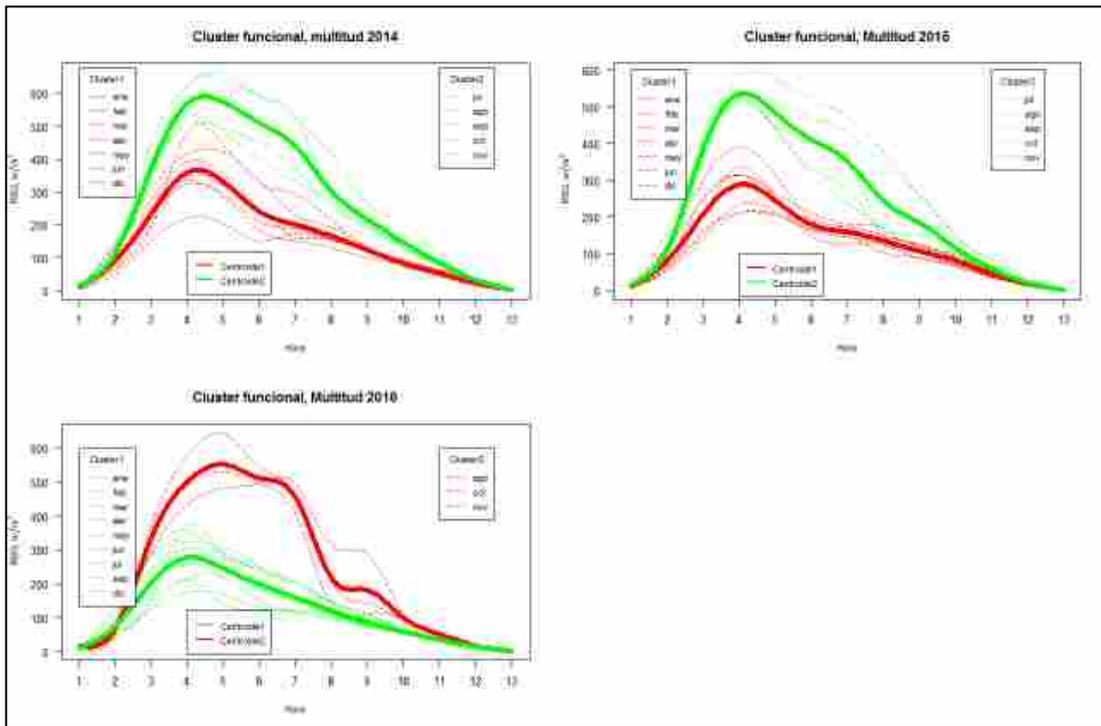


Gráfico 42-4: Análisis clúster funcional de la estación Multitud

Elaborado por: Taguacundo Oscar, 2019

El gráfico 42-4 indica el análisis clúster funcional de la estación Multitud, se observó una rsg que alcanza los $600 W/m^2$, los meses de agosto, octubre y noviembre presentan rsg más estable y alta en los tres años de estudio, mientras que los meses restantes no sobrepasan los $400 W/m^2$. Cabe destacar que esta estación muestra dos grupos bien marcados en el comportamiento de la rsg durante cada año.

4.4.4.10 Análisis clúster, Urbina

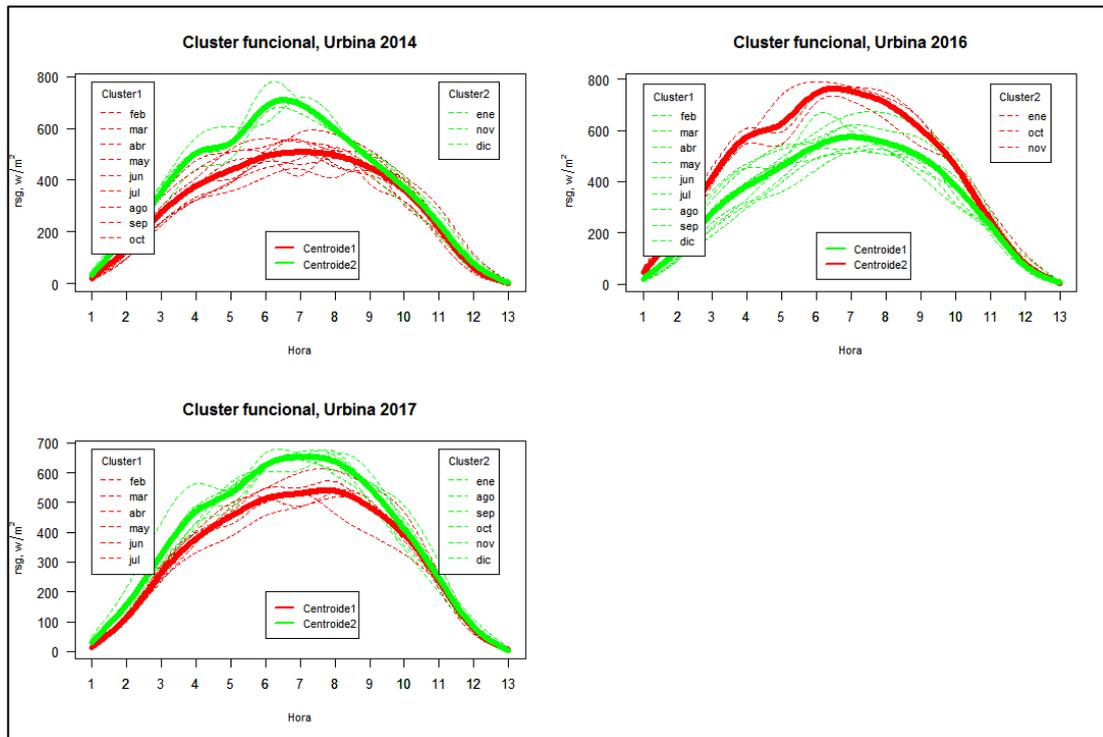


Gráfico 43-4: Análisis clúster funcional de la estación Urbina

Elaborado por: Taguacundo Oscar, 2019

El gráfico 43-4 indica el análisis clúster funcional de la estación Urbina, muestra que en año 2014 y 2016 llegan hasta los $800 W/m^2$, mientras que en el 2017 no sobrepasa los $700 W/m^2$, evidenciando así una reducción de rsg en dicho año. Los meses con mayor intensidad de rsg fueron: en el 2014 enero, noviembre y diciembre, en 2016 enero, octubre y noviembre, y en 2017 enero, julio, agosto, septiembre y octubre.

4.4.4.11 Análisis clúster, Quimiag

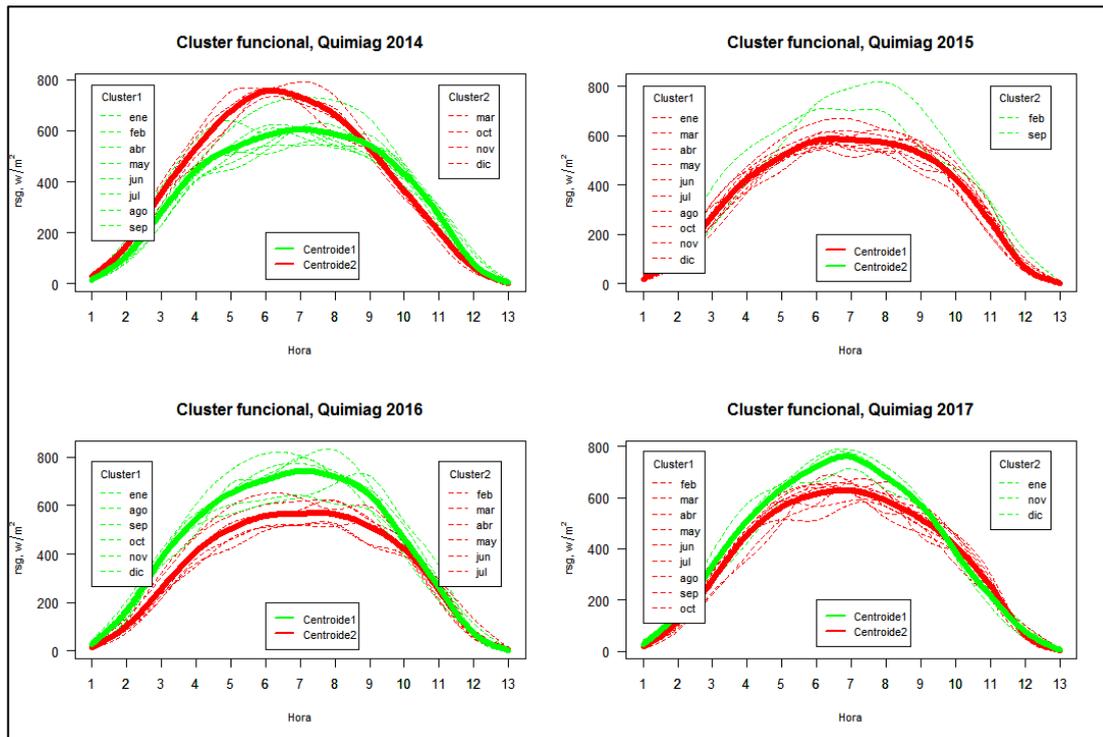


Gráfico 44-4: Análisis clúster funcional de la estación Quimiag

Elaborado por: Taguacundo Oscar, 2019

El gráfico 61-4 indica el análisis clúster funcional de la estación Quimiag, muestra que la rsg alcanza los $800 W/m^2$, en el 2014 y 2016 se observó dos comportamientos de rsg bien marcados durante el año, manteniendo alta rsg en los meses de octubre, noviembre y diciembre, y las bajas son los meses de febrero, abril, mayo, junio y julio; mientras que para el 2015 y 2017 en comportamiento de rsg muestra mayor homogeneidad en todos los meses.

4.4.5 Análisis de Componentes principales por estación

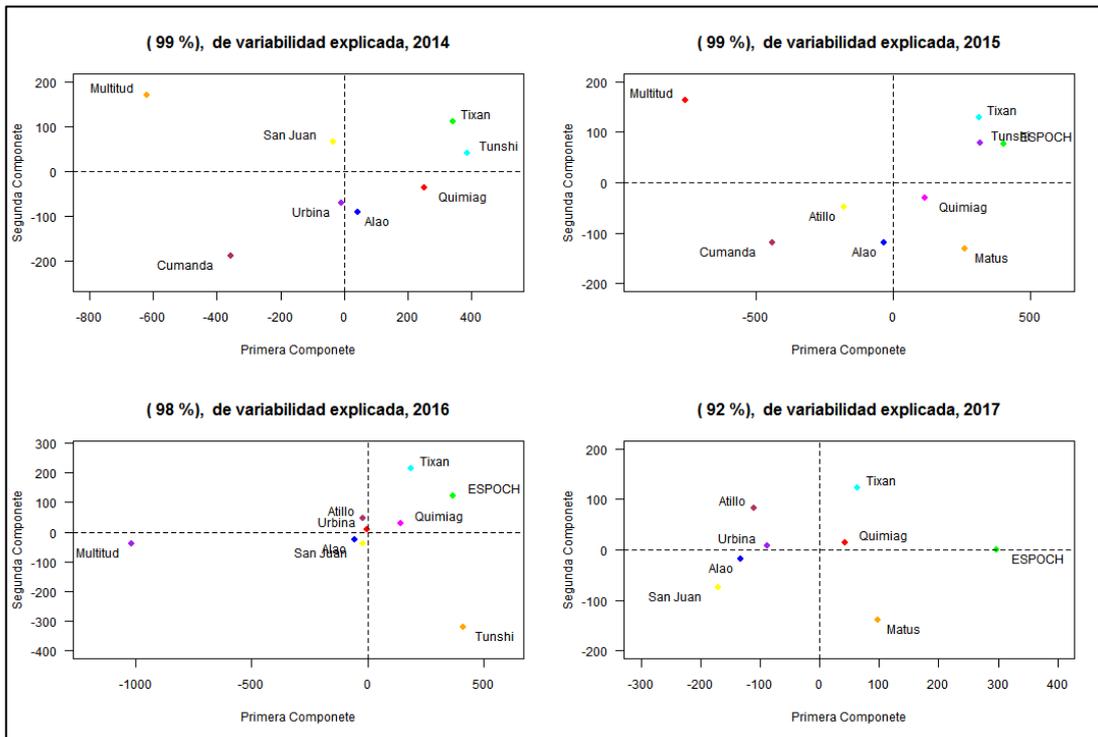


Gráfico 45-4: Componentes principales funcionales por estación y año de estudio

Elaborado por: Taguacundo Oscar, 2019.

El gráfico muestra las estaciones que tienen rsg similar, sin embargo, en 2014 se observó que la estación de Multitud y Cumanda son las más alejadas de las demás estaciones, además las más comunes en el comportamiento de la rsg son: San Juan, Urbina, Alao, Quimiag, Tunshi y Tixan. En 2015 Multitud sigue muy alejado, siendo las demás estaciones homogéneas en su comportamiento. En 2016 Multitud y Tunshi están alejados de las demás estaciones, por tanto, Tixan, ESPOCH, Quimiag, Attilo, Urbina, Alao, San Juan son las más comunes. En 2017 San Juan, Alao, Urbina, Attilo presentan un comportamiento de rsg similar, mientras que las demás estaciones tienen comportamientos diferentes.

4.4.6 Análisis Clúster Funcional por Estación

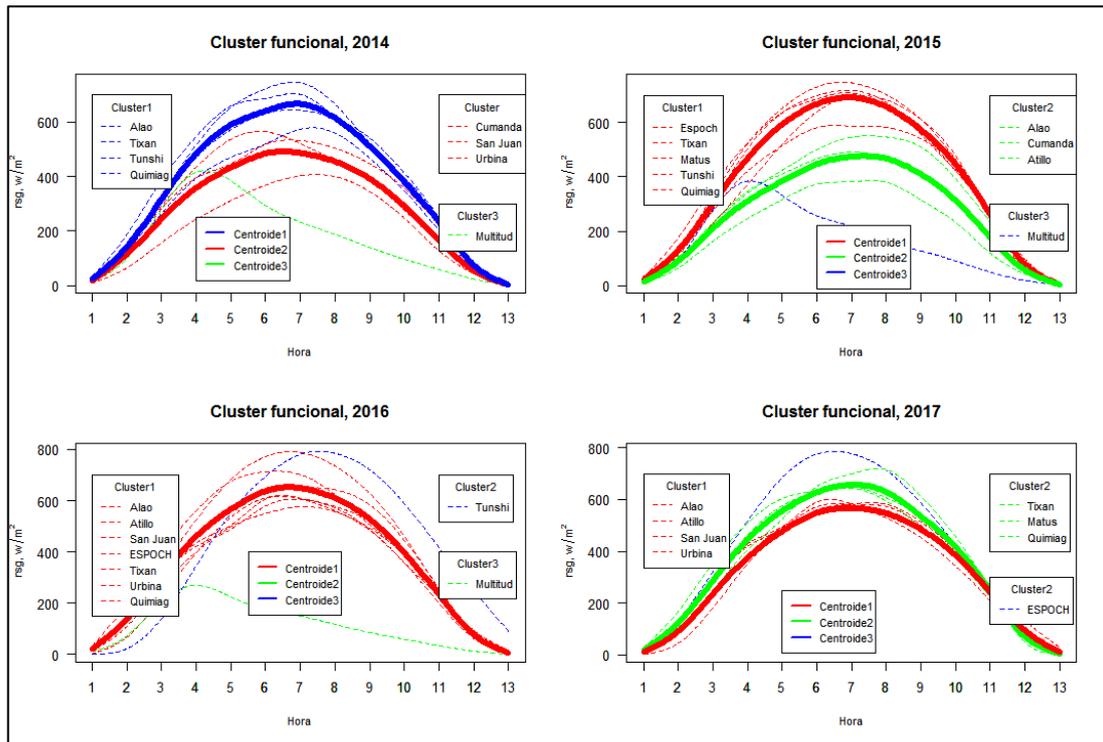


Gráfico 46-4: Análisis clúster funcional por estación y año de estudio

Elaborado por: Taguacundo Oscar, 2019.

El gráfico muestra que las estaciones meteorológicas en el año 2014 y 2015 son inferiores a los 800 W/m^2 , mientras que en 2016 y 2017 alcanza los 800 W/m^2 . La estación de Multitud tiene un comportamiento de rsg único y alcanza hasta 400 W/m^2 durante el 2014 y 2015, mientras que en el 2016 no sobrepasa los 300 W/m^2 . En 2014 Alao, Tixan, Tunshi y Quimiag, en 2015 ESPOCH, Tixan, Matus, Tunshi y Quimiag, En 2016 Alao, Atillo, San Juan, ESPOCH, Tixan, Urbina y Quimiag, y en 2017 ESPOCH presentan rsg altas con respecto a las demás estaciones.

4.4.7 Análisis de componentes principales y Clúster funcional general por Estación

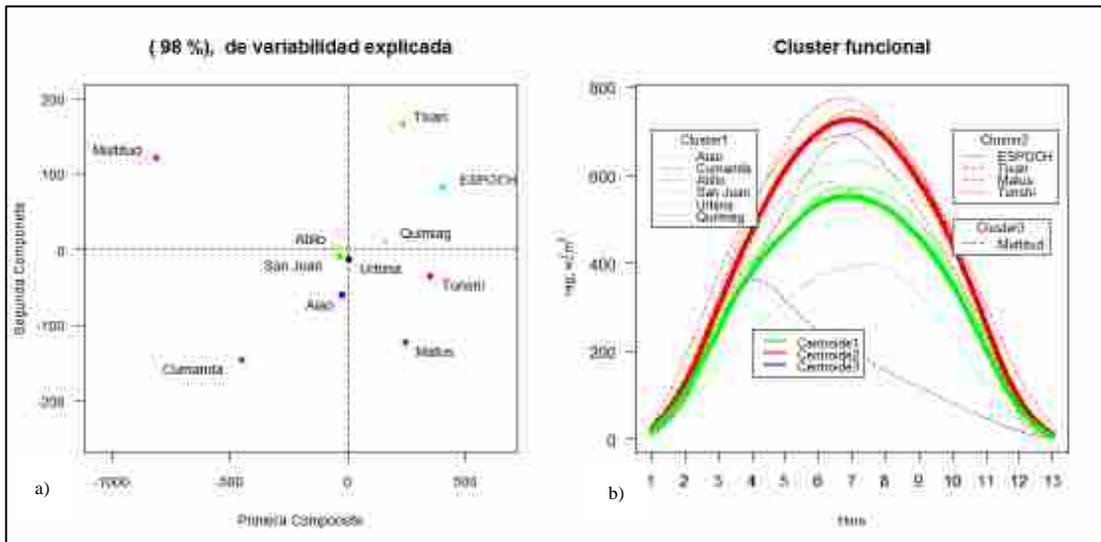


Gráfico 47-4: a) Gráfico de la primera y segunda componentes principal funcional; b) Análisis clúster funcional

Elaborado por: Taguacundo Oscar, 2019.

En el gráfico 64-4 se identifica tres grupos de comportamiento de rsg similares en la provincia de Chimborazo. El primer grupo conformado por: Alao, Cumanda, Atillo, San Juan, Urbina y Quimiag, el segundo grupo por: ESPOCH, Tixan, Matus y Tunshi, y el grupo 3 por: Multitud. La característica principal que diferencia un grupo de otro es la intensidad de rsg, es decir el grupo 2 presenta mayor intensidad de rsg (no inferior a 680 W/m^2 y alcanza hasta los 800 W/m^2), el grupo 1 intensidad intermedia (no inferior a 550 W/m^2 y alcanza hasta los 680 W/m^2) y el grupo 3 con baja intensidad (no sobrepasa los 400 W/m^2). La estación de multitud se caracteriza por alcanzar una rsg máxima alrededor de las 10h00.

CONCLUSIONES

Con la información proporcionada por el GEAA se identificó la rsg desde el 2014 hasta el 2017 considerando una matriz de datos por cada año en cada una de las estaciones armando así 34 base de datos debido a que no en todas las estaciones se dispuso de los 4 años de datos. Las filas representaron la rsg en cada hora y las columnas cada día. Los datos proporcionados por el GEAA fueron imputados mediante regresión múltiple, por lo que en esta investigación no fue necesario ningún tipo de imputación de datos.

Se define como dato funcional la rsg diaria tomada las 12 horas (luz) del día, es decir que se contó con 365 datos funcionales por año. Y mediante correlación cruzada y la función *min.basis()* de R se identificó que el suavizado B-spline se ajustó mejor a la rsg que el de Fourier, dando así 15 bases B-spline cúbico para el suavizado de curvas con un error cuadrático medio en promedio de $4.273653e-06$.

Se realizó la descripción del comportamiento de las 11 estaciones meteorológicas durante cada año disponible, mediante el cálculo de: las curvas medias anuales con su respectiva banda de confianza en base a la media global, curvas de desviación estándar y de nivel con el fin de analizar la variabilidad de la rsg en función de las horas estudiadas, identificando variabilidad entre las 10h00 hasta 15h00, debido a que en dichas horas la intensidad de rsg es elevada. También se graficó modelos de superficie raster para visualizar la intensidad de la variabilidad de la rsg, denotada con rojo la máxima radiación si la superficie de color rojo.

El análisis de componentes principales funcionales ayudo a reducir la dimensionalidad de las curvas mensuales de cada una de las estaciones meteorológicas, con el objetivo de identificar características similares de cada una de ellas, para lo cual se redujo a dos componentes principales debido a que en general explican más del 74% de la variabilidad total. La variabilidad total explicada por las dos primeras componentes principales funcionales para la estación Alao fue de 94%, 93%, 96%, 89%, en Tixan se obtuvo el 88%, 92%, 93%, 94% y en Quimiag se obtuvo 84%, 84%, 92%, 80% para los años 2014, 2015, 2016, 2017, en Cumanda tuvo 95%, 90% para los años 2014, 2015, para Atillo 92%, 96%, 86% y en ESPOCH se tuvo 80%, 85%, 80% para los años 2015, 2016, 2017, en San Juan se obtuvo el 85%, 84%, 75% y en Urbina se obtuvo 89%, 92%, 88% para 2014, 2016, 2017, en Matus se observó el 89%, 84% para los años 2015, 2017, en Tunshi se tiene el 89%, 91%, 90% y en Multitud el 97%, 98%, 97%, para los años 2014, 2015, 2016.

Se realizó análisis de clúster funcional no supervisado con el fin de identificar grupos con patrones de comportamiento de rsg similares al 95% de confiabilidad, con la función *kmean.fd()* se obtuvo dos grupos de importancia en cada año, las estaciones meteorológicas en el año 2014 y 2015 tienen intensidad de rsg inferior a los $800 W/m^2$, mientras que en 2016 y 2017 alcanza los $800 W/m^2$, Multitud tiene un comportamiento de rsg único y alcanza hasta $400 W/m^2$ durante el 2014 y 2015, mientras que en el 2016 no sobrepasa los $300 W/m^2$. En 2014 Alao, Tixan, Tunshi y Quimiag, en 2015 ESPOCH, Tixan, Matus, Tunshi y Quimiag, En 2016 Alao, Atillo, San Juan, ESPOCH, Tixan, Urbina y Quimiag, y en 2017 ESPOCH presentan rsg altas con respecto a las demás estaciones.

De las 11 estaciones meteorológicas estudiadas en la provincia de Chimborazo, se identificó tres formas de comportamiento de rsg. El primer grupo conformado por: Alao, Cumanda, Atillo, San Juan, Urbina y Quimiag, el segundo grupo por: ESPOCH, Tixan, Matus y Tunshi, y el grupo 3 por: Multitud. La característica principal que diferencia un grupo de otro es la intensidad de rsg, es decir el grupo 2 presenta mayor intensidad de rsg (no inferior a $680 W/m^2$ y alcanza hasta los $800 W/m^2$), el grupo 1 intensidad intermedia (no inferior a $550 W/m^2$ y alcanza hasta los $680 W/m^2$) y el grupo 3 con baja intensidad (no sobrepasa los $400 W/m^2$). La estación de multitud se caracteriza por alcanzar una rsg máxima alrededor de las 10h00, mientras que las otras a las 12h30.

RECOMENDACIONES

El GEAA intensifique esfuerzos en el almacenamiento de la información de las variables meteorológicas con el fin de reducir el problema de datos faltantes.

Intensificar el análisis estadístico de datos de la radiación solar global y demás variables meteorológicas mediante datos funcionales, con el fin de indagar el comportamiento diario, mensual o anual de dichos fenómenos, así, como estudio de curvas atípicas, análisis de varianza, entre otros.

Los docentes e investigadores promuevan la investigación a los estudiantes mediante la realización de prácticas preprofesionales y trabajo de titulación con datos funcionales debido a su aplicabilidad y optimización en la presentación de resultados en cualquier área de estudio.

En la carrera de Estadística se mantenga la electiva Análisis Estadístico de Datos Funcionales con el fin ampliar los conocimientos de los estudiantes en cuanto a nuevas tecnologías y metodologías para el tratamiento estadístico de datos.

GLOSARIO

1. **AEDF:** Análisis estadístico de datos funcionales
2. **rsg:** Radiación dólar global
3. **FM:** Fraiman y Muniz
4. **GCV:** Generalized cross-validation
5. **CV:** Cross-validation
6. **ACPF:** Análisis de componentes principales funcionales
7. **CEEA:** Centro de Investigación de Energías Alternativas y Ambiente
8. **GEEA:** Grupo de Energías Alternativas y Ambiente

BIBLIOGRAFÍA

ARENAS, D. Y ZAPATA, H. *Libro interactivo sobre energía solar y sus aplicaciones* (tesis) (Grado). Universidad Tecnológica de Pereira. Colombia. Pereira. 2011. pp.18-30 [Consulta: 12 marzo 2018]. Disponible en: <http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/2369/62131244A681.pdf?sequence=1>

ARRIAZA, A., et al. "Estadística Básica con R y R-Commander" [en línea] *Servicio de Publicaciones de la Universidad de Cádiz*. Cádiz. 2008 [Consulta: 19 marzo 2018]. Disponible en: <http://knuth.uca.es/repos/ebrcmdr/pdf/13marzo/ebrcmdr.pdf>

BENKO, M. *Functional data analysis with applications in finance* (tesis) (Doctorado) Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, Berlin. 2006. [Consulta: 15 marzo 2018]. Disponible en: <https://edoc.hu-berlin.de/bitstream/handle/18452/16237/benko.pdf?sequence=1>

BOMBAL, F. "Los orígenes del análisis funcional". *Real Academia de Ciencias de Madrid* [en línea], 1997, (España) pp.35-56. [Consulta: 10 febrero 2018] Disponible en: <https://studylib.es/doc/5414599/los-origenes-del-analisis-funcional>

BRUZUAL, R. Y DOMÍNGUEZ, M. *Espacios de Hilbert*. Caracas-Venezuela, Universidad Central de Venezuela. 2005, p.1. Disponible en: <http://www.matematica.ciens.ucv.ve/labfg/anfun/afh-t.pdf>

CARREÑO, S. *Aplicaciones en meteorología de modelos funcionales de predicción en componentes principales* (tesis) (Master), Universidad de Granada, Granada. 2013. [Consulta: 10 febrero 2018] [en línea], Disponible en: http://masteres.ugr.es/moea/pages/tfm-1213/tfm_carrenosheila/

Centro de Investigación de Energías Alternativas y Ambiente [en línea]. CEEA, Riobamba, 2018 [consulta: 19 febrero 2018]. Disponible en: http://190.15.135.179/ceaa_espoch/index.php/investigacion

CHÁVEZ, C., et al. "Análisis de componentes principales funcionales en series de tiempo económicas". *Revista Internacional de Gestión del Conocimiento y la Tecnología* [en línea], 2015, (España), vol. 3, no. 2, pp. 13-25. [Consulta: 10 febrero 2018] ISSN 2255-5684, Disponible en: <http://hdl.handle.net/11117/2757>

ESCUADERO, A. *Implementación del tratamiento de datos funcionales en los datos de velocidad de viento, temperatura, humedad, precipitación tomados en la estación meteorológica de la Facultad de Ciencias período 2007-2012* (tesis) (Grado). Escuela Superior Politécnica de Chimborazo, Ciencias, Física y Matemática, Riobamba, Ecuador, 2012. pp.9-38.

ESCUADERO, A. *Modelos funcionales para el tratamiento de datos de Radiación Solar Global* (tesis) (Maestría) [en línea] Universidad de Granada, Departamento de Estadística e I.O. 2016, pp.43-44 [Consulta: 10 febrero 2018] Disponible en: http://masteres.ugr.es/moea/pages/curso201516/tfm1516/TFM_Escudero_Villa

ESCUADERO, A., et al. "Spline Cúbico para el Tratamiento Funcional de la Radiación Solar Global". *Scielo* [en línea], 2016, (Ecuador), vol. 27, pp. 153-162. [Consulta: 10 febrero 2018], DOI 10.4067/S0718-07642016000200018 Disponible en: https://scielo.conicyt.cl/scielo.php?script=sci_abstract&pid=S0718-07642016000200018&lng=es&nrm=iso

FEBRERO, M. Y OVIEDO, M., "Statistical Computing in Functional Data Analysis: The R Package fda.usc" *Journal of Statistical Software* [en línea], 2012, España, vol. 51, no. 4. pp.2-28 [Consulta: 10 febrero 2018], Disponible en: https://scielo.conicyt.cl/scielo.php?script=sci_abstract&pid=S0718-07642016000200018&lng=es&nrm=iso

FEBRERO, M., et al. "Package fda.usc" *CRAN* [en línea] 2013 [Consulta: 10 febrero 2018] Disponible en: <http://www.jstatsoft.org/v51/i04/>

FERNÁNDEZ, A. "Actualidad y aplicaciones de la topología y el análisis funcional" *Facultad de CC Económicas y Empresariales de la UCM* [en línea], 2017, España, ISSN 2255-5471, Disponible en: <http://eprints.ucm.es/40910/>

FRAIMAN, R. Y MUNIZ, G. "Trimmed means for functional data" *Sociedad de Estadística e Investigación Operativa* [en línea], 2001, (Argentina), vol. 10, no. 2, pp.419-440. [Consulta: 10 febrero 2018] Disponible en: <https://link.springer.com/article/10.1007/BF02595706>

GONZÁLEZ, S. *Marcos en Espacios de Hilbert* (tesis) [en línea] Universidad Nacional de la Plata, Facultad de Ciencias exactas, Departamento de Matemática. 2016, Argentina, pp.4-10. [Consulta: 10 febrero 2018] Disponible en: <http://www.mate.unlp.edu.ar/~demetrio/Monografias/Materias/AF/25.%20Marcos%20en%20Espacios%20de%20Hilbert%20-%20Santiago%20Gonzalez%20Zerbo.pdf>

HOOKER, G. *Functional Data Analysis*, p.20, 2017, [Consulta: 10 febrero 2018] Disponible en: <http://faculty.bscb.cornell.edu/~hooker/ShortCourseHandout.pdf>

JACQUES, J. Y PREDA, C. "Functional data clustering: a survey". *Informatics Mathematics INRIA* [en línea], 2014, Francia, vol. 8, no. 3, pp. 24. [Consulta: 10 febrero 2018] ISSN 0249-6399. Disponible en: <https://hal.inria.fr/hal-00771030>

KRINER, A., et al. "El adelgazamiento de la capa de ozono: algunos obstáculos para su aprendizaje". *Revista electrónica de enseñanza de las ciencias* [en línea], 2003, Argentina, vol. 2, pp. 136-154. [Consulta: 10 febrero 2018]. Disponible en: http://reec.uvigo.es/volumenes/volumen2/REEC_2_2_4.pdf

LOJANO, D. Y RÍOS, O. *Análisis técnico-económico para la generación de energía solar fotovoltaica en el Ecuador y su conexión a la red pública, basada en el artículo 63 de la regulación no. CONELEC – 004/11* (tesis) (Grado) [en línea]. Universidad de Cuenca, Facultad de Ingeniería, Escuela de Ingeniería Eléctrica. Loja, Ecuador, 2013, p.32. [Consulta: 10 febrero 2018]. Disponible en: <http://dspace.ucuenca.edu.ec/jspui/bitstream/123456789/5022/1/Tesis.pdf>.

MARRERO, I. *Espacios de Hilbert* (tesis) [en línea] Universidad de La Laguna, Departament de anàlisi matemàtic, 2011, pp.9-10 [Consulta: 10 febrero 2018] Disponible en: https://campusvirtual.ull.es/ocw/pluginfile.php/3756/mod_resource/content/3/tema1/1-ehilbert.pdf

MILLÁN, L. *Outliers de datos funcionales para la detección de caudales anómalos en el sector hidráulico* (tesis) (Master) Universitat Jaume, España, 2017, pp.50-56 [Consulta: 10 febrero 2018] Disponible en: <http://repositori.uji.es/xmlui/handle/10234/174477>

NAULA, I. *Análisis de datos funcionales aplicado a la distribución de la población ecuatoriana* (tesis) (Magister), Escuela Politécnica Nacional, Facultad de Ciencias, Quito, Ecuador, 2016, p.61. [Consulta: 10 febrero 2018] Disponible en: <http://bibdigital.epn.edu.ec/handle/15000/15776>

NAVARRO, V. *Análisis de datos funcionales. implementación y aplicaciones* (tesis) (Licenciatura). Universidad Politécnica de Catalunya, Facultad de matemáticas y Estadística, España, 2004, p.7 [Consulta: 10 febrero 2018] Disponible en: <https://upcommons.upc.edu/handle/2099.1/13315?locale-attribute=es>

RAMÍREZ, M., et al. "Análisis de componentes principales funcionales para el estudio de material particulado (pm10) en la ciudad de Bogotá" *Ingeniería Y Competitividad* [en línea] 2016 (Colombia), vol. 88, no. 2, pp. 75-88. [Consulta: 10 febrero 2018] Disponible en: http://revistaingenieria.univalle.edu.co/index.php/ingenieria_y_competitividad/%20article%20/view%20/2155

RAMSAY, J. Y SILVERMAN, B. *Functional data Analysis*. 1ª ed. New York-United State of America, Springer Series in Statistics. 1997. ISBN 0-387-94956-9.

RAMSAY, J. Y SILVERMAN, B. *Functional data Analysis*. 2ª ed. New York-United State of America Springer Series in Statistics. 2005, pp.45-56. ISBN 9780387400808.

RODRÍGUEZ, R., et al. *Meteorología y Climatología*. 1ª ed. España, Fundación Española para la Ciencia y la Tecnología, p.6 ISBN 8468885355.

STRANDBERG, J. *Cluster analysis for functional data* (tesis) (Master) [en línea] Umeå University, Department of Mathematics and Mathematical Statistics, Suecia, 2013, pp15-16, [Consulta: 10 febrero 2018] Disponible en: <https://www.diva-portal.org/smash/get/diva2:691473/FULLTEXT01.pdf>

TORRECILLA, J. *Análisis de Datos Funcionales, Clasificación y Selección de Variables* (tesis) (Master) Universidad Autónoma de Madrid, España, 2010, p.13, [Consulta: 10 febrero 2018] Disponible en: https://repositorio.uam.es/bitstream/handle/10486/12556/60789_Torrecilla_Noguerales_Jose_Luis.pdf?sequence=1&isAllowed=y

URDIALES, L. Y ESPINOZA, J. *Energía solar en el Ecuador* 1ª ed. Cuenca-Ecuador, Universidad de Cuenca, 2015, pp.8-25 ISBN 9789978143179.

WACKERLY, D., et al. *Estadística Matemática con Aplicaciones*. 7ª ed. Santa Fé-Mexico, Cengage Learning. pp.1-2. ISBN 139786074813999.

ANEXOS

Anexo A: Lista de Librerías de R utilizadas

```
library(lubridate)
library(reshape2)
library(dplyr)
library(fda)
library(fda.usc)
library(dplyr)
library(readxl)
library(latex2exp)
library(calibrate)
```

Anexo B: Importación datos a Rstudio

Contiene los códigos utilizados para realizar el llamado de la información contenida en un archivo con extensión .xlsx hacia la Rstudio, y seguidamente analizar las matrices de datos obtenidos de las cuales se realizó la identificación de diferentes características estadísticas mediante el análisis de datos funcionales.

ALAO

```
alao2014 <- read_xlsx(path = "1. ALAO-2014.xlsx", sheet = "alao2014")
alao2015 <- read_xlsx(path = "2. ALAO-2015.xlsx", sheet = "alao2015")
alao2016 <- read_xlsx(path = "3. ALAO-2016.xlsx", sheet = "alao2016")
alao2017 <- read_xlsx(path = "4. ALAO-2017.xlsx", sheet = "alao2017")
```

Anexo C: Limitación de las horas de estudio

En esta parte delimitamos la información a las horas de 07h00 hasta las 19h00, esta delimitación se realizó porque cada hora contiene los promedios de radiación solar obtenidas, por ejemplo, la magnitud de radiación solar global a las 19h00 contiene la magnitud promedio de radiación obtenido desde las 18h01 hasta 19h00

```
alao2014mt <- as.matrix(t(alao2014[,8:20]))
alao2015mt <- as.matrix(t(alao2015[,8:20]))
alao2016mt <- as.matrix((alao2016[8:20,]))
alao2017mt <- as.matrix(t(alao2017[,8:20]))
```

Anexo D: Suavizado de curvas y optimización de número de bases B-spline y Fourier

El código `min.basis()` ayudo a optimizar el número de bases necesarias para detallar de mejor manera a la radiación solar global de cada una de las estaciones, además la función `plotfit.fd()` apporto en la elección de las bases B-spline con el error cuadrático medio mínimo.

```
nbopt <- min.basis(rsgestdia, type.CV = GCV.S, verbose = T, type.basis = "bspline")
```

```
The minimum GCV (GCV.OPT=27905.31) is achieved with  
the number of basis (numbasis.opt=23)  
and lambda value (lambda.opt=0)
```

```
argm <- 1:nrow(alao2014mt)
```

```
bsrsg <- create.bspline.basis(rangeval=c(min(argm), max(argm)),  
nbasis= opt$numbasis.opt)
```

La función `Data2fd()` ajusta a la información de rsg de los diferentes años de estudio obteniendo curvas que serán utilizadas en el tratamiento estadístico.

```
fdalao2014 <- Data2fd(alao2014mt, argvals = argm, bsrsg)
```

```
fdalao2015 <- Data2fd(alao2015mt, argvals = argm, bsrsg)
```

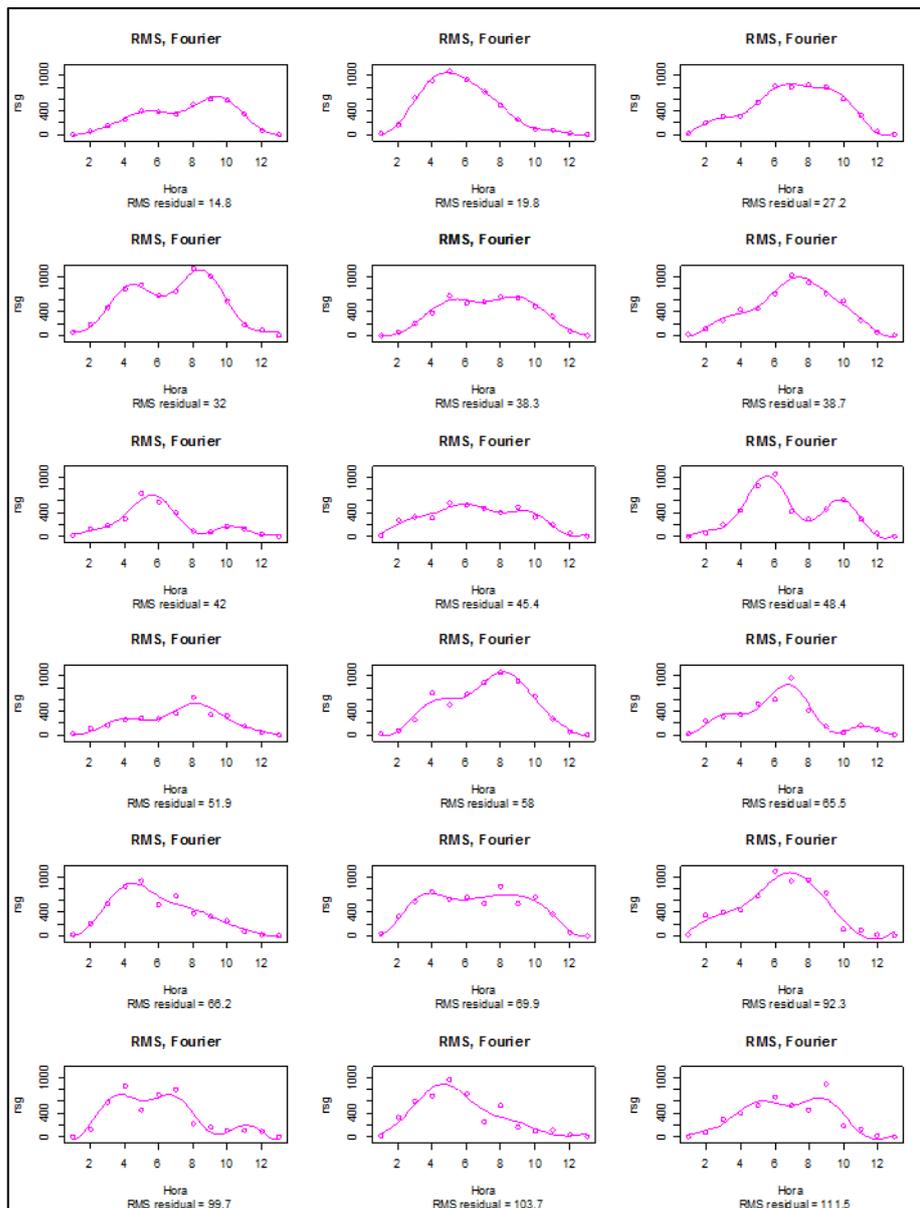
```
fdalao2016 <- Data2fd(alao2016mt, argvals = argm, bsrsg)
```

```
fdalao2017 <- Data2fd(alao2017mt, argvals = argm, bsrsg)
```

La siguiente lista de gráficos muestra los ajustes B-spline y Fourier, según el número de bases con su respectivo RMS.

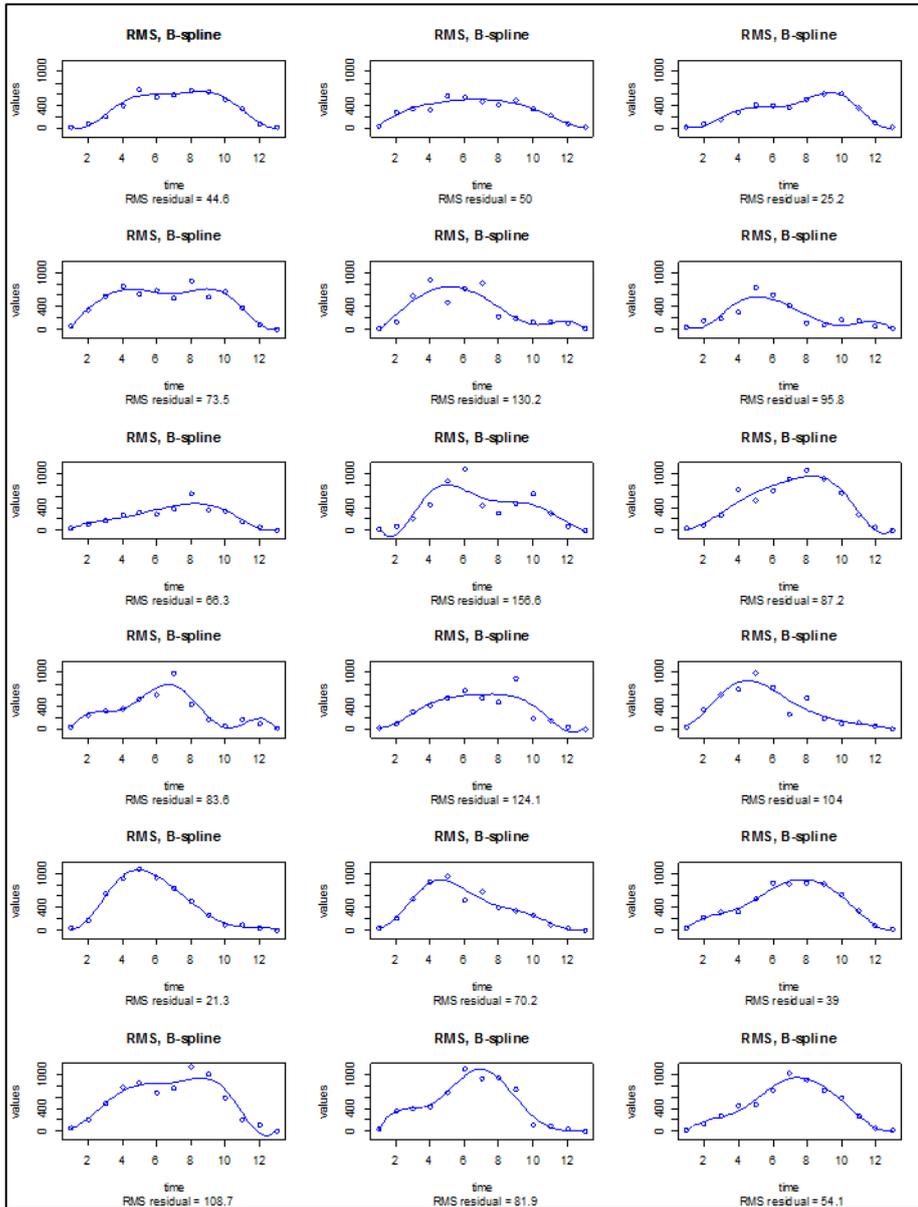
```
plotfit.fd(y = rsgestdia[,1000:1017], argvals = argm, fdoj =  
fd7rsgestdia[1000:1017], rng = fd7rsgestdia$basis$rangeval, index =  
NULL, nfine = 101, residual = F, sortwr = T, titles=NULL,  
ask=TRUE, type=c("p", "l")[1+residual],  
col = 6, lwd=1, ylab = 'rsg', xlab = 'Hora',  
cex.pch=1, axes=NULL, main = 'RMS, Fourier')
```

Con 7 bases Fourier



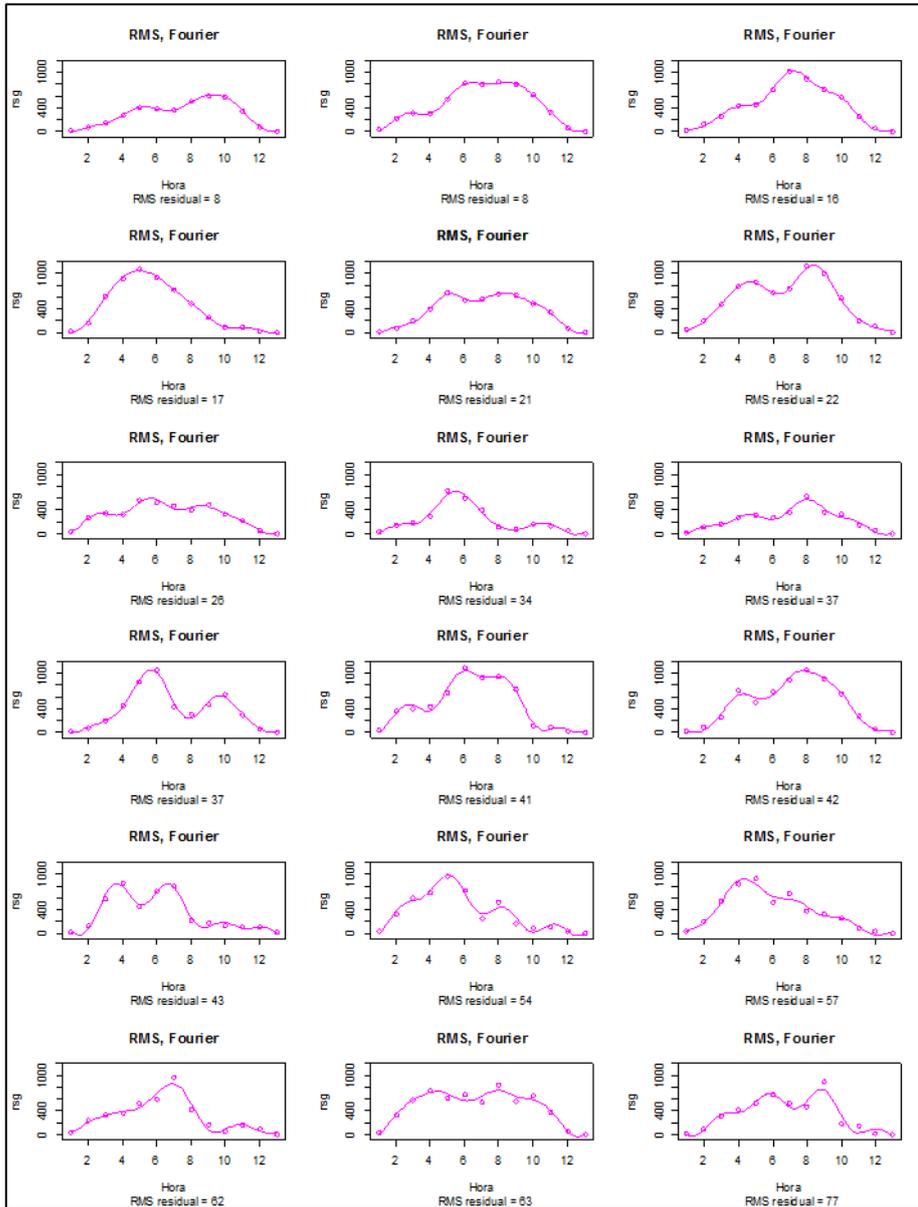
Elaborado por: Taguacundo Oscar, 2019

Con 7 bases B-spline



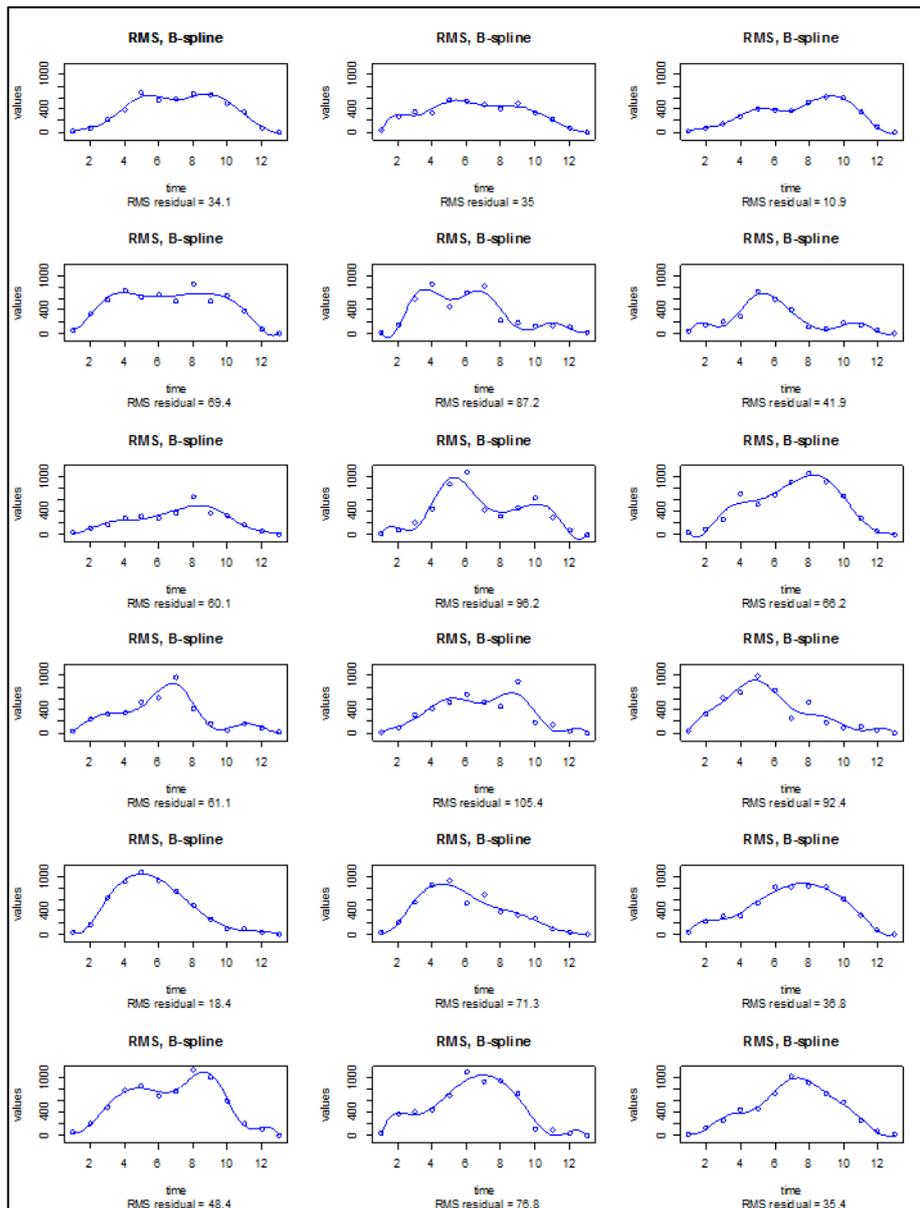
Elaborado por: Taguacundo Oscar, 2019

Con 9 bases Fourier



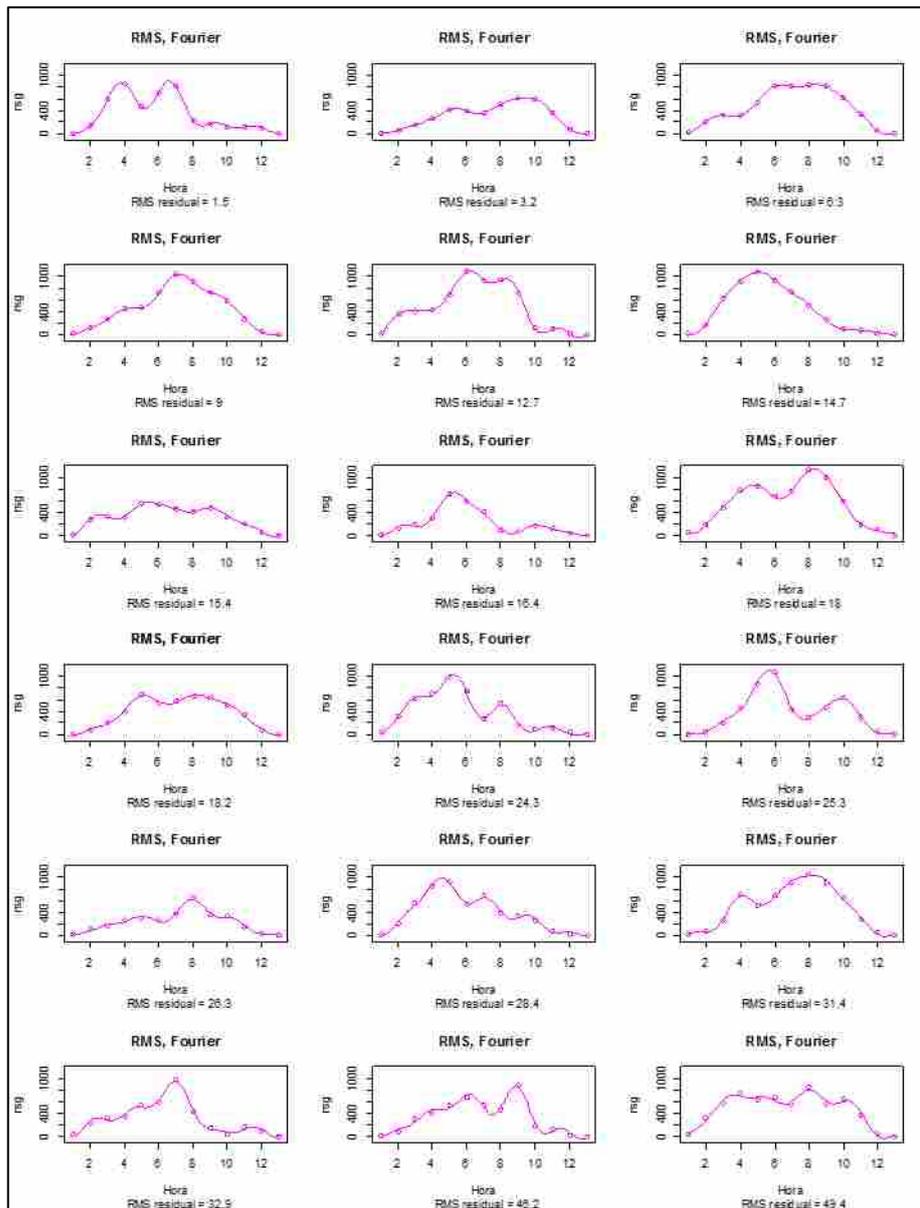
Elaborado por: Taguacundo Oscar, 2019

Con 9 bases B-spline



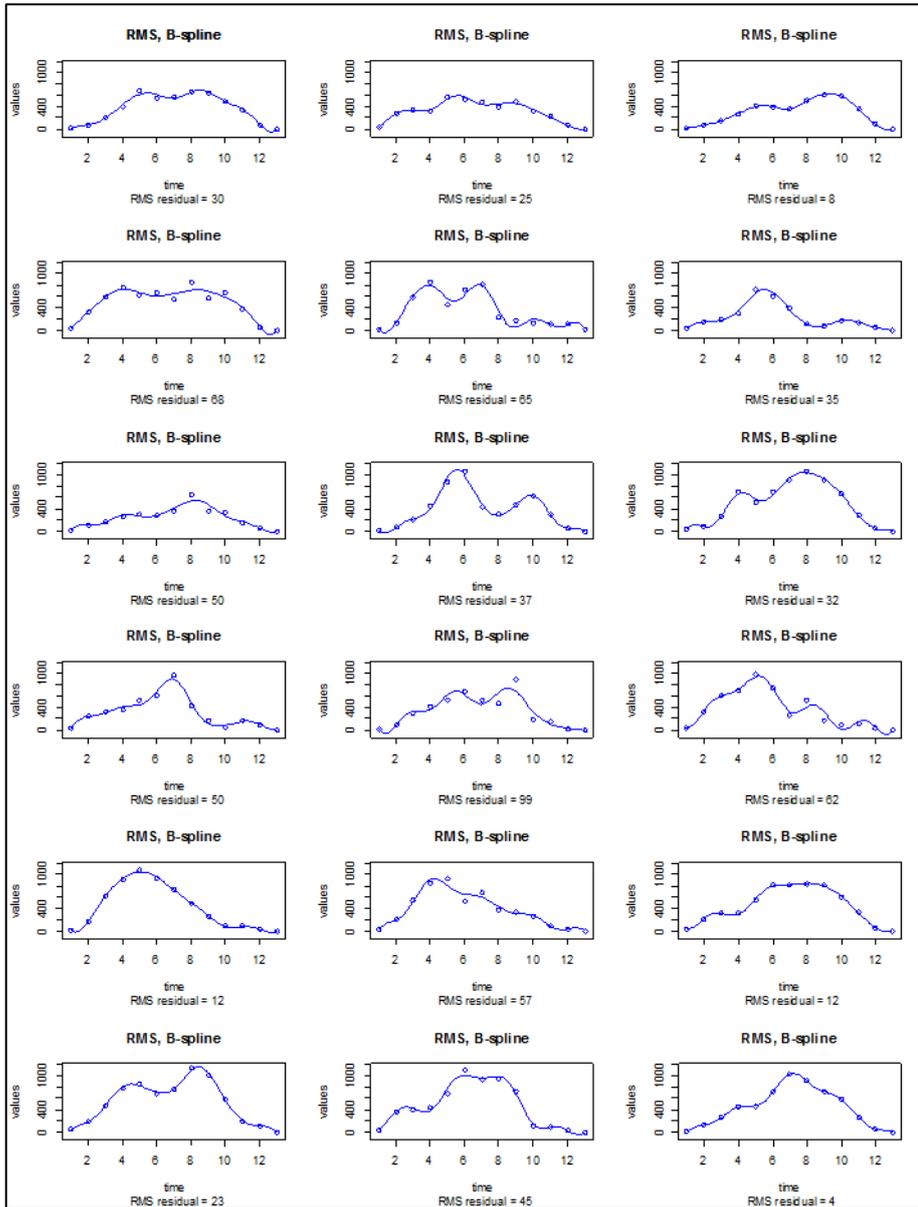
Elaborado por: Taguacundo Oscar, 2019

Con 11 bases Fourier



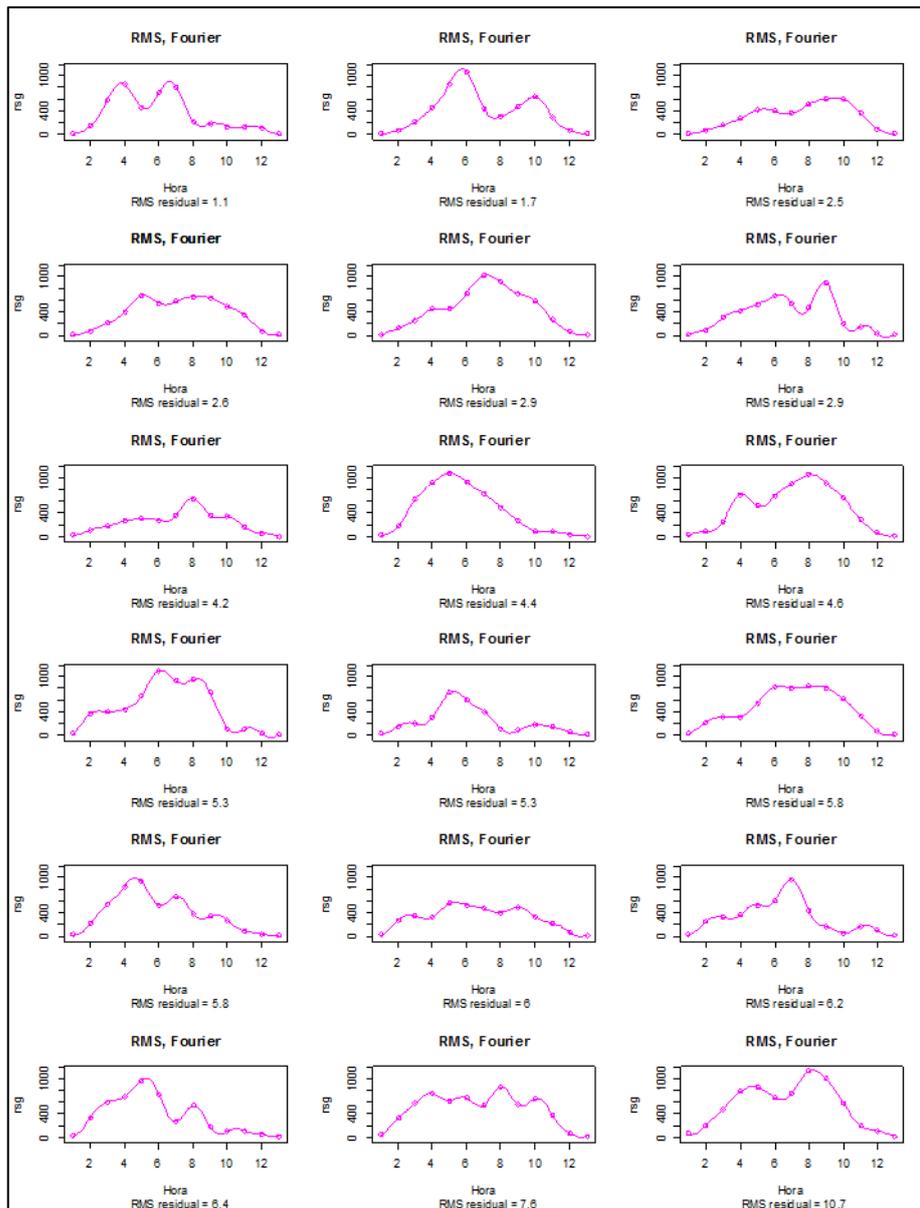
Elaborado por: Taguacundo Oscar, 2019

Con 11 bases B-spline



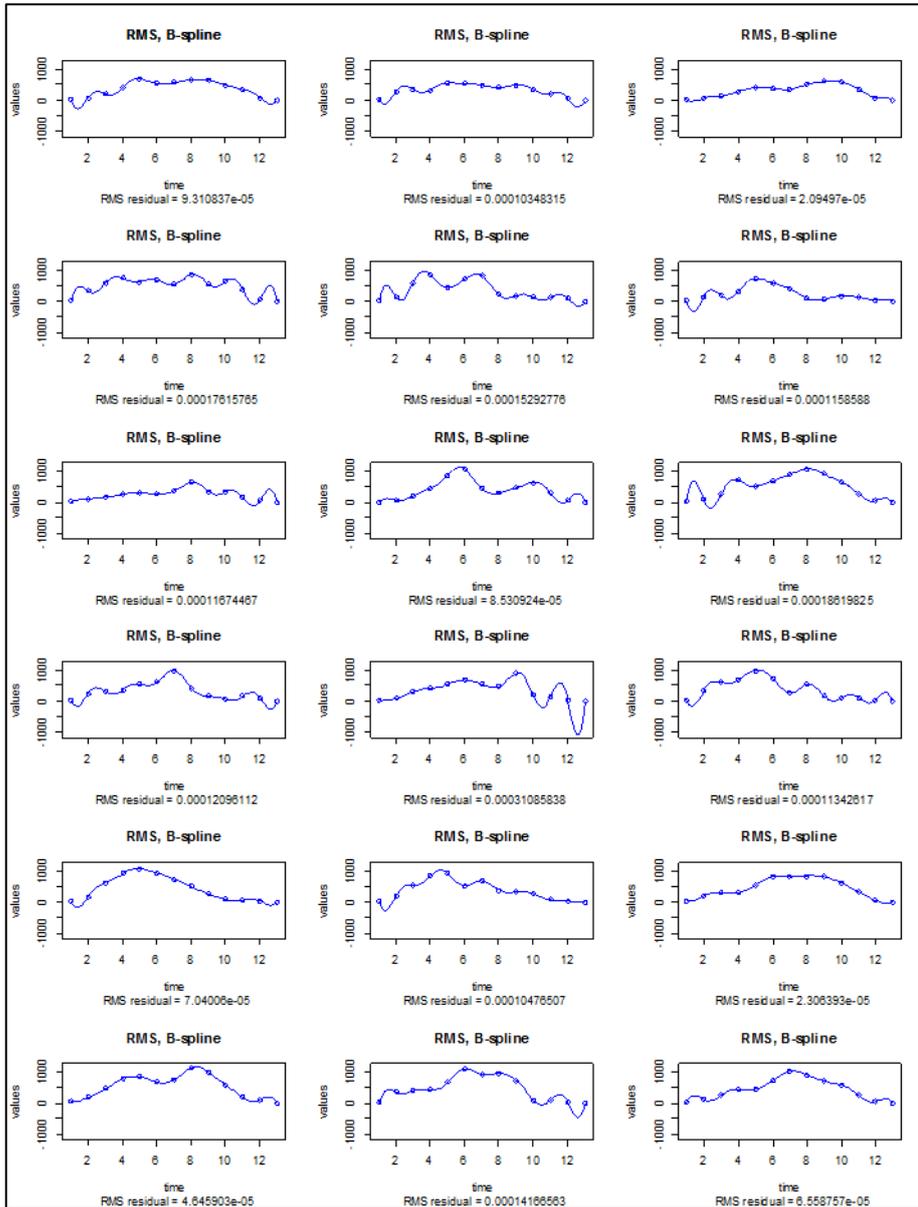
Elaborado por: Taguacundo Oscar, 2019

Con 13 bases Fourier



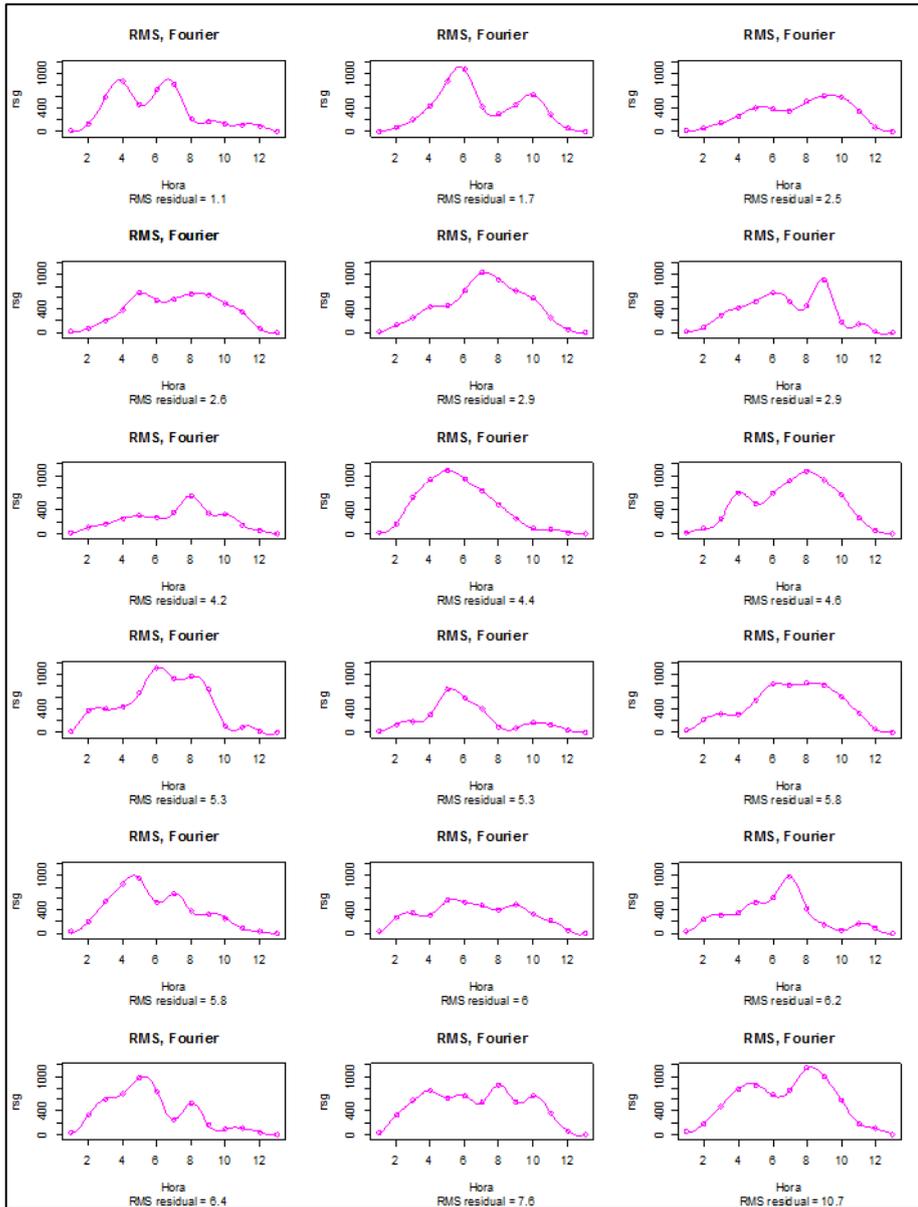
Elaborado por: Taguacundo Oscar, 2019

Con 13 bases B-spline



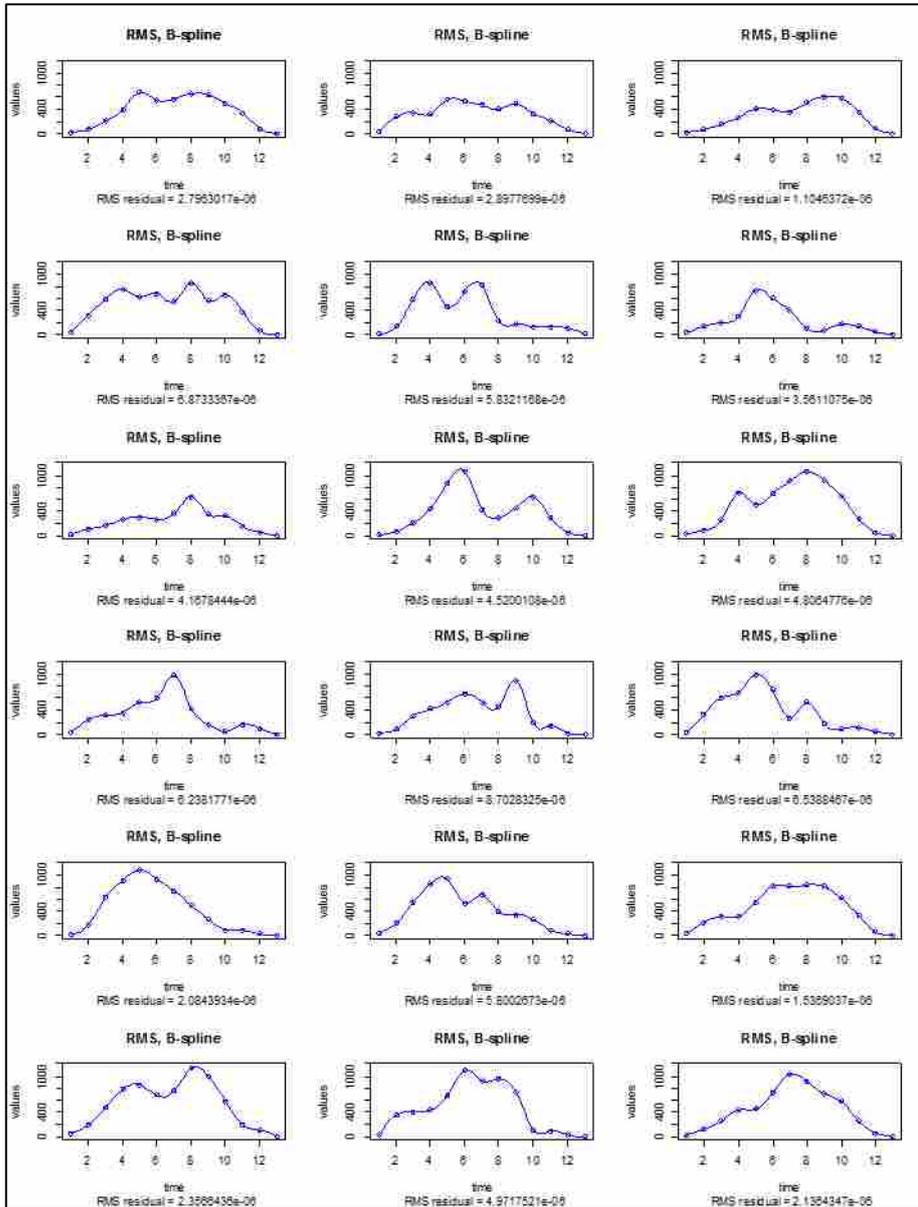
Elaborado por: Taguacundo Oscar, 2019

Con 15 bases Fourier



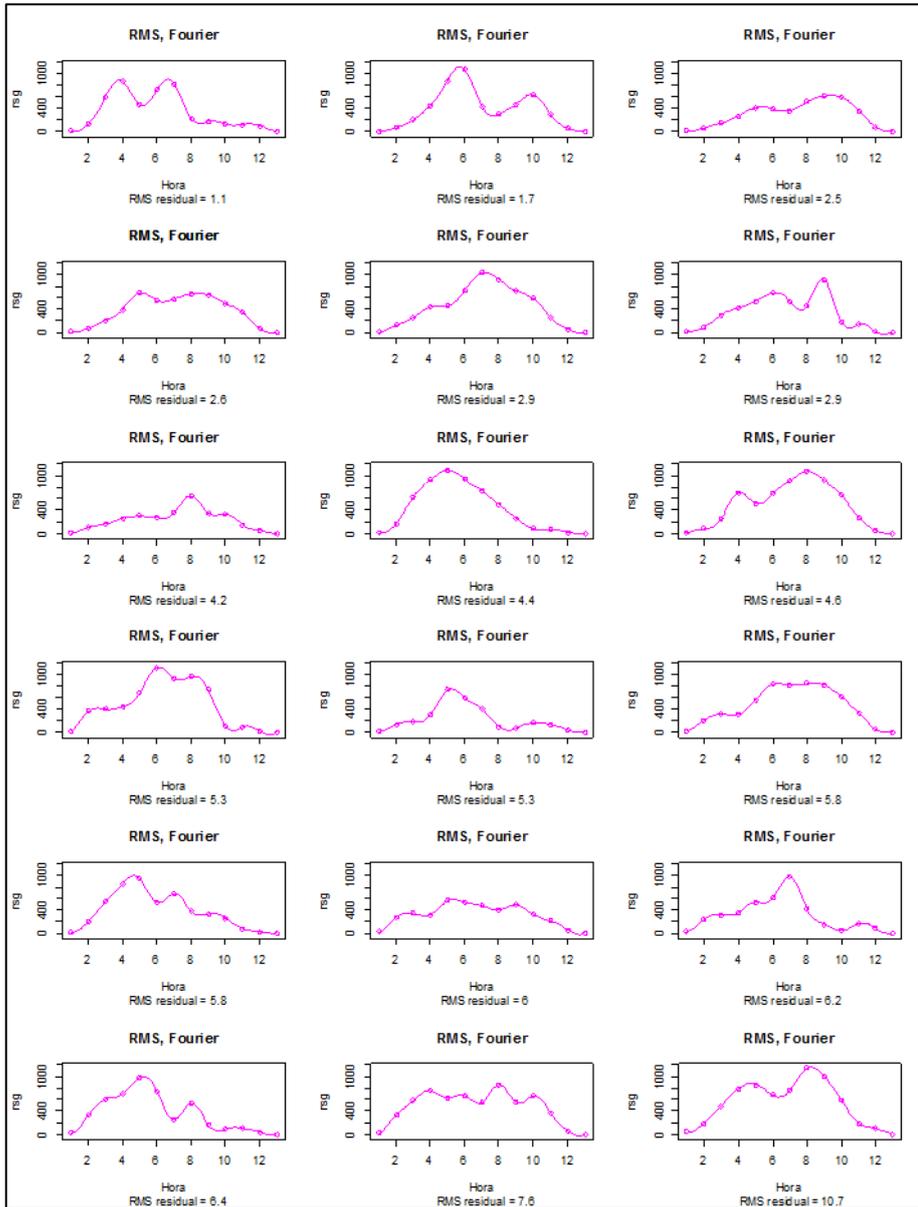
Elaborado por: Taguacundo Oscar, 2019

Con 15 bases B-spline



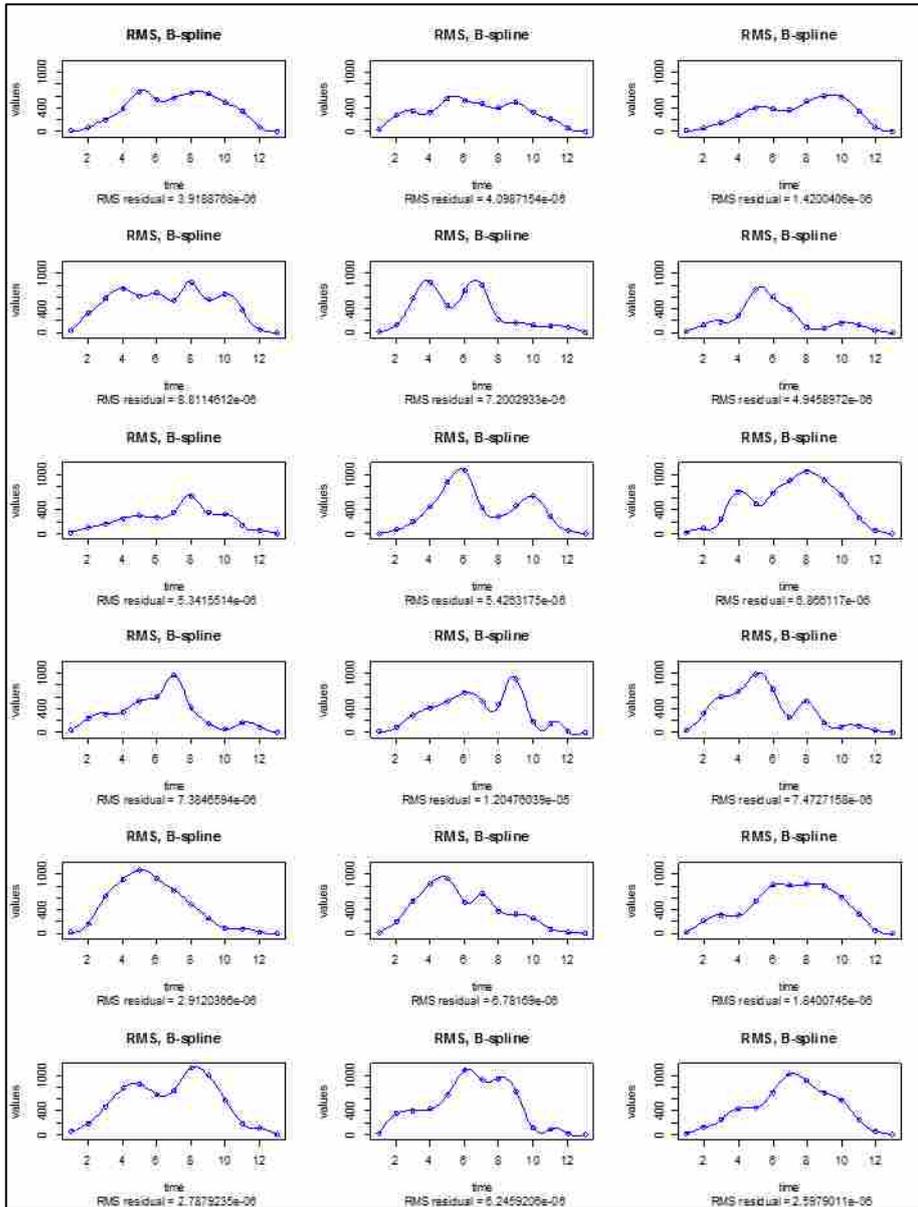
Elaborado por: Taguacundo Oscar, 2019

Con 17 bases Fourier



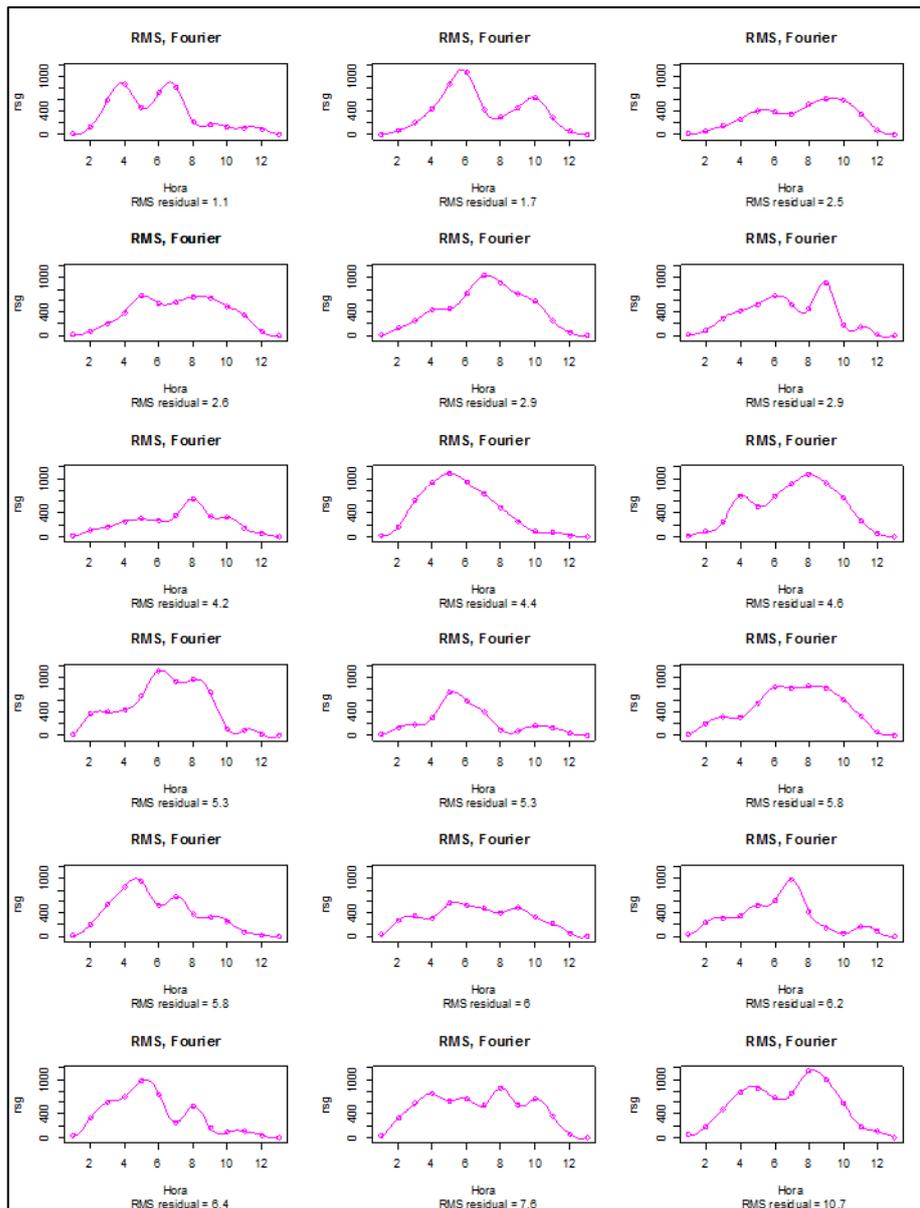
Elaborado por: Taguacundo Oscar, 2019

Con 17 bases B-spline



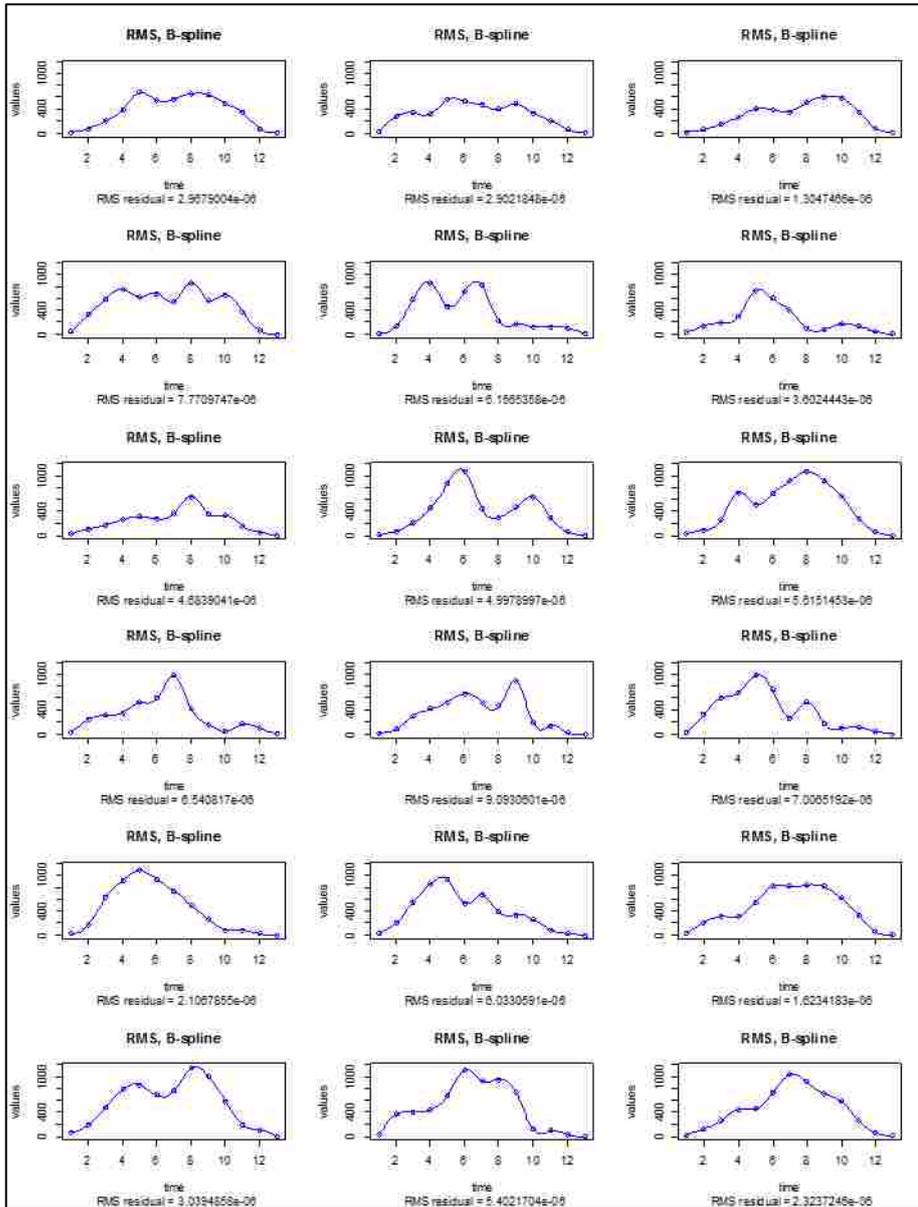
Elaborado por: Taguacundo Oscar, 2019

Con 19 bases Fourier



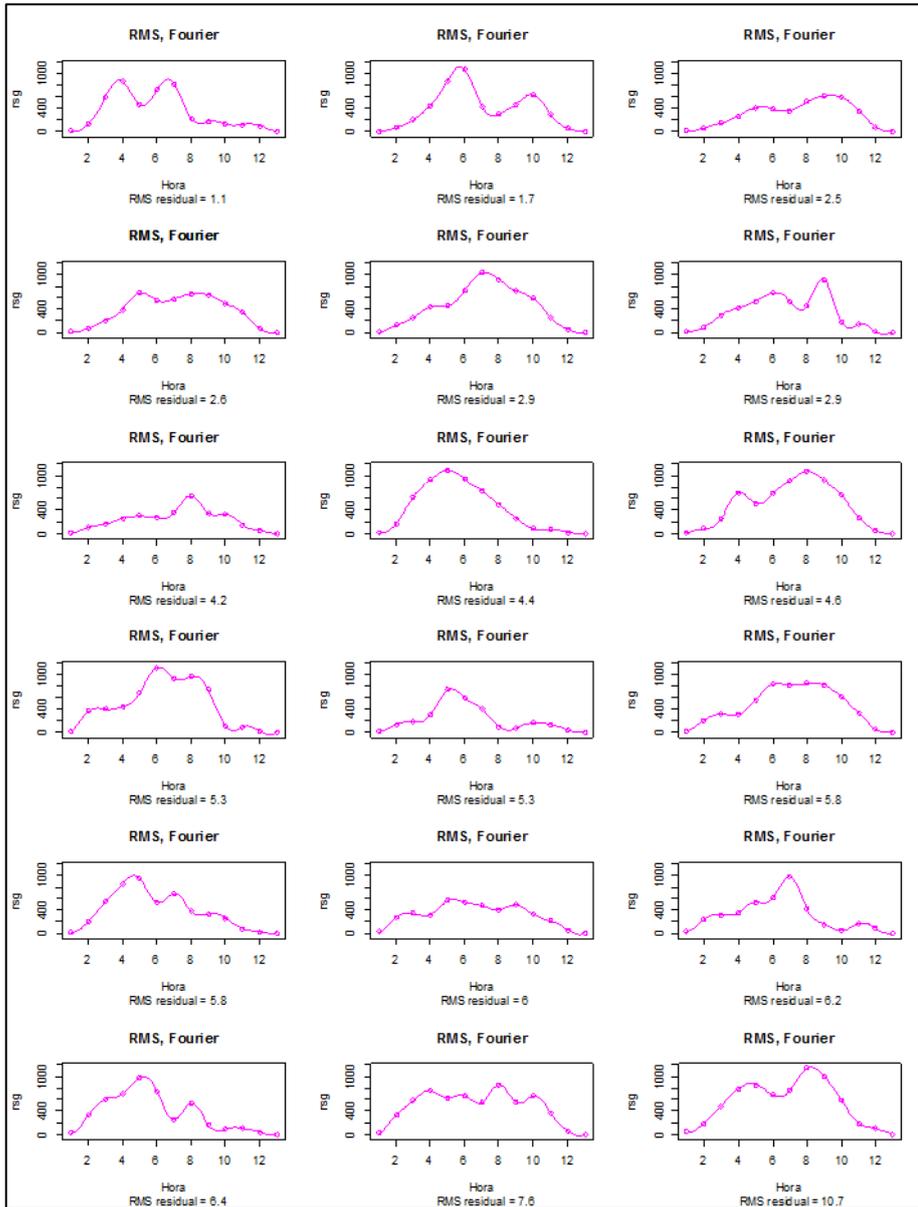
Elaborado por: Taguacundo Oscar, 2019

Con 19 bases B-spline



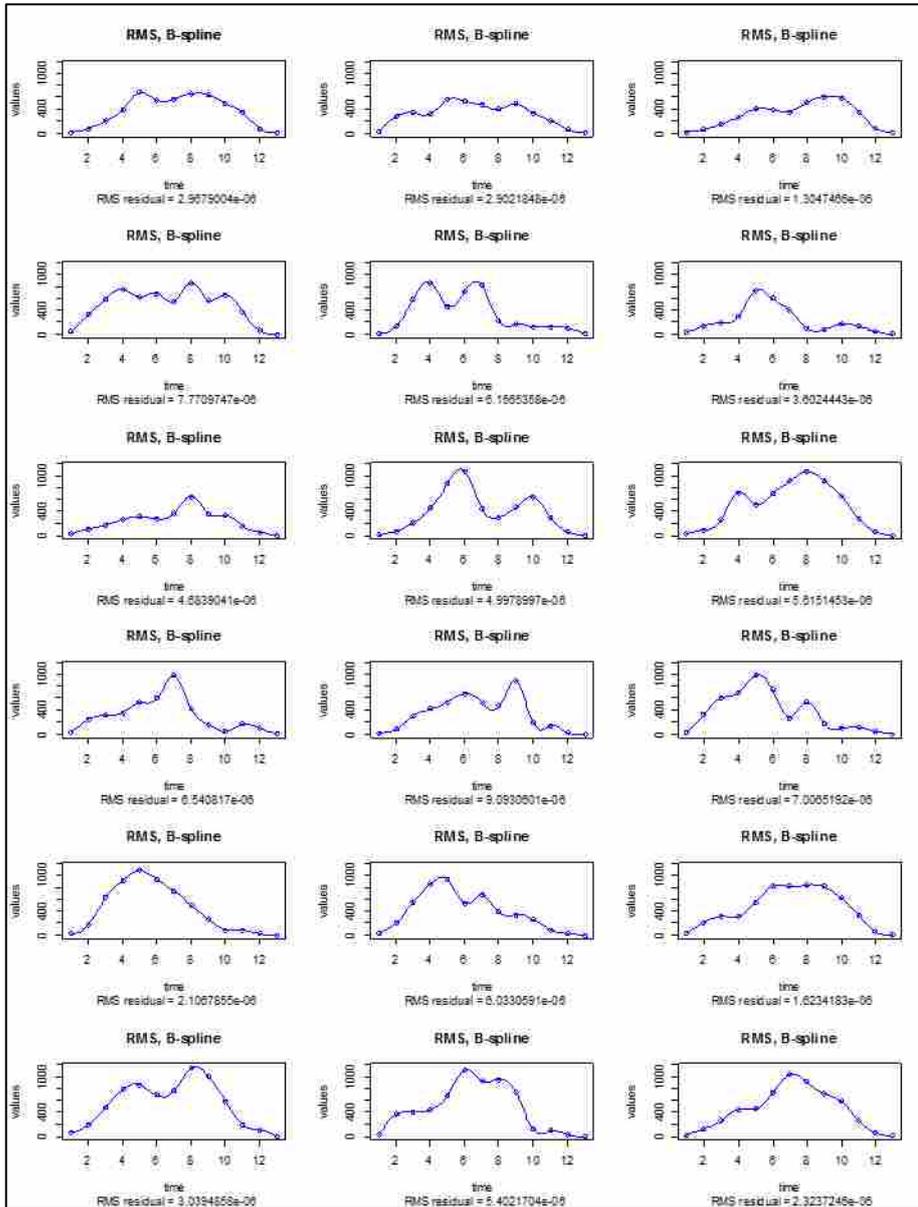
Elaborado por: Taguacundo Oscar, 2019

Con 21 bases Fourier



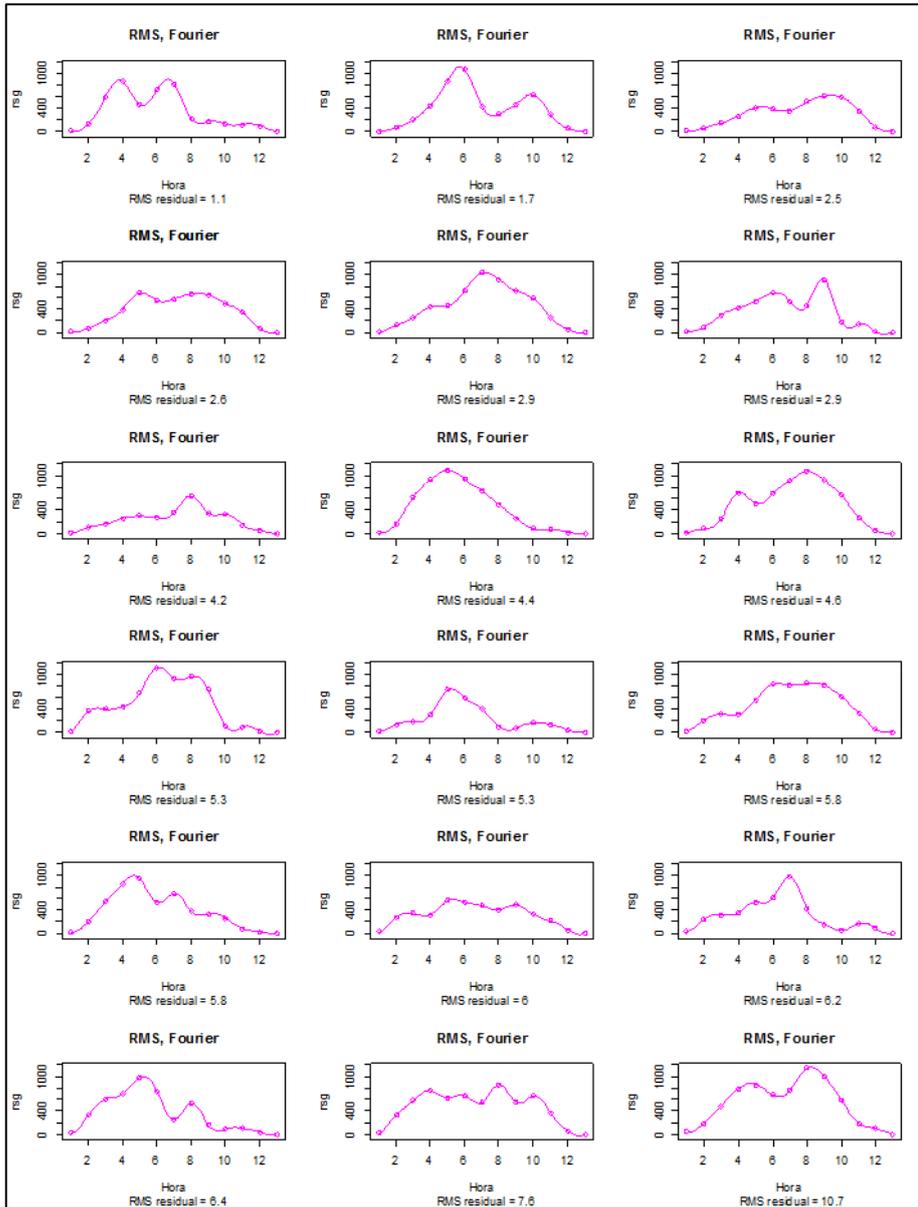
Elaborado por: Taguacundo Oscar, 2019

Con 21 bases B-spline



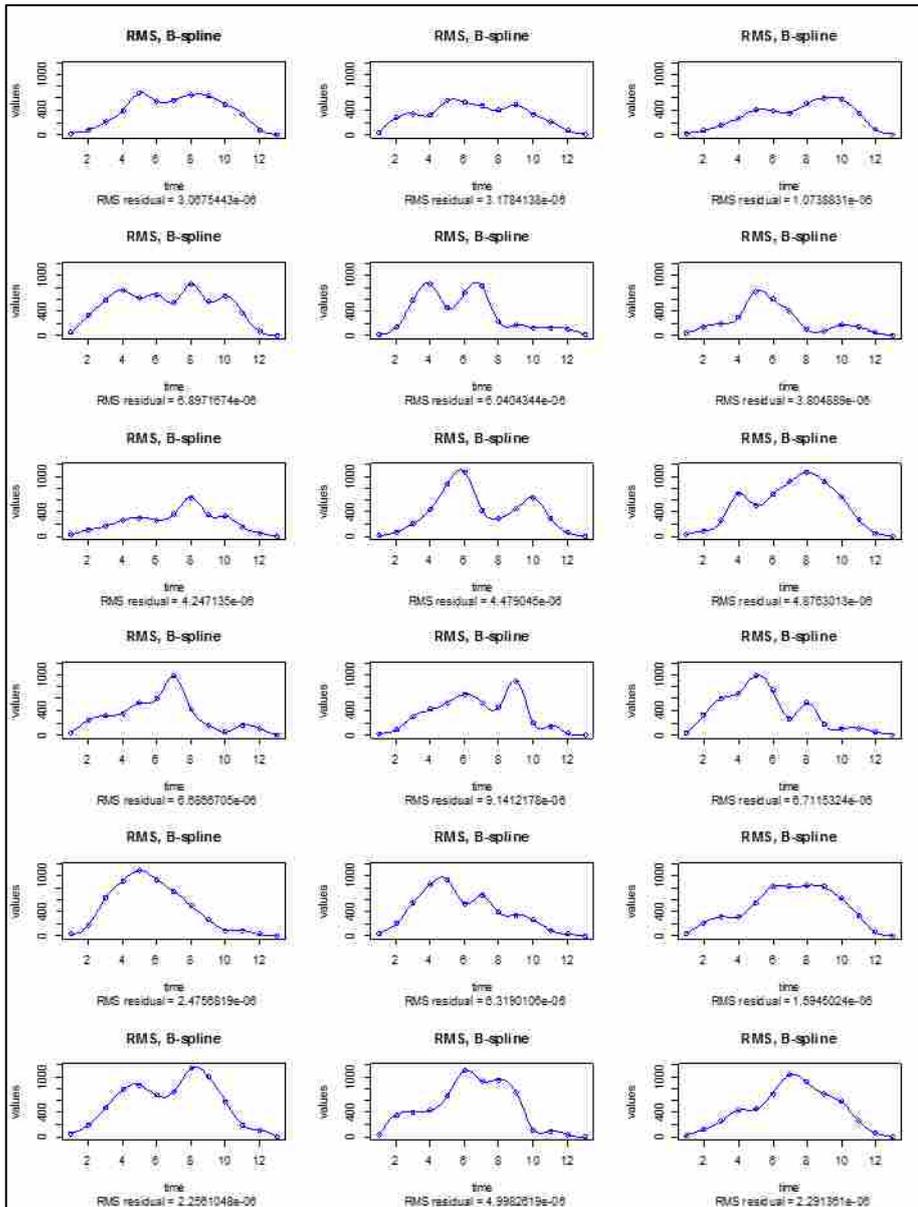
Elaborado por: Taguacundo Oscar, 2019

Con 23 bases Fourier



Elaborado por: Taguacundo Oscar, 2019

Con 23 bases B-spline



Elaborado por: Taguacundo Oscar, 2019

Anexo E: Detección de curvas atípicas

La función *set.seed()* propone una semilla aleatoria de esta manera la función *outliers.depth.trim()* realizar el mismo procedimiento y calculara el número de atípicos en cada caso, pues si no se implanta una semilla los resultados de atípicos cambia pues las remuestras utilizadas realizan cambios es los resultados.

```
set.seed(1234)
```

Alao

```
outalao2014 <- outliers.depth.trim(fdataobj = fdalao2014, nb = 200,  
smo = 0.05, trim = 0.01, quan = 0.5,  
                                dfunc=depth.mode, ns = 0.01)  
oalao2014 <- outalao2014$outliers  
sepalao2014 <- strsplit(oalao2014, split = "rep" )  
colalao2014 <- as.integer(unlist(sepalao2014))  
opalao2014 <- na.omit(colalao2014)
```

Anexo F: Análisis de estadísticas descriptivas funcionales

La función *mean.fd()*, ayudo a analizar la curva promedio diaria de cada una de las estaciones meteorológicas, además identificar su comportamiento en el transcurso del día.

Gráfico media y desviación estándar funcional

```
par(mfrow = c(1,2))  
  
plot(sofalaogl, lty = 1, col = 8, xlab = "Horas", ylab = (TeX("Radiación Solar Global,  $\$\\w\$/\$\\m^2\$"$ ))), cex.axis = 1, cex.lab = 1,  
      main = "Radiación Solar Global, Alao", cex.lab = 0.8, las = 1)  
axis(1, at = c(1:20))  
  
lines(bcsalaogl, col = "yellow")  
lines(mean.fd(sofalao2014), col = "red")  
lines(mean.fd(sofalao2015), col = "green")  
lines(mean.fd(sofalao2016), col = "Blue")  
lines(mean.fd(sofalao2017), col = "cyan")  
lines((mean.fd(sofalaogl)), col = "black")  
  
legend(x = 11,y = 1250,legend = c ("RSG", "2014", "2015", "2016",  
"2017", "G. mean"),  
       fill = c("gray","red", "green", "blue", "cyan", "black"),cex =  
0.6, text.font = 8, bg = 'white')
```

Curvas de nivel

```
contour(x = tanalao, y = tanalao, z = tvarmatalao,xlab = "Hora",  
ylab = "Hora", main = "Curvas de Nivel, Alao", col = rainbow(9))
```

Imagen Superficial raster

```
image.plot(x = tanalao, y = tanalao, z = tvarmatalao,xlab = "Hora",  
ylab = "Hora", main = "Raster, Alao")
```

Anexo G: Análisis de componentes principales funcionales

Para poder realizar el Análisis de Componentes Principales Funcionales se utilizó la función `pca.fd()`, con el fin de reducir la dimensionalidad de los meses analizados y crear grupos que contengan meses homogéneos entre sí y heterogéneos entre los grupos, además se analiza los `scatplot` de cada una de las estaciones en los respectivos años de estudio.

```
pcalao2014 <- pca.fd(fdmalao2014, nharm = 4)
vefdalao2014 <- cumsum(pcalao2014$varprop); vefdalao2014
```

Comparación de Scatplot

```
cm <- c('blue', 'blue', 'blue', 'blue', 'blue', 'orange', 'orange',
'orange', 'orange', 'blue', 'blue', 'blue')
```

```
plot(pcalao2014$scores[,1:2], col = cm, main =
paste("(", round(vefdalao2014[2], 2)*100), "%),", " de variabilidad
explicada, Alao 2014"),
  xlab = "Primera Componete", ylab = "Segunda Componete", xlim =
c(-450, 560), ylim = c(-120,170), las = 1, cex.lab = 0.9, pch = 19,
  cex.axis = 0.9)
xline( 0, lty = 2)
yline( 0, lty = 2)
```

```
pchalao2014 <- pcalao2014$scores[,1]
schalao2014 <- pcalao2014$scores[,2]
```

```
textxy(X = pchalao2014, Y = schalao2014, labs = colnames(malao2014),
  cex = 0.9)
```

Anexo H: Análisis clúster

El análisis clúster se realizó con el fin de corroborar los resultados del ACPF, con la ayuda de la función `kmeans.fd()` se logró clasificar los meses según la profundidad de Fraiman y Muniz la cual determina la distancia de las curvas mensuales a los centroides.

```
set.seed(1012)
```

```
kmalao2014 <- kmeans.fd(fdmalao2014, ncl = 2, metric = metric.Lp,
  dfunc = func.trim.FM, max.iter = 100, par.metric = NULL, par.dfunc =
list(trim=0), par.ini = list(method = "sample"), draw = F)
```

Gráficos Clústers

```
par(mfrow = c(2,2))
```

```

calao2014 <- Data2fd(argvals = kmalao2014$centers$argvals,
t(kmalao2014$centers$data), bsrsg)

clalao2014 <- kmalao2014$cluster
clalao2014[clalao2014 == 1] = "red"
clalao2014[clalao2014 == 2] = "green"

plot(fdmalao2014, lty = 2, las = 1, cex.lab = 0.8, xlab = "Hora", ylab
= (TeX(" RSG,  $w/m^2$  ")), col = clalao2014, main = "Cluster
funcional, Alao 2014") + axis(1, at = c(1:20))

lines(calao2014, lwd = 6, lty = 1, col = c("red", "green") )

Legend(x = 1,y = 700, title = "Cluster1", legend = c ('feb', 'abr',
'may', 'jun', 'jul', 'ago', 'sep'),
col = clalao2014[c(2, 4, 5, 6, 7, 8, 9)],cex = 0.8, text.font =
1, bg = 'white', lty = 2)
Legend(x = 11,y = 700, title = "Cluster2", legend = c ('ene', 'mar',
'oct', 'nov', 'dic'),
col = clalao2014[-c(2, 4, 5, 6, 7, 8, 9)],cex = 0.8, text.font
= 1, bg = 'white', lty = 2)
Legend(x = 6,y = 200, legend = c ('Centroide1', 'Centroide2'),
col = c("green", "red") ,cex = 0.8, text.font = 1, bg =
'white', lty = 1, lwd = 2)

```