



**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**  
**FACULTAD DE CIENCIAS**  
**CARRERA ESTADÍSTICA**

**ANÁLISIS ESTADÍSTICO EXPLORATORIO FUNCIONAL DE LA  
RADIACIÓN GAMMA EN LA PROVINCIA DE CHIMBORAZO**

**Trabajo de Titulación**

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

**INGENIERO EN ESTADÍSTICA INFORMÁTICA**

**AUTOR:**

**YUNGAICELA SAGÑAY VÍCTOR ANDRÉS**

Riobamba – Ecuador

2022



**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**

**FACULTAD DE CIENCIAS**

**CARRERA ESTADÍSTICA**

**ANÁLISIS ESTADÍSTICO EXPLORATORIO FUNCIONAL DE LA  
RADIACIÓN GAMMA EN LA PROVINCIA DE CHIMBORAZO**

**Trabajo de Titulación**

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

**INGENIERO EN ESTADÍSTICA INFORMÁTICA**

**AUTOR:** YUNGAICELA SAGÑAY VÍCTOR ANDRÉS

**DIRECTORA:** Ing. AMALIA ISABEL ESCUDERO VILLA MSc.

Riobamba – Ecuador

2022

© 2022, Yungaicela Sagnay Víctor Andrés

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, VÍCTOR ANDRÉS YUNGAICELA SAGÑAY, 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, 24 de marzo de 2022

A handwritten signature in purple ink, reading "Victor AU", with a stylized flourish at the end.

**Víctor Andrés Yungaicela Sagñay**

**060470846-1**

**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**  
**FACULTAD DE CIENCIAS**  
**CARRERA ESTADÍSTICA**

El Tribunal del Trabajo de Titulación, certifica que: El Trabajo de Titulación; Tipo: Proyecto de Investigación, **ANÁLISIS ESTADÍSTICO EXPLORATORIO FUNCIONAL DE LA RADIACIÓN GAMMA EN LA PROVINCIA DE CHIMBORAZO**, realizado por el señor: **VÍCTOR ANDRÉS YUNGAICELA SAGÑAY**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal autoriza su presentación.

	<b>FIRMA</b>	<b>FECHA</b>
Ing. Johanna Enith Aguilar Reyes MSc. <b>PRESIDENTE DEL TRIBUNAL</b>	 _____	2022-03-24
Ing. Amalia Isabel Escudero Villa MSc. <b>DIRECTORA DEL TRABAJO DE TITULACIÓN</b>	 _____	2022-03-24
Dra. Jheny Del Carmen Orbe Ordoñez PhD. <b>ASESORA DEL TRABAJO DE TITULACIÓN</b>	 _____	2022-03-24

## **DEDICATORIA**

Este trabajo se lo dedico a Dios, a mi familia y a mi madre.

Víctor

## **AGRADECIMIENTO**

A mi familia que siempre me apoya incondicionalmente a lo largo de la vida estudiantil. A mis maestros que han sido una gran influencia, en especial a la Ing. Isabel Escudero y la Dra. Jenny Orbe quienes tutelaron este trabajo de investigación. Al Grupo de Investigación-Desarrollo para el Ambiente y Cambio Climático “GIDAC” de la Escuela Superior Politécnica de Chimborazo, por su aporte constante a los estudiantes.

Víctor

## ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS.....	ix
ÍNDICE DE ILUSTRACIONES.....	x
ÍNDICE DE ANEXOS.....	xiii
RESUMEN.....	1
ABSTRACT.....	2
INTRODUCCIÓN.....	3

## CAPÍTULO I

<b>1. MARCO TEÓRICO REFERENCIAL.....</b>	<b>8</b>
<b>1.1. Bases teóricas.....</b>	<b>8</b>
1.1.1. <i>Análisis de datos funcionales ADF.....</i>	8
1.1.2. <i>Dato funcional.....</i>	8
1.1.3. <i>Variable Funcional.....</i>	8
1.1.4. <i>Datos funcionales atípicos.....</i>	9
1.1.5. <i>Estadísticas descriptivas funcionales.....</i>	9
1.1.6. <i>Bases.....</i>	10
1.1.7. <i>Bases fourier.....</i>	10
1.1.8. <i>Splines.....</i>	11
1.1.9. <i>B-Splines.....</i>	12
1.1.10. <i>Suavizado.....</i>	12
1.1.11. <i>Remuestreo de bootstrap.....</i>	13
1.1.12. <i>Análisis funcional de la varianza.....</i>	13
1.1.13. <i>Razones por la cual clasificar.....</i>	14
1.1.14. <i>Análisis clúster para datos funcionales.....</i>	14
1.1.15. <i>Pasos para un correcto análisis de clusters.....</i>	15
1.1.16. <i>K-means.....</i>	16
1.1.17. <i>Algoritmo k-means para análisis de datos funcionales.....</i>	16
1.1.18. <i>Software estadístico R.....</i>	17
1.1.19. <i>Radiación.....</i>	17
1.1.20. <i>Radiación gamma.....</i>	17
1.1.21. <i>Efectos de la radiación en los ecosistemas.....</i>	18



1.1.22.	<i>Radiación ionizante</i> .....	18
1.1.23.	<i>Espectrometría gamma</i> .....	19
1.1.24.	<i>Espectrómetro</i> .....	19
1.1.25.	<i>Dosis absorbida</i> .....	19
1.1.26.	<i>Dosis equivalente</i> .....	19
1.1.27.	<i>Dosis efectiva</i> .....	19
1.1.28.	<i>Límite de dosis</i> .....	20
1.1.29.	<i>Síntomas inmediatos en la salud</i> .....	20
1.1.30.	<i>Detección de la radiación gamma</i> .....	20
1.1.31.	<i>Interacciones de rayos gamma con la materia</i> .....	21
1.1.32.	<i>Absorción fotoeléctrica</i> .....	21
1.1.33.	<i>Dispersión compton</i> .....	21
1.1.34.	<i>Producción de pares</i> .....	22
1.1.35.	<i>Efectos biológicos de las radiaciones ionizantes</i> .....	22
1.1.36.	<i>Relación dosis-efecto</i> .....	22
1.1.37.	<i>Efectos determinísticos</i> .....	22
1.1.38.	<i>Los efectos no determinísticos</i> .....	23
1.1.39.	<i>Efecto de la radiación a nivel celular</i> .....	23

## CAPÍTULO II

2.	<b>MARCO METODOLÓGICO</b> .....	24
2.1.	<b>Tipo de investigación</b> .....	24
2.2.	<b>Localización de estudio</b> .....	24
2.3.	<b>Muestra</b> .....	25
2.4.	<b>Identificación de variables</b> .....	26
2.5.	<b>Modelo estadístico</b> .....	26
2.6.	<b>Método para la obtención de los datos funcionales</b> .....	26
2.7.	<b>Método de realización del análisis estadístico</b> .....	27
2.7.1.	<i>Detección de outliers</i> .....	27
2.7.2.	<i>Análisis descriptivo y exploratorio funcional</i> .....	27
2.7.3.	<i>Análisis de varianza funcional</i> .....	27
2.7.4.	<i>Variabilidad funcional</i> .....	27
2.7.5.	<i>Remuestreo de bootstrap</i> .....	27
2.7.6.	<i>Análisis de clúster funcional</i> .....	28

## CAPÍTULO III

<b>3.</b>	<b>RESULTADOS Y DISCUSIÓN</b> .....	29
<b>3.1.</b>	<b>Construcción de la base de datos</b> .....	29
<b>3.2.</b>	<b>Selección de base y número de funciones</b> .....	30
<b>3.3.</b>	<b>Calidad de ajuste</b> .....	31
<b>3.4.</b>	<b>Determinación de outliers funcionales</b> .....	39
<b>3.4.1.</b>	<i>Alausí</i> .....	39
<b>3.4.2.</b>	<i>Chambo</i> .....	41
<b>3.4.3.</b>	<i>Chunchi</i> .....	42
<b>3.4.4.</b>	<i>Colta</i> .....	44
<b>3.4.5.</b>	<i>Cumandá</i> .....	46
<b>3.4.6.</b>	<i>Guano</i> .....	47
<b>3.4.7.</b>	<i>Pallatanga</i> .....	49
<b>3.4.8.</b>	<i>Penipe</i> .....	51
<b>3.4.9.</b>	<i>Riobamba</i> .....	52
<b>3.5.</b>	<b>Análisis descriptivo funcional</b> .....	54
<b>3.6.</b>	<b>Variabilidad funcional</b> .....	59
<b>3.7.</b>	<b>Remuestreo de bootstrap</b> .....	64
<b>3.8.</b>	<b>Análisis clúster funcional</b> .....	69
<b>3.9.1.</b>	<i>Número óptimo de clusters</i> .....	69
<b>3.9.2.</b>	<i>Composición de los clusters</i> .....	73
<b>3.9.</b>	<b>ANOVAF entre clústers</b> .....	75
<b>3.10.</b>	<b>ANOVAF de los clusters</b> .....	75
<b>3.11.</b>	<b>Caracterización de los clusters</b> .....	77
<b>3.11.1.</b>	<i>Clúster en función de la altitud del cantón</i> .....	79
	<b>CONCLUSIONES</b> .....	82
	<b>RECOMENDACIONES</b> .....	84
	<b>BIBLIOGRAFÍA</b>	
	<b>ANEXOS</b>	

## ÍNDICE DE TABLAS

<b>Tabla 1-3:</b>	Construcción de la base de datos de la tasa de radiación gamma en Chimborazo	29
<b>Tabla 2-3:</b>	Cantones y puntos de muestra .....	30
<b>Tabla 3-3:</b>	Comparativa de B-splines.....	33
<b>Tabla 4-3:</b>	Outliers encontrados con cada una de las técnicas en Alausí.....	39
<b>Tabla 5-3:</b>	Outliers encontrados con cada una de las técnicas en Chambo .....	41
<b>Tabla 6-3:</b>	Outliers encontrados con cada una de las técnicas en Chunchi.....	42
<b>Tabla 7-3:</b>	Outliers encontrados con cada una de las técnicas en Colta.....	44
<b>Tabla 8-3:</b>	Outliers encontrados con cada una de las técnicas en Cumandá.....	46
<b>Tabla 9-3:</b>	Outliers encontrados con cada una de las técnicas en Guano.....	47
<b>Tabla 10-3:</b>	Outliers encontrados con cada una de las técnicas en Pallatanga.....	49
<b>Tabla 11-3:</b>	Outliers encontrados con cada una de las técnicas en Penipe .....	51
<b>Tabla 12-3:</b>	Outliers encontrados con cada una de las técnicas en Riobamba.....	52
<b>Tabla 13-3:</b>	Composición por clusters .....	73
<b>Tabla 14-3:</b>	Composición por cantones .....	74
<b>Tabla 15-3:</b>	Composición de clusters por color .....	74
<b>Tabla 16-3:</b>	Composición de clusters.....	79
<b>Tabla 17-3:</b>	Composición por altitud .....	79
<b>Tabla 18-3:</b>	Composición de clusters k=3.....	79

## ÍNDICE DE ILUSTRACIONES

<b>Ilustración 1-1:</b>	Campo eléctrico y campo magnético .....	18
<b>Ilustración 2-1:</b>	Muestra la emisión de rayos gamma a la cara del detector.....	21
<b>Ilustración 1-2:</b>	Mapa de la provincia de Chimborazo con sus puntos de .....	25
<b>Ilustración 1-3:</b>	Número de Bases B-Spline óptimos .....	30
<b>Ilustración 2-3:</b>	Calidad de ajuste por RMS .....	31
<b>Ilustración 3-3:</b>	RMS con varios Bsplines.....	32
<b>Ilustración 4-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma .....	33
<b>Ilustración 5-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	34
<b>Ilustración 6-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	34
<b>Ilustración 7-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	35
<b>Ilustración 8-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	35
<b>Ilustración 9-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	36
<b>Ilustración 10-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	36
<b>Ilustración 11-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	37
<b>Ilustración 12-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	37
<b>Ilustración 13-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	38
<b>Ilustración 14-3:</b>	Curvas suavizadas de la tasa de dosis de rayos gamma del cantón .....	38
<b>Ilustración 15-3:</b>	Detección de outliers con técnica Bagplot.....	39
<b>Ilustración 16-3:</b>	Detección de outliers con técnica Rainbow para el cantón.....	40
<b>Ilustración 17-3:</b>	Detección de outliers con técnica HDR para el cantón.....	40
<b>Ilustración 18-3:</b>	Detección de outliers con técnica .....	41
<b>Ilustración 19-3:</b>	Detección de outliers con técnica Rainbow .....	41
<b>Ilustración 20-3:</b>	Detección de outliers con técnica HDR para .....	42
<b>Ilustración 21-3:</b>	Detección de outliers con técnica Bagplot.....	43
<b>Ilustración 22-3:</b>	Detección de outliers con técnica Rainbow .....	43
<b>Ilustración 23-3:</b>	Detección de outliers con técnica HDR .....	44
<b>Ilustración 24-3:</b>	Detección de outliers con técnica Bagplot.....	44
<b>Ilustración 25-3:</b>	Detección de outliers con técnica Rainbow .....	45
<b>Ilustración 26-3:</b>	Detección de outliers con técnica HDR .....	45
<b>Ilustración 27-3:</b>	Detección de outliers con técnica Bagplot.....	46
<b>Ilustración 28-3:</b>	Detección de outliers con técnica Rainbow .....	46
<b>Ilustración 29-3:</b>	Detección de outliers con técnica HDR para el cantón Cumandá .....	47

<b>Ilustración 30-3:</b>	Detección de outliers con técnica Bagplot para el cantón Guano .....	48
<b>Ilustración 31-3:</b>	Detección de outliers con técnica Rainbow para el cantón Guano .....	48
<b>Ilustración 32-3:</b>	Detección de outliers con técnica HDR .....	49
<b>Ilustración 33-3:</b>	Detección de outliers con técnica Bagplot.....	49
<b>Ilustración 34-3:</b>	Detección de outliers con técnica Rainbow .....	50
<b>Ilustración 35-3:</b>	Detección de outliers con técnica HDR .....	50
<b>Ilustración 36-3:</b>	Detección de outliers con técnica Bagplot para .....	51
<b>Ilustración 37-3:</b>	Detección de outliers con técnica Rainbow .....	52
<b>Ilustración 38-3:</b>	Detección de outliers con técnica HDR .....	52
<b>Ilustración 39-3:</b>	Detección de outliers con técnica Bagplot.....	53
<b>Ilustración 40-3:</b>	Detección de outliers con técnica Rainbow .....	53
<b>Ilustración 41-3:</b>	Detección de outliers con técnica HDR para .....	54
<b>Ilustración 42-3:</b>	Media y desviación funcional en el cantón Alausí .....	55
<b>Ilustración 43-3:</b>	Media y desviación funcional en el cantón Colta .....	55
<b>Ilustración 44-3:</b>	Media y desviación funcional en el cantón Cumandá.....	56
<b>Ilustración 45-3:</b>	Media y desviación funcional en el cantón Guamote .....	56
<b>Ilustración 46-3:</b>	Media y desviación funcional en el cantón Guano .....	57
<b>Ilustración 47-3:</b>	Media y desviación funcional en el cantón Pallatanga .....	57
<b>Ilustración 48-3:</b>	Media y desviación funcional en el cantón Penipe .....	58
<b>Ilustración 49-3:</b>	Media y desviación funcional en el cantón Riobamba .....	58
<b>Ilustración 50-3:</b>	Variabilidad funcional en el cantón Alausí.....	59
<b>Ilustración 51-3:</b>	Variabilidad funcional en el cantón Chambo.....	60
<b>Ilustración 52-3:</b>	Variabilidad funcional en el cantón Chunchi.....	60
<b>Ilustración 53-3:</b>	Variabilidad funcional en el cantón Colta .....	61
<b>Ilustración 54-3:</b>	Variabilidad funcional en el cantón Cumandá.....	61
<b>Ilustración 55-3:</b>	Variabilidad funcional en el cantón Guamote.....	62
<b>Ilustración 56-3:</b>	Variabilidad funcional en el cantón Guano.....	62
<b>Ilustración 57-3:</b>	Variabilidad funcional en el cantón Pallatanga.....	63
<b>Ilustración 58-3:</b>	Variabilidad funcional en el cantón Penipe .....	63
<b>Ilustración 59-3:</b>	Variabilidad funcional en el cantón Riobamba.....	64
<b>Ilustración 60-3:</b>	Remuestreo Bootstrap para el cantón Alausí .....	65
<b>Ilustración 61-3:</b>	Remuestreo Bootstrap para el cantón Alausí .....	65
<b>Ilustración 62-3:</b>	Remuestreo Bootstrap para el cantón Chunchi .....	66
<b>Ilustración 63-3:</b>	Remuestreo Bootstrap para el cantón Colta.....	66
<b>Ilustración 64-3:</b>	Remuestreo Bootstrap para el cantón Cumandá .....	67

<b>Ilustración 65-3:</b>	Remuestreo Bootstrap para el cantón Guamote .....	67
<b>Ilustración 66-3:</b>	Remuestreo Bootstrap para el cantón Guano .....	68
<b>Ilustración 67-3:</b>	Remuestreo Bootstrap para el cantón Pallatanga .....	68
<b>Ilustración 68-3:</b>	Remuestreo Bootstrap para el cantón Penipe.....	69
<b>Ilustración 69-3:</b>	Remuestreo Bootstrap para el cantón Riobamba .....	69
<b>Ilustración 70-3:</b>	Número óptimo de Clúster $k=4$ .....	70
<b>Ilustración 71-3:</b>	Número óptimo de Clúster $k=3$ .....	71
<b>Ilustración 72-3:</b>	Mapa clúster para la provincia de Chimborazo .....	72
<b>Ilustración 73-3:</b>	ANOVAF entre clústers.....	75
<b>Ilustración 74-3:</b>	ANOVAF del clúster 1 .....	75
<b>Ilustración 75-3:</b>	ANOVAF del clúster 2 .....	76
<b>Ilustración 76-3:</b>	ANOVAF del clúster 3 .....	77
<b>Ilustración 77-3:</b>	Caracterización de los clusters.....	78
<b>Ilustración 78-3:</b>	La media funcional de cada clúster.....	78
<b>Ilustración 79-3:</b>	Mapa Clúster de la altitud de los cantones.....	81

## **ÍNDICE DE ANEXOS**

**ANEXO A:** CÓDIGO EN R

**ANEXO B:** AVAL

## RESUMEN

En este trabajo de tesis se analizó el comportamiento de la tasa de dosis de radiación gamma presente en el ambiente, en la provincia de Chimborazo. Se utilizó metodología funcional mediante R, tanto para identificar las observaciones faltantes, atípicas y suavizado óptimo, como para la caracterización de su comportamiento. Los datos fueron muestreados por el Grupo de Investigación y Desarrollo para el Ambiente y Cambio Climático (GIDAC) de la Escuela Superior Politécnica de Chimborazo (ESPOCH) en: Guano, Alausí, Chambo, Chunchi, Colta, Cumandá, Guamote, Guano, Pallatanga, Penipe y Riobamba durante junio, julio y agosto del año 2021. La validación de datos se realizó mediante una intensa exploración de atípicos, a través de 4 métodos, se identificó curvas anómalas en todos los cantones, a excepción de Guamote. En análisis descriptivo de datos funcionales mostró que la tasa de dosis de radiación gamma en Colta, Alausí, Guamote, Guano; y Riobamba, poseen una curva media más amplia que el resto de cantones. El cantón Cumandá presentó un intervalo más bajo de dosis que el resto de cantones. Se concluye que la desviación típica introdujo una idea clara en el comportamiento de la tasa de dosis de radiación gamma, identificando poca variabilidad en todos los cantones a excepción de Guamote. El análisis de clusters permitió identificar características similares de cada cantón. Colta, Guano, Pallatanga y Riobamba presentaron niveles de  $[0.06, 0.08]$ Sv, siendo estos los más altos, mientras que los más bajos fueron de  $[0.02, 0.04]$ Sv en Cumandá y Guamote. Cumandá y Pallatanga mostraron dosis. El ANOVA funcional al 95% de confiabilidad, identificó diferencias significativas, entre las medias del clúster 1 (Cumandá y Pallatanga) y los dos siguientes, lo que se presume que es debido a las características geográficas y climáticas de estos dos cantones.

**Palabras clave:** <ESTADÍSTICA>, <ANÁLISIS DE DATOS FUNCIONALES>, <CHIMBORAZO>, <RADIACIÓN GAMMA>, <CURVAS>, <VARIABILIDAD>.

0122-DBRA-UPT-2023





## ABSTRACT

This research work sought to analyze the behavior of the gamma radiation dose rate present in the environment, in the province of Chimborazo. Functional methodology through R was used, both to identify missing and atypical observations alongside optimal smoothing, as well as to characterize their behavior. The data were sampled by the Research and Development Group for the Environment and Climate Change (GIDAC) of the Escuela Superior Politécnica de Chimborazo (ESPOCH) in: Guano, Alausí, Chambo, Chunchi, Colta, Cumandá, Guamote, Guano, Pallatanga, Penipe and Riobamba during June, July and August of the year 2021. The data validation was done through an intense exploration of atypical data, through 4 methods, anomalous curves were identified in all the cantons, with the exception of Guamote. In descriptive analysis of functional data, it was shown that the gamma radiation dose rate in Colta, Alausí, Guamote, Guano and Riobamba, have a wider average curve than the rest of the cantons. The Cumandá canton presented a lower dose interval than the rest. It is concluded that the standard deviation introduced a clear idea in the behavior of the gamma radiation dose rate, identifying slight variability in all the cantons except Guamote. The cluster analysis allowed to identify similar characteristics of each canton. Colta, Guano, Pallatanga and Riobamba presented levels of [0.06, 0.08] Sv, these being the highest, while the lowest were [0.02, 0.04] Sv in Cumandá and Guamote. Cumandá and Pallatanga showed doses. The functional ANOVA at 95% reliability identified significant differences between the means of cluster 1 (Cumandá and Pallatanga) and the next two, which is assumed to be due to the geographical and climatic characteristics of these two cantons.

**Keywords:** <STATISTICS>, <FUNCTIONAL DATA ANALYSIS>, <CHIMBORAZO>, <GAMMARADIATION>, <CURVES>, <VARIABILITY>.



Edgar Mesias Jaramillo Moyano  
0603497397

## INTRODUCCIÓN

Muchas veces al escuchar la terminología “rayos gamma”, rápidamente se viene a la mente la palabra peligro, debido a su relación con la energía nuclear y los lamentables accidentes que han ocurrido a través de la historia, como uno de los ejemplos más sonados, se tiene el de la central nuclear de Chernóbil. Por otro lado, nuestro planeta Tierra posee una radioactividad natural, que ha sido fundamental para el desarrollo de la vida tal cual se la conoce.

Los rayos gamma son considerados radiaciones electromagnéticas producidas por la desintegración radiactiva de los núcleos atómicos, ondas, que acompañan a la emisión de partículas alfa o beta. En la Tierra se observan fuentes naturales de este tipo de destellos tanto en la desintegración de los radionucleidos como en las interacciones provenientes del cosmos con la atmosfera (Rangel, 2016, p. 2).

A través de la estadística se canalizan grandes volúmenes de información mediante distintos métodos y técnicas para la toma de decisiones. El análisis de datos funcionales es una rama de la estadística, que estudia y analiza la información tomada como curvas, superficies o cualquier elemento que varíe sobre un espacio continuo. Los objetivos que se esperan son esencialmente los mismos que en cualquier otra área de la estadística, algunos que se mencionan son: la obtención de una representación de los objetos de estudio de tal manera que facilite su entendimiento y resalte las distintas características, el estudio de posibles patrones de variabilidad, la explicación del porqué fluctúa entre resultados, o bien la comparación con respecto a cierto tipo de variaciones (Plazola, 2013, p. 4).

En esta investigación se caracterizó el comportamiento de la tasa de dosis de radiación gamma en la provincia de Chimborazo, a través de un análisis exploratorio de datos funcionales, identificación de atípicos, y mapas clusters, se usó R especialmente las librerías `fda.usc` y `fda`. 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 los rayos gamma y las técnicas y métodos estadísticos de manera funcional. El capítulo III detalla la metodología aplicada a la información. El capítulo IV los resultados obtenidos. Y finalmente las conclusiones, recomendaciones, referencias bibliográficas y Anexos.

## **Antecedentes**

El análisis funcional de datos, a pesar de que en las últimas dos décadas ha cobrado gran importancia el campo científico, es una herramienta de la estadística relativamente nueva. Inició sus pasos en los años 60 del siglo XX, y como muchas otras ciencias, se ha ido optimizando con la velocidad y manejo de grandes volúmenes de información proporcionados por las computadoras desde la década de 1990, hasta la popularidad del Big Data en la actualidad. Los padres de esta técnica, por así llamarlos, son los profesores Ramsey y Silverman (Ramsay y Silverman, 2005, p.1).

Introdujeron una nueva herramienta para el tratado de los datos en el análisis multivariante, que consiste en dejar de utilizar las observaciones en forma de vectores, y a su vez utilizarla en forma de una función. A su vez, adaptar técnicas de la estadística clásica y modelarlas para su correcto funcionamiento con el manejo de las funciones (Giraldo, 2007, p.2). Como cualquier otro instrumento estadístico, se busca una forma de representación que sea entendible, tanto para el investigador y para el lector (Millán, 2017, p.23).

Existen muchos estudios utilizando este conjunto de técnicas, pero se debe destacar el de Giraldo et al. (2010, p.6) que basó la aplicación de este análisis con ciencias ambientales, llegando a realizar una predicción basada en la correlación de curvas, en estaciones meteorológicas a través de Canadá. Peñafiel (2021, p.71) realizó un estudio con la finalidad de identificar el comportamiento de cultivos de quinua en la provincia de Chimborazo, con la ayuda de imágenes satelitales, vio similitudes y encontró relaciones altamente proporcionales con el clima de la zona y las distintas parcelas del cultivo, concluyendo que la planta se adapta mejor a climas más húmedos.

Escudero et al. (2016, p.9) estudiaron los distintos comportamientos de los periodos climáticos en la región interandina, a través de curvas funcionales, en el año 2011, donde a través de splines cúbicos analiza la radiación solar global. Asimismo, recomienda la detección de outliers para una mayor certeza en los resultados del estudio. El autor Taguacundo (2019, p.60) aplica la metodología de herramientas funcionales con el fin de identificar distintos comportamientos de la radiación solar, durante los años 2014 al 2017, donde a través de un análisis de clúster encontró comportamientos similares en diferentes estaciones meteorológicas.

En el trabajo de “Análisis geoestadístico de datos funcionales de temperatura del aire en la provincia de Chimborazo” (Checa, 2020, p.85), se aplicó esta metodología en el campo

geoestadístico. se estimó la temperatura del aire en distintas áreas no muestreadas en la provincia de Chimborazo, en los años 2014-2017, obteniendo sumas de cuadrados del error.

En el trabajo investigativo “Análisis de varianza funcional para medir el efecto de recubrimientos revestibles de almidón en características físico-químicas de papayas durante las dos primeras semanas del período de maduración post-cosecha” (Galarza, 2013, p.4), se realizó una varianza funcional con el fin de determinar la efectividad del uso de los varios tipos de películas comestibles en frutas al preservar el producto durante su periodo de maduración pos-cosecha. Los autores recomiendan el disponer de recursos, para la efectividad de los resultados con un mayor número de réplicas.

La unidad básica es la curva funcional, en general. Por ejemplo, un conjunto de imágenes de alta resolución es en un dominio de dos dimensiones. En la práctica, estos sucesos son recogidos por máquinas que toman muestras de una determinada variable en distintos puntos del continuo que se considere. (Ramsey y Silverman, 2013: p.31). El otro estudio denominado como: “Aplicación del análisis estadístico de datos funcionales al estudio del patrón de movimiento del raquis lumbar y de la pelvis durante la flexión y la extensión del tronco en personas sanas” (Fuster, 2021, p.9) se estudia el patrón de movimiento lumbopélvico en un grupo de personas con distintos padecimientos. En el estudio Sriharsha (2008, p.86) se ha intentado monitorear los niveles de radiación de fondo (gamma) en templos, histórico monumentos, escuelas y colegios, albergues, bodas, eventos y teatros en los distritos Mysore y Chamaraj Nagar, y evaluar los posibles efectos sobre el medio ambiente y la salud humana.

El autor Uyttenhove (2006, p.43) analiza la disponibilidad durante la última década de grandes detectores portátiles ha mejorado en gran medida el campo de aplicación de la espectrometría. se utilizó inicialmente para medir las consecuencias de una bomba nuclear atmosférica. Ha sido particularmente útil en el seguimiento de las consecuencias del accidente de Chernóbil en Europa. Debido al alto costo de los equipos, la recolección de información sobre tasa de dosis de radiación no es algo muy frecuente en el Ecuador, pero gracias a la ESPOCH y al Grupo de Investigación-Desarrollo para el ambiente y Cambio Climático (GIDAC), se logró monitorear la información a lo largo de la provincia.

La gran mayoría, al pensar en emisión radioactiva, asocia netamente a los laboratorios y centros de salud que están preparados con sus respectivos equipos. Sin embargo, existe la forma natural de este fenómeno proviene del espacio. No todos los lugares en la Tierra tienen un mismo nivel de radioactividad. Un claro ejemplo es la India que la radioactividad es 10 veces mayor que

algunos países de Europa. Se considera que Los Alpes y zonas que se encuentren en cordilleras presentan un nivel relativamente alto de radioactividad, por diversos factores, como la composición de su suelo y granito. Como se conoce, el continente sudamericano posee una cadena montañosa, la cual atraviesa 7 países distintos, entre uno de ellos Ecuador, denominada la Cordillera de los Andes.

## **Planteamiento del problema**

### **➤ Enunciado del problema**

La globalización y el avance de la industria ha ocasionado cambios en los factores ambientales. El Ecuador, al tener regiones tan biodiversas debe buscar una manera de conservar su medio ambiente. La radiación gamma, en caso de que su tasa de dosis de exposición sea la no adecuada, ocasiona daños en cultivos, fauna y también en la población humana, causando todo tipo de alteraciones genéticas y enfermedades. Este estudio pretende evidenciar el comportamiento de este fenómeno en la provincia de Chimborazo, a través de la herramienta estadística de análisis funcional de datos, aprovechando la masiva recolección de información a través de toda la provincia.

### **➤ Formulación (Incógnita)**

¿Cómo se comporta la radiación gamma en la provincia de Chimborazo?

## **Justificación**

La aplicación de herramientas estadísticas permite una mejor explicación del funcionamiento de distintos tipos de fenómenos, todo esto para una mejor toma de decisiones. Este comportamiento electromagnético, que podría afectar desde la cosecha, el ganado y la misma salud de los habitantes de una zona, es más que necesario sea estudiado de la manera adecuada. Por tal motivo la Escuela Superior Politécnica de Chimborazo (ESPOCH) mediante el Grupo de Investigación-Desarrollo para el ambiente y Cambio Climático (GIDAC) realizo la toma de muestras en campo y proporcione la información necesaria para el análisis estadístico funcional. El gran beneficiado de esta investigación será la sociedad en general, especialmente en la provincia de Chimborazo, por conocer el proceder de este tipo de radiación electromagnética e impacto que tienen en su vida común.

## **OBJETIVOS**

### **Objetivo general**

- Analizar el comportamiento de los rayos gamma en el tiempo en la provincia de Chimborazo mediante metodología con datos funcionales.

### **Objetivos específicos**

- Estructurar una base de datos según los requerimientos del análisis.
- Identificar el suavizado óptimo para la base funcional.
- Realizar un análisis estadístico exploratorio funcional de los datos.
- Analizar comportamientos semejantes mediante Clúster con datos funcionales.

# CAPÍTULO I

## 1. MARCO TEÓRICO REFERENCIAL

### 1.1. Bases teóricas

#### 1.1.1. *Análisis de datos funcionales ADF*

En el ADF, también conocido como Functional Data Analysis (FDA), ha experimentado un gran desarrollo en la era actual, sin importar haber aparecido a mediados del siglo pasado. Se consideran unidades primas de información. En pocas palabras, cualquier tipo de observación que cambia constantemente se la puede considerar como dato funcional (Plazola, 2013, p.19).

La filosofía básica es pensar en términos de variables observables. Las secuencias de casos deben tratarse como entidades individuales que consisten en notas. El término significa una función en relación con los sujetos observados y la estructura interna da lugar a su forma explícita (Ramsay y Silverman, 2005, p.38). La definición universal es un poco engañosa.

Como definición más sencilla, se dice que es una de las variables o unidades de importancia transformadas en un conjunto de observaciones, que se visualizan normalmente como una curva. Se consideran como un análisis estadístico de curvas de muestra (Kokoszka y Reimherr, 2017, p.1).

#### 1.1.2. *Dato funcional*

Una variable de función aleatoria  $X$  es una variable aleatoria que tiene valores en un espacio dimensional infinito. Es la representación funcional en un conjunto de observaciones  $\{X_1, \dots, X_n\}$  de  $X$  según Ferraty (2006, p.6) también son conocidos como objetos de curva.

#### 1.1.3. *Variable Funcional*

Una variable aleatoria  $X$  que toma valores en el espacio funcional  $\varepsilon$ , se la conoce como variable funcional. Comúnmente se presentan de manera discreta en un conjunto de observaciones no necesariamente equidistantes (Castillo, et al., 2015, p.6).

#### 1.1.4. *Datos funcionales atípicos*

Como en la estadística descriptiva clásico, se debe detectar individuos atípicos, aunque las técnicas de detección de outliers funcionales es un campo relativamente nuevo. Existen varias formas para que se presenten, como los errores de medición o la mala calibración de los equipos usados. Las observaciones se presentan como aislados, que se muestran de manera poco usual durante un breve momento, y como persistentes, que se comporta de manera inusual por un tiempo más extenso Fuster (2021, p.65). También se consideran aquellas puntuaciones “significativamente” pequeñas en profundidad, que difieren en magnitud o forma del resto de las curvas. La mediana funcional de la muestra tiene el mayor valor de profundidad. Por esta razón, el comportamiento de profundidad es uno de los métodos más eficaces a la hora de la detección. Actualmente existen múltiples instrumentos para visualizar objetos inusuales, como el gráfico de arcoíris, bagplots y boxplots. Son muy útiles ya que facilitan el hallazgo de características que podrían no ser vistos con el uso de estadísticos, con gran velocidad de cálculo y una gran precisión (Millán, 2017, pp. 49-53).

#### 1.1.5. *Estadísticas descriptivas funcionales*

Para la representación de las principales características del conjunto de datos se inicia con un análisis descriptivo y exploratorio, algunos descriptores funcionales (Ramsay & Silverman, 2005, p. 22) se presentan a continuación.

Sea  $x_i(t)$   $i = 1, 2, \dots, N$  un conjunto de curvas funcionales, la media, varianza, covarianza y correlación funcional están proporcionadas por:

➤ **Media funcional**

$$\bar{x}(t) = N^{-1} \sum_i x_i(t),$$

➤ **Varianza funcional:**

$$Var_x(t) = (N - 1)^{-1} \sum_i (x_i(t) - \bar{x}(t))^2$$



➤ **Función de covarianzas**

$$Cov(t_1, t_2) = (N - 1)^{-1} \sum_i [(x_i(t_1) - \bar{x}(t_1))] [x_i(t_2) - \bar{x}(t_2)]$$

➤ **Función de correlación**

$$Corr_x(t_1, t_2) = \frac{Cov_x(t_1, t_2)}{\sqrt{Var_x(t_1)Var_x(t_2)}}$$

➤ **Función de correlación cruzada**

$$Corr_{x,y}(t_1, t_2) = \frac{Cov_{x,y}(t_1, t_2)}{\sqrt{Var_x(t_1)Var_x(t_2)}}$$

### 1.1.6. Bases

En el campo de acción los datos funcionales se recopilan y registran de manera intermitente, en la mayoría de veces es necesario convertir sin procesar en funciones viables, posiblemente utilizando varios tipos de modificaciones. Dado lo anterior, la representación como a través de una base ortogonal ofrece ventajas teóricas y prácticas, se utilizó esta base para transformar los valores discretos en curvas funcionales. Por lo cual se dice, como concepto sencillo de bases, como un grupo de funciones conocidas  $\{\phi_k\}_{k \in \mathbb{N}}$  al punto que cada función logra ser aproximada, tal como demande, a través de una combinación lineal de  $K$  con un  $K$  suficientemente grande. A través de este método la observación funcional llega a ser  $X(t) \approx \sum_{k=1}^K c_k \phi_k(t)$ . A veces, simplemente usando interpolación simple, que consiste en información contigua entre sí por segmentos de línea, pero esta estimación se transforma insuficiente, si se quiere obtener información de derivadas de funciones (Plazola, 2013, p.20).

### 1.1.7. Bases Fourier

Las reglas de Fourier rara vez se usan, son funciones que se usan para estimar datos periódicos y sus funciones base principales son funciones de seno y coseno. La serie de Fourier como función periódica está definida por:

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

$$w = \frac{2\pi}{T}$$

**Dónde:** T define el periodo de oscilación

Por su estructura de adición se usa como una expansión de bases, definido como:

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

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

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

para  $r = 1, \dots, L/2$ , donde  $L$  es un entero par. La frecuencia señala 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 con anterioridad es una base ortonormal. La gran popularidad de esta base se basa parcialmente en la posibilidad de un cálculo veloz del coeficiente por el algoritmo de transformación de Fourier rápida (Fast Fourier Transformation FFT) (Plazola, 2013, p.28).

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

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

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

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

Para  $r = 1, \dots, L/2$

### 1.1.8. Splines

La función Spline se usa muy a menudo cuando desea aproximar una serie acíclica. Estas funciones polinómicas se organizan en un período de tiempo o espacio T. El primer paso para

definir un segmento es dividir el intervalo de tiempo en el que se aproxima la función de interés en  $L$  sub intervalos separados por valores. Estos valores se denominan nudos. En cualquier intervalo, la rebanada es un polinomio de grado  $m$  (Plazola, 2013, p.22).

De manera muy frecuente, una función Spline es explícita por el orden de los segmentos polinomiales y el orden de la secuencia de los nodos  $\tau$  (Ramsay, 2005, p.49).

La base más utilizada para aproximar funciones Spline es la B-Splines desarrollada por Boor. Estas clases de bases están disponibles en una gran variedad de software incluyendo R, otras bases de funciones Splines son las funciones potencias, de base truncada, M-Spline y Spline naturales (Boor, 2001, p.34).

### **1.1.9. B-Splines**

Según Boor (2001, p.6) el método recursivo para construir los B-Splines es el siguiente: al principio se inicia con una secuencia de nodos, es decir una secuencia que no decrece  $t = (t_i)$  de  $N+2$  valores reales llamados “nodos” tal que:

$$t_0 \leq t_1 \leq \dots \leq t_{N+1}$$

Se define el conjunto de todos los nodos adicionales

$$t_{-k} = \dots = t_0 \leq t_1 \leq \dots \leq t_N \leq t_{N+1} = \dots = t_{(N+1)+k},$$

Donde  $k$  veces, donde  $k$  señala el grado de la base B-Spline. El índice  $i$  de los nodos se puede reajustar, desde el ultimo  $t_{-k}$ , es cuando, los  $N + 2(k + 1)$ , los nodos aumentados  $t_i$  son identificados por  $i = 0, \dots, N + 2k + 1$ .

Aunque existan diversos métodos de suavización, este es el método más común, por lo que los diferentes softwares optan por implementarlo, tal como lo hace R, que optimiza la técnica para el gran volumen de datos (Morillo, 2009, p.10).

### **1.1.10. Suavizado**

El primer inconveniente a la hora de aplicar el AFD, es encontrar un modelo adecuado. En el campo de acción la mayoría de información recolectada presentan muchos errores, por lo que se debe adaptar a su forma de curva funcional, por lo que es fundamental el correcto suavizado (Fuster, 2021, p.10).

### 1.1.11. Remuestreo de Bootstrap

Gran Parte de la simulación matemática trabaja con un método de remuestreo denominado Bootstrap, el cual se encarga de generar un masivo conjunto de muestras obtenidas mediante la aleatorización. Esta técnica no requiere ningún tipo de supuesto sobre la forma generadora, por ende, aproxima cualquier distribución en el muestreo de un estadístico (Meneses, et al., 2016, p.6).

### 1.1.12. Análisis funcional de la varianza

La mayoría de las técnicas estadísticas del análisis univariado han sido implementadas al ámbito funcional. La inferencia estadística para determinar diferencias en las curvas promedio o medias funcionales cuando se contrasta la hipótesis nula de igualdad de medias funcionales; esto permite probar, en diseños experimentales funcionales, con  $k$  grupos de curvas, el efecto que tiene un determinado tratamiento sobre una muestra y así inferir sobre una población determinada.

Cada dato funcional se puede representar como con siendo el subíndice que indica el nivel del factor y el número de réplica ( $i, j$ ), la media correspondiente a cada nivel o muestra independiente tiene la forma  $\bar{X}_{i.}$ , mientras la covarianza entre dos instantes en una curva en el caso clásico (homocedástico) es común y para el caso general (heterocedástico) no necesariamente iguales.

El estadístico de prueba asociado para el contraste es análogo al  $F$  clásico para el ANOVA de una vía, ya que en el ANOVA funcional se puede ver que mientras el numerador mide la variabilidad externa entre las muestras, el denominador mide la variabilidad interna dentro de las muestras, la expresión:

$$F_n = \frac{\sum_{i=1}^k \frac{n_i \|\bar{X}_i - \bar{X}_{..}\|^2}{k-1}}{\sum_{i,j} \frac{\|\bar{X}_{ij} - \bar{X}_{i.}\|^2}{n-k}}$$

Donde

$$\bar{X}_{i.}(t) = \sum_{j=1}^{n_i} \frac{X_{ij}(t)}{n_i}$$
$$\bar{X}_{..}(t) = \sum_{j=1}^k \frac{n_i X_{i.}(t)}{n}$$
$$n = \sum_{i=1}^k n_i$$

$$\|x\| = \left( \int_a^b x^2(t) dt \right)^2$$

### **1.1.13. Razones por la cual clasificar**

Un proceso de clasificación es un método conveniente para organizar un gran conjunto de individuos a estudiar, para que pueda ser entendido con mayor facilidad. Si se logran resumir correctamente mediante un pequeño número de grupos de individuos, entonces las denominaciones del grupo proporcionan una óptima descripción concisa de patrones de similitudes y diferencias. En la investigación de mercados, por ejemplo, es beneficioso ser útil agrupar a un gran número de encuestados según sus particularidades por determinados productos. La necesidad de resumir información recolectada es cada vez más importante debido al creciente número de grandes bases disponibles en muchas áreas de la investigación y la ciencia, la exploración de análisis de clusters y otras técnicas de análisis multivariante ahora, muy comúnmente llamado minería. En el siglo XXI, la minería de datos se ha vuelto de particular interés donde el objetivo es extraer información útil o conocimiento para la correcta toma de decisiones (Everitt, et al., 2011, p.3).

### **1.1.14. Análisis Clúster para datos funcionales**

Actualmente el manejo y entendimiento de la información se ha convertido en procesos mucho más complejos, por lo que el desarrollo de técnicas mucho más avanzadas se ha vuelto normal en estos días. Es normal encontrar nuevas técnicas, de manera que una herramienta para los clusters se vuelve fundamental. Además, es una contribución transcendental utilizar la incertidumbre que tiene un individuo y que es representada bajo el enfoque funcional (Venegas, 2017, p.2).

Sin embargo, la elección de una técnica apropiada de agrupamiento es siempre más difícil en la presencia de outliers, por lo que se debe realizar una limpieza de la información, puesto que unas pocas observaciones anómalas deben ser suficientes para crear serios problemas en la búsqueda de clusters. Por ejemplo, grupos claros se juntan involuntariamente y detectan conjuntos poco interesantes formados por pocas observaciones atípicas (Escudero y Gordaliza, 2005, p. 3).

Muchos algoritmos de agrupamiento de uso común se basan en las diferencias por pares entre los objetos para guiar el agrupamiento. Hay muchas maneras diferentes de definición de (des)similitud entre objetos, y la elección de la medida de disimilitud depende en gran medida del tipo de datos con los que se trabaja (discretos, continuos, binario, etc). Por lo general, cuando los elementos se agrupan, la diferencia entre dos elementos se indica mediante algún tipo de distancia. Algunas medidas comunes de la distancia entre dos vectores de datos multivariados incluyen la

distancia euclidiana ordinaria o la distancia de Mahalanobis. Con datos funcionales, una disimilitud más adecuada medida para dos funciones  $y_i(t)$  y  $y_j(t)$  en algún dominio, por ejemplo,  $[0, T]$ , es la distancia  $L_2$  al cuadrado entre las dos curvas:

$$d(i, j) = \int_0^T [y_i(t) - y_j(t)]^2 dt$$

Una clase importante de métodos de agrupamiento son los análisis de conglomerados jerárquicos. Hay dos tipos principales de métodos de agrupamiento jerárquico: aglomerativo y divisivo. Este estudio se centrará en los métodos aglomerativos. El método comienza con cada objeto como su propio grupo. Luego se fusiona sucesivamente la mayoría de los clústeres similares juntos hasta que todo el conjunto de datos se convierte en un grupo. Para determinar que grupos deben fusionarse, se pueden usar varios métodos de vinculación.

#### ***1.1.15. Pasos para un correcto análisis de clusters***

Los pasos que se deben seguir para un correcto análisis clúster son los siguientes. Lo primero que se debe hacer para llevar a cabo será realizar una correcta elección de las variables que van a representar a las diferentes observaciones (curvas). Una de los temas importantes a tener en cuenta sobre la elección de variables es ver si realmente son notables para el tipo de clasificación a la que se quiere llegar. El investigador no tiene ninguna información sobre los posibles grupos a formar. Por lo que es beneficioso saber con anterioridad que tipo de clasificación se desea como resultado final e intentar recolectar la información que se ajuste a este. El segundo paso consiste en llevar a cabo el mejor sistema de asociación entre las observaciones, es decir, una habilidad que permita medir el acercamiento entre los individuos. Habitualmente esta medida de proximidad es considerada por distancias, aunque está bien utilizar medidas de similitud. En el paso a seguir se desarrollarán con más detalle las medidas más utilizadas. Consistirá en indicar el clúster funcional óptimo para el estudio. Debido a la existencia de numerosos y diversos métodos, es importante elegir un método acorde a la representación que se tiene para realizar el estudio. La elección del modelo dependerá del comportamiento de los sujetos y el objetivo final al que se quiere llegar. Por último, habrá que aprobar e interpretar los resultados. Esta etapa es de las más importantes, ya que va a ser donde se expongan las conclusiones y recomendaciones concluyentes del estudio. Existen diferentes métodos para validar los resultados dependiendo de los métodos llegados a cabo (Ibañez, 2018, p.9).

### 1.1.16. *K-means*

Es el método más común que se aplica para llevar a cabo el análisis de clusters. Esta técnica es muy adecuada a la hora de analizar cómo se agrupan de forma reiterativa. Consiste básicamente en asignar cada uno de los casos a su clúster correspondiente, de los  $K$  que ya se han fijado con anterioridad disminuyendo la distancia con el centroide asignado a cada clúster. Tiene una ventaja significativa, y es que los centroides de cada clúster se van calculando cada vez tras una asignación y no al final del proceso (Ibañez, 2018, p.14).

### 1.1.17. *Algoritmo k-means para análisis de datos funcionales*

Si  $X = \{X_1, \dots, X_n\}$  denota una muestra aleatoria de funciones aleatoria, cada una distribuida como la función genérica  $X$  y apoyada en el intervalo. El objetivo de este algoritmo de agrupamiento consiste en agrupar el conjunto  $X$  en dos partes separadas  $X_1$  y  $X_2$ . Para  $k = 1, 2$ , se deja que denote la  $i$ -ésima curva que se clasifica en el  $k$ -ésimo grupo,  $X_k = \{X_{k1}, \dots, X_{knk}\}$ . Donde  $n_1 + n_2 = n$ . En el contexto funcional, este algoritmo se aplica de varias maneras. A primera posibilidad es reemplazar la norma de vectores por la norma de funciones o una de sus derivadas (Delaigle & Tung, 2016, p.87).

Como paso inicial se elige aleatoriamente  $k$  representantes de conglomerados  $\varphi_1(t), \dots, \varphi_k(t)$  de las curvas  $x_1(t), \dots, x_n(t)$ . Para todas las curvas, se asignó y alineo las curvas al grupo donde la similitud entre la curva alineada y el representante del grupo es el más alto. Así, se asigna la curva  $x_i(t)$  y alineado con el grupo  $j$ -ésimo sí.

$$\max_{i=1, \dots, k} \rho(x_i(h_{il}), \varphi_i) = \rho(x_i(h_{ij}), \varphi_j),$$

Donde,  $(h_{ij}(t))$ , es la función de deformación óptima que maximiza la medida de similitud  $\rho(x_i(h_{ij}), \varphi_j)$ . Para evitar que los grupos se separen, sea  $j = 1, \dots, k$ , sea  $N_j$ , el número de curvas, y  $\Gamma_j$ , el conjunto de todas las curvas  $i = 1, \dots, n$  perteneciente al grupo  $j$ . Para  $j = 1, \dots, k$ , todas las curvas  $N_j$  que fueron asignadas y alineadas a el grupo  $j$ , en el paso anterior ahora debe formarse a lo largo de la función de deformación,  $h_j^{-1}(t)$ , donde:

$$h_j^{-1}(t) = \frac{1}{N_j} \sum_{i \in \Gamma_j} h_{ij}(t)$$

Para todas las curvas, se calcula la medida de similitud entre las curvas alineadas y su grupo representante utilizando  $\rho(x_i(h_{ij}), \varphi_j)$ . Si los incrementos de las medidas de similitud son todo por debajo de un axioma fijo y luego se detiene. Se calcula los nuevos clusters representativos usando las curvas asignadas y alineadas por la media de las curvas.

#### **1.1.18. Software estadístico R**

El software estadístico R es una herramienta computacional de software libre que se utiliza para el análisis de datos a través de varias librerías y paquetes, que aporta cada miembro de la comunidad sin ningún tipo de costo. La gran variedad de técnicas estadísticas hace de R, un instrumento indispensable para el manejo de información en la actualidad, debido a la inmensa información de procesamiento.

#### **1.1.19. Radiación**

Según Salcedo (2013, p. 2) la radiación es la emisión y difusión de energía en forma de partículas y ondas a través del espacio. En simples palabras, se les conocen como 'paquetes de energía' que circula en caminos rectos. Este fenómeno electromagnético tiene una característica particular, que son las variaciones de energía eléctrica y magnética atravesando campos en el espacio y el tiempo. Un tipo muy especial es la denominada 'rayos cósmicos', que son partículas energéticas, conocidas por penetrar el espacio interestelar (Diehl, 2000, p.2).

Los distintos efectos secundarios a una exposición severa, son las diversas enfermedades a órganos vitales y estructura ósea, alterando su integridad molecular. Se ocasionan a largo o a corto plazo (Garzón, et al., 2008, p.5).

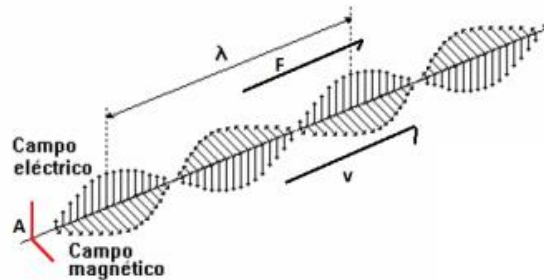
#### **1.1.20. Radiación gamma**

Considerada como la muestra de energía a través de ondas electromagnéticas de muy poco alcance a un nivel de frecuencia mayor con relación a la luz UV. Tiene una naturaleza similar a la luz perceptible, con la única diferencia de que presenta una longitud de onda muy pequeña y, por ende, un nivel de destello más alto que la luz presenta un comportamiento muy parecido a los rayos X ya que para ambos le es sencillo atravesar a la materia viva. También conocidos como "fotones", logran viajar a la velocidad de la luz. Los fenómenos gamma tienen una gran capacidad para ionizar la materia y, por lo tanto, perjudican a las células vivas. El daño producido en una



célula es proporcional al número de ionizantes que produzca el material que lo esté recibiendo (Stalter, 2012, p.15).

Nurilmala, et al. (2016, p.33) indica que una utilización es incitar mutación proveniente de la genética que da como resultado el mutante requerido.



**Ilustración 1-1:** Campo eléctrico y campo magnético

Fuente: (Rojas Katherine, 2009)

#### ***1.1.21. Efectos de la radiación en los ecosistemas***

Desde principios de la década de 1960 se han realizado numerosos estudios sobre el efecto de las ondas gamma en los ecosistemas. Estos estudios fueron impulsados por la carrera armamentista entre la Unión Soviética y los Estados Unidos. A mediados del siglo pasado se ha popularizado las aplicaciones y su efecto en los seres humanos y ecosistemas. Una gran razón de este interés, fue la guerra fría y la necesidad de armas y nuevas aplicaciones para el beneficio de cada potencia mundial. A través de diversos experimentos realizados en zonas naturales, se vio el gran daño que ocasiona una exposición en distintos tipos de vegetación, y como es capaz de afectar las células vivas (Stalter, 2012, p.4).

#### ***1.1.22. Radiación ionizante***

El fenómeno ionizante expulsa su energía por medio de ondas electromagnéticas con un mayor alcance de frecuencia y una onda de longitud corta comparada con la no ionizante, siendo capaz de romper enlaces covalentes en las distintas moléculas. La radiación gamma presenta un nivel más alto de energía, debido a lo poco extensas que son sus ondas por lo que le es más fácil la penetración en distintos elementos. Las radiaciones ionizantes se utilizan en gran parte del campo de la medicina, principalmente en la medicina nuclear y laboratorios (Rojas, 2009, p.3).

### **1.1.23. Espectrometría gamma**

Considerada una de las técnicas más eficaces para el estudio de la propagación de radionúclidos en el ecosistema. Su principal objetivo es la determinación cuantitativa de los radioelementos emisores gamma presentes en los sujetos antes de estudio (Barrera, et al., 2008, p.22).

Entre las metodologías analíticas nucleares, es de las más importantes, por realizar un análisis simultáneo de muchos elementos y no secuencial como la absorción atómica (Rodríguez, et al., 2004, p.7).

### **1.1.24. Espectrómetro**

Costa de tres partes: cristal de centelleo, que incide en el cristal que van a interaccionar con el mismo, creando un fotón de un específico tamaño de onda, dependiente del tipo de cristal usado. El analizador de pulsos, un pulso de tensión que es amplificado por esta etapa y detecta los pulsos no compatibles. Y un sistema de registro, que traslada la información a un procesador computacional, entregando una base de datos con la cual se inicie el análisis (Delahay, 1970, p.17).

### **1.1.25. Dosis absorbida**

Una vez que la radiación penetra un material, una cantidad de energía se invierte por cada electrón. Se considera a la energía con la que fue penetrado el material y a la vez lo que este haya absorbido. Se mide con Julio/Kilogramo (Carrasco, 2003, p.12).

### **1.1.26. Dosis equivalente**

Es el producto de la dosis absorbida por un factor de calidad mayor. La unidad de medida es el Sievert (uSv/h), que es igual a 100 rems en el sistema cegesimal. Es una unidad muy grande, por eso en protección radiología se utilizan sus submúltiplos, el milisievert y el microsievert. También se mide en Julio/Kilogramos (Carrasco, 2003, p.13).

### **1.1.27. Dosis efectiva**

Se produce cuando la irradiación afecta más a un órgano q a otro debido a los efectos radioinducidos. Se denota con la letra H. Para su cálculo se multiplica la dosis equivalente por un factor de ponderación de tejido. También se mide en uSv/h (Carrasco, 2003, p.14).

### ***1.1.28. Límite de dosis***

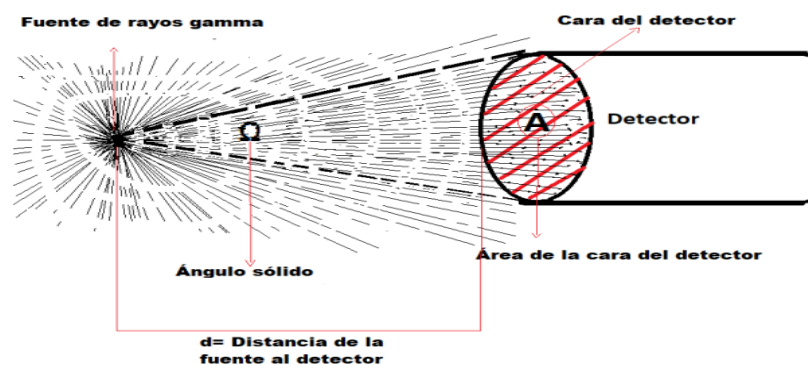
Los límites de dosis son medidas calculadas por profesionales de la radiología a lo largo de los años, que recomiendan que nunca deben ser sobrepasados. Están referidos al periodo de un año oficial, y están contempladas tanto las personas expuestas constantemente y los ciudadanos comunes (Carrasco, 2003, p.15).

### ***1.1.29. Síntomas inmediatos en la salud***

La radiación ionizante tiene un gran alcance para afectar a las células vivas. Aunque nuestro cuerpo tenga un gran sistema de defensa, posee vulnerabilidades y presentar diferentes enfermedades como cáncer. Una dosis gigantesca de ondas radioactivas produce enfermedades mortales y daños al instante. Una dosis pequeña no produce efectos en el ser humano de una manera inmediata, pero las complicaciones a largo plazo se presentan de manera toxica al organismo (Carrasco, 2003, p.22).

### ***1.1.30. Detección de la radiación gamma***

En la organización de la capa de electrones del átomo, el núcleo presenta características por los niveles muy discretos de energía. La transformación entre estos niveles se realiza mediante la emisión o impregnación de radiación electromagnética de la correcta energía. Las interacciones que esparcen estos fotones, que tienen un rango de los pocos cientos de keV hasta una cantidad similar de MeV, determinan la potencia tan alta de los núcleos. A los fotones de alta energía se han designado por lo largo de la historia, rayos gamma. La mayoría de principios se ubican en estados excitados como consecuencia de la desintegración beta, sin embargo, los estados nucleares excitados se originan en reacciones nucleares. La detección es una serie de procesos que trabajan de manera ambigua, los fotones no poseen carga intrínseca, por lo tanto, la ionización no es posible, lo que hace que su paso sea directo, la correcta manera de medición es cuando este, experimenta algún tipo de convivencia, donde deja toda su constitución en el electrón perteneciente al material de interacción. Para un detector los fotones son incorpóreos, por lo que, realiza un proceso basado en dos principales características, ser un medio donde la oportunidad de incidencia de un destello sea de altas probabilidades, y la segunda, localizar con gran rapidez la interacción de los electrones (Raya, 2021, p.4).



**Ilustración 2-1:** Muestra la emisión de rayos gamma a la cara del detector

Fuente: (Raya, 2021)

### **1.1.31. Interacciones de rayos gamma con la materia**

Existen tres posibles maneras para que un rayo gamma y la materia realicen una interacción. La dispersión Compton, la producción de pares, y la absorción fotoeléctrica. Para que se produzca una asignación a cualquiera de los tres métodos depende netamente de la energía que posea el electrón y el número atómico del material penetrado. Este tipo de radiación es mucho más perspicaz en la materia que en las partículas cargadas (Raya, 2021, p.5).

### **1.1.32. Absorción fotoeléctrica**

La absorción fotoeléctrica, también conocida como efecto fotoeléctrico, es el proceso que se lleva a cabo cuando un electrón atómico absorbe un fotón, lo que conlleva a la expulsión del electrón del átomo. Cuando los rayos gamma consiguen suficiente energía, el fotoelectrón es el electrón K de un átomo superiormente ligado al núcleo. Este efecto prevalece para destellos de energía comparativamente pequeños, y es más acertado en materia absorbente con un número atómico elevado “Z” (Raya, 2021, p.6).

### **1.1.33. Dispersión Compton**

Descubierto en el año de 1923 por A.H.Compton, cuando observó radiación aleatoriamente dispersa con mucha menor frecuencia que la radiación incidente. Por lo que propuso un nuevo mecanismo de interacción de electrones con la radiación electromagnética, considerando como un choque elástico de partículas (Morales, 1973, p.4).

#### ***1.1.34. Producción de pares***

Un fotón que recorre junto a un núcleo, y desaparece naturalmente, y luego vuelve a regresar como un positrón y un electrón. Este proceso de energía en masa debe tener lugar contiguo a la partícula, tal como un núcleo, para que el impulso se mantenga (Díaz, 2017, p.24).

#### ***1.1.35. Efectos biológicos de las radiaciones ionizantes***

El efecto dañino, del alcance de estas ondas, en la vida humana ha sido estudiado a lo largo de los años, especialmente en el siglo 20, cuando comenzó la revolución energética. Dependiendo de muchas medidas complicadas, cuando la radiación ionizante traslada energía a un ser vivo, provocará varios tipos de resultados finales. El daño ocasionado depende totalmente de la dosis absorbida. Para el ser humano, hay dos tipos de efectos afines con la dosis: somáticos y genéticos. el daño genético se refiere al daño que afectará a las post generaciones y el daño somático se refiere al daño sucedido en los tejidos del individuo irradiado (Nuñez, 2008, p.1).

#### ***1.1.36. Relación dosis-efecto***

El efecto somático se revelará en el individuo que absorbe este tipo de ondas electromagnéticas. Por lo general se divide en efectos con clara seguridad de que va a ocurrir conocidos como efectos determinísticos, y los que ocurren completamente al azar o destino llamados efectos no determinísticos.

#### ***1.1.37. Efectos determinísticos***

Los efectos determinísticos implican dosis elevadas en el cuerpo humano. Los efectos determinísticos se dividen en dos: efectos tempranos y efectos tardíos. En el primer caso, se presentan por lo general en el primer año de la exposición, están afines con el número de células muertas, y el resarcimiento del daño ocasionado. Algunos síntomas que se manifiestan son la caída del pelo, y varios tipos de neumonitis. Los efectos tardíos son los que se manifiestan después del año uno, y tienen una relación con los daños ocasionados durante el primer año. Se presentan por el deterioro de los órganos, y el avance de la enfermedad inicial. Produce cataratas y fibrosis pulmonar, entre otros (Nuñez, 2008, p.2).

### ***1.1.38. Los efectos no determinísticos***

Suelen ocurrir a niveles reducidos de exposición, en cuyo caso el daño será estocástico en naturaleza: es posible pronosticar la proporción de una población dada de individuos expuestos que será afectada, pero improbable predecir el sujeto exacto que perecerá. No existe una dosis origen justificable y el daño se exhibe como un ligero aumento en el suceso directo y se expresa luego de un largo ciclo de latencia (Nuñez, 2008, p.2).

### ***1.1.39. Efecto de la radiación a nivel celular***

La radio sensibilidad del tejido depende de varios elementos. La refutación del tejido a la radiación se basa en el número de células particulares en el tejido, el número de células mitóticas, el periodo de tiempo en que las células persisten activas en difusión. Ha sido demostrado que las células en proceso de diferencia son destruidas por estas ondas con una gran facilidad. Cuanto más tiempo las células persisten en propagación activa, más grande es la sensibilidad a la desintegración nuclear (Nuñez, 2008, p.2).

## CAPÍTULO II

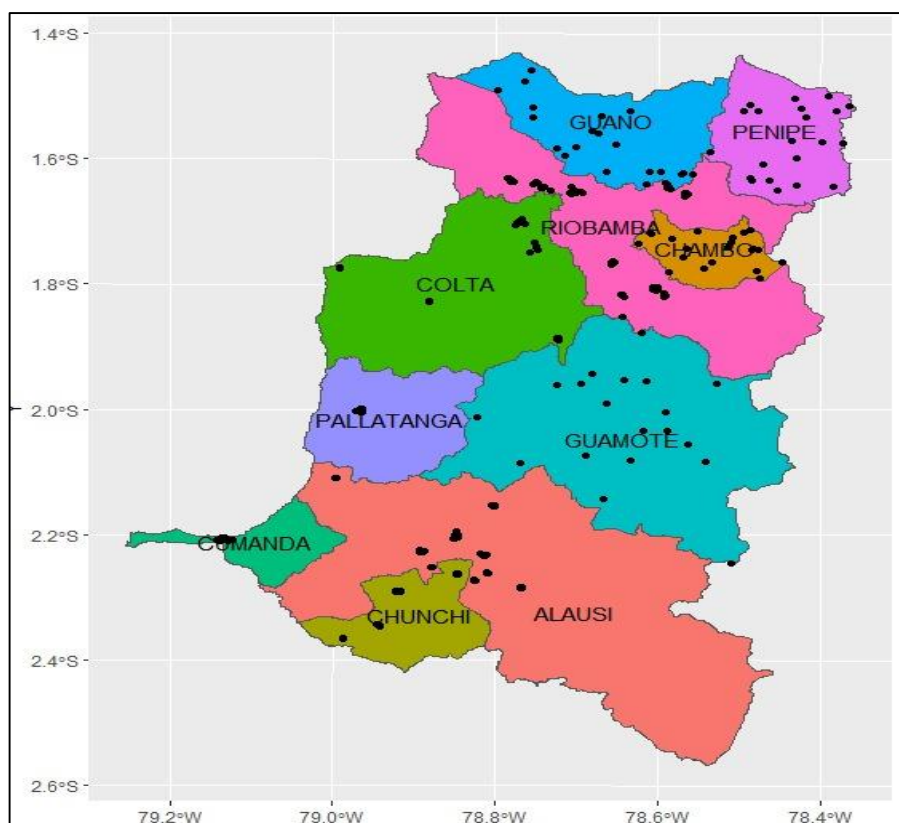
### 2. MARCO METODOLÓGICO

#### 2.1. Tipo de investigación

Según el tipo de investigación es cuantitativa debido a que se estudia las variable tasa de dosis de radiación gamma y cómo se comporta en la provincia de Chimborazo, el objetivo de estudio es aplicada ya que se centra en la caracterización de un fenómeno en el ecosistema; los valores que se van a manipular son no experimentales debido a que la matriz de información procede de una fuente secundaria; dado el tipo de inferencia es inductiva ya que se utilizó un análisis de curvas funcionales y según el periodo temporal es transversal dado que la medición fue por un intervalo de tiempo determinado.

#### 2.2. Localización de estudio

La provincia de Chimborazo se encuentra ubicado en el centro del Ecuador, en la región interandina. Con una extensión de  $6500.66 \text{ km}^2$ . Lugar de grandes montañas y volcanes, como el volcán Chimborazo, considerado el nevado más alto del Ecuador y del mundo. Con una población total de 501.584 habitantes según una proyección realizada por el Instituto Ecuatoriano de Censos (INEC, 2010). Su principal fuente de recursos es la ganadería y la agricultura, ya que su ubicación permite que el ecosistema sea abundante en productos como papas, cebollas, etc. Su división consta de 10 cantones, constituidos por 61 parroquias, 45 rurales y 16 urbanas. Se recolectó la información en cada una de las parroquias, para su debido análisis.



**Ilustración 1-2:** Mapa de la provincia de Chimborazo con sus puntos de muestreo.

**Realizado por:** Yungaicela, Víctor, 2022.

### 2.3. Muestra

El estudio se realizó con los datos de tasa de dosis de radiación gamma, en distintos puntos en la provincia de Chimborazo durante el año 2021, gracias a la labor de recolección de información del Grupo de Investigación-Desarrollo para el ambiente y Cambio Climático (GIDAC). Se clasificó a la provincia en cantones, a lo cual se subdividió en parroquias, y en cada una respectivamente, se señaló un numeroso grupo de puntos posibles de muestreo. Dado que se rebasó la capacidad y los recursos para llegar a toda la población o universo de puntos por muestrear, se decidió optar por un muestreo aleatorio simple.

A cada punto de muestra seleccionado, se consideró un total de 500 observaciones, cada una represento un dato obtenido por segundo, con el fin de sintetizar el análisis y el manejo de las curvas funcionales.



## **2.4. Identificación de variables**

En el presente trabajo de investigación, la variable de interés para el respectivo análisis de datos funcional es la tasa de dosis de radiación gamma. La variable Altitud del punto de monitoreo, se consideró en la realización de los mapas clúster

## **2.5. Modelo estadístico**

Para el análisis de datos funcionales se utilizó el software de licencia gratuita R (versión 64.3.4.2) que es un lenguaje de alto nivel y un entorno para el análisis de datos y gráficos. Para el desarrollo del código se utilizó Rstudio (versión 1.2.5033). Se aplicó el análisis funcional con las librerías disponibles, además se implementó detección de outliers y comparación mediante: Estadísticas descriptivas y técnicas de clúster, según la naturaleza de los datos.

## **2.6. Método para la obtención de los datos funcionales**

Los datos de la tasa de dosis de radiación gamma se registraron en un ordenador personal, se trasladaron a una hoja de cálculo para la facilidad de la manipulación de datos con el fin de prepararlos para el análisis estadístico funcional.

Para la obtención de los datos funcionales se utilizan las librerías `readr`, `fda`, y `fda.usc` del software libre R, con la herramienta Rstudio; el procedimiento para la obtención de datos se realizó de la siguiente manera:

- a) Los datos exportados en formato csv, se leen en Rstudio con la función `read.csv2` y se transforman en una matriz de datos con la función `data.frame`.
- b) Se define el vector de los instantes de tiempo con la función `seq`.
- c) Posteriormente se realiza un suavizado con funciones B-splines, para convertir los datos puntuales en curvas, usando la función `create.bsplines.basis`. También se comprobó la calidad de ajuste.
- d) Finalmente se obtienen los datos funcionales utilizando la función `Data2fd`.

## **2.7. Método de realización del análisis estadístico**

### **2.7.1. Detección de outliers**

Para encontrar posibles curvas atípicas (outliers) se utilizaron distintos métodos gráficos (Functional Bagplot, Raimbow y HDR Boxplot), para crear una tabla comparativa. Se aplicó también una función de R no grafica titulada outliers.depth.trim que ha sido usada por múltiples autores. En el plot arco iris y el funcional bagplot se representa en color negro la curva de la mediana, y en distintos colores los atípicos. Se determinaron las curvas atípicas a través de técnicas graficas de profundidad de datos funcionales las cuales fueron posteriormente eliminadas de la muestra y de los análisis subsecuentes. A través de la función fboxplot, se seleccionó la técnica correcta a su utilización, entre depth, functional y hdr.

### **2.7.2. Análisis descriptivo y exploratorio funcional**

A través de la función boxplot.fd , se calculó la media y la desviación típica funcional de los cantones de la provincia, y se analizó su comportamiento y diferencias entre ellos. Se utilizo para para observar tendencias y variabilidad entre sus datos.

### **2.7.3. Análisis de varianza funcional**

Para determinar si los valores de tasa de dosis de radiación gamma son estadísticamente iguales o si son diferentes, se realiza un ANOVA funcional, a través de la función anova.fd, con el objetivo de identificar si hay diferencias significativas entre los cantones.

### **2.7.4. Variabilidad funcional**

Para calcular la variabilidad de la tasa de dosis de radiación gamma se realiza un gráfico de curvas de contornos y de niveles mediante los comandos contour y filled.contour para que resalte tipos con mayor énfasis en tasas de dosis bajas y altas.

### **2.7.5. Remuestreo de bootstrap**

Para identificar la banda de confianza de la media, se usó remuestreo de bootstrap para cada cantón de la provincia de Chimborazo, se utilizó el paquete fda.usc del software R. A través del

comando `fdatabootstrap`, se obtuvo la banda de confianza de Bootstrap, un estimador de la media funcional y el porcentaje total de absorción.

#### ***2.7.6. Análisis de clúster funcional***

Con el fin de profundizar en el estudio de la variabilidad de los datos, se realizó un análisis clúster en la muestra total de la provincia de Chimborazo. Usando el paquete `funfem()` del software R, se obtuvo un mapa de dispersión que indican la relación de cada uno de las curvas muestrales. También se obtuvo el número óptimo de clúster para su óptimo desempeño, a cada curva funcional se le asignó un número, que representa al cantón de procedencia para saber su localización. Se observó gráficamente la relación que presentan los puntos de muestreo en la provincia de Chimborazo. La categorización de la variable altura se basó en la altura al nivel del mar. Esta variable da una mejor vista al comportamiento de la correlación entre la tasa de dosis y la altura en donde se tomó su muestreo.

## CAPÍTULO III

### 3. RESULTADOS Y DISCUSIÓN

#### 3.1. Construcción de la base de datos

Con la información proporcionada, primero se ordenó las bases de datos, ya que, como toda tecnología y equipamiento de recolección, presenta ciertos inconvenientes, que el investigador es encargado de resolverlas. La base consta de 7 variables cualitativas informativas que son cantón, parroquia, coordenada X, coordenada Y, fecha, punto de muestreo, y la numeración de los segundos tomados para el análisis, dando como resultado la tasa de dosis de radiación de 500 segundos.

**Tabla 1-3:** Construcción de la base de datos de la tasa de radiación gamma en Chimborazo

N	Cantón	Parroquia	X	Y	Fecha	Punto de muestreo	Hora de Inicio	1	2	3	...	500
1	ALAUÍS	Achupallas	748143	9747552	12/7/2021	5	8:40:40	0,11	0,01	0,10	...	0,07
...	...	...	...	...	...	...	...	...	...	...	...	...
87	CHAMBO	Cabecera cantonal	7259441	9572737	17/7/2021	13	10:27:12	0,04	0,06	0,05	...	0,05
...	...	...	...	...	...	...	...	...	...	...	...	...
100	CHUNCHI	Chunchi	731412	9746848	6/7/2021	54	8:56:53	0,00	0,04	0,03	...	0,04
...	...	...	...	...	...	...	...	...	...	...	...	...
135	COLTA	Cañi	723314	9803773	4/8/2021	3	10:25:12	0,03	0,07	0,03	...	0,06
...	...	...	...	...	...	...	...	...	...	...	...	...
172	CUMANDÁ	Cabecera cantonal	708539	9755730	5/7/2021	218	7:59:36	0,01	0,02	0,03	...	0,04
...	...	...	...	...	...	...	...	...	...	...	...	...
186	GUAMOTE	Cebadas	710777	9377189	8/8/2021	23	10:34:24	0,03	0,06	0,05	...	0,05
...	...	...	...	...	...	...	...	...	...	...	...	...
195	GUANO	San José de Chazo	60789	63105	15/8/2021	10	11:47:19	0,07	0,06	0,08	...	0,07
...	...	...	...	...	...	...	...	...	...	...	...	...
245	PALLATANGA	Pallatanga	722722	9766645	27/6/2021	11	7:16:59	0,05	0,04	0,05	...	0,06
...	...	...	...	...	...	...	...	...	...	...	...	...
257	PENIPE	Matus	7259432	9574456	14/8/2021	1	10:17:58	0,05	0,05	0,05	...	0,04
...	...	...	...	...	...	...	...	...	...	...	...	...
301	RIOBAMBA	Calpi	752334	9817489	3/8/2021	72	8:48:42	0,04	0,07	0,08	...	0,07

Realizado por: Yungaicela, Víctor, 2022.

La tabla 1-3 muestra un resumen general de los cantones y su totalidad de muestras tomadas en cada uno de ellos.

**Tabla 2-3:** Cantones y puntos de muestra

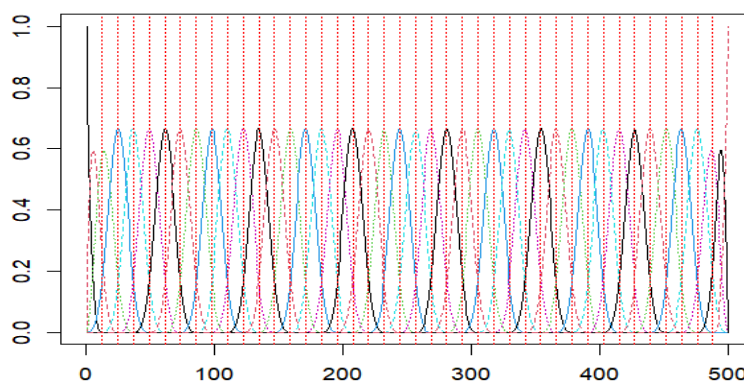
N°	Cantones	Puntos
1	Alausí	84
2	Chambo	14
3	Chunchi	35
4	Colta	37
5	Cumandá	14
6	Guamote	4
7	Guano	55
8	Pallatanga	12
9	Penipe	44
10	Riobamba	101

Realizado por: Yungaicela, Víctor, 2022.

No se tomó el resto de información ya que la toma de datos no fue uniforme para todos los puntos de muestreo. Es importante identificar este tipo de situaciones a la hora de tratar la base de datos, ya que esto aumenta la probabilidad de distorsión en los resultados finales.

### 3.2. Selección de base y número de funciones

El primer paso en el análisis funcional es hallar el método de base funcional que mejor se ajuste a las observaciones, y eliminar el ruido producido por la calibración de los equipos. Gracias al paquete *fda.usc*, con la función *optim.basis*, del software R se pudo optimizar el número de bases necesarias para el suavizado, el cual dio como resultado un total de 43 bases funcionales. Para la información de tasa de dosis de radiación gamma se utilizó las bases de funciones B-splines, debido a que es uno de los métodos más comunes.

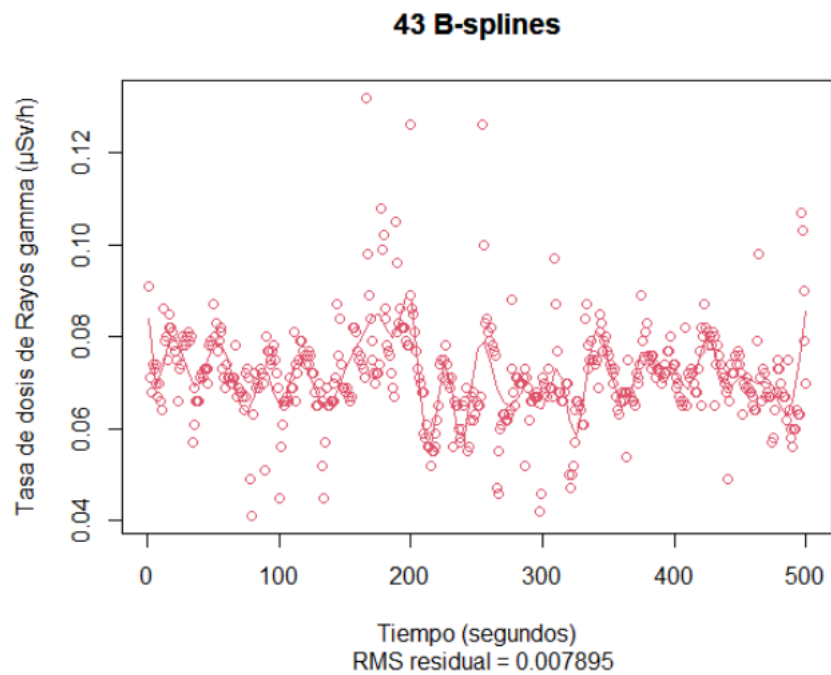


**Ilustración 1-3:** Número de Bases B-Spline óptimos

Realizado por: Yungaicela, Víctor, 2022.

### 3.3. Calidad de ajuste

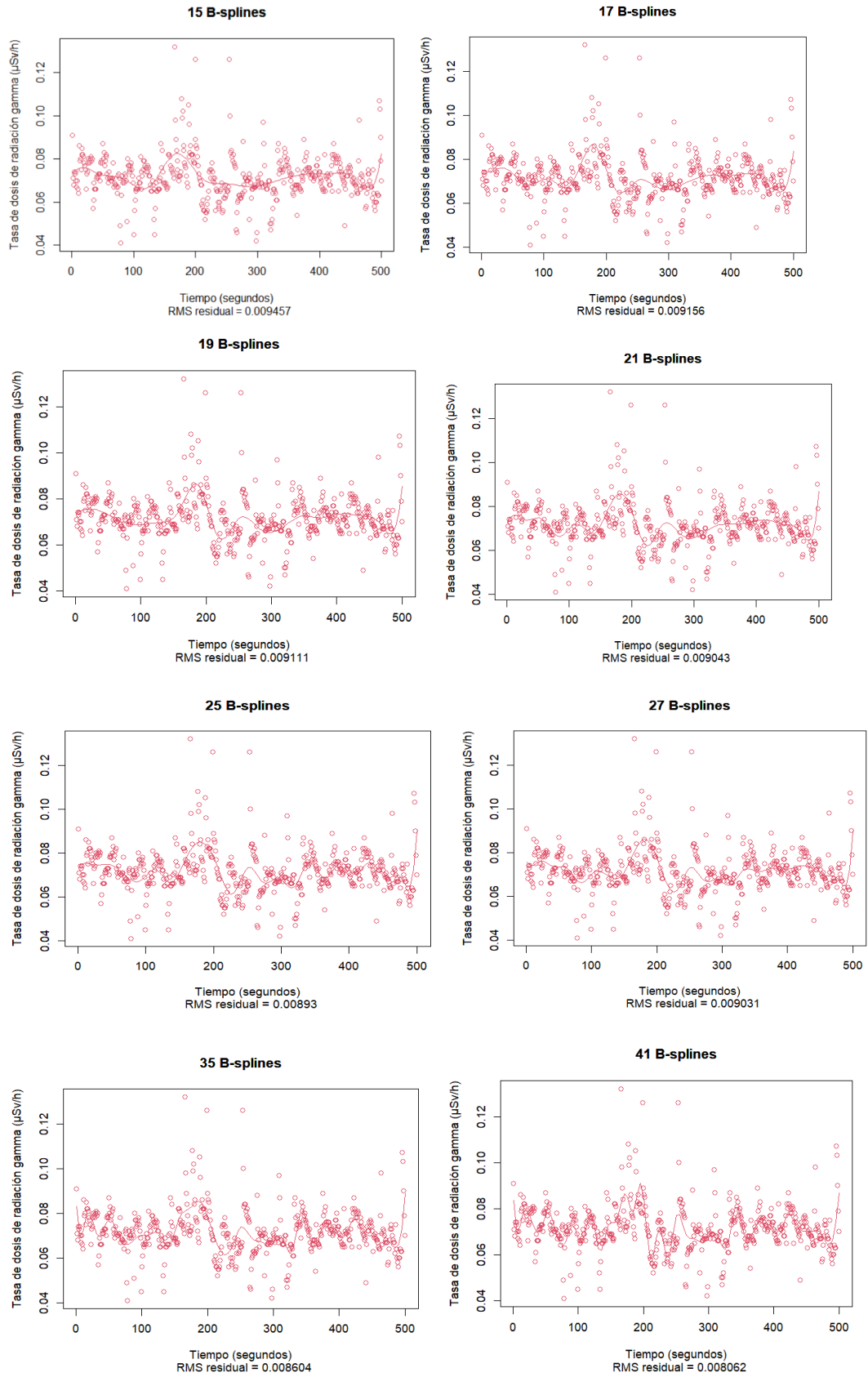
Al contar con los datos funcionales, a través de un proceso por B-Spline, a través de la función `plotfit.fd`, se obtuvo el RMS (root mean square), que es una medida de las diferencias entre los valores predichos por un modelo y las observaciones. Esta medida tiene como característica que siempre va a presentar un valor positivo y su ajuste perfecto es 0.



**Ilustración 2-3:** Calidad de ajuste por RMS

**Realizado por:** Yungaicela, Víctor, 2022.

En el gráfico 2.3 se concluye que el RMS con un ajuste B-Spline de 43 bases es el óptimo con un  $\text{RMS}=0,007$ . A partir de los satisfactorios resultados, se nota una clara presencia de valores atípicos, por lo que se procedió a localizarlos y separarlos.



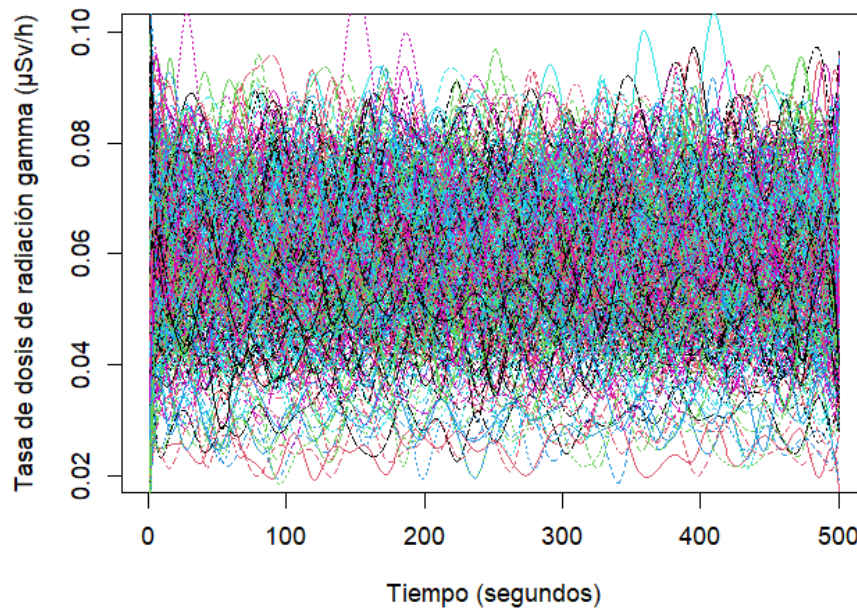
**Ilustración 3-3:** RMS con varios Bsplines  
 Realizado por: Yungaicela, Víctor, 2022.

En la tabla 3-3, se observa, con la ayuda de la función *plotfit. f()*, algunos ejemplos de RMS, con b-spline 15, 17, 19, 21, 25, 27, 35 y 41, el mejor RMS sigue siendo el número de bases optimo igual a 43.

**Tabla 3-3:** Comparativa de B-splines

B-splines	RMS
15	0.009457
17	0.009156
19	0.009111
21	0.009043
25	0.00893
27	0.009031
35	0.008604
41	0.008062
43	0.007794

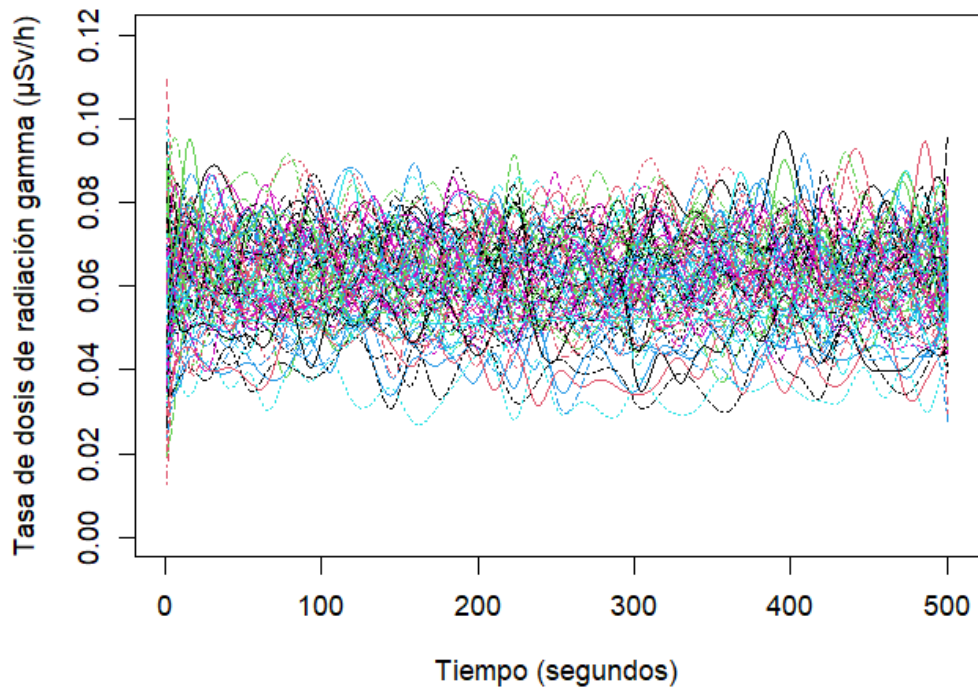
Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 4-3:** Curvas suavizadas de la tasa de dosis de rayos gamma en Chimborazo

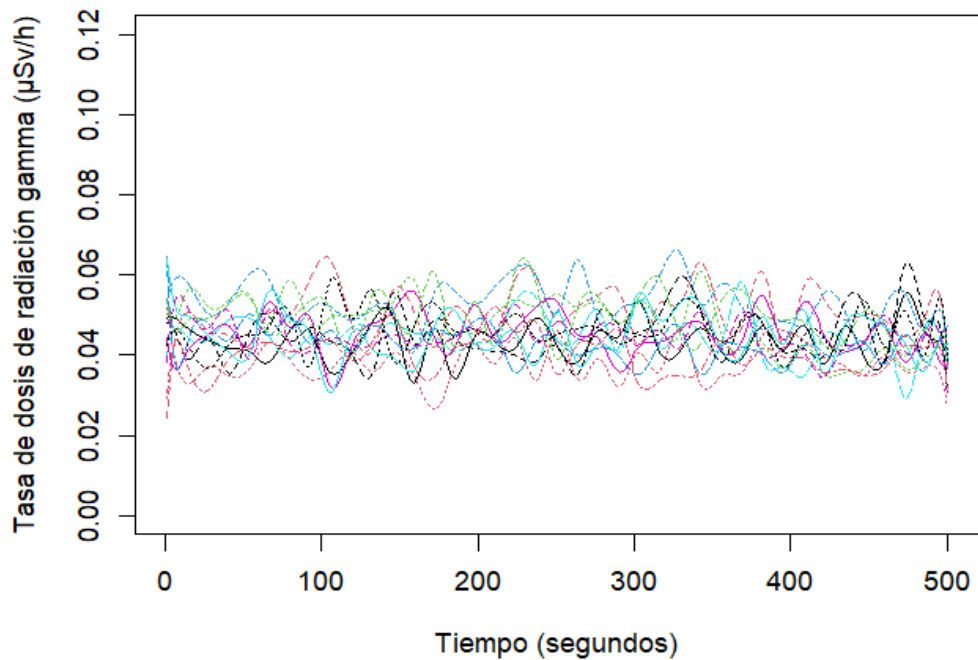
Realizado por: Yungaicela, Víctor, 2022.





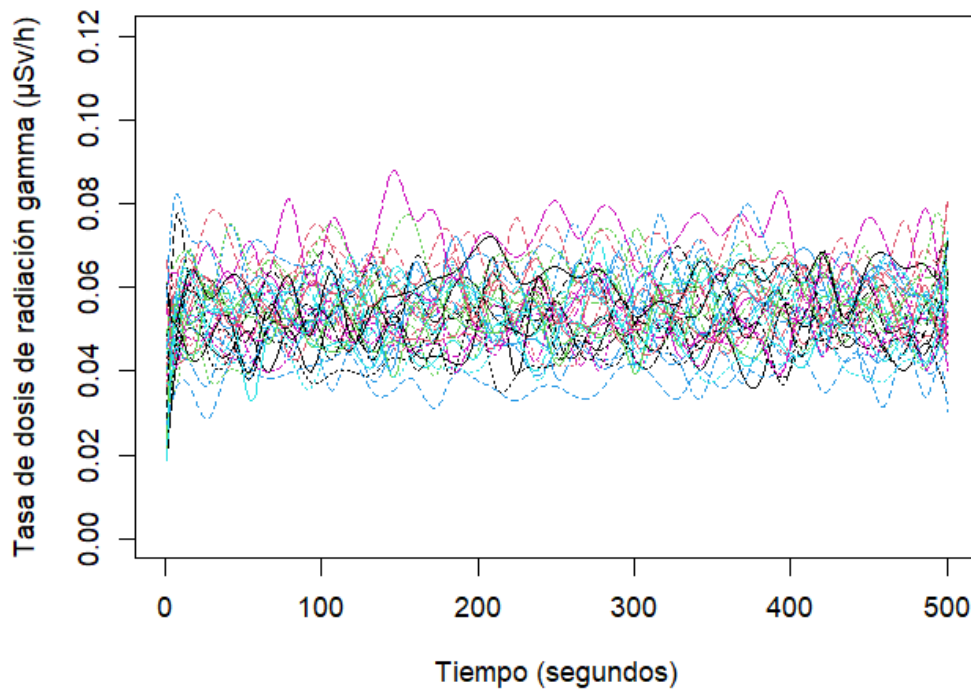
**Ilustración 5-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Alausí

Realizado por: Yungaicela, Víctor, 2022.



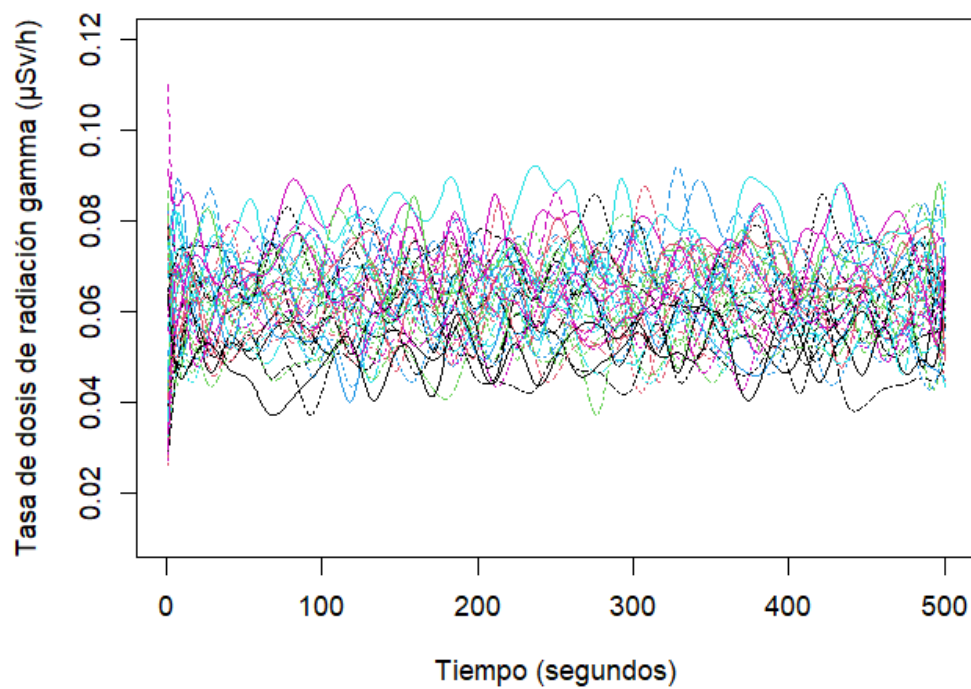
**Ilustración 6-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Chambo

Realizado por: Yungaicela, Víctor, 2022.



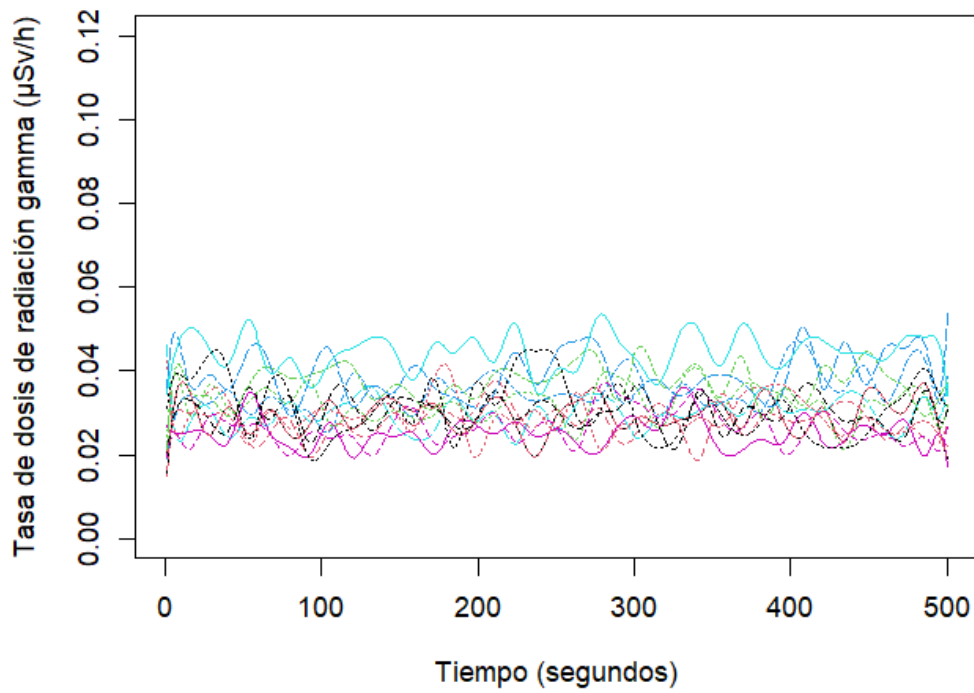
**Ilustración 7-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Chunchi

Realizado por: Yungaicela, Víctor, 2022.



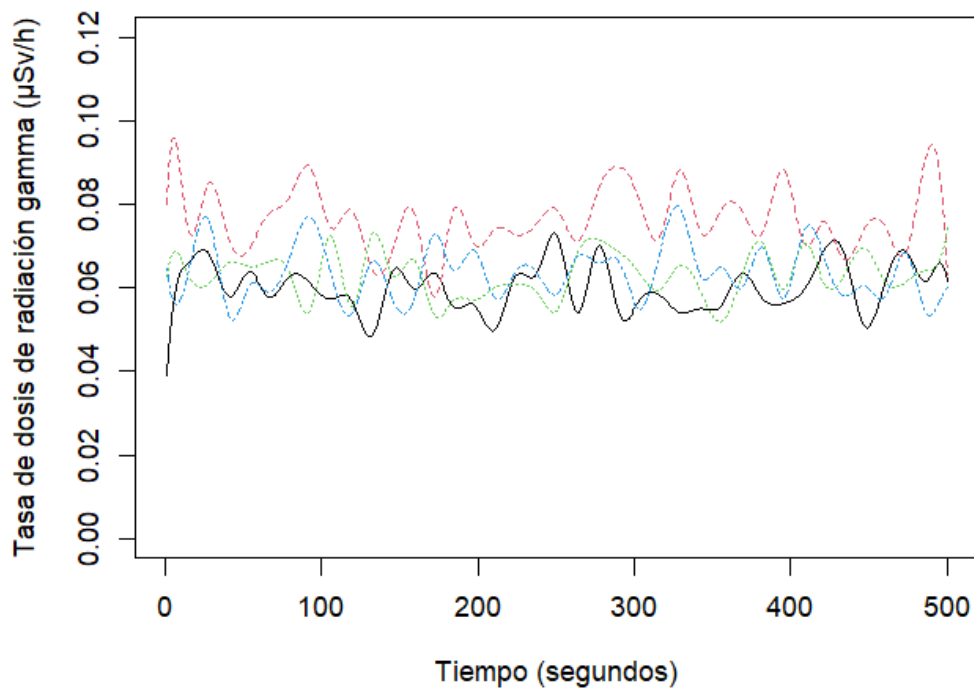
**Ilustración 8-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Colta

Realizado por: Yungaicela, Víctor, 2022.



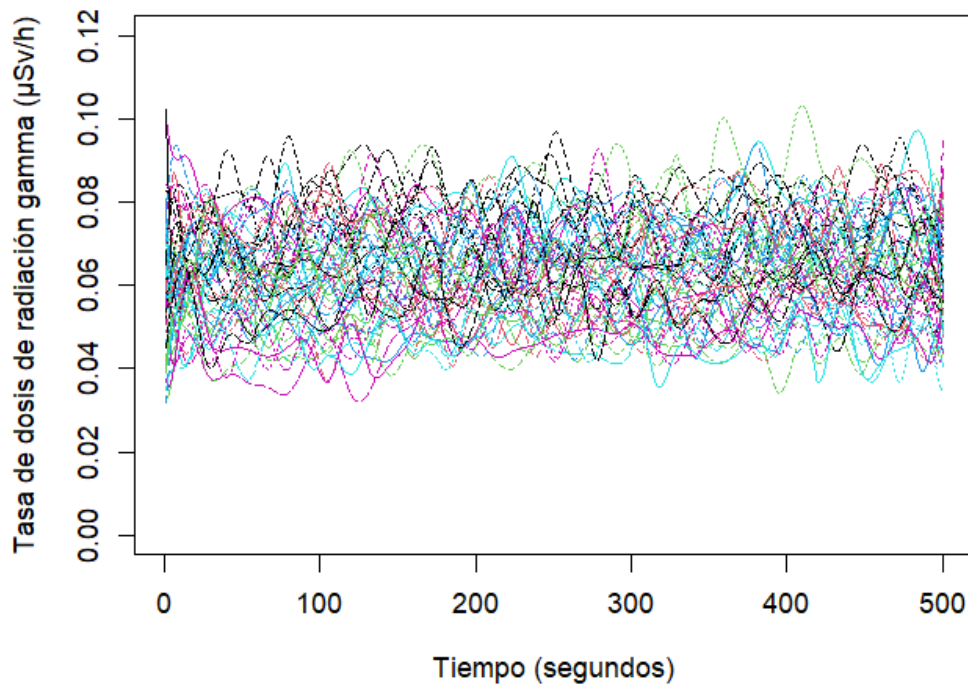
**Ilustración 9-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Cumandá

Realizado por: Yungaicela, Víctor, 2022.



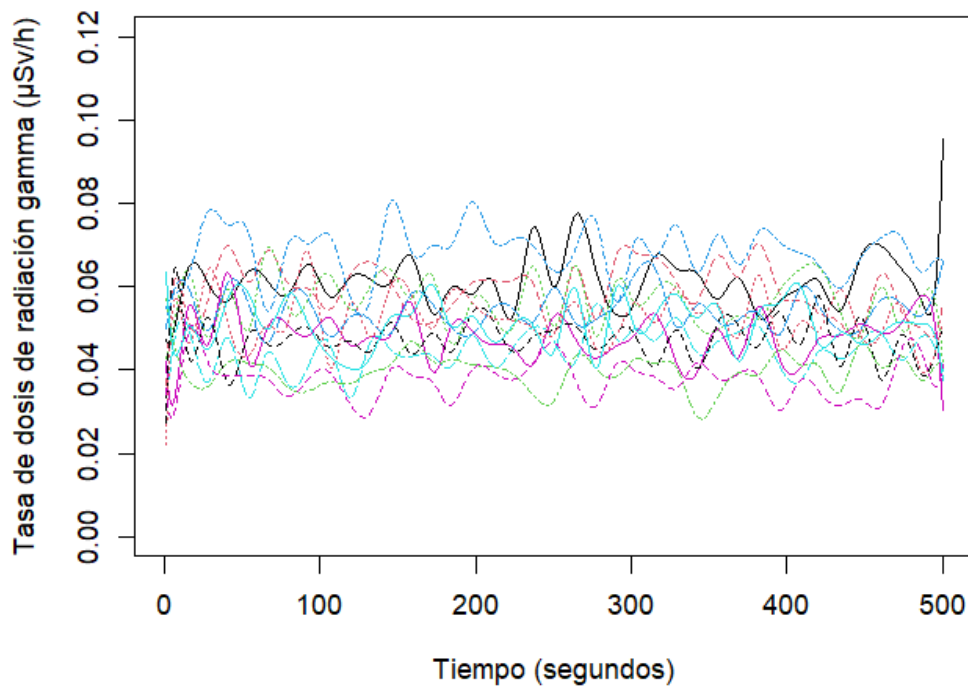
**Ilustración 10-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Guamate

Realizado por: Yungaicela, Víctor, 2022.



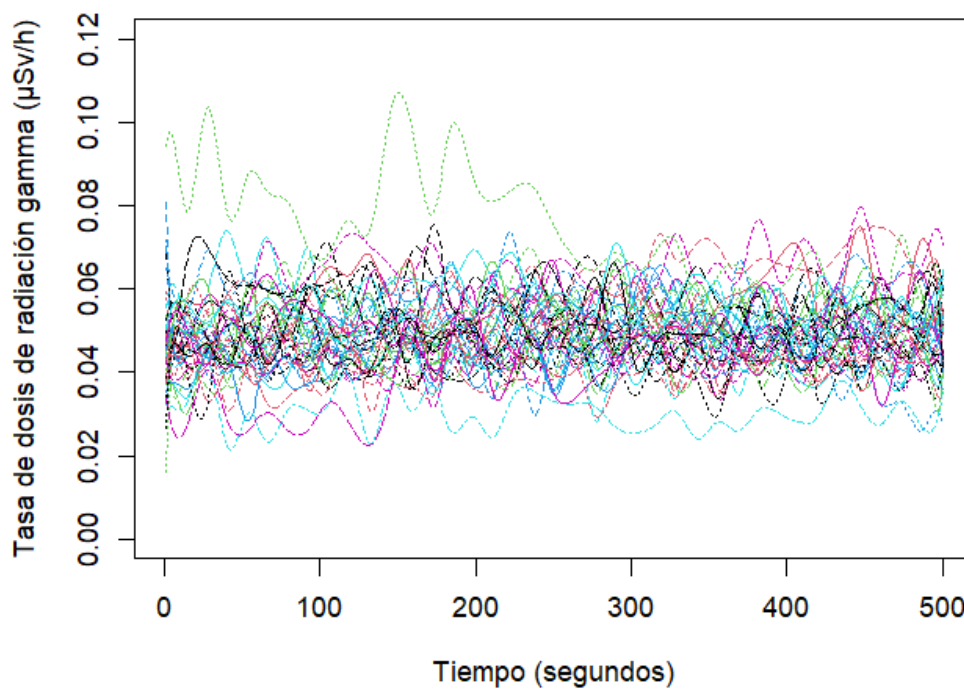
**Ilustración 11-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Guano

Realizado por: Yungaicela, Víctor, 2022.



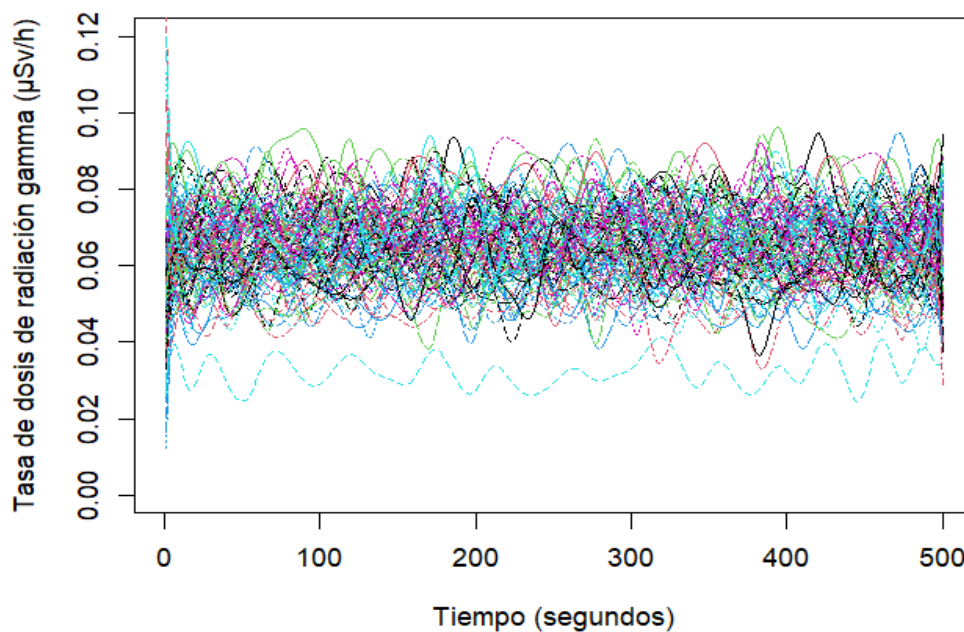
**Ilustración 12-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Pallatanga

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 13-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Penipe

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 14-3:** Curvas suavizadas de la tasa de dosis de rayos gamma del cantón Riobamba

Realizado por: Yungaicela, Víctor, 2022.

### 3.4. Determinación de outliers funcionales

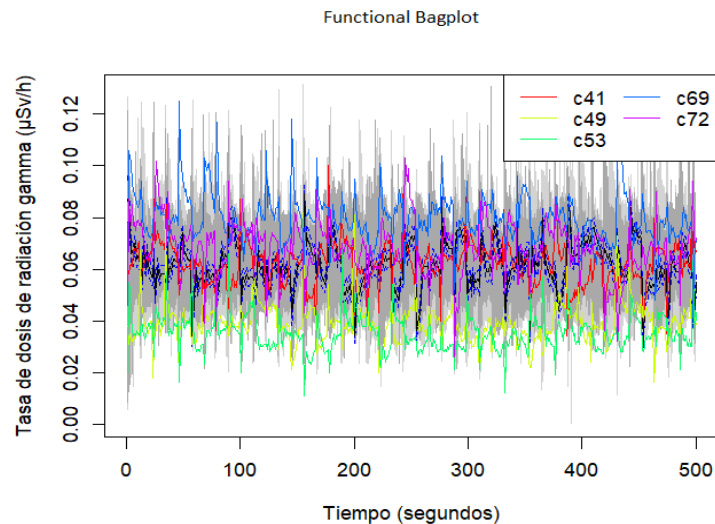
En el HDR boxplot se visualiza en negro la moda funcional. En Alausí se decidió prescindir de dos atípicos funcionales, la observación 49 y 53, perteneciente al punto de muestreo 2 y 6, ambas de Multitud. En Chambo se encontró una única curva atípica (150), de la parroquia con su mismo nombre. En Chunchi, se identificó el sujeto atípico 30, punto 21, de Gonzol. En Colta se halló un outliers funcional en Columbe, lugar 10. Cumandá, se separó la manzana 59, de la parroquia Cumandá. En Guano, Ilapo detectó el muestreo 3 como no factible. En Pallatanga, el outlier característico 35, de Pallatanga. En Penipe se decidió eliminar 4 atípicos, los lugares muestreados 2 y 3 de Matus, 7 de Puela, y 4 de La Candelaria. En Riobamba, Quimiag (zona 31 y 40), y Púngala (punto 39). Finalmente, en Guamote no se presentaron datos atípicos.

#### 3.4.1. Alausí

**Tabla 4-3:** Outliers encontrados con cada una de las técnicas en Alausí

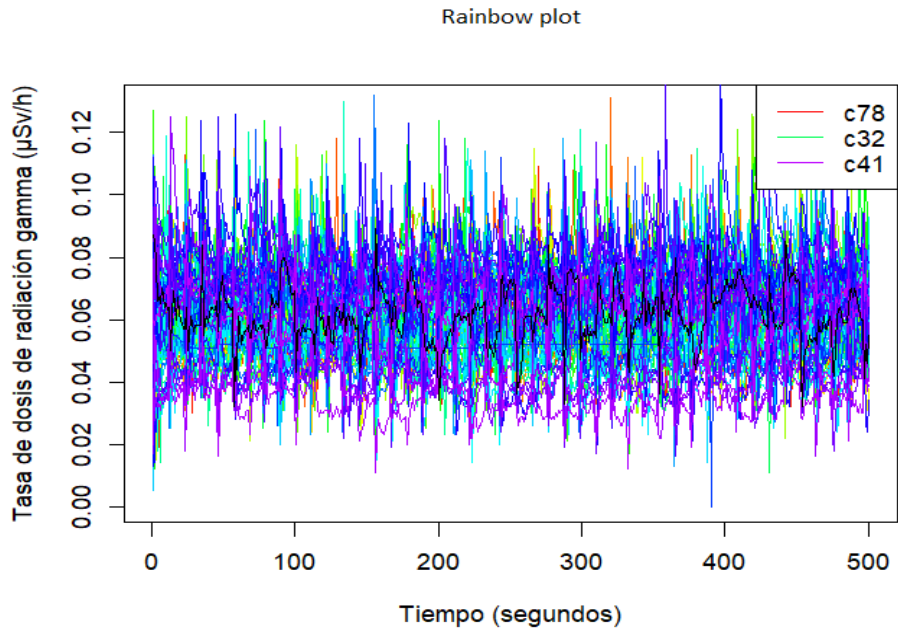
Técnica	Puntos de muestreo Outliers
Functional Bagplot	41,49,53,68,69,72
Rainbow plot	76,32,41
HDR plot	41,49,53,69,72

Realizado por: Yungaicela, Víctor, 2022.



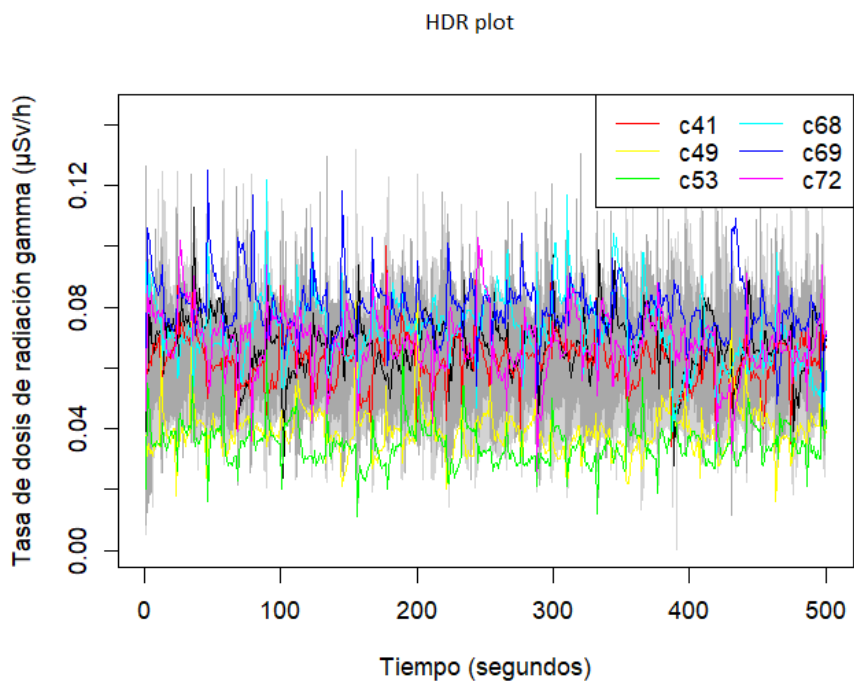
**Ilustración 15-3:** Detección de outliers con técnica Bagplot para el cantón Alausí

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 16-3:** Detección de outliers con técnica Rainbow para el cantón Alausí

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 17-3:** Detección de outliers con técnica HDR para el cantón Alausí

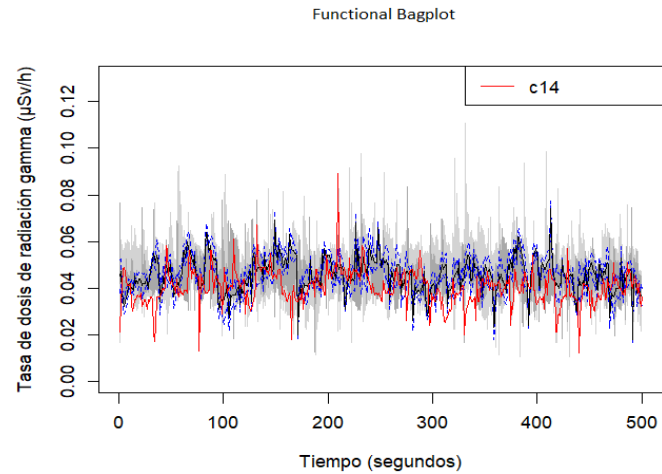
Realizado por: Yungaicela, Víctor, 2022.

### 3.4.2. Chambo

**Tabla 5-3:** Outliers encontrados con cada una de las técnicas en Chambo

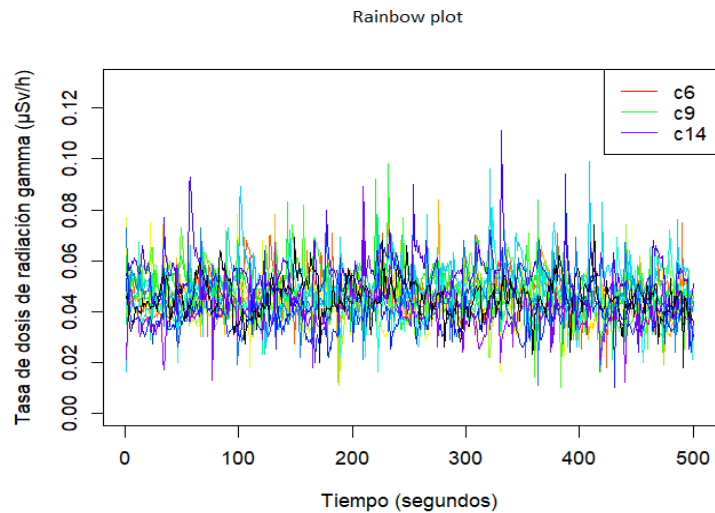
Técnica	Puntos de muestreo Outliers
Funcional Bagplot	14
Rainbow plot	6,9,14
HDR plot	14

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 18-3:** Detección de outliers con técnica Bagplot para el cantón Chambo

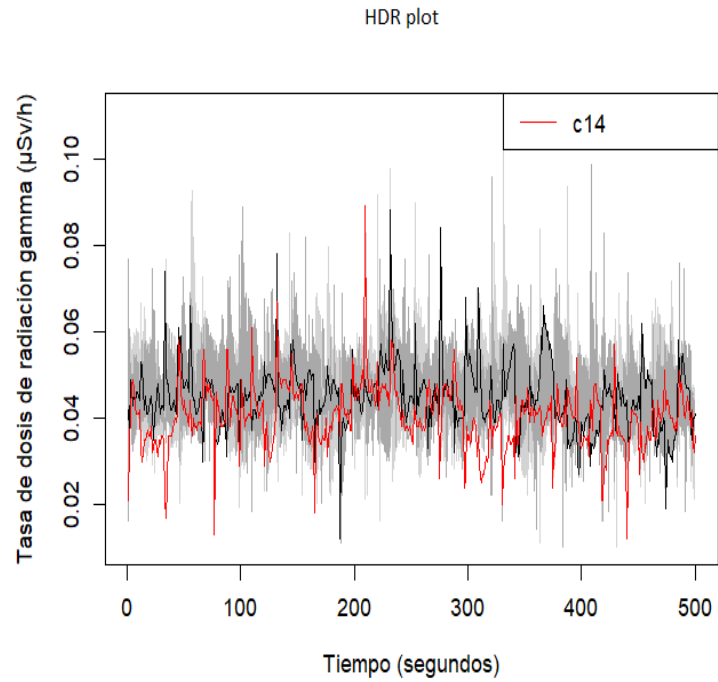
Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 19-3:** Detección de outliers con técnica Rainbow para el cantón Chambo

Realizado por: Yungaicela, Víctor, 2022.





**Ilustración 20-3:** Detección de outliers con técnica HDR para el cantón Chambo

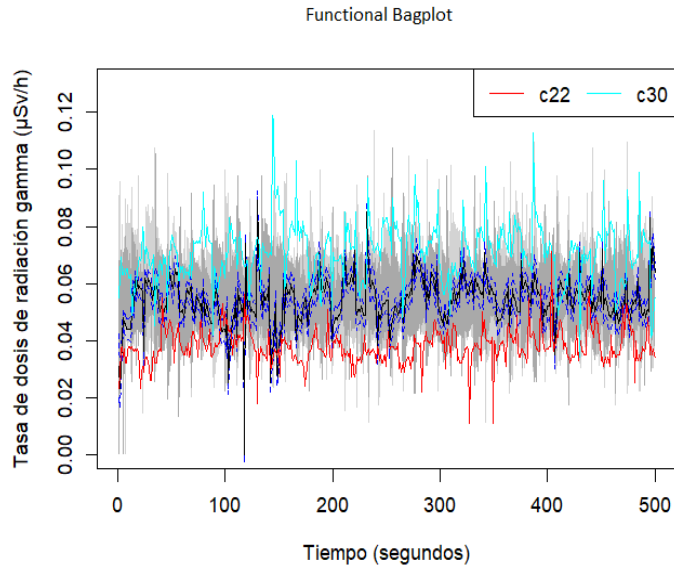
Realizado por: Yungaicela, Víctor, 2022.

### 3.4.3. Chunchi

**Tabla 6-3:** Outliers encontrados con cada una de las técnicas en Chunchi

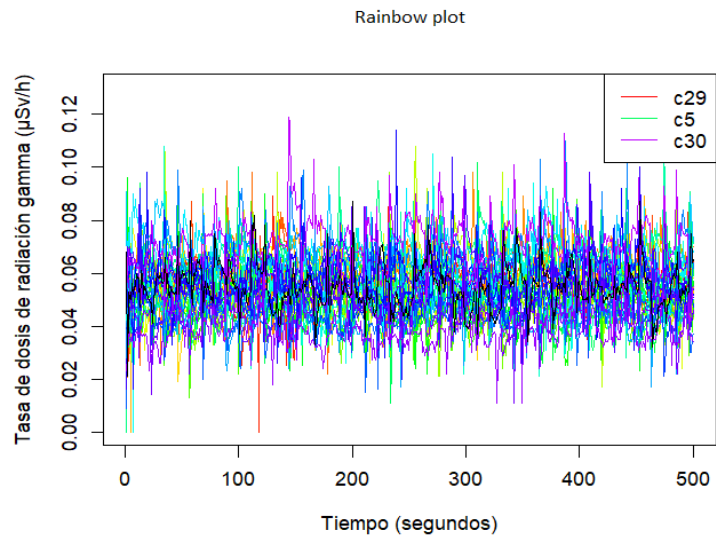
Técnica	Puntos de muestreo Outliers
Functional Bagplot	22,30
Rainbow plot	29,5,30
HDR plot	4,30,22

Realizado por: Yungaicela, Víctor, 2022.



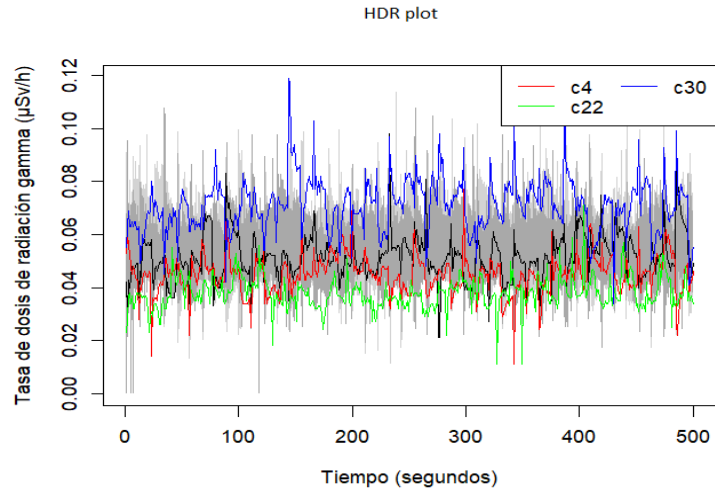
**Ilustración 21-3:** Detección de outliers con técnica Bagplot para el cantón Chunchi

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 22-3:** Detección de outliers con técnica Rainbow para el cantón Chunchi

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 23-3:** Detección de outliers con técnica HDR para el cantón Chunchi

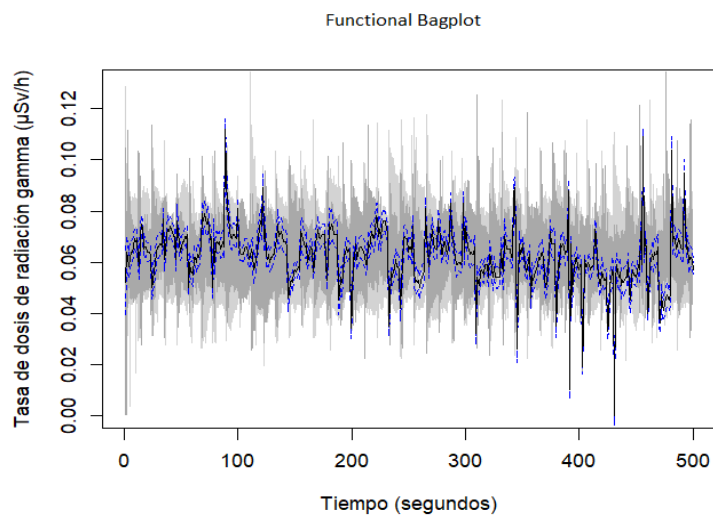
Realizado por: Yungaicela, Víctor, 2022.

### 3.4.4. Colta

**Tabla 7-3:** Outliers encontrados con cada una de las técnicas en Colta

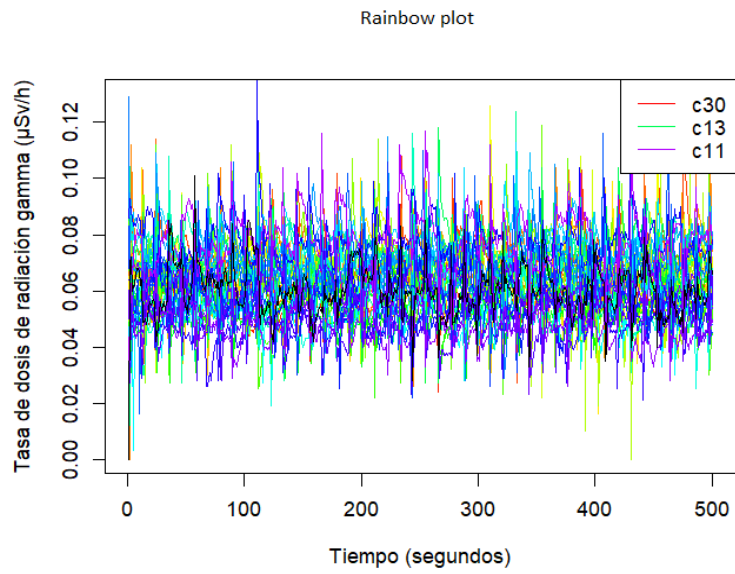
Técnica	Puntos de muestreo Outliers
Funcional Bagplot	-----
Rainbow plot	30,13,11
HDR plot	11,24,31

Realizado por: Yungaicela, Víctor, 2022.



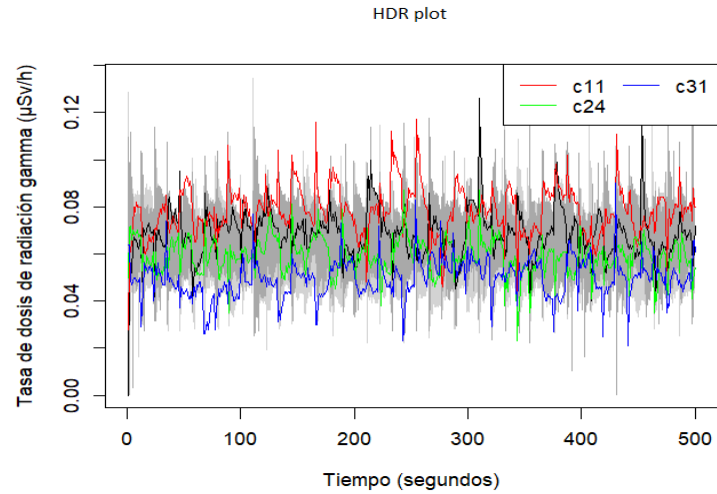
**Ilustración 24-3:** Detección de outliers con técnica Bagplot para el cantón Colta

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 25-3:** Detección de outliers con técnica Rainbow para el cantón Colta

**Realizado por:** Yungaicela, Víctor, 2022.



**Ilustración 26-3:** Detección de outliers con técnica HDR para el cantón Colta

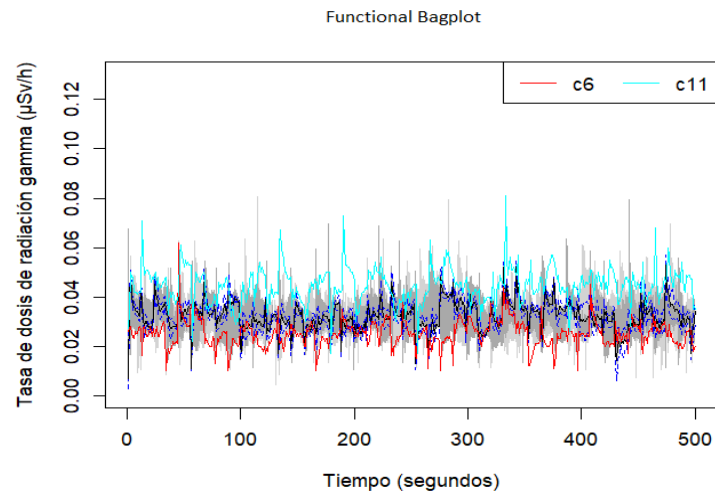
**Realizado por:** Yungaicela, Víctor, 2022.

### 3.4.5. Cumandá

**Tabla 8-3:** Outliers encontrados con cada una de las técnicas en Cumandá

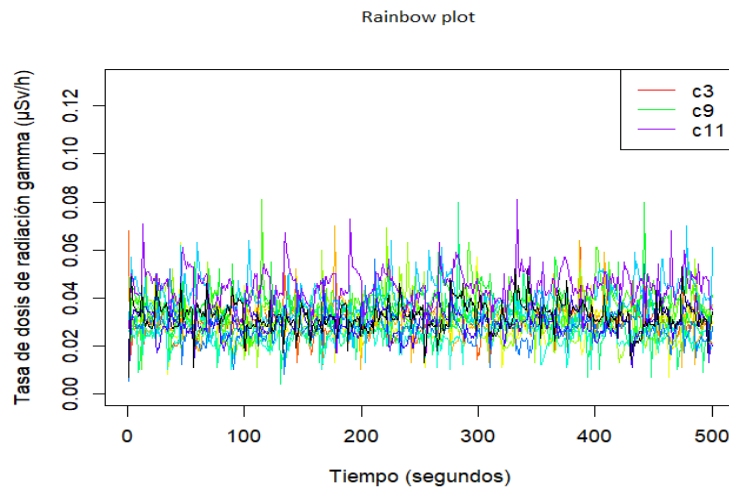
Técnica	Puntos de muestreo Outliers
Functional Bagplot	6,11
Rainbow plot	3,9,11
HDR plot	11

Realizado por: Yungaicela, Víctor, 2022.



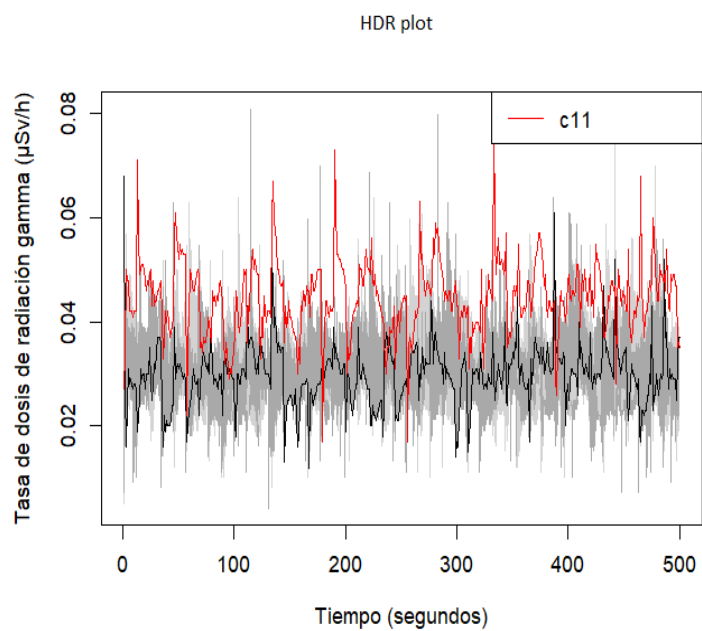
**Ilustración 27-3:** Detección de outliers con técnica Bagplot para el cantón Cumandá

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 28-3:** Detección de outliers con técnica Rainbow para el cantón Cumandá

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 29-3:** Detección de outliers con técnica HDR para el cantón Cumandá

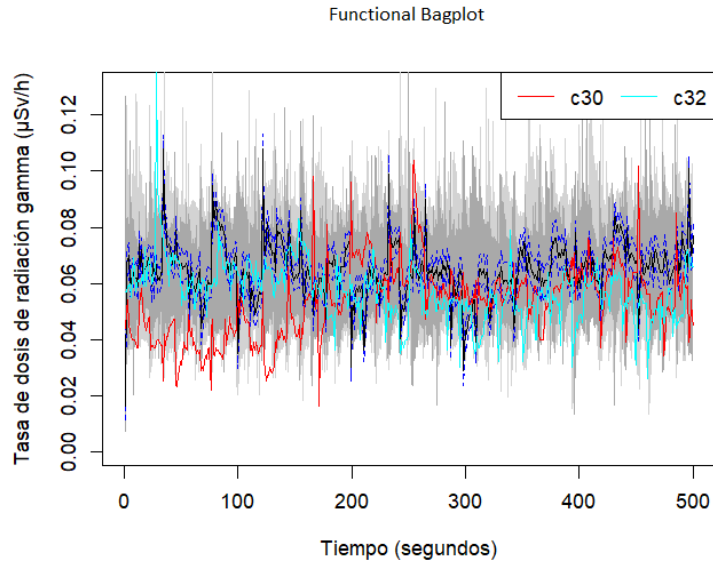
Realizado por: Yungaicela, Víctor, 2022.

### 3.4.6. Guano

**Tabla 9-3:** Outliers encontrados con cada una de las técnicas en Guano

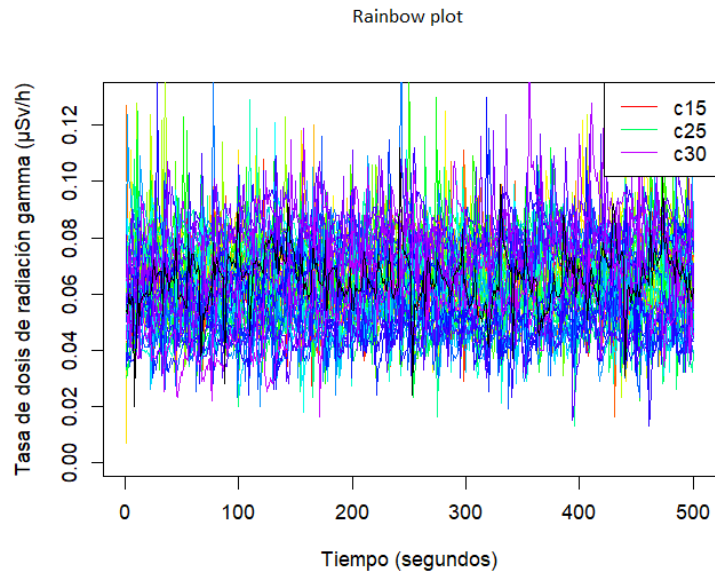
Técnica	Puntos de muestreo Outliers
Functional Bagplot	30,32
Rainbow plot	15,25,30
HDR plot	3,5,30,32
deph.trim.out	-----

Realizado por: Yungaicela, Víctor, 2022.



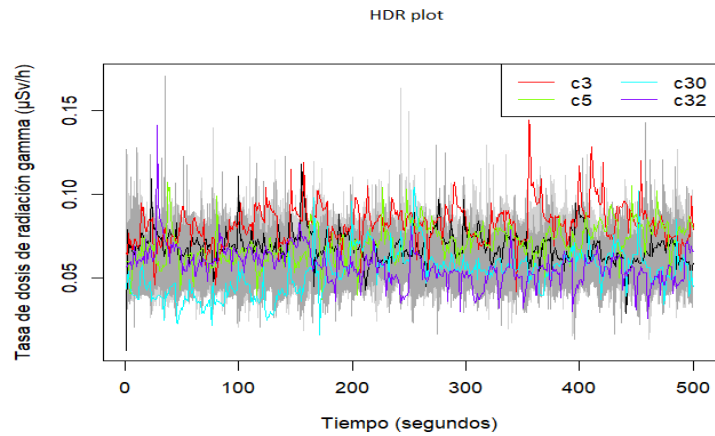
**Ilustración 30-3:** Detección de outliers con técnica Bagplot para el cantón Guano

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 31-3:** Detección de outliers con técnica Rainbow para el cantón Guano

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 32-3:** Detección de outliers con técnica HDR para el cantón Guano

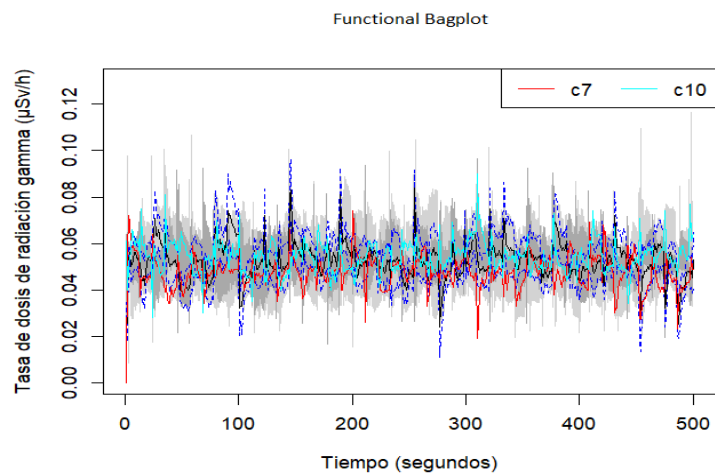
Realizado por: Yungaicela, Víctor, 2022.

### 3.4.7. Pallatanga

**Tabla 10-3:** Outliers encontrados con cada una de las técnicas en Pallatanga

Técnica	Puntos de muestreo Outliers
Funcional Bagplot	7,10
Rainbow plot	11,2,10
HDR plot	10
deph.trim.out	4

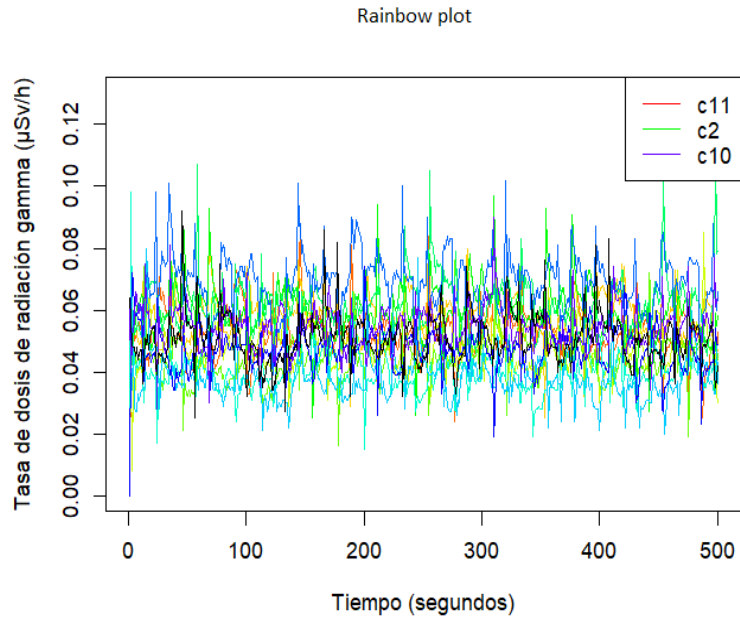
Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 33-3:** Detección de outliers con técnica Bagplot para el cantón Guano

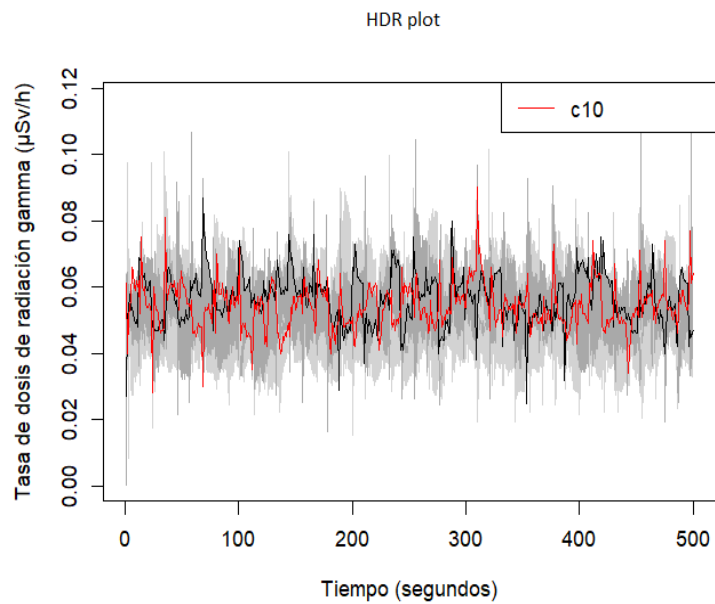
Realizado por: Yungaicela, Víctor, 2022.





**Ilustración 34-3:** Detección de outliers con técnica Rainbow para el cantón Guano

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 35-3:** Detección de outliers con técnica HDR para el cantón Guano

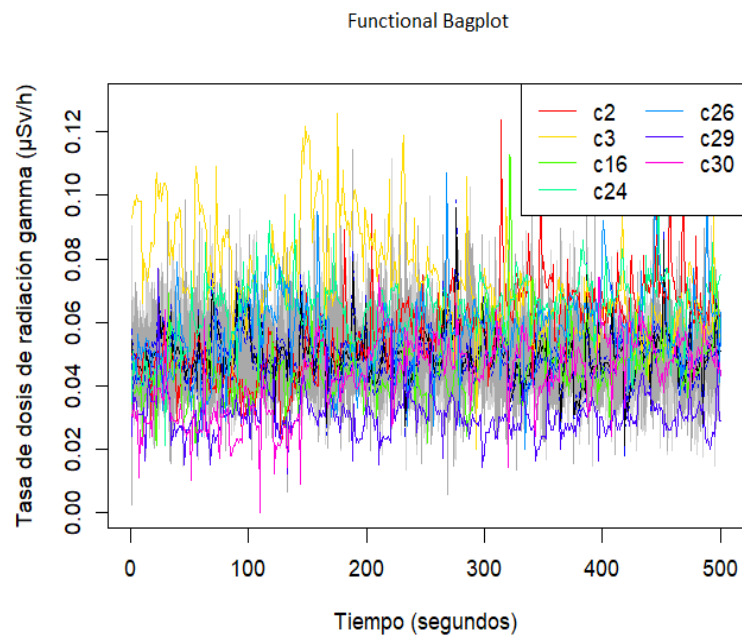
Realizado por: Yungaicela, Víctor, 2022.

### 3.4.8. Penipe

**Tabla 11-3:** Outliers encontrados con cada una de las técnicas en Penipe

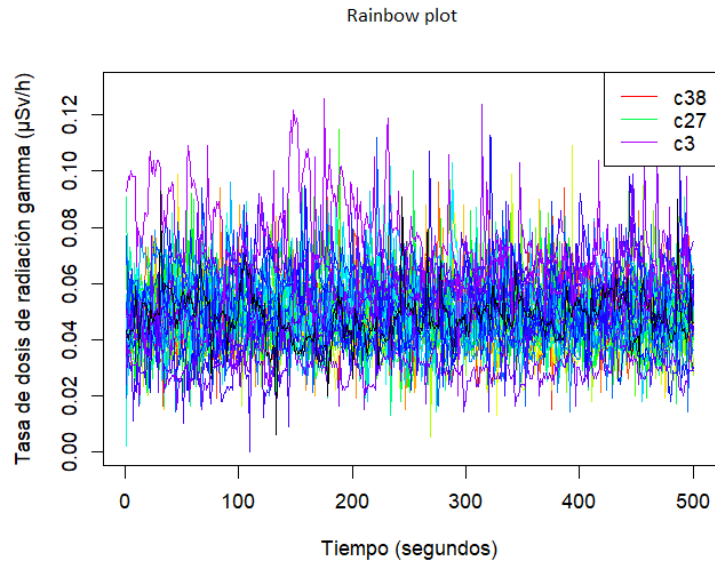
Técnica	Puntos de muestreo Outliers
Functional Bagplot	2,3,16,24,25,29,30
Rainbow plot	38,27,3
HDR plot	2,3,24,29
deph.trim.out	4

Realizado por: Yungaicela, Víctor, 2022.



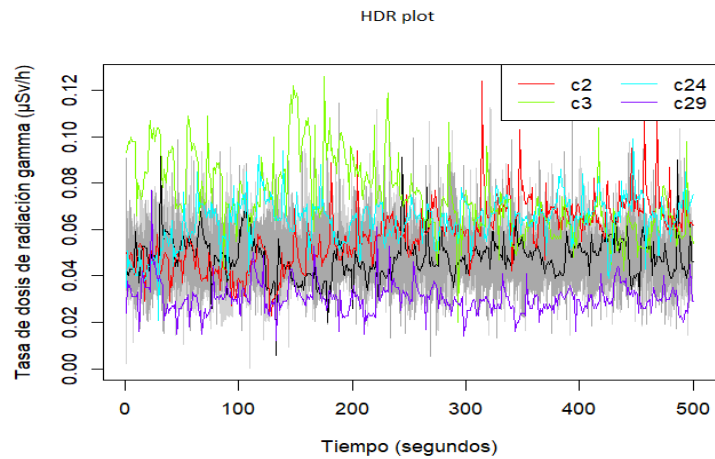
**Ilustración 36-3:** Detección de outliers con técnica Bagplot para el cantón Penipe

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 37-3:** Detección de outliers con técnica Rainbow para el cantón Penipe

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 38-3:** Detección de outliers con técnica HDR para el cantón Penipe

Realizado por: Yungaicela, Víctor, 2022.

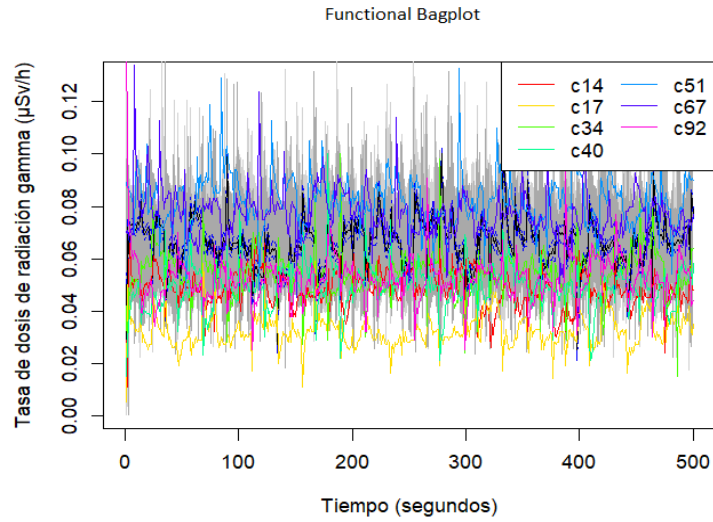
### 3.4.9. Riobamba

**Tabla 12-3:** Outliers encontrados con cada una de las técnicas en Riobamba

Técnica	Puntos de muestreo Outliers
Funcional Bagplot	14,17,34,40,51,67,92
Rainbow plot	79,89,17

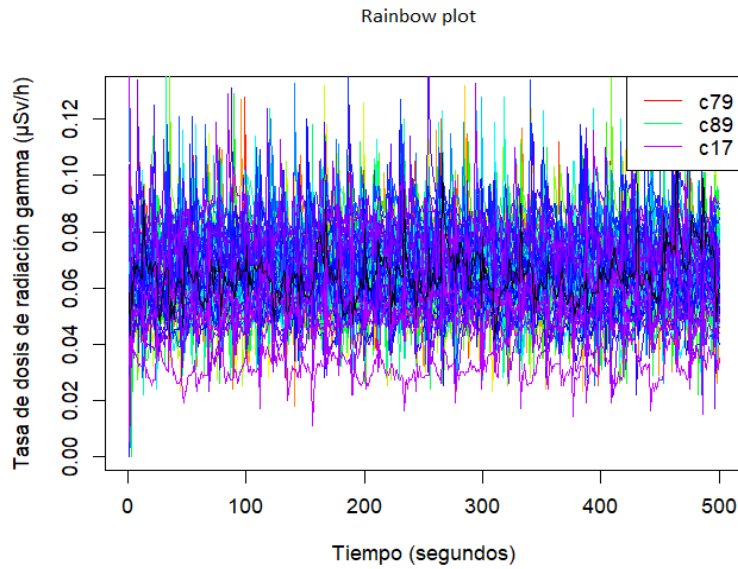
HDR plot	14,17,34,51,67,88,92
deph.trim.out	17,51,14,40

Realizado por: Yungaicela, Víctor, 2022.



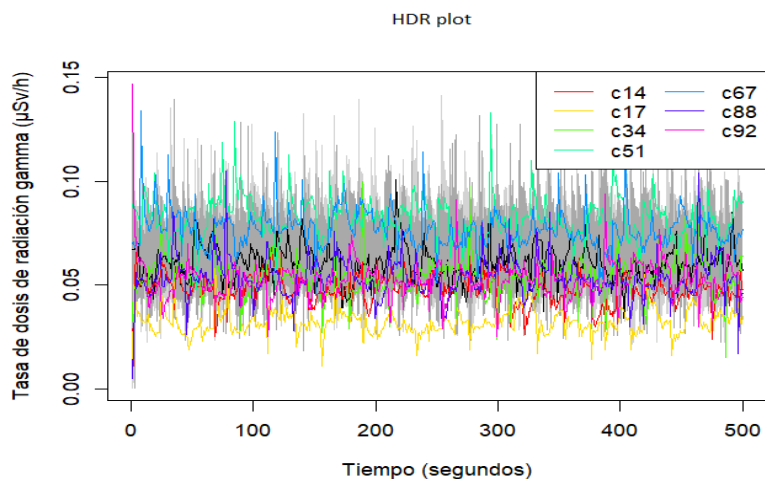
**Ilustración 39-3:** Detección de outliers con técnica Bagplot para el cantón Riobamba

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 40-3:** Detección de outliers con técnica Rainbow para el cantón Riobamba

Realizado por: Yungaicela, Víctor, 2022.



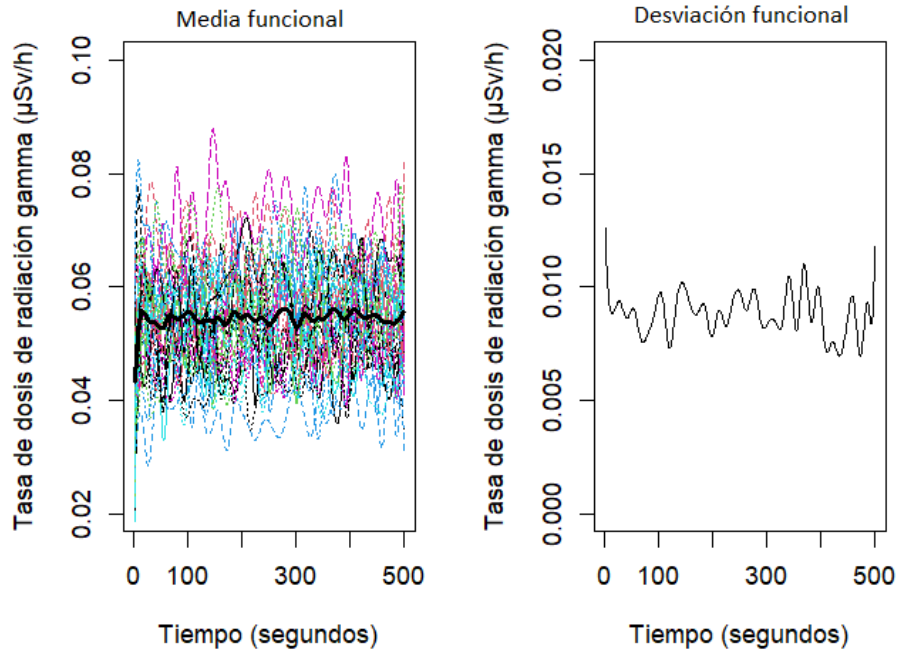
**Ilustración 41-3:** Detección de outliers con técnica HDR para el cantón Riobamba

Realizado por: Yungaicela, Víctor, 2022.

### 3.5. Análisis Descriptivo Funcional

En los diferentes cantones de provincia de Chimborazo se observa una media funcional de 0,04 uSv/h a 0,06 uSv/h en los cantones de Chambo, Chunchi, Pallatanga y Penipe. En cambio, Alausí, Guamote, Colta, Guano, y Riobamba se mantiene en 0,06 uSv/h. Cumandá presenta la menor tasa de dosis en promedio, con 0,03 uSv/h. La desviación estándar funcional presenta poca variabilidad en la mayoría de cantones, con un valor de 0.01 uSv/h, a excepción de Guamote, Chambo, y Chunchi, que su intervalo de variabilidad se establece en un rango de 0.05 uSv/h a 0.01 uSv/h.

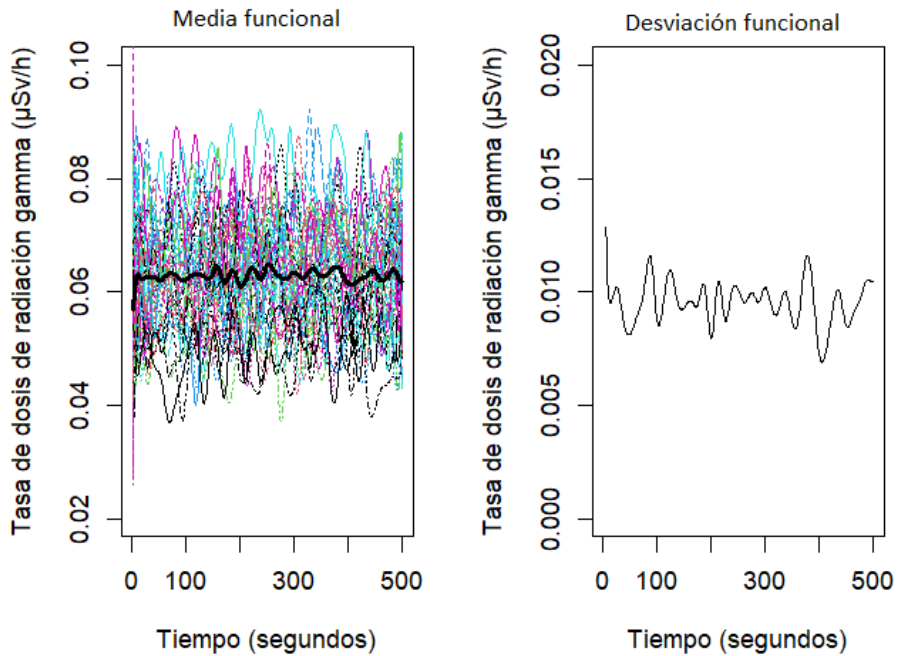
ALAUSI



**Ilustración 42-3:** Media y desviación funcional en el cantón Alausí

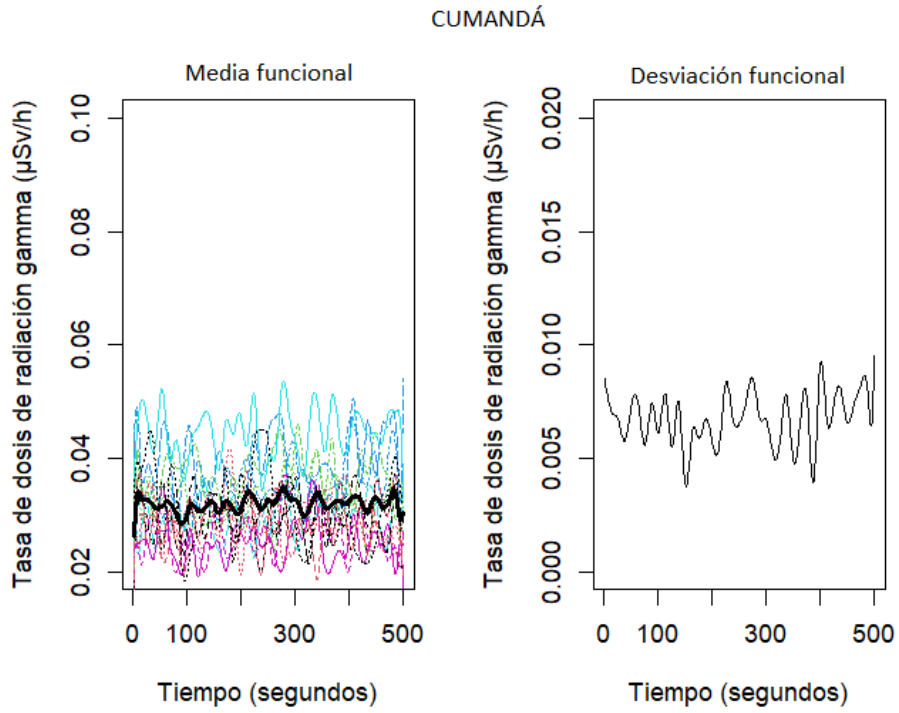
Realizado por: Yungaicela, Víctor, 2022.

COLTA



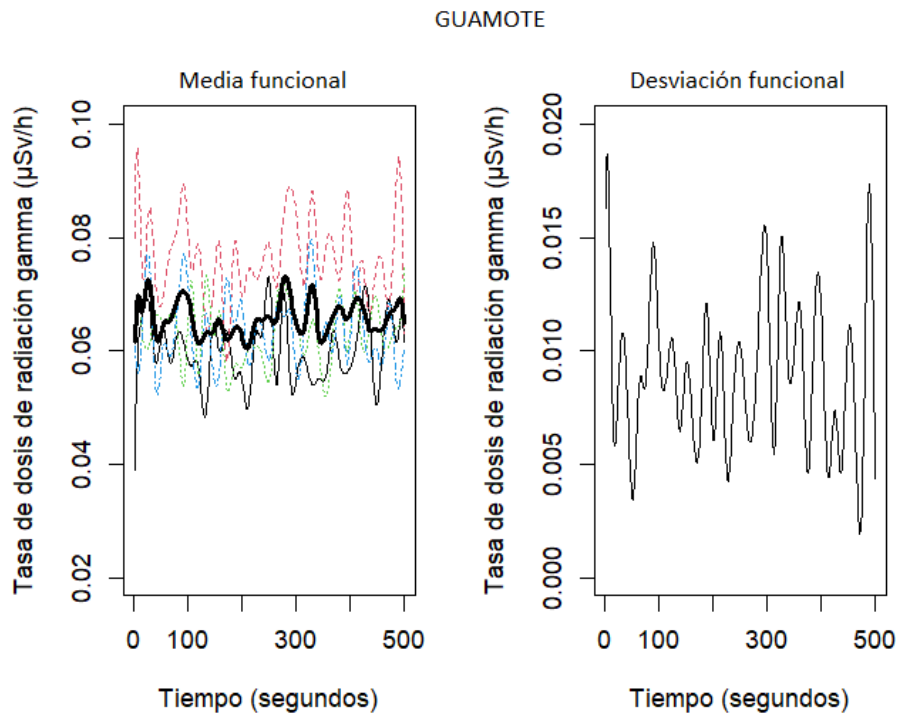
**Ilustración 43-3:** Media y desviación funcional en el cantón Colta

Realizado por: Yungaicela, Víctor, 2022.



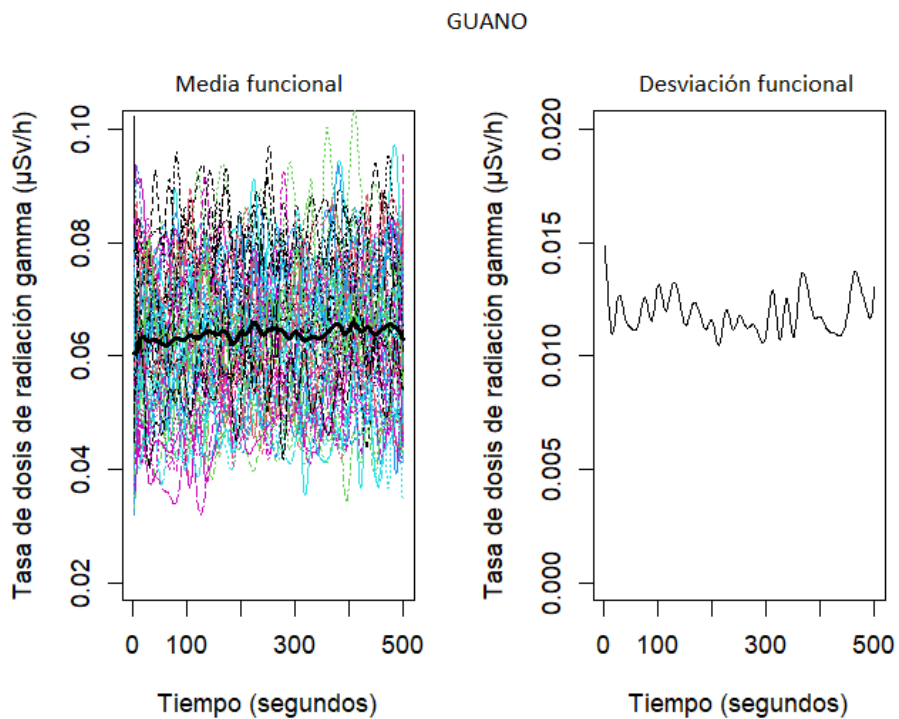
**Ilustración 44-3:** Media y desviación funcional en el cantón Cumandá

Realizado por: Yungaicela, Víctor, 2022.



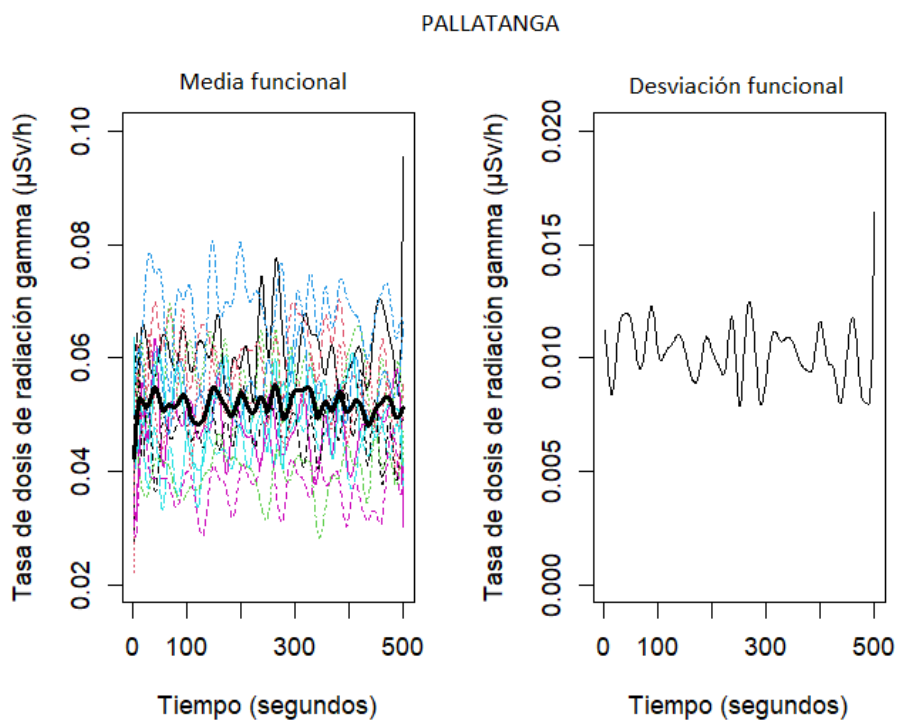
**Ilustración 45-3:** Media y desviación funcional en el cantón Guamote

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 46-3:** Media y desviación funcional en el cantón Guano

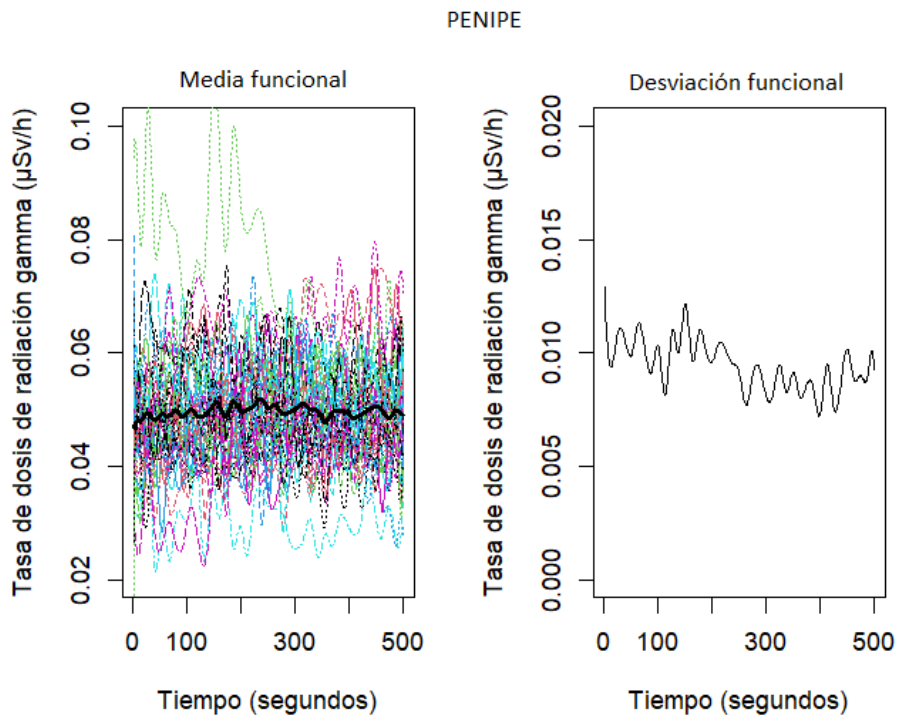
Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 47-3:** Media y desviación funcional en el cantón Pallatanga

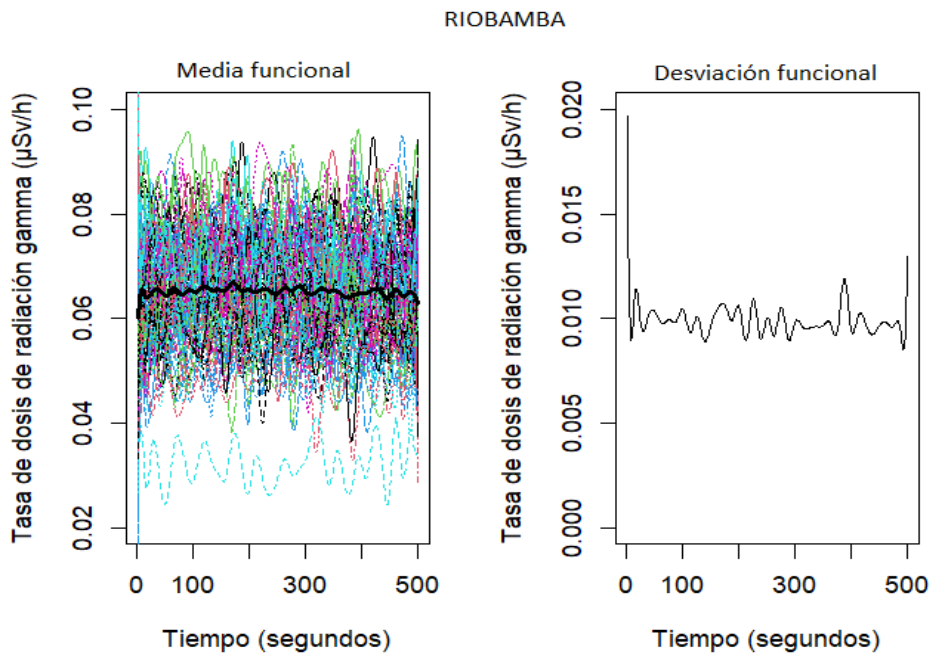
Realizado por: Yungaicela, Víctor, 2022.





**Ilustración 48-3:** Media y desviación funcional en el cantón Penipe

Realizado por: Yungaicela, Víctor, 2022.

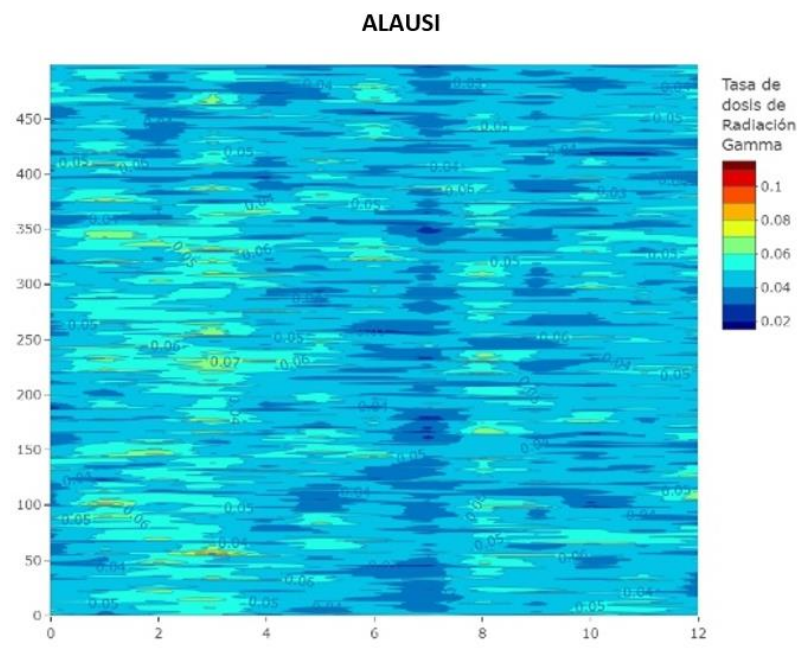


**Ilustración 49-3:** Media y desviación funcional en el cantón Riobamba

Realizado por: Yungaicela, Víctor, 2022.

### 3.6. Variabilidad funcional

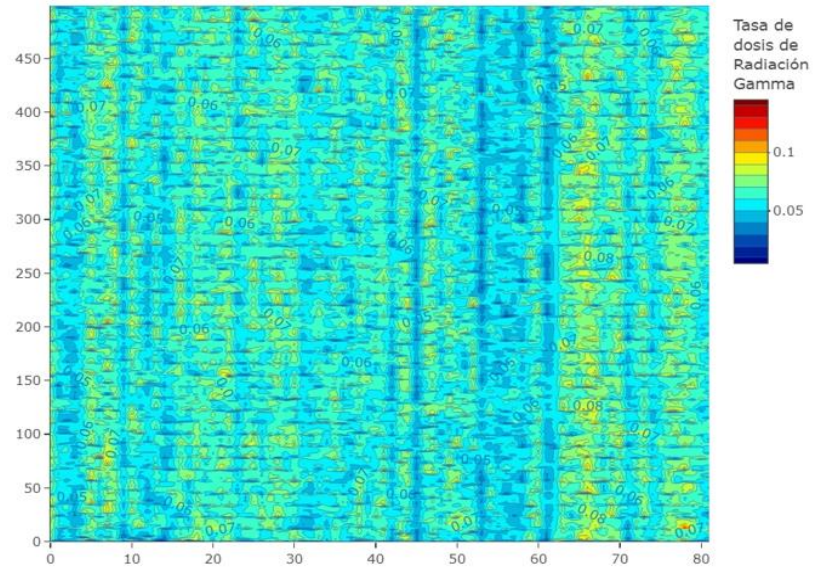
En los gráficos se observa la variación de la tasa de dosis de radiación gamma de cada uno de los cantones de la provincia de Chimborazo, los colores en azul presentan una tasa baja de dosis de radiación, contrario a los colores anaranjados y rojos que muestran un crecimiento de la variable de estudio. Se observó que Alausí presento valores considerablemente bajos en su tasa de dosis, caso opuesto de Guamote, Riobamba, Chunchi que se notó una gran influencia de valores de tasa de radiación altos. El resto de los cantones presentan una equidad en la variable de estudio. Se debe tener presente que esta técnica actúa como una radiografía de la matriz de datos, para representar de forma gráfica la distribución de los valores de la tasa de radiación, más adelante a través de mapa de clúster se vera la interacción de los cantones.



**Ilustración 50-3:** Variabilidad funcional en el cantón Alausí

**Realizado por:** Yungaicela, Víctor, 2022.

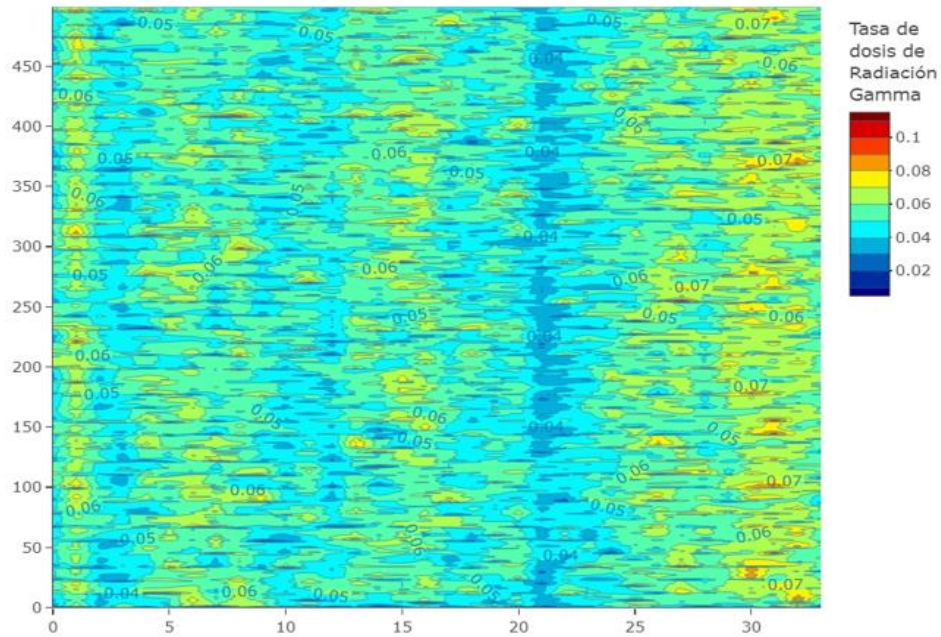
### CHAMBO



**Ilustración 51-3:** Variabilidad funcional en el cantón Chambo

Realizado por: Yungaicela, Víctor, 2022.

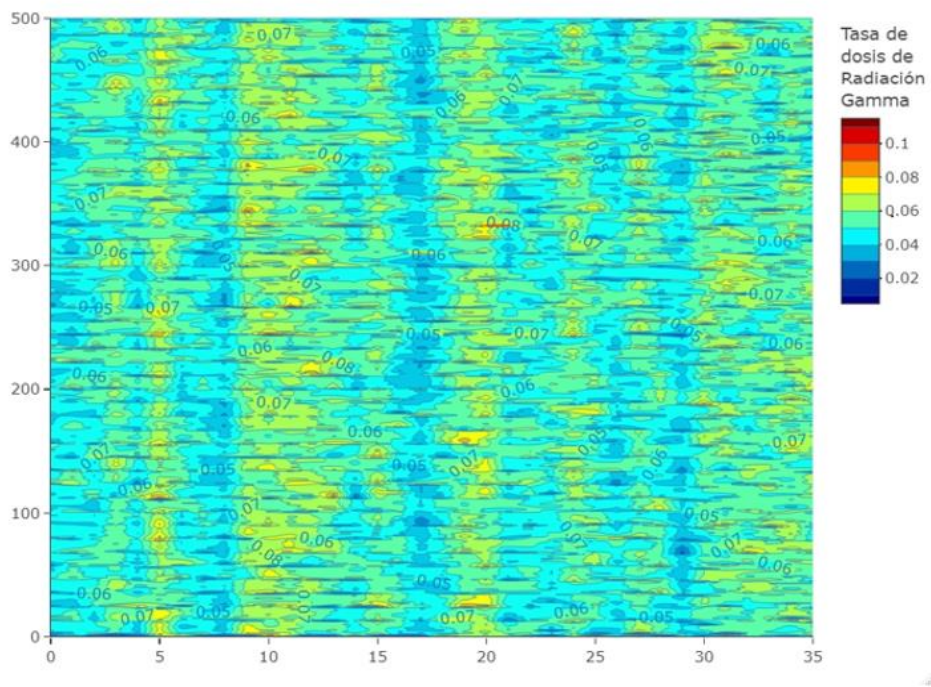
### CHUNCHI



**Ilustración 52-3:** Variabilidad funcional en el cantón Chunchi

Realizado por: Yungaicela, Víctor, 2022.

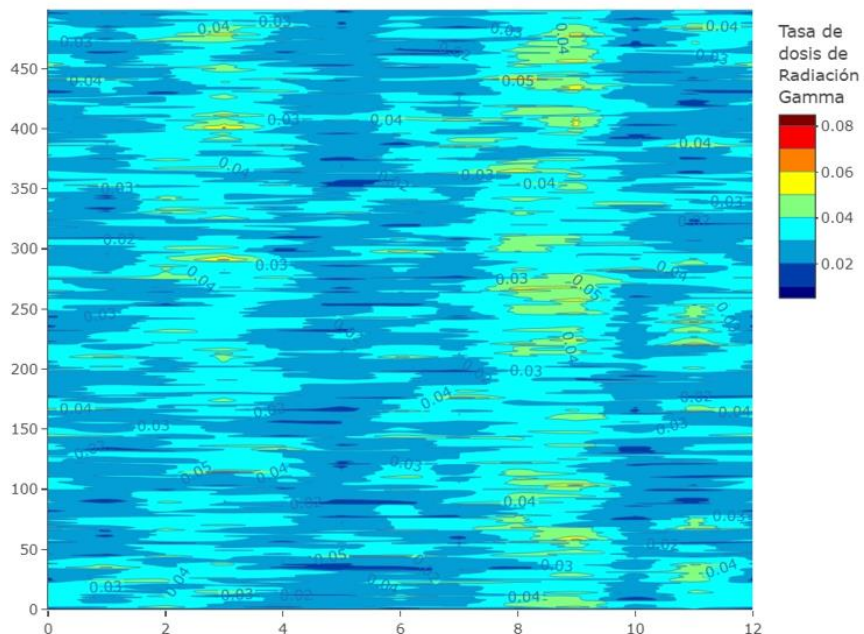
### COLTA



**Ilustración 53-3:** Variabilidad funcional en el cantón Colta

Realizado por: Yungaicela, Víctor, 2022.

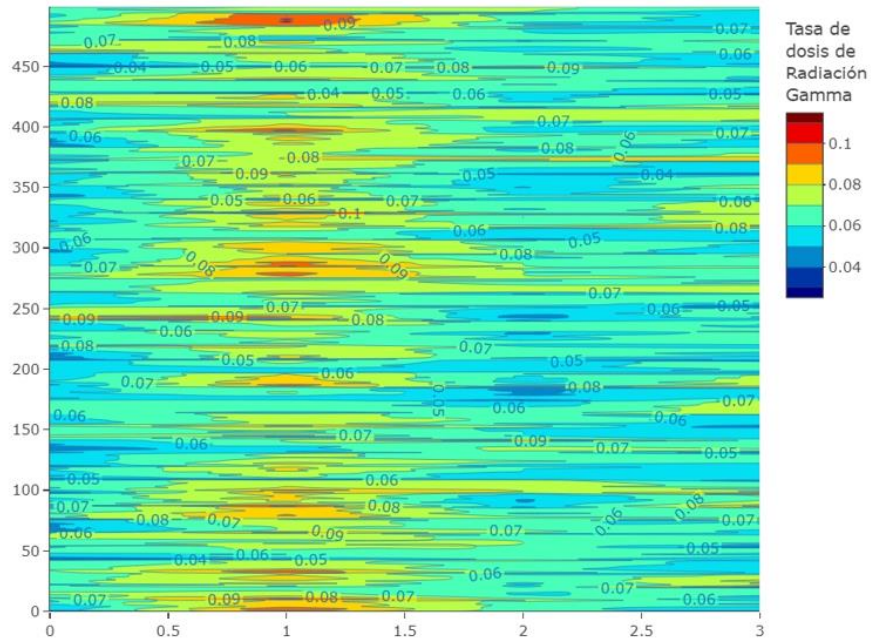
### CUMANDÁ



**Ilustración 54-3:** Variabilidad funcional en el cantón Cumandá

Realizado por: Yungaicela, Víctor, 2022.

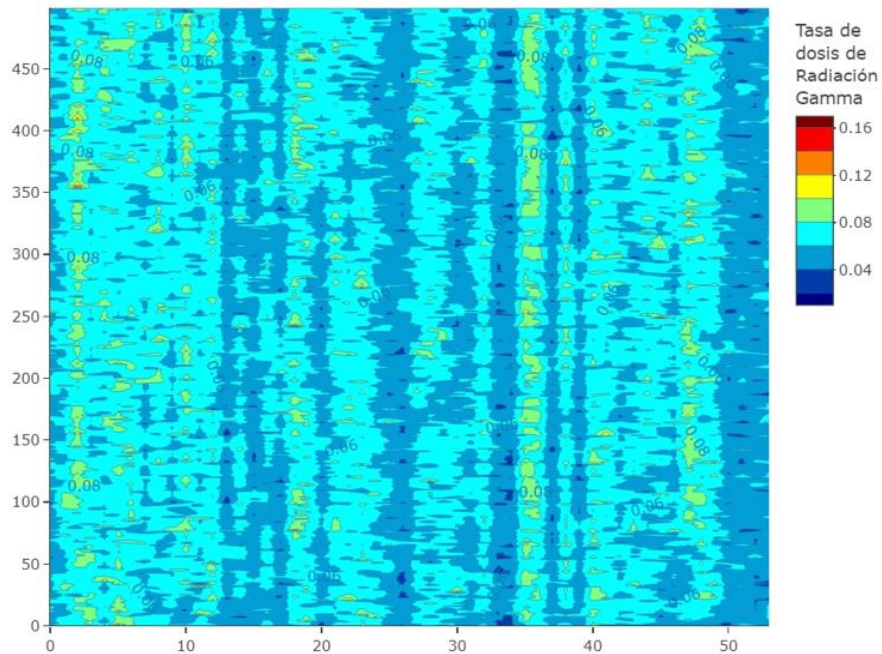
### GUAMOTE



**Ilustración 55-3:** Variabilidad funcional en el cantón Guamate

Realizado por: Yungaicela, Víctor, 2022.

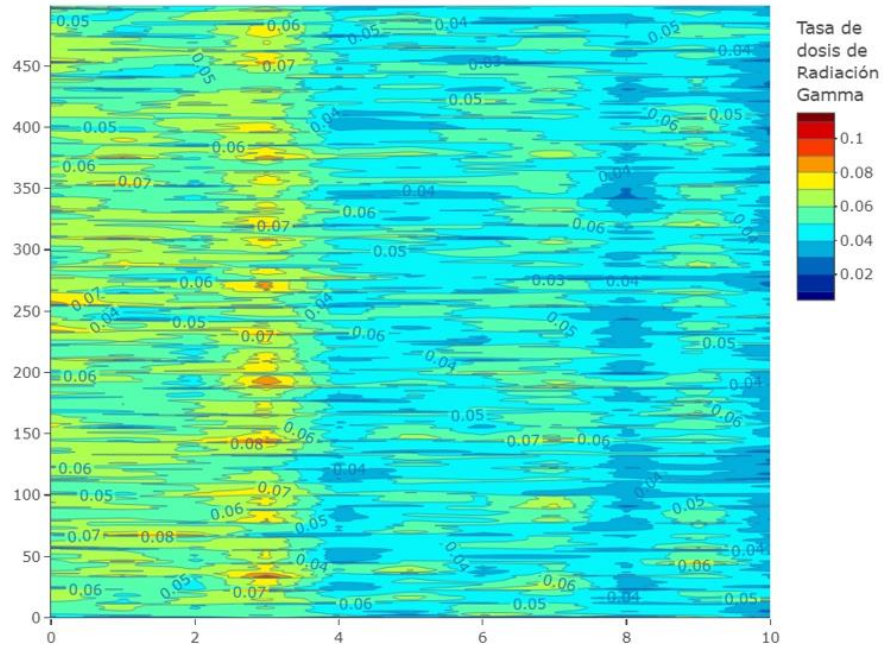
### GUANO



**Ilustración 56-3:** Variabilidad funcional en el cantón Guano

Realizado por: Yungaicela, Víctor, 2022.

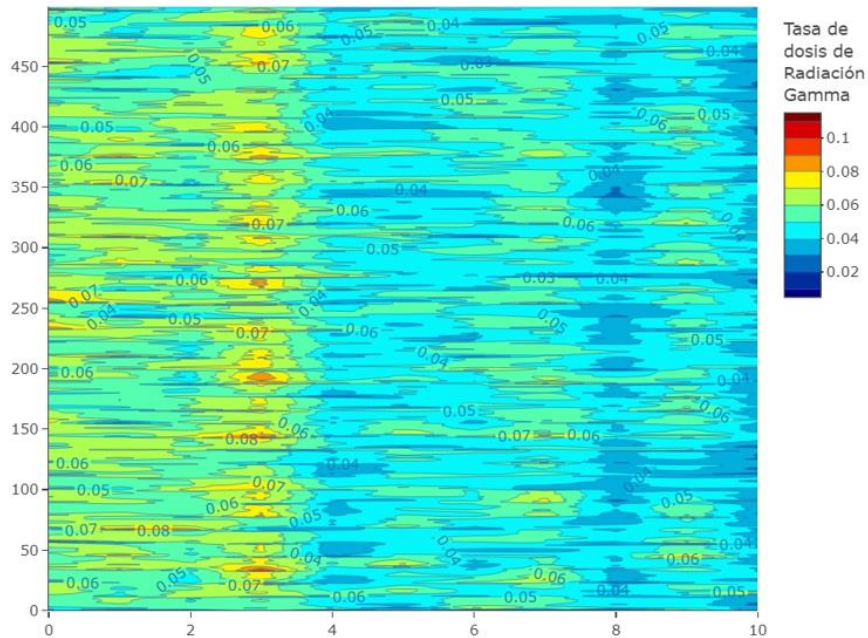
### PALLATANGA



**Ilustración 57-3:** Variabilidad funcional en el cantón Pallatanga

Realizado por: Yungaicela, Víctor, 2022.

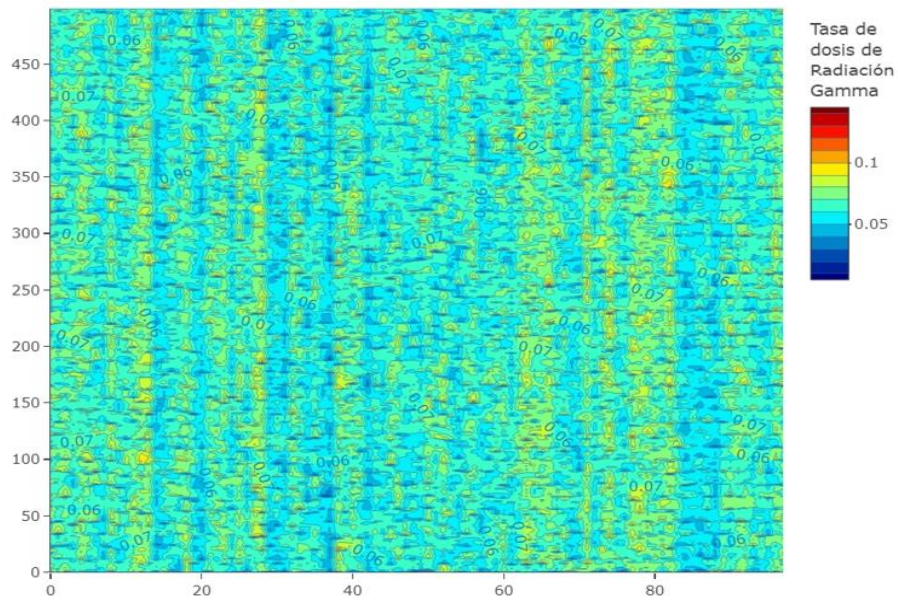
### PENIPE



**Ilustración 58-3:** Variabilidad funcional en el cantón Penipe

Realizado por: Yungaicela, Víctor, 2022.

## RIOBAMBA

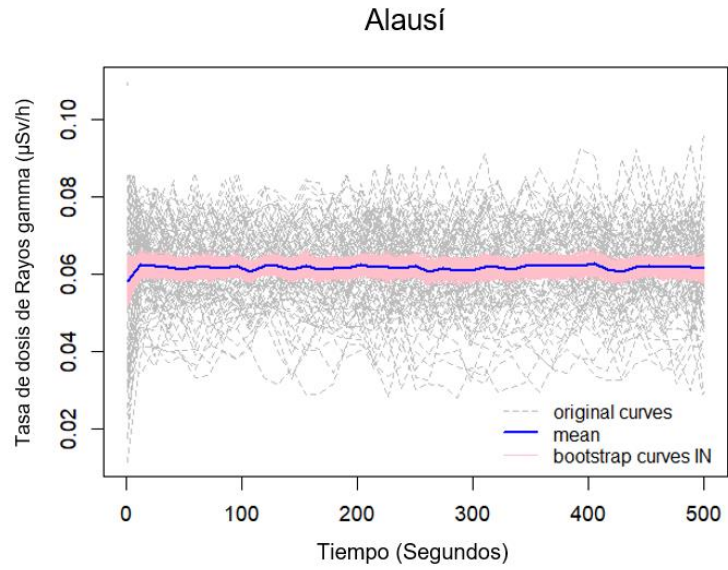


**Ilustración 59-3:** Variabilidad funcional en el cantón Riobamba

**Realizado por:** Yungaicela, Víctor, 2022.

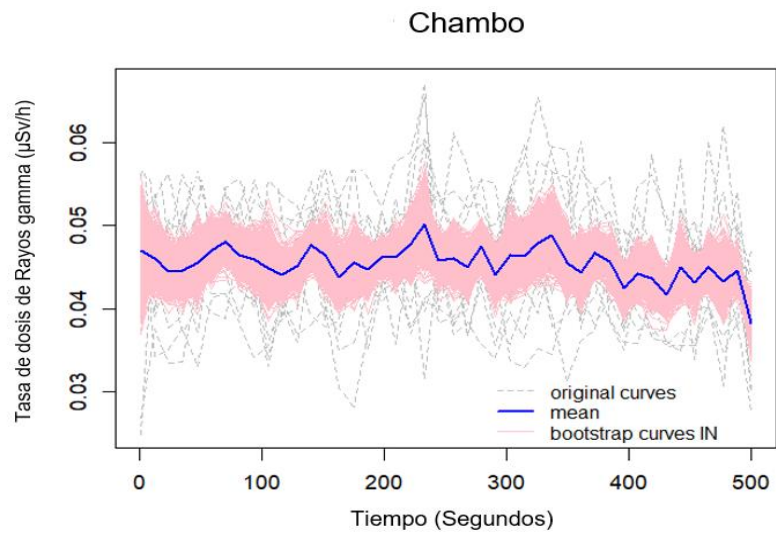
### 3.7. Remuestreo de Bootstrap

Para identificar la banda de confianza de la media, se usó remuestreo de bootstrap para cada cantón de la provincia de Chimborazo, se utilizó el paquete `fda.usc` del software R. A través del comando `fdatabootstrap`, se obtuvo la banda de confianza de Bootstrap, un estimador de la media funcional y el porcentaje total de absorción.



**Ilustración 60-3:** Remuestreo Boostrap para el cantón Alausí

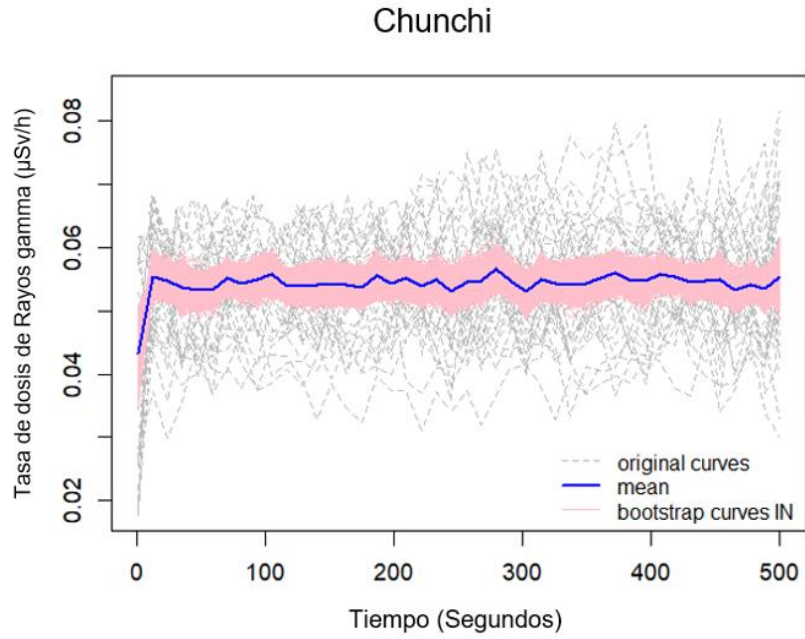
Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 61-3:** Remuestreo Boostrap para el cantón Alausí

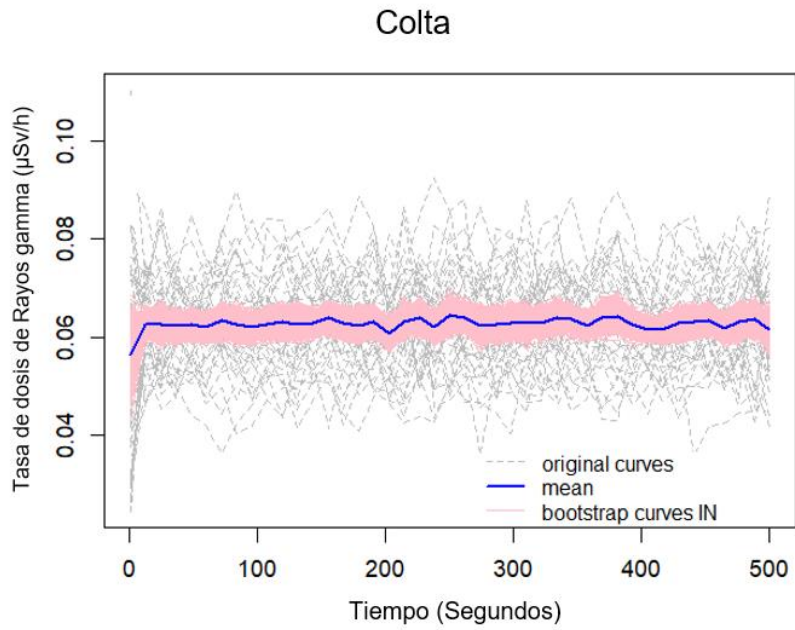
Realizado por: Yungaicela, Víctor, 2022.





**Ilustración 62-3:** Remuestreo Bootstrap para el cantón Chunchi

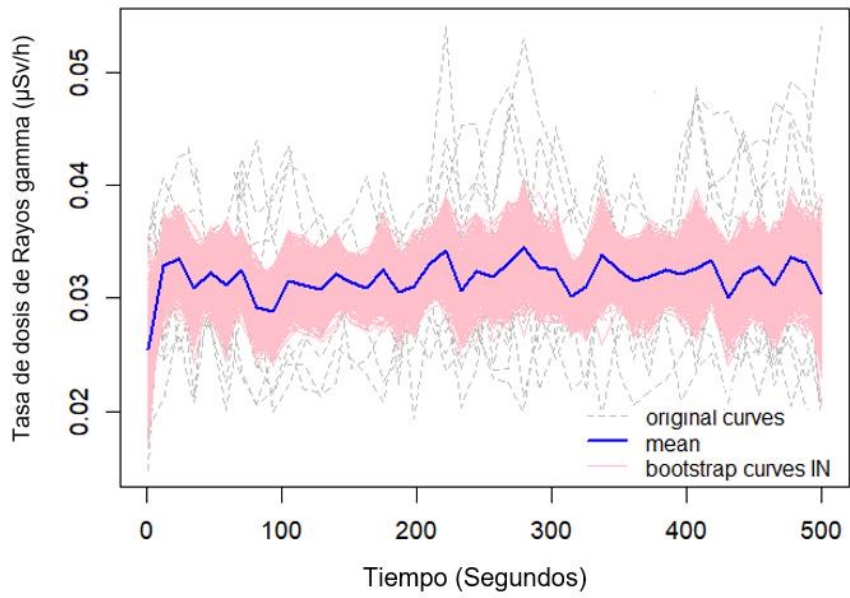
Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 63-3:** Remuestreo Bootstrap para el cantón Colta

Realizado por: Yungaicela, Víctor, 2022.

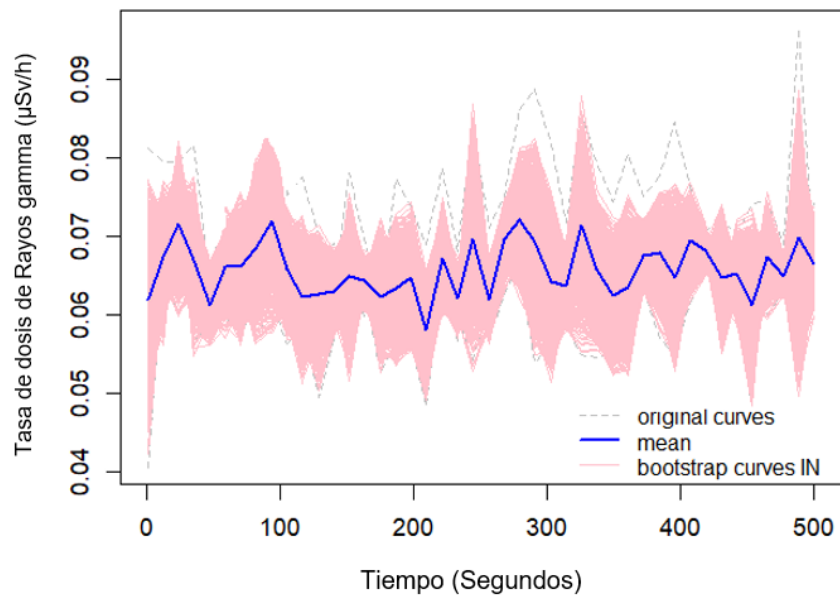
### Cumandá



**Ilustración 64-3:** Remuestreo Bootstrap para el cantón Cumandá

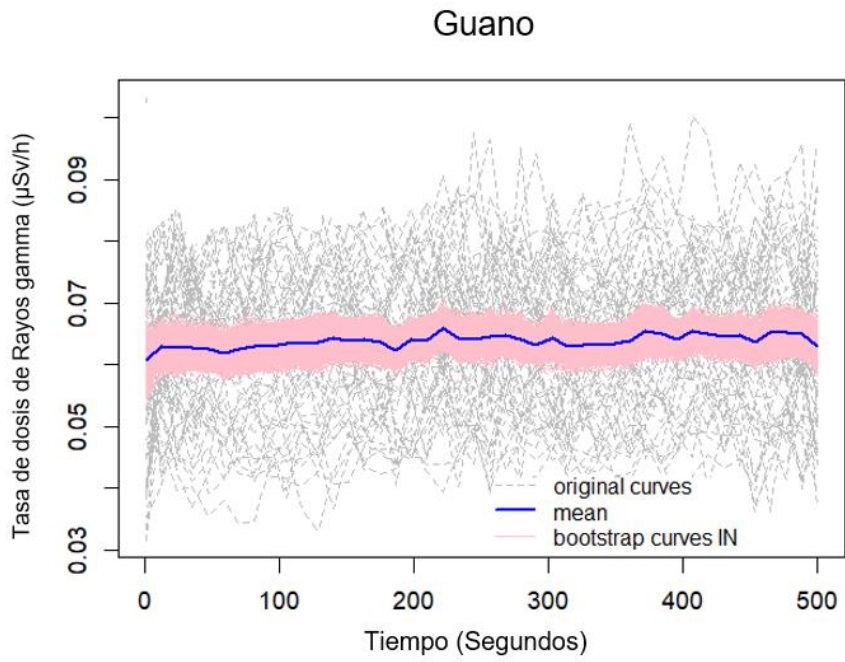
Realizado por: Yungaicela, Víctor, 2022.

### Guamote

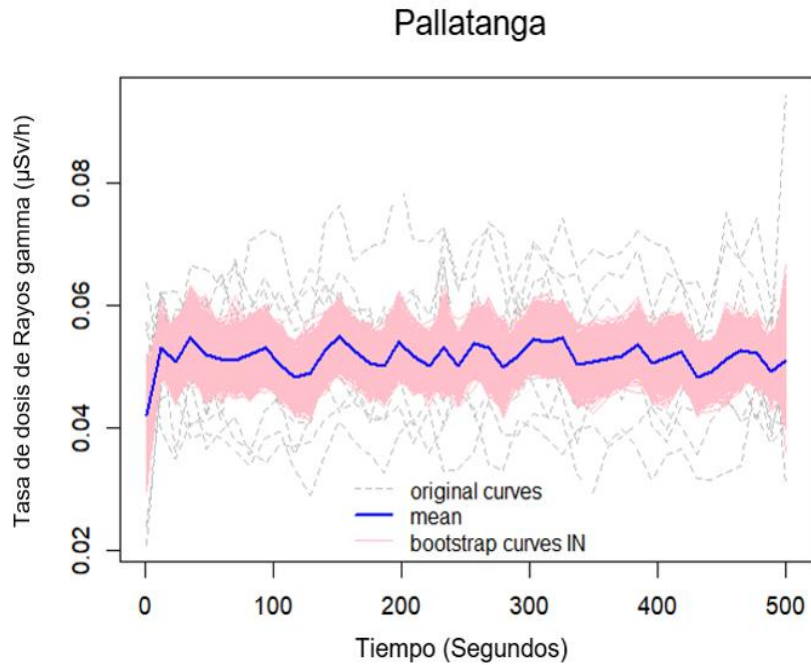


**Ilustración 65-3:** Remuestreo Bootstrap para el cantón Guamote

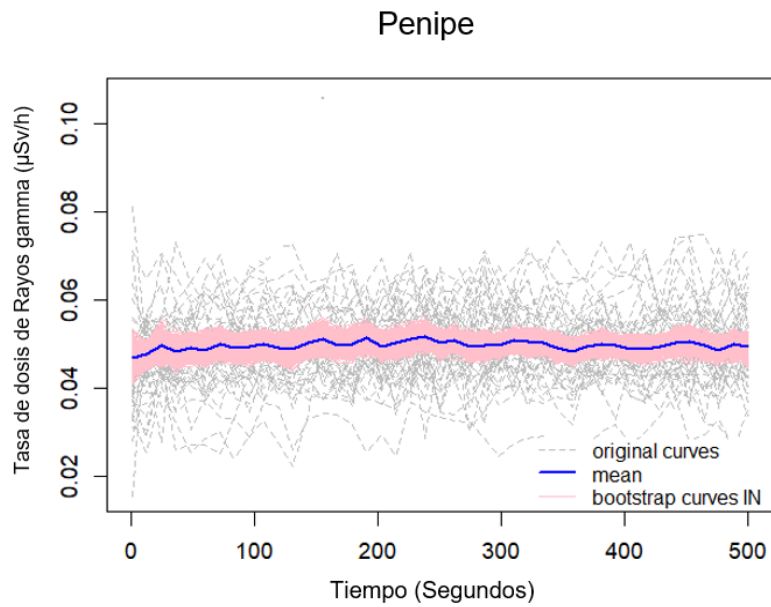
Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 66-3:** Remuestreo Bootstrap para el cantón Guano  
 Realizado por: Yungaicela, Víctor, 2022.

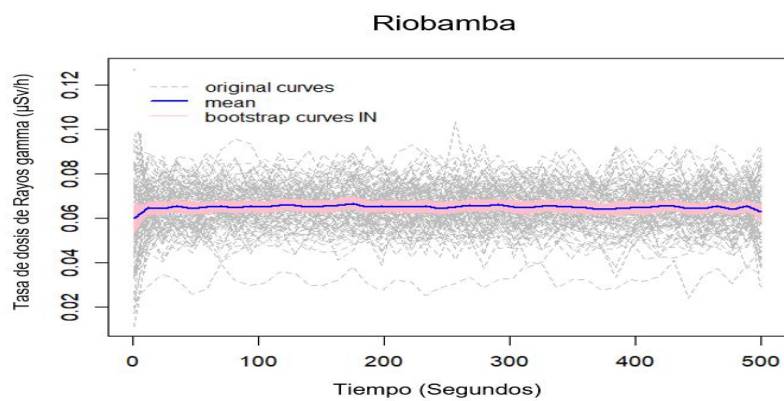


**Ilustración 67-3:** Remuestreo Bootstrap para el cantón Pallatanga  
 Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 68-3:** Remuestreo Bootstrap para el cantón Penipe

Realizado por: Yungaicela, Víctor, 2022.



**Ilustración 69-3:** Remuestreo Bootstrap para el cantón Riobamba

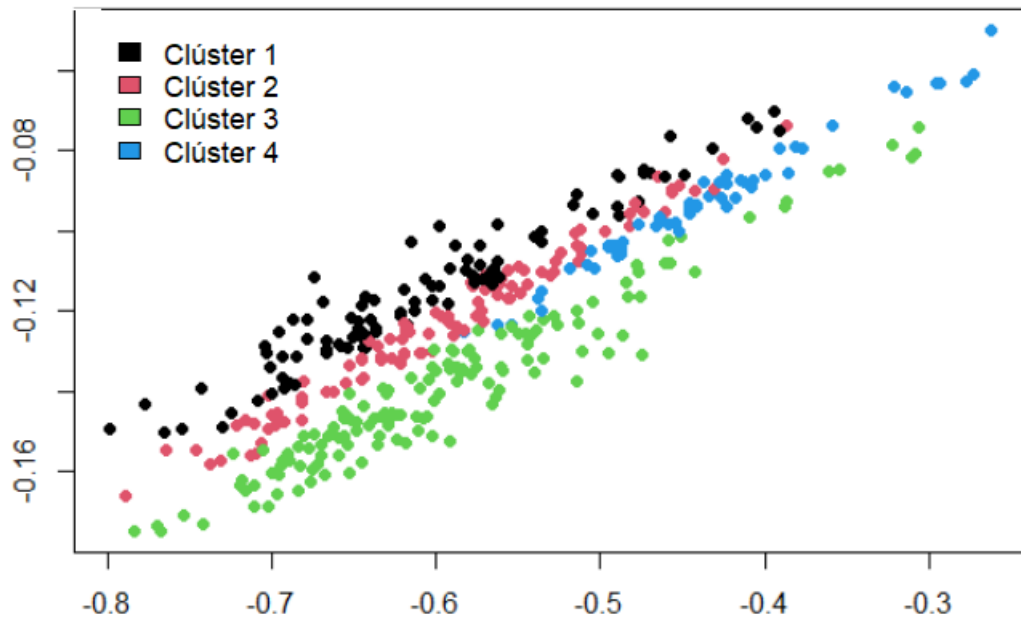
Realizado por: Yungaicela, Víctor, 2022.

### 3.8. Análisis Clúster funcional

#### 3.9.1. Número óptimo de clusters

En este análisis, se utilizó el método gráfico, con el que se encontró conclusiones para hallar el número óptimo de clusters haciendo una comparativa con diferentes números  $k$ . El gráfico 71-3

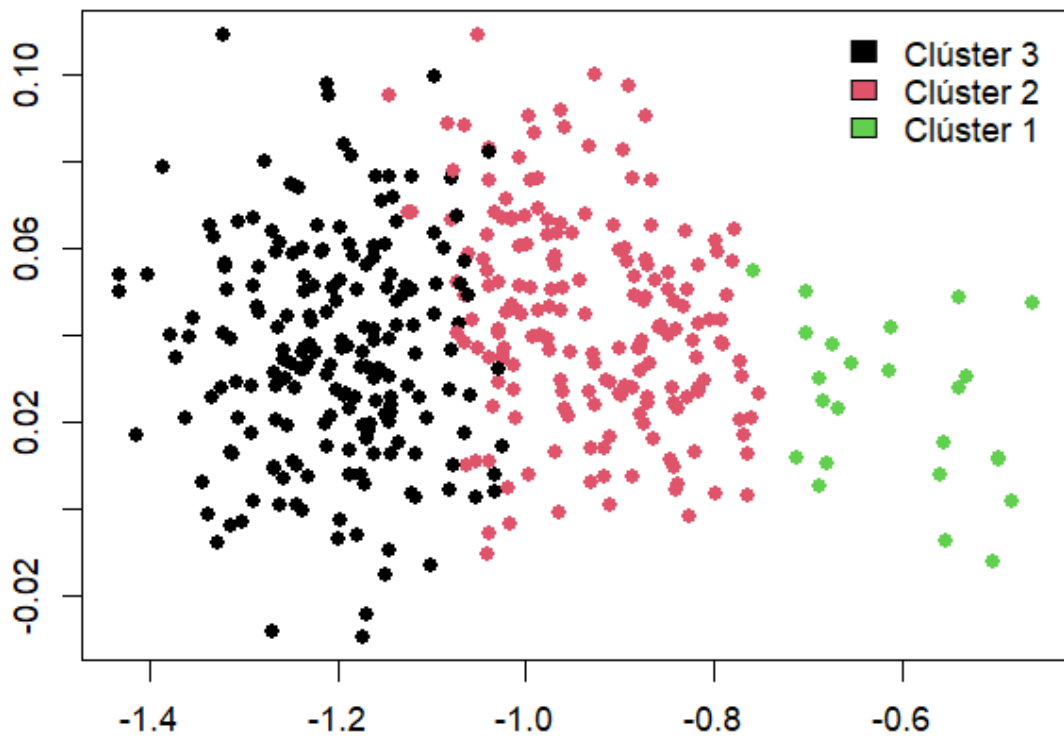
se realizó con  $k = 4$  clusters. Se observó cómo los datos tienden a formar una recta, descartando cualquier tipo de agrupación, por lo que se decidió reducir el número de clusters a  $k = 3$ .



**Ilustración 70-3:** Número óptimo de Clúster  $k=4$

**Realizado por:** Yungaicela, Víctor, 2022.

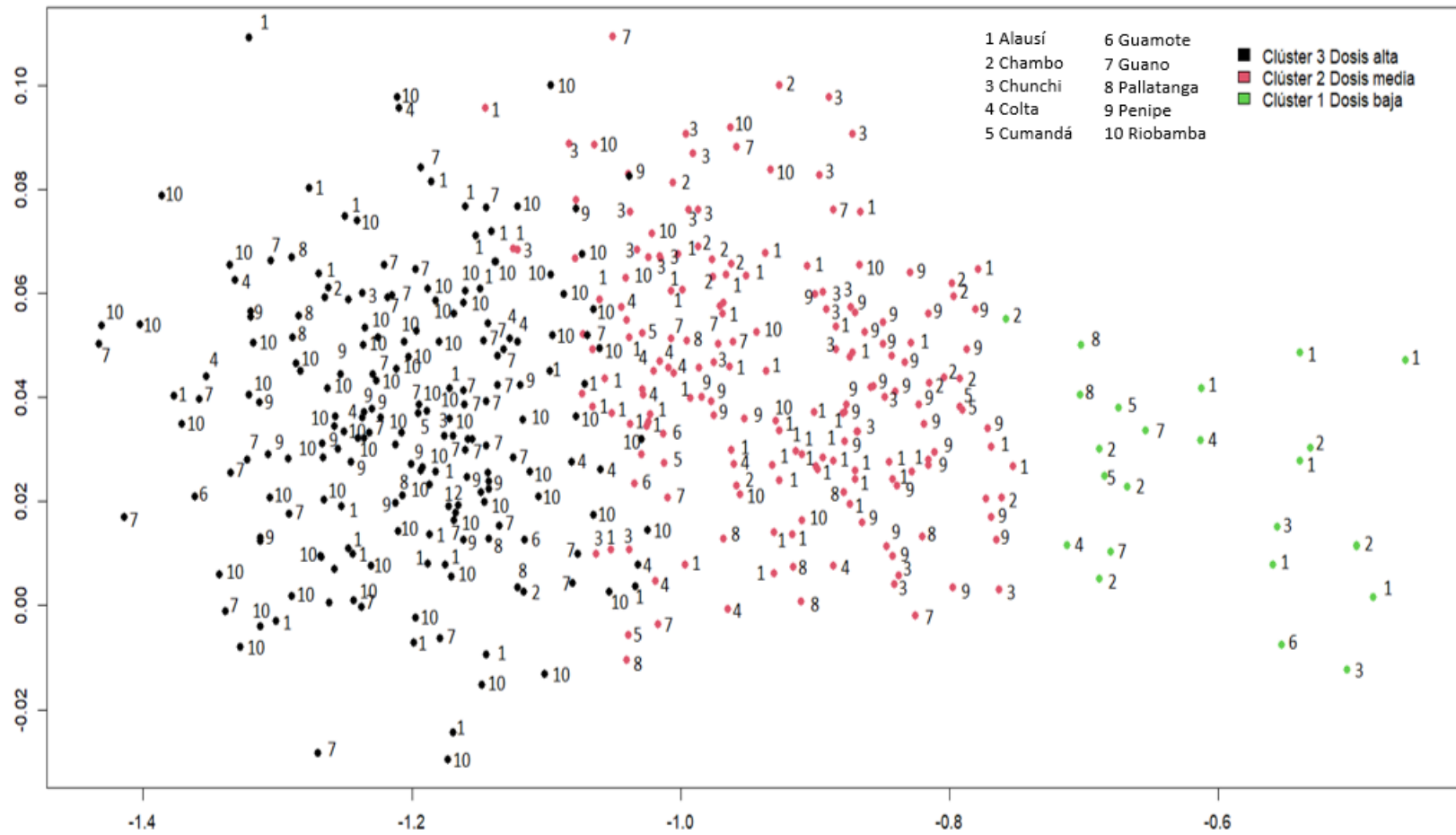
En la ilustración 70-3 se realizó con  $k = 3$  clusters. Se observó como los clusters son más heterogéneos entre ellos. Se hace evidente que es más coherente llevar a cabo el estudio del comportamiento de las observaciones agrupándolas en tres clusters solamente, ya que quedan mejor explicadas.



**Ilustración 71-3:** Número óptimo de Clúster  $k=3$

**Realizado por:** Yungaicela, Víctor, 2022.

En la ilustración 71-3, se representa el mapa clúster con 3 grupos de asociación. En clúster 1 se encuentran 22 curvas funcionales, el 2 contiene 107, y, por último, el 3, con una gran mayoría de 257 observaciones.



**Ilustración 72-3:** Mapa clúster para la provincia de Chimborazo

Realizado por: Yungaicela, Víctor, 2022.

### 3.9.2. Composición de los clusters

A continuación, la tabla 13-3 muestra la composición de cada clúster, con el fin de identificar si los clusters discriminan teniendo en cuenta algún aspecto de las observaciones.

**Tabla 13-3:** Composición por clusters

Tipificación en el mapa	Cantón	Clúster 1 Verde	Clúster 2 Rojo	Clúster 3 Negro
1	Alausí	14%	29%	19%
2	Chambo	5%	11%	0%
3	Chunchi	9%	15%	6%
4	Colta	0%	7%	11%
5	Cumandá	27%	1%	2%
6	Guamote	9%	2%	0%
7	Guano	9%	7%	17%
8	Pallatanga	0%	3%	3%
9	Penipe	23%	12%	9%
10	Riobamba	5%	12%	33%

Realizado por: Yungaicela, Víctor, 2022.

Como muestra la tabla 13-3 se formaron tres clusters. El primero (verde) se encuentra en un intervalo de rayos gama de entre 0,03 uSv/h y 0,04 uSv/h, está formado por el 6% del total de puntos muestreados, de los cuales Cumandá aporta con un 27%, Penipe con el 23%, Alausí con el 14%, y el 37% restante los demás cantones. El segundo grupo (Rojo) está entre 0,04 uSv/h y 0,06 uSv/h, lo conforman el 28% de los cuáles Alausí, Chunchi y Chambo aportan el 29%, 11% y 15% respectivamente. Penipe y Riobamba aportan un 12% cada uno. Y el Tercero y último grupo (negro) contiene el 67% del total de puntos muestreados de los cuales los únicos que no aportan a este son Chambo y Guamote. Riobamba, Guano, Colta y Alausí aportan el 80% del clúster, su intervalo esta entre está 0,06 uSv/h a 0,07 uSv/h.



**Tabla 14-3:** Composición por cantones

Tipificación en el mapa	Cantón	Clúster 1 Verde	Clúster 2 Rojo	Clúster 3 Negro
1	Alausí	4%	38%	59%
2	Chambo	8%	92%	0%
3	Chunchi	6%	47%	47%
4	Colta	0%	22%	78%
5	Cumandá	46%	8%	46%
6	Guamote	50%	50%	0%
7	Guano	4%	15%	81%
8	Pallatanga	0%	27%	73%
9	Penipe	12%	32%	56%
10	Riobamba	1%	13%	86%

Realizado por: Yungaicela, Víctor, 2022.

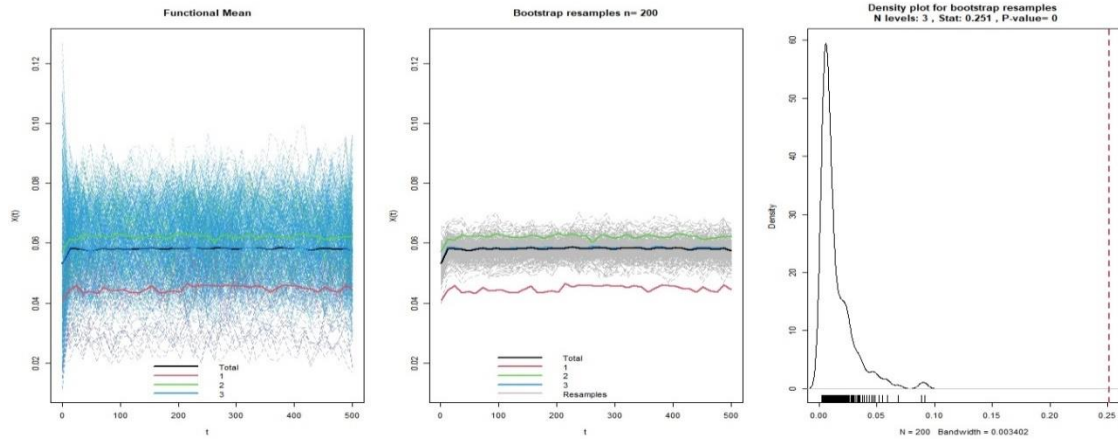
En la tabla 14-3 se muestra la composición por cantones. Alausí se comporta en un 97% en el clúster 2 y 3. Chambo se compone en un 92% en el grupo 2. Chunchi, en un 94%, en el conglomerado 2 y 3. Colta no presenta ninguna similitud con el grupo 1, y un 78% en el clúster 3. Cumandá se divide en el grupo 1 y 3 con el 92%. Guamote no presenta ninguna similitud con el conglomerado 3, lo contrario de Guano, Pallatanga, Penipe, y Riobamba, que su mayoría se presenta en el grupo 3, con el 81%, 73%, 56%, y 86% respectivamente. La tabla 11-3, sirve para una representación de lo ya antes dicho, pero de una manera gráfica, para facilitar al lector.

**Tabla 15-3:** Composición de clusters por color

Cantones	Clúster 1	Clúster 2	Clúster 3
Alausí			
Chambo			
Chunchi			
Colta			
Cumandá			
Guamote			
Guano			
Pallatanga			
Penipe			
Riobamba			

Realizado por: Yungaicela, Víctor, 2022.

### 3.9. ANOVAF entre clústers

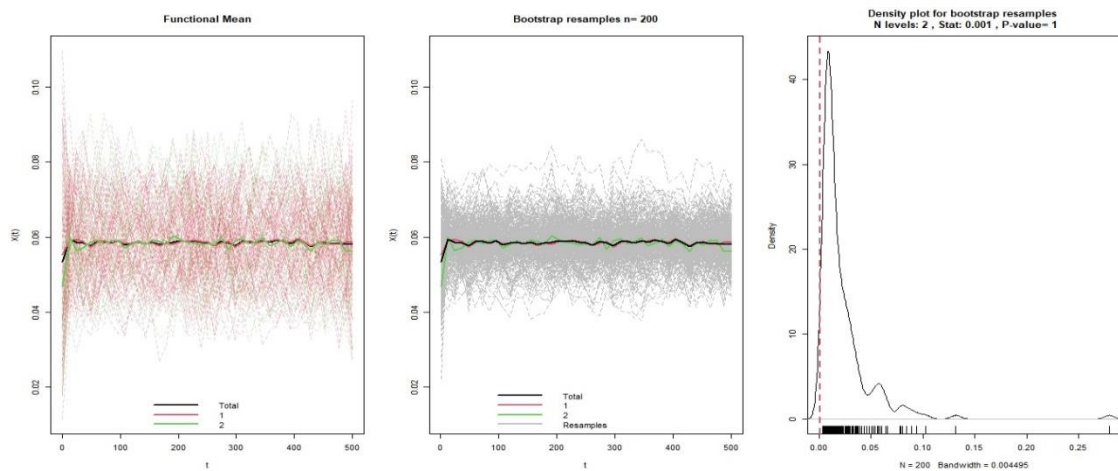


**Ilustración 73-3:** ANOVAF entre clústers

**Realizado por:** Yungaicela, Víctor, 2022.

En la ilustración 73-3, se muestra el ANOVA funcional entre los 3 clústers, este análisis mostro características similares en los valores de tasa de dosis de radiación gamma, es decir que existe evidencia suficiente para rechazar la hipótesis nula igual de la media con un nivel de confianza del 98%, por lo tanto, la diferencia entre medias del valor de tasa de dosis de radiación gamma de los clusters respectivos, es significativa. Se observa que el primer clúster presento un comportamiento distintivo con respecto a los otros dos, la principal causa es la altura con respecto al nivel del mar, del cantón Cumandá. Su localización es la más cercana posible a la costa ecuatoriana, y es el clúster con menos tasa de dosis de radiación gamma.

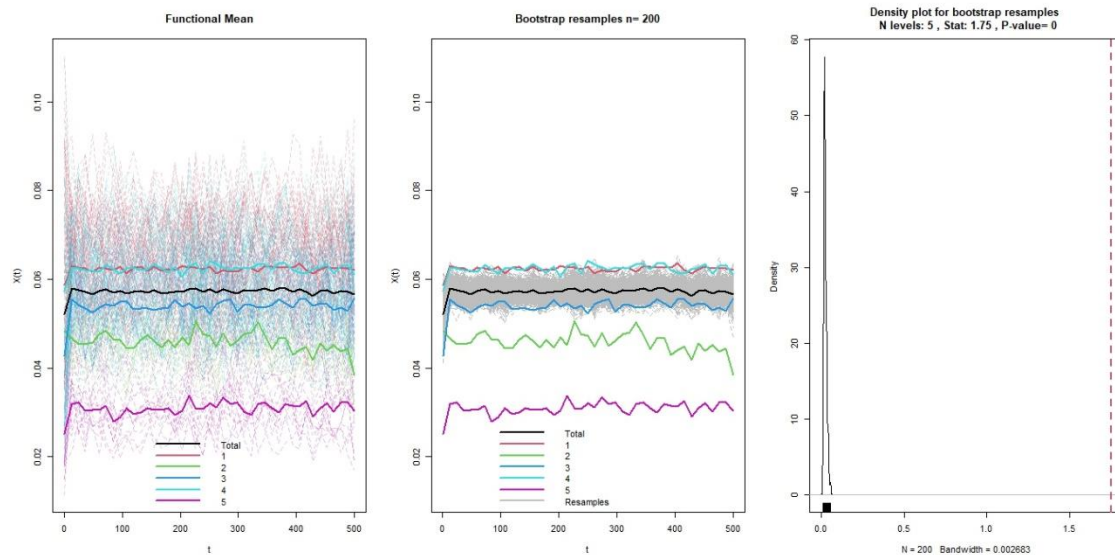
### 3.10. ANOVAF de los clusters



**Ilustración 74-3:** ANOVAF del clúster 1

**Realizado por:** Yungaicela, Víctor, 2022.

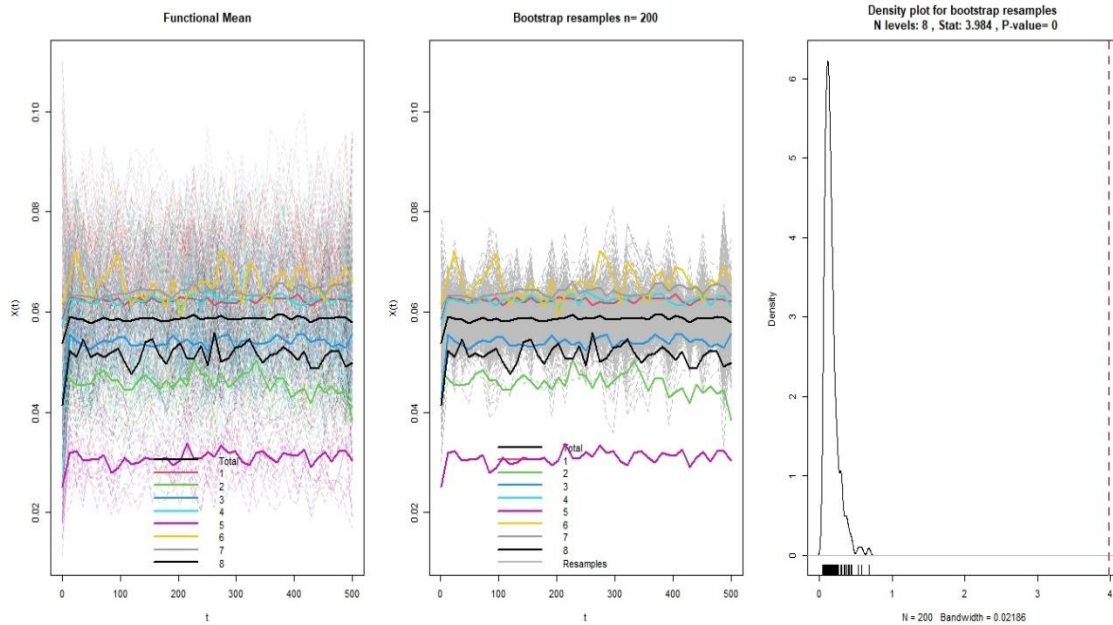
En la ilustración 74-3, se muestra el ANOVA funcional entre los cantones Cumandá y Guamote pertenecientes al clúster 1, este análisis mostro características similares en los valores de tasa de dosis de radiación gamma, es decir que existe evidencia suficiente para rechazar la hipótesis nula igual de la media con un nivel de confianza del 98%, por lo tanto, la diferencia entre medias del valor de tasa de dosis de radiación gamma de los cantones respectivos, es significativa.



**Ilustración 75-3:** ANOVA F del clúster 2

**Realizado por:** Yungaicela, Víctor, 2022.

En la ilustración 75-3, se muestra el ANOVA funcional entre los cantones Alausí, Chambo, Chunchi, Guamote y Penipe, pertenecientes al clúster 2, este análisis mostró características similares en los valores de tasa de dosis de radiación gamma, es decir que existe evidencia suficiente para rechazar la hipótesis nula igual de la media con un nivel de confianza del 98%, por lo tanto, la diferencia entre medias del valor de tasa de dosis de radiación gamma de los cantones respectivos, es significativa.



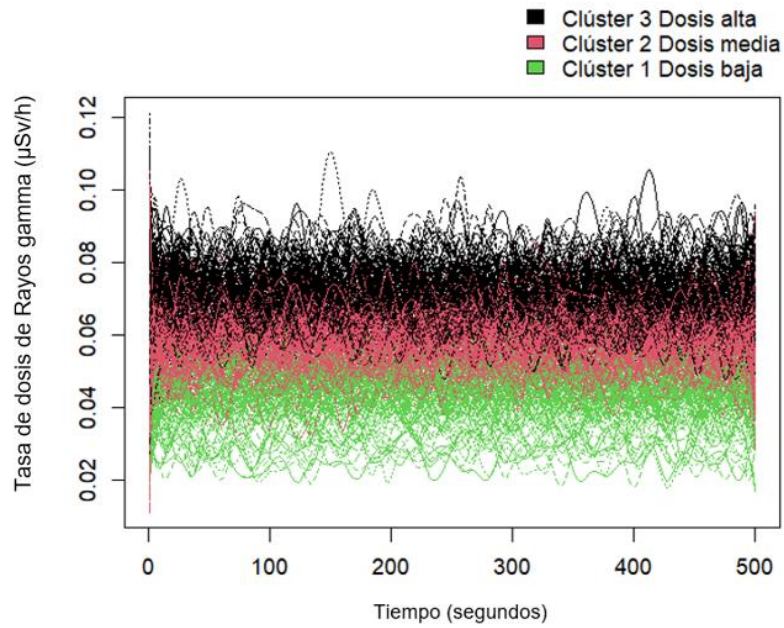
**Ilustración 76-3:** ANOVA del clúster 3

**Realizado por:** Yungaicela, Víctor, 2022.

En la ilustración 76-3, se muestra el ANOVA funcional entre los cantones Alausí, Colta, Chunchi, Cumandá, Guano, Pallatanga, Penipe y Riobamba, pertenecientes al clúster 3, este análisis mostro características similares en los valores de tasa de dosis de radiación gamma, es decir que existe evidencia suficiente para rechazar la hipótesis nula igual de la media con un nivel de confianza del 98%, por lo tanto, la diferencia entre medias del valor de tasa de dosis de radiación gamma de los cantones respectivos, es significativa.

### 3.11. Caracterización de los clusters

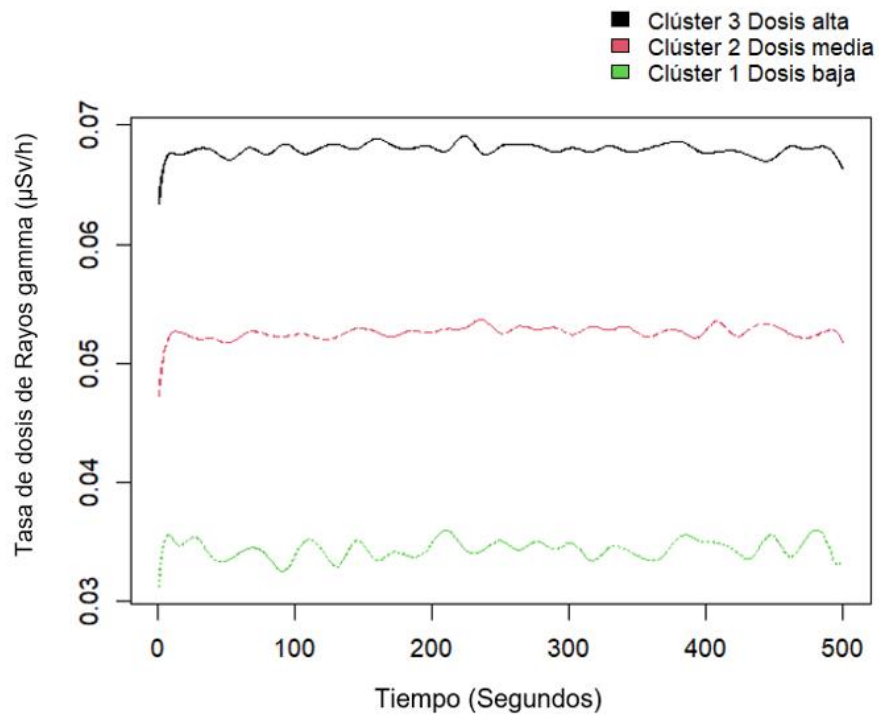
A través de la función FUNHDDC, del paquete fda.usc se obtuvo el Gráfico 74-3. Para su elaboración se necesitó encontrar el mejor modelo. Para ello en el parámetro *model*, se eligió la opción “All”, que tiene por objetivo buscar el modelo óptimo al cual se ajuste nuestros datos funcionales. El modelo seleccionado fue el “ABQKDK”,  $k = 3$ .



**Ilustración 77-3:** Caracterización de los clusters

Realizado por: Yungaicela, Víctor, 2022.

La ilustración 77-3, representa cada curva mediana funcional de la cada agrupación dividida en los 3 clústeres, detallando así su tipo de comportamiento en general.



**Ilustración 78-3:** La media funcional de cada clúster

Realizado por: Yungaicela, Víctor, 2022.

### 3.11.1. Clúster en función de la altitud del cantón

La ilustración 78-3, se puede visualizar la influencia de la variable Altitud, para observar la presencia de alguna relación al momento de la agrupación de los clusters. Se realizó un promedio de las alturas de los puntos de muestreo dentro de cada cantón. La tabla 12-3 detalla el proceso de codificación.

**Tabla 16-3:** Composición de clusters

CANTÓN	ALTITUD	CATEGORIZACIÓN
Guamote	4220	Alta
Colta	3212	Alta
Chambo	2780	Media
Riobamba	2754	Media
Chunchi	2548	Media
Alausí	2340	Media
Penipe	2335	Media
Guano	2000	Media
Cumandá	1600	Baja
Pallatanga	1520	Baja

Realizado por: Yungaicela, Víctor, 2022.

**Tabla 17-3:** Composición por altitud

Tipificación en el mapa	Altitud	Clúster 1 Verde	Clúster 2 Rojo	Clúster 3 Negro
1	Alta	5%	25%	70%
2	Media	4%	29%	67%
3	Baja	24%	20%	56%

Realizado por: Yungaicela, Víctor, 2022.

Los cantones que se encuentran a una altitud alta, en su mayoría pertenecen al clúster 2 y 3, con el 95%. La altitud media, presenta un comportamiento similar a la altitud anterior, con un 96%. Sin embargo, cuando se trata de la categoría baja, se observa un mayor porcentaje con el grupo 1, 24%, pero su agrupación mayoritaria sigue siendo la número 3, con el 56%.

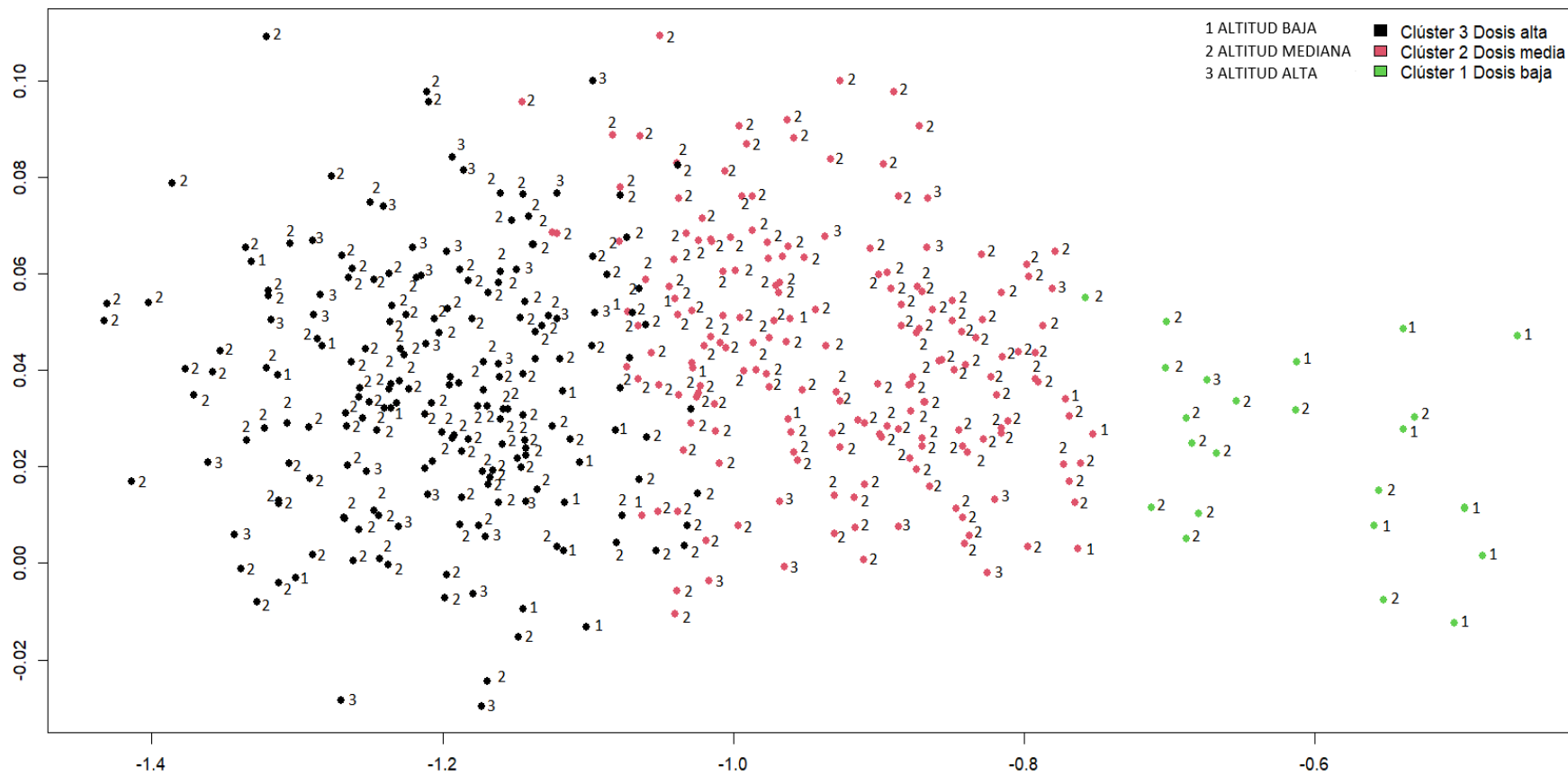
**Tabla 18-3:** Composición de clusters k=3

Tipificación en el mapa	Altitud	Clúster 1 Verde	Clúster 2 Rojo	Clúster 3 Negro
-------------------------	---------	-----------------	----------------	-----------------

1	Alta	9%	16%	11%
2	Media	64%	79%	84%
3	Baja	27%	4%	5%

**Realizado por:** Yungaicela, Víctor, 2022.

Como muestra la tabla 18-3 se formaron tres clústeres en relación con la altitud de los cantones. El primero (verde) se encuentra conformado por la altitud baja y media con un 91%, la altitud alta posee un porcentaje mínimo. El segundo grupo (rojo) se compone en su mayoría por la altitud media y alta con un 95%. Y el tercer y último grupo, se constituye en su mayoría por la altitud media con una 84%.



**Ilustración 79-3:** Mapa Clúster de la altitud de los cantones

Realizado por: Yungaicela, Víctor, 2022.



## CONCLUSIONES

- Se estructuró la base de datos, sin ningún dato faltante, y la ejecución del análisis con 500 segundos es factible.
- Se realizó el suavizado de las curvas con diferentes números de bases (15,17,19,21,25,27,35,41) y el mejor ajuste es el de 43 bases, dando un valor de RMS de 0.007895, el de más cercanía a cero comparado con el resto) en concordancia con el software estadístico R.
- Se detectaron los outliers para excluirlos del estudio, se hallaron las curvas atípicas a través de cuatro métodos (HDR plot, Rainbow plot, Functional bagplot, deph.trim.out), para luego proceder a su comparación y separarlas dependiendo de su mayor frecuencia, se evidenció que Guamote no presenta ningún outliers, mientras Penipe y Riobamba presentan el mayor número de la provincia
- Del análisis descriptivo de datos funcionales se observó que la tasa de dosis de radiación gamma de los 10 cantones de la provincia de Chimborazo, durante el año 2021, Pallatanga, Penipe, Chambo, y Chunchi, poseen una curva media en el intervalo 0,04 uSv/h a 0,06 uSv/h, Cumandá en el intervalo entre 0,02 uSv/h y 0,04 uSv/h, Colta, Alausí, Guamote, Guano, y Riobamba en el intervalo de 0,06 uSv/h a 0,07 uSv/h, por lo que se concluye que Cumandá presenta el intervalo más bajo de la provincia. La desviación típica introduce una idea clara del comportamiento de la tasa de dosis de radiación gamma, identificando poca variabilidad en todos los cantones. Todos poseen una baja variabilidad, a excepción de Guamote.
- La desviación típica introduce una idea clara del comportamiento de la tasa de dosis de radiación gamma, identificando poca variabilidad en todos los cantones. Todos poseen una baja variabilidad, a excepción de Guamote.
- El análisis de clusters, mediante la función *fun.fem()*, tiene como objetivo identificar características similares de cada cantón de la provincia. Se obtuvo 3 grupos con patrones de comportamientos similares. La principal característica que hace que un grupo difiera de otro, es el rango de absorción de la tasa de dosis este tipo de radiación. Se observó que los niveles más altos, entre 0,06 uSv/h a 0,08 uSv/h, se encuentran al noroeste de la provincia de Chimborazo, con los cantones Alausí, Guano, Colta, Chunchi y Riobamba, con el 86% de la composición del grupo. Los niveles medios, entre un intervalo de 0,04 uSv/h a 0,06 uSv/h, se presentan en los cantones de Alausí, Chunchi, Chambo, Riobamba y Penipe, con un 79%, también comparten un comportamiento parecido con el grupo 3. Los niveles bajos, entre 0,02 uSv/h a 0,04 uSv/h, se presentaron en los cantones de Cumandá, Alausí y Penipe, con el 64% de su constitución. Se hace hincapié que en los cantones de grandes poblaciones existe una mayor tasa de dosis de radiación. También se decidió analizar la altitud de cada cantón, que dio como resultado que la mayoría de curvas funcionales se sitúan en una altitud media. El

comportamiento del clúster 1, minimiza las alturas altas (9%), dándonos a entender que los cantones que más bajo están respecto al nivel del mar, no están sujetos a la dosis mayor. El resto de cantones en altitud media y alta, reducen la altitud baja, su predominio se centra en cantones de altitud media.

- Se realizó un ANOVA funcional intra clusters. En el cual se comparó los cantones que componen cada uno. Se presento un valor de p de 0.004, 0.002 y 0.02 respectivamente, rechazando la hipótesis nula en todos los casos. A su vez, en el ANOVA funcional entre clusters, se mostró un valor de p de 0.003 y un nivel de confianza del 98%, se mostró características similares entre clusters, a excepción del primero que varía su comportamiento por un factor principal como lo es la altura respecto al mar.

## **RECOMENDACIONES**

- El GIDAC siga brindando apoyo en proyectos relacionados en bien de la ciudadanía chimboracense.
- Tomar en cuenta los diferentes tipos de muestreo con el fin de aplicar mejoras en la recolección de muestras para los próximos estudios, eso depende de cierta medida en la generación de outliers, lo que conlleva a dar interpretaciones erróneas de los resultados.
- Motivar a los estudiantes a utilizar nuevas técnicas estadísticas, como el análisis funcional, ya que tiene herramientas muy eficaces para el procesamiento y velocidad de cómputo, y más cuando se habla de grandes cantidades de datos.
- Aplicar análisis de clusters para entender mejor el comportamiento de cualquier fenómeno, debido a que sus métodos gráficos permiten un mejor entendimiento al lector.

## BIBLIOGRAFÍA

**AREF, L.** "Nuclear Energy: the Good, the Bad, and the Debatable". Revista Niehs [en línea], [Consulta: 20 de marzo 2022.] Disponible en: [https://www.niehs.nih.gov/health/assets/docs\\_f\\_o/nuclear\\_energy\\_the\\_good\\_the\\_bad\\_and\\_the\\_debatable\\_508.pdf](https://www.niehs.nih.gov/health/assets/docs_f_o/nuclear_energy_the_good_the_bad_and_the_debatable_508.pdf).

**ANON, G.** "Investigación de emisión de radiación gamma en el rango de 1 TEV de la variable cataclísmica AE AQUARII". , 1995, [En línea] Disponible en: [https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis\\_n2808\\_Colombo.pdf](https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis_n2808_Colombo.pdf)

**AREF, L.** *Nuclear Energy: the Good, the Bad, and the Debatable* [blog]. [Consulta: 20 de marzo 2022]. Disponible en: [https://www.niehs.nih.gov/health/assets/docs\\_f\\_o/nuclear\\_energy\\_the\\_good\\_the\\_bad\\_and\\_the\\_debatable\\_508.pdf](https://www.niehs.nih.gov/health/assets/docs_f_o/nuclear_energy_the_good_the_bad_and_the_debatable_508.pdf)

**BARRERA, M. y VALIÑO, F.** *Puesta a punto de un sistema de espectrometría gamma para la determinación de Cs-137 en suelos españoles* [blog]. [Consulta: 20 de marzo 2022] Disponible en: [https://inis.iaea.org/collection/NCLCollectionStore/\\_Public/40/018/40018273.pdf](https://inis.iaea.org/collection/NCLCollectionStore/_Public/40/018/40018273.pdf)

**BOOR, C.** "A practical guide to splines". Revista Springer-Verlag New York: Springer.

**CARRASCO, J.** *Radiaciones ionizantes y no ionizantes* [blog]. [Consulta: 08 de marzo 2022]. Disponible en: [https://www.academia.edu/30308142/RADIACIONES\\_IONIZANTES\\_Y\\_NO\\_IONIZANTES\\_APLICACIONES\\_Y\\_RIESGOS](https://www.academia.edu/30308142/RADIACIONES_IONIZANTES_Y_NO_IONIZANTES_APLICACIONES_Y_RIESGOS)

**CASTILLO, S. y FLORES, M.** "Análisis de Datos Funcionales Aplicado al Riesgo Financiero: Un Caso de Estudio en Cooperativas de Ahorro y Crédito Ecuatorianas". Revista Latin American Journal of Computing [en línea], 2015, (Ecuador) 15(1), pp. 19-25.

**CHECA, M.** "Análisis geoestadístico de datos funcionales de temperatura del aire en la provincia de Chimborazo". Epoch [en línea]. [Consulta: 15 de marzo 2022]. Disponible en: <http://dspace.epoch.edu.ec/handle/123456789/14280>

**COLOMBO, E.** "Investigación de emisión de radiación gamma en el rango de 1 TEV de la variable cataclísmica AE AQUARII". Biblioteca digital [En línea], [Consulta: 5 de marzo 2022].

Disponible en: [https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis\\_n2808\\_Colombo.pdf](https://bibliotecadigital.exactas.uba.ar/download/tesis/tesis_n2808_Colombo.pdf)

**DELAHAY, P.** *Análisis instrumental*. 3ª ed. México: Paraninfo, 1970, pp. 12-100

**DELAIGLE, A.** *Clustering functional data into groups using projections*. [blog]. EEUU, 2016. [Consulta: 12 de marzo 2022]. Disponible en: <https://researchers.ms.unimelb.edu.au/~aurored/delaigle-hall-pham.pdf>

**DÍAZ, J.** *Desarrollo de un código para el transporte de fotones ionizantes en agua con la técnica evento-por-evento*. [blog], [Consulta: 11 de marzo 2022], 2017. Disponible en: <https://ri.uaemex.mx/bitstream/handle/20.500.11799/69412/Tesis%20JorgeDiaz.pdf?sequence=1&isAllowed=y>

**DIEHL, R.** *Gamma-Ray Production and Absorption*. [blog], [Consulta: 08 de marzo 2022], Disponible en: [http://integral.esac.esa.int/gamma\\_ray\\_astronomy.pdf](http://integral.esac.esa.int/gamma_ray_astronomy.pdf).

**ESCUADERO, A. Y HARO, S.** "Spline Cúbico para el Tratamiento Funcional de la Radiación Solar Global". *Revista Scielo* [en línea], 2020, (Ecuador) 2(1), ISSN 27(10.4067/S0718-07642016000200018), pp. 153-162.

**ESTRUCPLAN.** *Estrucplan* [blog], [Consulta: 05 de marzo 2022]. Disponible en: <https://estrucplan.com.ar/la-radiacion-gamma-y-sus-aplicaciones/>

**EVERITT, B. y LANDAU, S.** *Cluster Analysis*. 5ª ed. London: Wiley, 2011

**FERRATY, F. Y VIEU, P.** *Nonparametric Functional Data Analysis*. 1ª ed. New York: Springer, 2006

**FUSTER, M.** 2021. "Aplicación del análisis estadístico de datos funcionales al estudio del patrón de movimiento del raquis lumbar y de la pelvis durante la flexión y la extensión del tronco en personas sanas". [En línea], [Consulta: 08 de marzo 2022], Disponible en: <https://roderic.uv.es/bitstream/handle/10550/79981/Tesis%20completa%20definitiva%2026-06-21.pdf?sequence=1>

**GALARZA, C. Y VERA, F.** "Análisis de varianza funcional para medir el efecto de recubrimientos revestibles de almidón en características físico-químicas de papayas durante las dos primeras semanas del período de maduración post-cosecha". [En línea], [Consulta: 15 de marzo 2022]. Disponible en: <https://www.dspace.espol.edu.ec/handle/123456789/24485>.

**GARCÍA, A. y GORDALIZA, A.** "A Proposal for Robust Curve Clustering". [En línea], [Consulta: 10 marzo 2022], 2005, Disponible en: <https://doi.org/10.1007/s00357-005-0013-8>

**GARZÓN, D. y ROA, M.** "Factores que influyen en el crecimiento endocondral: experimentos y modelos". [En línea], [Consulta: 08 de marzo 2022], 2008, Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-215X2008000100006](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-215X2008000100006)

**GIRALDO, R.** "Análisis exploratorio de variables regionalizadas". [En línea], 2007, [Consulta: 07 de marzo 2022]. Disponible en: <https://www.redalyc.org/pdf/899/89930108.pdf>

**GIRALDO, R. y MATEU, J.** "Continuous Time-Varying Kriging for Spatial Prediction of Functional Data: An Environmental Application". [En línea], 2010, [Consulta: 07 de marzo 2022]. Disponible en: [https://www.researchgate.net/publication/33422356\\_Continuous\\_Time-Varying\\_Kriging\\_for\\_Spatial\\_Prediction\\_of\\_Functional\\_Data\\_An\\_Environmental\\_Application](https://www.researchgate.net/publication/33422356_Continuous_Time-Varying_Kriging_for_Spatial_Prediction_of_Functional_Data_An_Environmental_Application)

**IBAÑEZ, J.** "Análisis cluster aplicado al ambito agronomo para estudiar el comportamiento de nuevos tipos de compost". [En línea], 2018, [Consulta: 10 de marzo 2022]. Disponible en: [http://dspace.umh.es/bitstream/11000/6360/1/EST\\_TFG\\_Iba%C3%B1ez\\_Perea\\_Julio\\_Alberto.pdf](http://dspace.umh.es/bitstream/11000/6360/1/EST_TFG_Iba%C3%B1ez_Perea_Julio_Alberto.pdf)

**INEC.** "Instituto Nacional de Estadística y Censos". [En línea], 2010, [Consulta: 09 de marzo 2022]. Disponible en: [www.ecuadorencifras.gob.ec](http://www.ecuadorencifras.gob.ec).

**KOKOSZKA, P. y REIMHERR, M.** *Introduction to Functional Data Analysis*. 1ª ed. Boca Raton: Chapman and Hall/CRC.

**MENESES, A. y ZUÑIGA, L.** "Metodo Bootstrap propuesto para hipótesis concernientes a la diferencia de medias en variables independientes". [En línea], 2016, (Ecuador), [Consulta: 08 de marzo 2022]. Disponible en: <http://ceaa.esPOCH.edu.ec:8080/revista.perfiles/faces/Articulos/Perfiles17Art8.pdf>

**MILLÁN, L.** "Outliers de datos funcionales para la detección de caudales anómalos en el sector hidráulico". [En línea], 2017, [Consulta: 07 de marzo 2022]. Disponible en: [http://repositori.uji.es/xmlui/bitstream/handle/10234/174477/TFM\\_2017\\_Millan\\_Roures\\_Laura.pdf?sequence=1](http://repositori.uji.es/xmlui/bitstream/handle/10234/174477/TFM_2017_Millan_Roures_Laura.pdf?sequence=1)

**MORALES, A.** "Empleo de la técnica de coincidencias en experimentos nucleares". [En línea], 1973, [Consulta: 11 de marzo 2022], Disponible en: [https://inis.iaea.org/collection/NCLCollectionStore/\\_Public/05/134/5134123.pdf](https://inis.iaea.org/collection/NCLCollectionStore/_Public/05/134/5134123.pdf)

**MORILLO, M.** *Estimación penalizada con datos funcionales* [blog], 2009, [Consulta: 09 de marzo 2022]. Disponible en: <https://masteres.ugr.es/moea/pages/tfm0809/estimacin-penalizada-con-datos-funcionales>

**NUÑEZ, M.** *Efectos Biológicos de las Radiaciones - Dosimetría*. [blog], 2008, [Consulta: 10 de marzo 2022]. Disponible en: [https://www.alasbimn.net/comites/tecnologos/material/Efectos\\_biologicos\\_de\\_las\\_radiaciones.pdf](https://www.alasbimn.net/comites/tecnologos/material/Efectos_biologicos_de_las_radiaciones.pdf)

**NURILMALA, F. y PEMIMPIN, R.** *Somaclonal variation induction of Bogor taro (Colocasia esculenta) by gamma irradiation*. [blog], 2016, [Consulta: 08 de marzo 2022]. Disponible en: [https://www.researchgate.net/publication/312008263\\_Somaclonal\\_variation\\_induction\\_of\\_Bogor\\_taro\\_Colocasia\\_esculenta\\_by\\_gamma\\_irradiation](https://www.researchgate.net/publication/312008263_Somaclonal_variation_induction_of_Bogor_taro_Colocasia_esculenta_by_gamma_irradiation)

**PEÑAFIEL, P.** "Análisis estadístico funcional de la reflectancia obtenida mediante imágenes satelitales y espectroradiómetro del cultivo de quinua en Colta". [En línea], 2021, [Consulta: 07 de marzo 2022]. Disponible en: <http://dspace.esPOCH.edu.ec/handle/123456789/14809>

**PLAZOLA, R.** *Monitoreo de datos funcionales*. [blog], 2013, [Consulta: 10 de marzo 2022]. Disponible: [https://ciimat.repositorioinstitucional.mx/jspui/bitstream/1008/463/1/TE\\_1518.pdf](https://ciimat.repositorioinstitucional.mx/jspui/bitstream/1008/463/1/TE_1518.pdf)

**RAMSAY, O. y SILVERMAN, W.** *Functional Data Analysis*. 2005. 2ª ed. Segunda ed. New York: Springer.

**RANGEL, U.** *Aplicación de la irradiación gamma* [blog], 2016, [Consulta: 06 de marzo 2022]. Disponible en: [http://www.inin.gob.mx/Transparencia/irradiacion\\_gamma.cfm](http://www.inin.gob.mx/Transparencia/irradiacion_gamma.cfm)

**RAYA, I.** *Análisis de un espectro de rayos gamma utilizando detectores de germanio de alta pureza.* [blog], 2021, [Consulta: 09 de marzo 2022]. Disponible en: [http://bibliotecavirtual.dgb.umich.mx:8083/xmlui/bitstream/handle/DGB\\_UMICH/6369/FISMA-T-M-2021-1449.pdf?sequence=1](http://bibliotecavirtual.dgb.umich.mx:8083/xmlui/bitstream/handle/DGB_UMICH/6369/FISMA-T-M-2021-1449.pdf?sequence=1)

**RODRÍGUEZ, R. y LOZANO, A.** *Meteorología y Climatología.* 1ª ed. Madrid: Fundación Española para la Ciencia y la Tecnología, 2004

**ROJAS, K.** *Radiación Electromagnética.* [blog], 2009, [Consulta: 08 de marzo 2022]. Disponible en: [https://repository.upb.edu.co/bitstream/handle/20.500.11912/432/digital\\_17516.pdf?sequence=1](https://repository.upb.edu.co/bitstream/handle/20.500.11912/432/digital_17516.pdf?sequence=1)

**SALCEDO, C.** *Diseño de un detector de rayos x y rayos gamma para aplicaciones de radiografía industrial.* [blog], 2013, [Consulta: 06 de marzo 2022]. Disponible en: [https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/5307/SALCEDO\\_CARLOS\\_DISE%c3%91O\\_DETECTOR\\_RAYOS\\_X\\_GAMMA\\_APLICACIONES\\_RADIOGRAFIA\\_INDUSTRIAL.pdf?sequence=1&isAllowed=y](https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/5307/SALCEDO_CARLOS_DISE%c3%91O_DETECTOR_RAYOS_X_GAMMA_APLICACIONES_RADIOGRAFIA_INDUSTRIAL.pdf?sequence=1&isAllowed=y)

**SRIHARSHA, K.** *Study of gamma exposure rate in Mysore and Chamaraj Nagar district, Karnataka, India.* [blog], [Consulta: 12 de marzo 2022]. Disponible en: <http://ijrr.com/article-1-404-en.pdf> [Último acceso: 12 03 2022].

**STALTER, R. y HOWARTH, D.** "Gamma Radiation". [En línea], 2012, [Consulta: 08 de marzo 2022]. Disponible en: [https://www.researchgate.net/publication/221929415\\_Gamma\\_Radiation](https://www.researchgate.net/publication/221929415_Gamma_Radiation)

**TAGUACUNDO, O.**, 2019. Componentes principales funcionales de la radiación solar global de la provincia de Chimborazo, Dspace [en línea]. (Trabajo de titulación). (Ingeniería) Escuela Superior Politécnica de Chimborazo, Riobamba, Ecuador. 2012. pp. 120-153. [Consulta: 08 de marzo 2022]. Disponible en: <http://dspace.esPOCH.edu.ec/handle/123456789/13600>

**UYTTENHOVE, J.** *Measurement of natural gamma radiation in Belgium by means of high resolution in-situ spectrometry.* [blog], 2006, [Consulta: 06 de marzo 2022]. Disponible en: <https://www.irpa.net/irpa10/cdrom/00284.pdf>



**VENEGAS, G. y RÍOS, W.** *Método de Cluster Jerárquicos para Datos Funcionales Multivariados: Una Aplicación en Mercadeo.* [blog], 2017, [Consulta: 10 de marzo 2022].

Disponible

en:

<https://repository.usta.edu.co/jspui/bitstream/11634/12525/1/2018gustavovenegas.pdf>



## ANEXOS

### ANEXO A: CODIGO EN R

```
#####  
## DATOS FUNCIONALES DOSIS DE GAMMA ##  
#####  
## LIBRERIAS NECESARIAS ##  
library(readr)  
library(fda)  
  
## Loading required package: splines  
  
## Loading required package: Matrix  
  
## Loading required package: fds  
  
## Loading required package: rainbow  
  
## Loading required package: MASS  
  
## Loading required package: pcaPP  
  
## Loading required package: RCurl  
  
## Loading required package: deSolve  
  
##  
## Attaching package: 'fda'  
  
  
## Loading required package: nlme  
  
## This is mgcv 1.8-38. For overview type 'help("mgcv-package")'.  
  
## -----  
-----  
  
## Functional Data Analysis and Utilities for Statistical Computing  
## fda.usc version 2.0.2 (built on 2020-02-17) is now loaded  
## fda.usc is running sequentially usign foreach package
```

```

## Please, execute ops.fda.usc() once to run in local parallel mode
## Deprecated functions: min.basis, min.np, anova.hetero, anova.onefactor, anova.RPm
## New functions: optim.basis, optim.np, fanova.hetero, fanova.onefactor, fanova.RPm
## -----
-----

library(rainbow)
library(fdANOVA)
library(readxl)

data <- read_excel("C:/Users/59398/Documents/rgr/datos.xlsx")
datoschimborazo<-data.frame(lapply(data, as.numeric),stringsAsFactors
= FALSE)
FLEXTRK<-as.matrix(datoschimborazo)

## INSTANTES DE TIEMPO ##
Tiempo_segundos<-seq(1,500,1)

## CREACIÓN DE LAS FUNCIONES BSPLINE Y REPRESENTACIÓN ##
basisobjflex<-create.bspline.basis(c(min(Tiempo_segundos),max(Tiempo_segundos)),nbasis = 43)
plot(basisobjflex,main = "")

## OBTENCIÓN DE LOS DATOS FUNCIONALES Y REPRESENTACIÓN ##
## MUESTRA TOTAL ##
gammaChimborazo=Data2fd(Tiempo_segundos,FLEXTRK,basisobjflex)
plot.fd(gammaChimborazo,col="blue",main="",ylab="Radiación Gamma (USv/h)",xlab="Tiempo (segundos)",ylim = c(0.02,0.3))

#####
## CALIDAD DEL AJUSTE ##
#####
## ##
for (i in 1:400) {
  plotfit.fd(argvals=Tiempo_segundos,FLEXTRK[,i],fdoobj=gammaChimborazo[i],lty=1,lwd=1,col=2)
}

```

```

#####
## OUTLIERS ##
#####
## OBTENCIÓN DE ATÍPICOS DE FORMA GRÁFICA ##
## LECTURA DE LOS DATOS PARA LOS GRÁFICOS
datosChimborazo2<-read_excel("C:/Users/59398/Documents/rgr/datos.xlsx"
)
datosChimborazo2<-data.frame(lapply(datosChimborazo2, as.numeric), stringsAsFactors =
FALSE)

## PREPARACIÓN DE LOS DATOS##
datoschimborazofds=fds(Tiempo_segundos,datosChimborazo2)

## RAINBOW ##
plot(datoschimborazofds,plot.type="depth",plotlegend=TRUE,ylab="Radiación Gamma (USv/h)",xlab="Tiempo (segundos)")

## FUNCTIONAL BAGPLOT ##
fboxplot(datoschimborazofds,plot.type="functional",type="bag",ylab="Radiación Gamma (USv/h)",xlab="Tiempo (segundos)")

## FUNCTIONAL HIGHEST DENSITY REGION (HDR) BOXPLOT ##
fboxplot(datoschimborazofds,plot.type="functional",type="hdr",alpha=c(0.07,0.5))

## Warning in fts(x, centercurve): Please assign column name for the data matrix.

#####
## ESTADÍSTICOS DESCRIPTIVOS ##
#####
## CALCULO DE LAS MEDIAS Y DEUSV/HIACIONES TÍPICAS FUNCIONALES ##
## MUESTRA TOTAL ##
mediaaLausi=mean.fd(AFDALAU)

deuSv/htipalausi=sd.fd(AFDALAU)

```

**#Chambo**

**mediachambo=mean.fd(AFDCHAM)**

**deuSv/htipchambo=sd.fd(AFDCHAM)**

**#Chunchi**

**mediachunchi=mean.fd(AFDCHU)**

**deuSv/htipchunchi=sd.fd(AFDCHU)**

**#Colta**

**mediacolta=mean.fd(AFDCOL)**

**deuSv/htipcolta=sd.fd(AFDCOL)**

**#Cumanda**

**mediacumanda=mean.fd(AFDCUM)**

**deuSv/htipcumanda=sd.fd(AFDCUM)**

**#Guamote**

**mediaguamote=mean.fd(AFDGUA)**

**deuSv/htipguamote=sd.fd(AFDGUA)**

**#Guano**

**mediaguano=mean.fd(AFDGUAN)**

**deuSv/htipguano=sd.fd(AFDGUAN)**

**#PALLATANGA**

**mediapallatanga=mean.fd(AFDPAL)**

**deuSv/htippallatanga=sd.fd(AFDPAL)**

**#PENIPE**

```
mediapenipe=mean.fd(AFDPEN)
```

```
deuSv/htipenipe=sd.fd(AFDPEN)
```

```
#RIOBAMBA
```

```
mediariobamba=mean.fd(AFDRIO)
```

```
deuSv/htipriobamba=sd.fd(AFDRIO)
```

```
#
```

```
#####CLUSTER
```

```
library(funFEM)
```

```
res <- funFEM(gammaChimborazo, K=4, model = "DkB", crit = "bic", init  
= "kmeans", Tinit = c(), maxit = 500,
```

```
eps = 1e-06, disp = TRUE, lambda = 0, graph = TRUE)
```

```
# Visualization in the discriminative subspace (projected scores)
```

```
par(mfrow=c(1,1))
```

```
plot(t(gammaChimborazo$coefs) %*% res$U, col=res$cls, pch=19, main="Discr  
iminative space")
```

```
# Visualization in the discriminative subspace (projected scores)
```

```
par(mfrow=c(1,1))
```

```
ggplot(t(gammaChimborazo$coefs) %*% res$U, col=res$cls, pch=12, main="D  
iscriminative space")
```

```
text(t(gammaChimborazo$coefs) %*% res$U)
```

```
res2 = funFEM(gammaChimborazo,K=res$K,model='AkjBk',init='user',Tinit=
res$P,

                Lambda=0.01,disp=TRUE)

Library(funHDDC)

res.uni <-funHDDC(gammaChimborazo, K=3, model="all", threshold=0.2, it
ermax=200, eps=1e-6, init="kmeans",

                criterion="bic", algo='EM', d_select="Cattell", init.vector=TR
UE, show=TRUE,

                mini.nb=c(5, 10), min.individuals=2, mc.cores=1, nb.rep=1, kee
pALLRes=TRUE,

                kmeans.control = list(), d_max=100)

plot(gammaChimborazo,col=res.uni$class)

...

```

## ANEXO B: AVAL



# ESPOCH

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

Riobamba, 18 de octubre de 2021

Ingeniero  
Pablo Flores  
**PRESIDENTE DE LA UNIDAD DE INTEGRACIÓN CURRICULAR CARRERA DE ESTADÍSTICA**  
Presente

De mi consideración:

Reciba un atento saludo, con la finalidad de apoyar y colaborar en la ejecución del proyecto de investigación denominado **“Evaluación de elementos radiactivos de la serie del Uranio 238 en el ambiente en pacientes con cáncer”** del IDI que desarrolla el Grupo de Investigación GIDAC, con Fecha de Inicio: 06 de enero de 2020 y Fecha de Finalización: 31 de diciembre de 2021, se decide dar el AVAL al tesista **Víctor Andrés Yungaicela Sagñay** con C.I. 060470846-1 y Código Estudiantil 303, cuyo tema de tesis se titula **“ANÁLISIS ESTADÍSTICO EXPLORATORIO FUNCIONAL DE LA RADIACIÓN GAMMA EN LA PROVINCIA DE CHIMBORAZO”**.

Por la favorable atención, anticipo mi agradecimiento.

Atentamente,  
**“Saber para Ser”**

JHENY DEL  
CARMEN ORBE  
ORDONEZ

Firmado digitalmente por  
JHENY DEL CARMEN ORBE  
ORDONEZ  
Fecha: 2021.10.18 17:33:05  
+05'00'

Dra. Jheny Orbe Ordoñez  
**DIRECTORA PROYECTO URANIO 235**





**esPOCH**

**Dirección de Bibliotecas y  
Recursos del Aprendizaje**

**UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y  
DOCUMENTAL**

**REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA**

**Fecha de entrega:** 23-01-2023

<b>INFORMACIÓN DEL AUTOR/A (S)</b>
<b>Nombres – Apellidos:</b> Víctor Andres Yungaicela Sagñay
<b>INFORMACIÓN INSTITUCIONAL</b>
<b>Facultad:</b> Ciencias
<b>Carrera:</b> Carrera Estadística
<b>Título a optar:</b> Ingeniero en Estadística Informática
<b>f. Analista de Biblioteca responsable:</b> Ing. Rafael Inty Salto Hidalgo

0122-DRBA-UTP-2023



