

**UNIVERSIDAD DE EL SALVADOR  
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE  
DEPARTAMENTO DE MATEMÁTICA**



**TRABAJO DE GRADUACIÓN DENOMINADO:**  
“APLICACIÓN DE MODELOS ESTADÍSTICOS PARA LA INTERPRETACIÓN DE  
SEÑALES SÍSMICO-VOLCÁNICAS EN LOS PRINCIPALES COMPLEJOS  
VOLCÁNICOS DE EL SALVADOR”

**PARA OPTAR AL GRADO DE:**  
LICENCIADO EN ESTADÍSTICA

**PRESENTADO POR:**  
GUZMÁN MARTÍNEZ, EDWIN ALEXANDER

**DOCENTE DIRECTOR:**  
LICDO. JUAN HAROLDO LINARES MARTÍNEZ

**MARZO, 2016**

**SANTA ANA, EL SALVADOR, CENTROAMÉRICA**

**Universidad de El Salvador**  
**Autoridades Centrales**

Licdo. José Luis Argueta Antillón  
**Rector Interino**

Ing. Carlos Armando Villalta  
**Vice-Rector Administrativo Interino**

Dra. Ana Leticia Zavaleta de Amaya  
**Secretaria General**

Licda. Claudia María Melgar de Zambrana  
**Defensor de los derechos humanos universitarios**

Licda. Nora Beatriz Meléndez  
**Fiscal General Interina**

**Facultad Multidisciplinaria de Occidente**  
**Autoridades**

Ing. Jorge William Ortiz Sánchez  
**Decano Interino**

Licdo. David Alfonso Mata Aldana  
**Secretario Interino de la Facultad**

Licda. Medarda del Rosario Cáceres Aguilar  
**Jefa Interina del Departamento de Matemática**

Esta tesis, “Aplicación de modelos estadísticos para la interpretación de señales sísmico-volcánicas en los principales complejos volcánicos de El Salvador”, es una especialización en Geoestadística con la que se aprueba el cumplimiento parcial de los requerimientos para el grado académico de Licenciado en Estadística.

Departamento: Matemática

Jefa de  
Departamento

---

*Licda. Medarda Rosario Cáceres Aguilar*

Presidenta del  
Jurado Calificador

---

*Licda. Elia Elizabeth Pineda de Flores*

Secretario del  
Jurado Calificador

---

*Licdo. Rodolfo Antonio Olmos Guevara*

Vocal del Jurado  
Calificador

---

*Licdo. Juan Haroldo Linares Martínez*

Fecha y hora

*Viernes 18 de marzo del 2016; 2:00 P.M.*

Nota

*“Todo el conocimiento, la totalidad de preguntas y respuestas se encuentran en el  
perro.”*  
Franz Kafka

*“Hombres que pasan corriendo:  
Sale uno a caminar de noche y divisa desde lejos (porque la calle baja en pendiente ante  
nuestra vista y la luna ilumina con su claridad) a un hombre que viene corriendo hacia  
nosotros, como huyendo. Pero no lo atraparemos, por enclenque y débil que se lo vea,  
aunque alguien lo persigue a los gritos. No, dejaremos que continúe su carrera; es  
posible que estos individuos hayan organizado esta escena para su diversión, o quizá los  
dos están persiguiendo a un tercero que ya pasó, y no es imposible que el perseguido  
sea víctima de alguna injusticia, o el perseguidor un asesino de quién nos haríamos  
cómplices interceptando al fugitivo; o ninguno de los dos tiene conciencia de la  
existencia del otro, y ambos corren por propia cuenta; o que sean sonámbulos, o que el  
que se acerca tenga un arma...”*  
Franz Kafka

## Sumario

Esta tesis es una aplicación de modelos estadísticos para la interpretación de variables sísmico-volcánicas reales de los principales complejos volcánicos de El Salvador. Las bases de datos obtenidas por el trabajo vulcanológico y sismológico que conciernen a instituciones gubernamentales, son reales en lo absoluto; estas bases sirvieron para la modelización de variables de tipo sísmico-volcánicas durante esta investigación. Además incluye propuestas y sugerencias estadísticas innovadoras respecto a la vulcanología y sismología nacional.

*Palabras claves:* modelo, modelo estadístico, interpretación estadística, variable, variable sísmica, variable volcánica, complejos volcánicos, base de datos, vulcanología, sismología, modelización estadística, pronóstico, estadística, Kernel, series de Fourier, elipse, centroides, regresión no paramétrica, regresión Loess, regresión Kernel, interpolación probabilística, VSA, VSM, El Salvador.

## Agradecimientos

\*\*\*\*\*

*A Dios abstracto.*

*A mi amada y difunta madre Celia: Requiéscat In Pace.*

*A mi amada esposa Wendy: “Feliz Cumpleaños”, por si se me olvida algún día.*

*A mi amado bebé Marquitos.*

*A mis excelentes y apreciados suegros: Don Alfredo y Niña Marina.*

*A mis verdaderos amigos y amigas.*

*Agradecimientos especiales por sus altas contribuciones en esta investigación: Licda.*

*Celina Kattán (Directora del Observatorio Ambiental, MARN), Licdo. Eduardo*

*Gutiérrez (Vulcanólogo jefe, MARN), Licda. Rosario Cáceres (Jefa Departamento de Matemática, UES), Licdo. Rodolfo Olmos (Geofísico, UES), Licdo. Jaime Sermeño (Jefe*

*Departamento de Física), Licdo. Haroldo Linares (Estadístico, UES).*

*A mis excelentes y estimados profesores y profesoras matemáticos.*

*¡A la vida y a la naturaleza por su gentileza!*

*Con toda sinceridad y honestidad: ¡gracias!*

\*\*\*\*\*

## Contenido

Sumario .....	6
Agradecimientos .....	7
Índice de tablas .....	11
Índice de figuras .....	12
Presentación .....	14
Capítulo 1: Formulación del problema.....	18
1.1 Planteamiento del problema .....	18
1.2 Justificación .....	19
1.3 Planteamiento de objetivos .....	25
1.4 Planteamiento de las Hipótesis estadísticas .....	25
1.5 Delimitación .....	26
1.5.1 Delimitación espacial .....	26
1.5.2 Delimitación temporal .....	26
1.5.3 Delimitación estadística .....	27
Capítulo 2: Abordaje teórico .....	29
2.1 Enfoque de la investigación vulcanológica internacional .....	29
2.2 Introducción al modelamiento de datos.....	33
2.3 Introducción a la regresión de datos .....	36
2.4 Regresión no paramétrica.....	38
2.4.1 Introducción a la regresión Kernel .....	38
2.4.2 Estimación de la función de densidad univariante .....	39
2.4.3 Estimación de la función de densidad multivariante .....	44
2.4.4 Bandas de variabilidad y bandas de referencia normal .....	45
2.4.5 Estimación de la función de regresión .....	45
2.4.6 Regresión Loess.....	47
2.5 Curvas de Andrews.....	48
2.6 Introducción a las series temporales .....	52
2.7 Simulación MonteCarlo.....	55
2.7.1 Obtención de números aleatorios .....	56

2.7.2 Simulación de variables aleatorias continuas .....	58
2.8 Validación del modelo.....	59
Capítulo 3: Metodología .....	61
3.1 Confiabilidad de los datos .....	61
3.2 Descripción del complejo volcánico VSA.....	61
3.3 Descripción del complejo volcánico VSM .....	62
3.4 Descripción de los complejos volcánicos VIZA, VSS, VILO y VSVI.....	63
3.5 Descripción genérica de las variables sísmicas y volcánicas .....	66
3.6 Recursos .....	67
3.7 Muestra y población .....	68
3.8 Aparatos y dispositivos de la vulcanología y sismología .....	69
3.9 Método cuantitativo .....	69
3.10 Diseño de la investigación.....	70
3.11 Instrumentos de investigación.....	70
3.12 Procedimientos y especificaciones .....	71
Capítulo 4: Modelamiento estadístico VSA.....	73
Capítulo 5: Modelamiento estadístico VSM .....	106
Capítulo 6: Discusión de resultados .....	129
Capítulo 7: Propuestas concretas .....	141
Capítulo 8: Conclusiones .....	144
Capítulo 9: Recomendaciones.....	146
Bibliografía .....	148
Apéndice I: Consideraciones sobre Kernels Binarios .....	157
Apéndice II: Densidad y el estimador núcleo de los k vecinos más cercanos.....	160
Apéndice III: Densidad y series de funciones ortogonales.....	161
Apéndice IV: Principio de Dirichlet.....	162
Apéndice V: Scripts en R (correo: xxix.alexander@gmail.com) .....	164
Apéndice VI: Constancia de Tesis UES.....	168
Apéndice VII: Solicitud de información al Observatorio Ambiental .....	169
Apéndice VIII: Entrevista MARN 2015.....	170

Apéndice IX: Biografía de Jean Baptiste Fourier .....	175
Apéndice X: Biografía de Johann Peter Gustav Lejeune Dirichlet.....	178
Apéndice XI: Biografía de Johann Carl Friedrich Gauss.....	179
Apéndice XII: Biografía de James Cooley .....	182

## Índice de tablas

Tabla 1. <i>Especificaciones de las estaciones sísmicas</i> .....	27
Tabla 2. <i>Eficiencia relativa de las funciones Kernel más importantes.</i> .....	39
Tabla 3. <i>Detalle de la existencia de siete o menos clúster.</i> .....	50
Tabla 4. <i>Descomposición de series temporales</i> .....	53
Tabla 5. <i>Filtros que se aplican en la descomposición de series de tiempo.</i> .....	54
Tabla 6. <i>Coefficientes de autocorrelación que forman la Función de Autocorrelación.</i> ..	55
Tabla 7. <i>Frecuencia de las cifras de las tablas de números aleatorios.</i> .....	56
Tabla 8. <i>Idea intuitiva del método de los centros del cuadrado.</i> .....	57
Tabla 9. <i>Funciones de densidad, de distribución y distribución inversa VAU</i> .....	58
Tabla 10. <i>Funciones de densidad, de distribución y distribución inversa VAN</i> .....	59
Tabla 11. <i>Detalle de las variables más importantes en la modelización.</i> .....	67
Tabla 12. <i>Estaciones sísmicas de los principales complejos volcánicos ES.</i> .....	67
Tabla 13. <i>Límites y fronteras del RSAM por hora</i> .....	81
Tabla 14. <i>Escenario de Kernels I VSA</i> .....	84
Tabla 15. <i>Escenario de Kernels II VSA.</i> .....	88
Tabla 16. <i>Escenario de Kernels III VSA</i> .....	89
Tabla 17. <i>La elipse en coordenadas polares</i> .....	91
Tabla 18. <i>Geometría Analítica de la Elipse y la Circunferencia</i> .....	91
Tabla 19. <i>Mediciones químicas durante ebulliciones de la laguna VSA</i> .....	100
Tabla 20. <i>Índice IPPEV de cinco a diez años</i> .....	101
Tabla 21. <i>Escenario de Kernels I VSM</i> .....	116
Tabla 22. <i>Escenario de Kernels II VSM.</i> .....	118
Tabla 23. <i>Estadística aplicada a los complejos volcánicos ES.</i> .....	129
Tabla 24. <i>Índice de la modelización con Kernel Binario.</i> .....	130
Tabla 25. <i>Modelos estadísticos de pronóstico VSA.</i> .....	131
Tabla 26. <i>Modelos estadísticos encontrados para el RSAM VSA.</i> .....	133
Tabla 27. <i>Modelos estadísticos de pronóstico VSM.</i> .....	135
Tabla 28. <i>Modelos estadísticos encontrados para el RSAM VSM.</i> .....	136
Tabla 29. <i>Pendiente e inclinación de Kernels VSM y VSA.</i> .....	138

## Índice de figuras

<i>Ilustración 1.</i> Ruinas de Joya de Cerén. ....	22
<i>Ilustración 2.</i> Víctimas del año 72 y cráter del volcán Vesubio. ....	24
<i>Ilustración 3.</i> Descripción física del monte Vesubio. ....	25
<i>Ilustración 4.</i> Evolución española de la vulcanología. ....	30
<i>Ilustración 5.</i> Dinamismo del vulcanismo. ....	31
<i>Ilustración 6.</i> Ramas de la geofísica. ....	33
<i>Ilustración 7.</i> Laguna Verde y erupción del 01 10 05, VSA. ....	62
<i>Ilustración 8.</i> Cráter y erupción del 29 10 13, VSM. ....	63
<i>Ilustración 9.</i> Volcán Izalco y altas temperaturas en el cráter. ....	64
<i>Ilustración 10.</i> VSS, VILO y VSVI. ....	65
<i>Ilustración 11.</i> Ubicación geográfica de los complejos volcánicos, El Salvador. ....	66
<i>Ilustración 12.</i> Secuencia gráfica de los principales tipos de sismos, VSA. ....	75
<i>Ilustración 13.</i> Kernels profundidad-coordenadas geográficas VSA. ....	76
<i>Ilustración 14.</i> Kernels magnitud y coordenadas geográficas VSA. ....	77
<i>Ilustración 15.</i> RSAM por hora 2005, VSA. ....	80
<i>Ilustración 16.</i> Secuencia gráfica del RSAM, periodo post-eruptivo, VSA. ....	81
<i>Ilustración 17.</i> Función de densidad RSAM, 2005, VSA. ....	82
<i>Ilustración 18.</i> Kernels RSAM 2005 y julio del 2005, VSA. ....	83
<i>Ilustración 19.</i> Kernel Binario del RSAM, VILO ....	86
<i>Ilustración 20.</i> Kernels RSAM agosto 2005 y septiembre del 2005, VSA. ....	87
<i>Ilustración 21.</i> Kernels RSAM semana y horas previas a la erupción, 2005, VSA. ....	89
<i>Ilustración 22.</i> Regresión Kernel y regresión Loess, julio 2005, VSA. ....	93
<i>Ilustración 23.</i> Regresión Kernel, MM y SE, Loess, RSAM, agosto 2005, VSA. ....	94
<i>Ilustración 24.</i> ACF con 24 lags y espectro, RSAM, 2005, VSA. ....	95
<i>Ilustración 25.</i> Medias Móviles y Suavizado Exponencial, RSAM, julio 2005, VSA. ....	97
<i>Ilustración 26.</i> Regresión Kernel y regresión Loess, RSAM, agosto 2005, VSA. ....	98
<i>Ilustración 27.</i> Incandescencia del cráter en la erupción, 2005, VSA. ....	99
<i>Ilustración 28.</i> Gotas de Fourier 2005, VSA. ....	104
<i>Ilustración 29.</i> Gotas de Fourier, RSAM 2005, VSA. ....	105
<i>Ilustración 30.</i> Secuencia gráfica de los principales tipos de sismos, 2013, VSM. ....	107
<i>Ilustración 31.</i> Kernels, profundidad, magnitud y coordenadas geográficas VSM. ....	108
<i>Ilustración 32.</i> Kernels, profundidad, magnitud y coordenadas geográficas VSM. ....	109
<i>Ilustración 33.</i> RSAM por hora, VSM. ....	111
<i>Ilustración 34.</i> Función de densidad, RSAM, 2013 y 2014, VSM. ....	112
<i>Ilustración 35.</i> Kernels RSAM 2013, agosto 2013, noviembre 2013, VSM. ....	115
<i>Ilustración 36.</i> Kernels RSAM, diciembre 2013 y periodo post-eruptivo. ....	117
<i>Ilustración 37.</i> Regresión Kernel y regresión Loess, RSAM, nov. 2013, VSM. ....	119
<i>Ilustración 38.</i> Regresión Kernel y regresión Loess, RSAM, dic. 2013, VSM. ....	121
<i>Ilustración 39.</i> Medias Móviles y Suavizado Exponencial, RSAM, nov. 2013, VSM. ....	122
<i>Ilustración 40.</i> Medias Móviles y Suavizado Exponencial, RSAM, dic. 2013, VSM. ....	123
<i>Ilustración 41.</i> ACF y espectro en dB, RSAM, 2013, VSM. ....	124
<i>Ilustración 42.</i> Gotas de Fourier, RSAM, 2013, VSM. ....	126

<i>Ilustración 43.</i> Simulación Monte Carlo, 8699 observaciones, VSM. ....	127
<i>Ilustración 44.</i> Secuencia gráfica RSAM 2013, diferentes escalas, VSM.....	128
<i>Ilustración 45.</i> Representación de los modelos estadísticos VSA y VSM. ....	132
<i>Ilustración 46.</i> Modelos estadísticos VSA y VSM. ....	137
<i>Ilustración 47.</i> Detalle resumido del árbol de eventos de Newhall & Hoblitt, 2002.....	143

## Presentación

La implementación de modelos estadísticos en la vulcanología y la sismología es un hecho reciente. Actualmente existen esfuerzos en la búsqueda de patrones que expliquen de una mejor manera el comportamiento volcánico y sísmico de los volcanes, tanto a nivel nacional e internacional. El Salvador longitudinalmente, de poniente a oriente, incluye de forma parcial al cinturón de fuego, el cual comienza en el hemisferio norte y culmina en el hemisferio sur; en él están atrapados los principales complejos volcánicos: volcán Santa Ana, volcán Izalco, volcán San Salvador, caldera Ilopango, volcán San Miguel, volcán San Vicente, de los cuales se estudiara solamente al volcán Santa Ana y al volcán San Miguel debido a la actividad volcánica reciente.

Se considera que un volcán es activo si tiene reconocidos eventos explosivos por lo menos en los últimos cinco mil años (Saavedra, 1984), de lo contrario se considera inactivo. Clasificar un volcán en activo o inactivo no es sencillo, así que se dividen los volcanes de acuerdo a su actividad eruptiva en monogenética y poligenética, que significan una sola erupción y al menos dos erupciones respectivamente. Si un volcán tuvo una única erupción con o sin antecedentes desde hace dos mil años se considera inactivo en sentido restricto, ya que pueden existir condiciones eruptivas que lo lleve a clasificarse como poligenético; un volcán poligenético no se considera inactivo por la alta probabilidad de que suceda un nuevo escenario volcánico (Escobar, 2002).

La morfología, la geotectónica y el factor tiempo también influyen en la clasificación de los volcanes, los complejos volcánicos del país se agrupan en dos clústeres, los volcanes antiguos del terciario (al menos dos millones de años de antigüedad) y los volcanes jóvenes del cuaternario (menos de dos millones de años de antigüedad), (Escobar, 2002). Desde mil novecientos sesenta y siete a mil novecientos setenta y uno, una misión geológica alemana descubrió más de setecientos centros eruptivos; hacia mil novecientos setenta y tres, se clasifican ciento setenta y siete volcanes catalogados como activos (Bohnenberger, 1973). Otra misión italiana ingresó en mil novecientos ochenta y ocho para estudiar la peligrosidad de los volcanes. Existen volcanes de ambas índoles, por lo que se ha decidido estudiar los volcanes activos en las dos décadas recientes.

En El Salvador se han realizado muchos estudios de vulcanología a nivel nacional e internacional, en mil ochocientos ochenta debido a la última erupción de la caldera Ilopango, el alemán Goodyear W. publicó “Temblores de Tierra y Fenómenos Volcánicos”; existen otras publicaciones como “Los Volcanes de América Central” de Karl Sapper, también alemán; Montessus de Ballore fue un investigador contemporáneo de Sapper. Las publicaciones entre mil novecientos veintitrés hasta mil novecientos cincuenta y dos del salvadoreño Jorge Lardé son importantes en Geología y Vulcanología salvadoreña.

A partir de la década de los cincuenta las publicaciones crecen virtiginosamente, científicos como Howell Williams, Helmut Meyer Abich, O. Bohenenberger, Richard Stoiber, Mike Carr, Willson Rose, Karl Pontier, Maximiliano Hernández Martínez, Darrel Soufield y Carlos Pullinger; han aportado considerablemente al estudio de este fenómeno natural. Esta investigación se denomina “Aplicación de modelos estadísticos para la interpretación de señales sísmico-volcánicas en los principales complejos volcánicos de El Salvador” y comprende exhaustivamente las dos recientes erupciones volcánicas, ocurriendo primero la del volcán Santa Ana en el dos mil cinco, y posteriormente la erupción del volcán San Miguel en el dos mil trece. La actividad sísmica y volcánica del volcán San Miguel, es pronunciada y hasta estas fechas se encuentra inestable.

Predicción y pronóstico suelen utilizarse con error; la predicción corresponde a una afirmación inequívoca de que un evento sucederá, mientras que el pronóstico, produce resultados más flexibles en cuanto a la ocurrencia de estos eventos. La importancia de este trabajo radica en la búsqueda de modelos estadísticos que expliquen el fenómeno del vulcanismo con enfoque a la proyección social. La razón de ser de estos modelos tiene su eje en la reducción de daños materiales y económicos, además de la reducción de muertes de habitantes que viven en los alrededores de estos volcanes; la historia eruptiva del país sugiere que no han ocurrido cantidades exorbitantes de muertes por vulcanismo, mientras que sí han habido pérdidas materiales considerables (Major, Schilling, Pullinger, Escobar, Chesner & Howell, 2001).

El Salvador a través del Ministerio de Medio Ambiente y Recursos Naturales considera importante la navegación estadística en el campo de la vulcanología y sismología, los esfuerzos por la modelización estadística principalmente de las cantidades de fluidos volcánicos prevalecen. Los factores principales en el modelamiento estadístico son, en primer lugar, el complejo volcánico en estudio; y en segundo lugar sus variables, es decir sus características medibles. Estos factores son importantes para un modelamiento óptimo e integral, sin dejar a un lado aquellas variables cuantitativas que son altas premisas en los procesos volcánicos y sísmicos.

La orientación de esta navegación científica se rige en el planteamiento de los objetivos con los que se trabajó durante toda la investigación. El objetivo general se fundamenta en la inferencia de la ocurrencia de erupciones volcánicas y terremotos de origen volcánico, dadas las condiciones y la interrelación de las premisas importantes. Aunque el factor aleatorio podría cambiarlo todo, este objetivo fue la guía en la búsqueda de los modelos estadísticos más certeros. En el planteamiento de los objetivos las variables sísmico-volcánicas se interpretan en función de la construcción del modelo, es necesaria e interesante la comparación de la eficacia de los modelos estadísticos empleados. Referente a las hipótesis de trabajo, se plantea la existencia de un porcentaje menor de señales no explicadas en relación al resto, el otro tipo de hipótesis vinculadas a los modelos estadísticos van incluidas implícitamente en pro del rigor que los validan.

El método cuantitativo ayudó a indagar científicamente el problema de investigación, ya que la información cuantitativa en materia de vulcanismo y sismología, se convierte en conocimiento sustantivo y profundo. Este método de investigación se aplicó para conocer la esencia del modelamiento estadístico de las variables de los fenómenos del vulcanismo y la sismicidad. A continuación se hace una breve descripción de los contenidos de cada uno de los capítulos desarrollados en el presente trabajo.

#### *Capítulo 1: Formulación del problema*

El problema en estudio se describe explicando la situación en que se desarrolla y enfocándolo a su contexto, los factores generales y específicos que lo caracterizan; la importancia de esta investigación; el planteamiento de objetivos; y las hipótesis.

## *Capítulo 2: Abordaje teórico*

Este capítulo está constituido por el historial sísmico a nivel nacional así como también el historial volcánico de las erupciones más potentes en la historia reciente, al final, el soporte matemático de la modelización estadística empleada.

## *Capítulo 3: Metodología*

Descripción del método utilizado, su aplicación, las variables y su medición, la descripción de las unidades de estudio; las técnicas apropiadas para la recopilación y análisis de datos referentes a la vulcanología y sismología; el problema de la muestra y la población; el diseño de la investigación, y las especificaciones y procedimientos.

## *Capítulo 4: Modelamiento estadístico VSA*

Experimentaciones y modelizaciones estadísticas de las variables sísmico–volcánicas del complejo volcánico Santa Ana.

## *Capítulo 5: Modelamiento estadístico VSM*

Experimentaciones y modelizaciones estadísticas de las variables sísmico–volcánicas del complejo volcánico Santa Ana.

## *Capítulo 6: Discusión de resultados*

Incluye los resultados obtenidos y los análisis e interpretaciones de los modelos encontrados. El instrumento administrado a los especialistas en materia de vulcanología y sismología es analizado. Finalmente, se establecen las propuestas de la investigación.

## *Capítulo 7: Conclusiones*

Se detalla con firmeza, consistencia y efectividad, todo lo expuesto a lo largo del proyecto.

## *Capítulo 8: Recomendaciones*

Se presentan las recomendaciones que los interesados pueden tomar en cuenta para la mejora continua del modelado estadístico en la vulcanología y sismología.

La sección final incluye las demostraciones matemáticas de las técnicas estadísticas utilizadas en la investigación, código R, además de la evidencia de los hechos que ocurrieron en este tiempo, se concluye con la bibliografía de la que se valió.

## Capítulo 1: Formulación del problema

### 1.1 Planteamiento del problema

“El vulcanismo es la manifestación sobre la superficie de la tierra de los procesos que se producen en el interior de un cuerpo celeste plenamente vivo y en continuo proceso de evolución” (González, E., Escobar, E., Becerra, R., Gosálvez, R. U., & Dónis, J., 2010). Este concepto abarca el detalle concreto de un evento volcánico, más sin embargo, próximamente se ampliará dicho concepto a uno más integral; el que incluye las formas de respuesta y planificación creadas por la humanidad para hacer frente a una erupción volcánica, en términos más sencillos, incluye la preocupación de la sociedad por la protección civil. La protección civil es la razón de ser de muchas instituciones gubernamentales y no gubernamentales.

La vulcanología es la rama de la geología que se encarga de este fenómeno natural que no puede pasar desapercibido. El término vulcanismo viene del latín *Vulcānus*, cuyo origen pertenece a la cultura griega, el cual es utilizado para nombrar al Dios del fuego, Dios cuyo taller se encuentra en el centro de la tierra. “El seísmo es un movimiento vibratorio que se origina en zonas internas de la Tierra y se propaga en forma de ondas elásticas por los materiales rocosos de la corteza terrestre” (La Enciclopedia, 2004); en contraste la sismología, rama de la geología, estudia la frecuencia e intensidad de estos movimientos. Un sinónimo de seísmo es terremoto.

El vulcanismo y la sismicidad generalmente, son fenómenos naturales de intensidad pronunciada y causantes de grandes catástrofes. A pesar de que en El Salvador el vulcanismo no era muy frecuente, relativamente, hasta hace un lustro de años; el seísmo sí lo es, a veces por causas volcánicas y otras por subducción. Actualmente, la tendencia de aparición de erupciones volcánicas y sismos tiene un rango temporal similar de ocurrencia de al menos doce años (Major et al., 2001), sin excluir la coincidencia. El Salvador es conocido como el “país de las hamacas” debido a la fuerte actividad sísmica. De lo anterior, surge la necesidad de búsqueda de modelos estadísticos certeros y con soporte estadístico, para la inferencia en el campo del vulcanismo y la sismicidad, con el

único fin de pronosticar, prevenir y dar seguimiento (prepararse ante futuros eventos) a estos fenómenos naturales.

La manipulación de las herramientas estadísticas aplicadas a las variables relacionadas con el vulcanismo es necesaria, con el fin de analizar e interpretar el comportamiento de las variables de tipo sísmico-volcánico de los principales complejos volcánicos de El Salvador. La naturaleza y la estadística están íntimamente relacionadas en más de un sentido. La estadística es una ciencia indispensable para estudiar a la naturaleza en todas sus direcciones, proporciona información sobre los elementos de un “conjunto” llamado “naturaleza”; además, permite buscar modelos que se adapten a la realidad de interés para dar respuesta a las grandes interrogantes planteadas por un ser pensante: el hombre; quien en su vida cotidiana tiene, instintivamente, la curiosidad de responder a todo lo que le acontece.

El modelamiento estadístico, no importando su campo de aplicación, está en su apogeo gracias a los avances tecnológicos como producto del gran interés por interpretar y analizar el comportamiento de los fenómenos naturales. Finalmente, la certeza de la siguiente proposición es muy interesante, ilustra el afán en modelar las características volcánicas y sísmicas para la obtención de esa medida. “La probabilidad es una medida que procura un grado de creencia racional de que algo pueda acontecer” (Salguero, 2006).

## **1.2 Justificación**

El modelamiento estadístico es importante para el estudio de fenómenos naturales y sociales; se decide abordar una búsqueda científica de interpretación y análisis de señales sísmico-volcánicas de los principales complejos volcánicos de El Salvador. La realización de esta investigación es fundamental para describir y analizar científicamente el comportamiento del vulcanismo en particular, que afecta determinadas zonas del país, sino es que al territorio en general, convirtiéndose en catástrofes y desastres naturales. En realidad, ¿qué se logra con este tipo de procedimientos estadísticos? Se observa y determina, en cierto grado, el comportamiento de las variables de interés para inferir y ofrecer puntos de vistas alternos y aceptables científicamente para la toma de decisiones

con bases estadísticas. En otras palabras, se brinda soporte estadístico de alto nivel a las instituciones salvadoreñas que se encargan de prevenir y emitir alertas sobre posibles catástrofes en zonas vulnerables; instituciones dedicadas al rubro de la protección civil.

Esta investigación amplía la aplicación de modelos estadísticos. Brinda las herramientas necesarias, teóricas y prácticas, para la futura implementación y comprobación de las aplicaciones a fenómenos naturales en un contexto real. El investigador o los investigadores podrían estar interesados en analizar y aplicar los modelos utilizados en este documento. No obstante, el investigador no debe limitarse únicamente a este material ya que debe cumplir el rol de descubridor y presentar a la sociedad sus avances alcanzados. La sociedad actual refleja una evidente necesidad de ampliar las investigaciones científicas que combatan con los aprietos naturales y sociales que la debilitan cada vez más. La realización de ésta investigación tiene como finalidad la reducción de las pérdidas materiales debidas a la inminente actividad volcánica y sísmica característica del país en los últimos diez años.

Un hecho histórico es el cambio de coordenadas geográficas de la Villa de San Salvador, capital actual de El Salvador, debido a la inestabilidad de los suelos en que se fundó en tiempos de la colonia. En mil quinientos veinticinco, el objetivo de Don Gonzalo de Alvarado era crear una villa de españoles alrededor de lo que hoy es Antiguo Cuscatlán a orillas de la laguna o cráter llamado “Puerta de La Laguna”, una encomienda de su hermano Don Pedro de Alvarado; se estableció alrededor del primero de abril de mil quinientos veinticinco más sobre “los lomos de las cabalgaduras que sobre el inestable suelo es estas comarcas”, pero luego fue trasladada en el año de mil quinientos veintiocho al actual departamento de San Salvador, siendo esta su ubicación definitiva cuyo primer alcalde es el señor Diego de Holguín (Lardé, 1984).

En la historia de El Salvador hay fuertes terremotos que han azotado y destruido la capital y la mayoría de los doscientos sesenta y dos municipios, los más catastróficos se detallan a continuación: terremoto de fecha quince de septiembre del año de mil novecientos dos con ocho punto uno en la escala Richter y nueve en la escala Mercalli, con epicentro en Ahuachapán a veinticinco metros de profundidad, la destrucción se

concentra en la Barra de Santiago, se totalizan cuatrocientos muertos. Una pareja de terremotos de fecha siete de junio del año mil novecientos diecisiete, el primero con seis punto siete y el segundo con seis punto cinco en la escala Richter, ambos con ocho en la escala Mercalli, a diez kilómetros de profundidad del volcán San Salvador, causan mil cincuenta muertos además de la posterior erupción volcánica con quince segundos de diferencia; erupción cuyo flujo de lava se extendió y petrificó en un gran perímetro del volcán, incluyendo las ciudades de Armenia, Ilopango y Quezaltepeque.

El peor terremoto en la historia sísmica de El Salvador de fecha diez de octubre de mil novecientos ochenta y seis, con escasos cinco punto siete puntos en la escala Richter y nueve en la escala Mercalli, cinco punto cuatro kilómetros de profundidad y San Salvador como epicentro, específicamente Los Planes de Renderos; fue una tragedia que destruyó la capital en su totalidad dejando mil quinientos muertos, se sintió desde México Distrito Federal hasta la República de Colombia. Y por último, el terremoto del trece de enero del año dos mil uno, con ocho punto cero en la escala Richter y nueve en la escala de Mercalli, con treinta y ocho kilómetros de profundidad, el epicentro fue ubicado frente a las costas de Usulután, zona oriente del país.

Un alud de la Cordillera del Bálsamo se desprendió y yació sobre una residencial construida recientemente en la zona sur de Santa Tecla departamento de La Libertad sin los requerimientos mínimos del estudio de vulnerabilidad de la zona. Sólo en ese evento se estiman setecientas víctimas, siendo un total de más de mil. Dejó muchas pérdidas materiales, parte de la red de carreteras se dañó severamente, la intensidad fue registrada desde la República de Colombia hasta los Estados Unidos Mexicanos. Exactamente a un mes del terremoto, se origina otro con epicentro en San Pedro Nonualco, La Paz; con una profundidad de once kilómetros; con una intensidad de seis punto seis en la escala de Richter y ocho en la escala de Mercalli, dejando con más miseria y pánico al territorio salvadoreño (Wikipedia, 2016).

Por otro lado, el vulcanismo provoca caos tanto en las pérdidas materiales como en las humanas. El Salvador, coyunturalmente se excluye de tanta pérdida humana; los registros de daños emitidos por las autoridades competentes reflejan escasas pérdidas

humanas, nulas en su mayoría. Por el contrario, las pérdidas materiales ascienden vertiginosamente las cifras en dólares, el vulcanismo afecta principalmente el rubro de la caficultura; históricamente la producción de café ha sido el elemento clave en la mejora de la economía nacional durante la primera mitad del siglo veinte, aunque en la actualidad no sea tan influyente como lo fue antes. En términos generales, ocasiona severos daños en la agricultura nacional (Major et al., 2001).

En la historia vulcanológica salvadoreña, es fundamental abordar el caso de la ciudad sepultada con un rango de cuatro a seis metros de cenizas, Joya de Cerén (ver ilustración 1); la Pompeya de América. Joya de Cerén es patrimonio de la humanidad, este templo se sitúa en el kilómetro treinta y cinco de la carretera panamericana en el municipio de San Juan Opico, departamento de La Libertad con coordenadas de trece grados, cuarenta y nueve minutos, treinta y nueve segundos con rumbo norte y; ochenta y nueve grados, veintiún minutos y veintidós segundos con rumbo oeste. Este sitio precolombino data desde el año cuatrocientos hasta el año seiscientos debido posiblemente a una fuerte erupción volcánica de la Laguna Caldera.



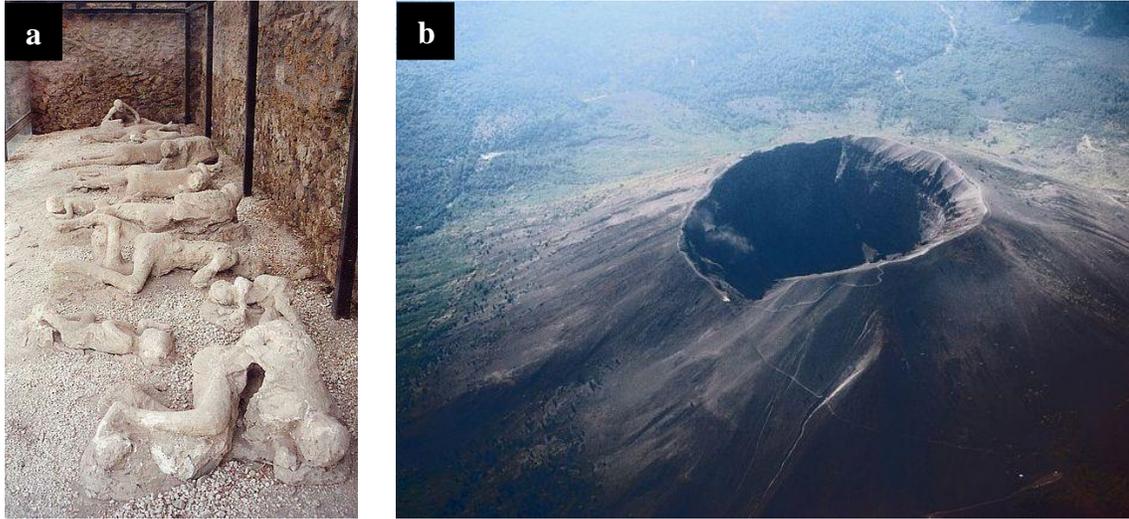
*Ilustración 1.* Ruinas de Joya de Cerén.  
Detalle de cuando ocurrió repentinamente la erupción del volcán Caldera ubicada en San Juan Opico, La Libertad, El Salvador. Cortesía del Ministerio de Turismo.

Fue una gran explosión volcánica, enterró la aldea con alrededor de catorce capas de cenizas con temperaturas entre cien y quinientos grados Celsius, sus habitantes lograron huir a tiempo ya que no se ha encontrado algún cuerpo, sólo sus utensilios, cerámicas y

alimentos. Un tipo muy peculiar de vulcanismo, una explosión en tan sólo segundos (Wikipedia, 2014). Al igual que Joya de Cerén existen otras regiones en el mundo donde pasó exactamente lo mismo. En la ciudad italiana de Pompeya hubo un número arbitrario de pérdidas humanas en contraste a Joya de Cerén. Pompeya y Joya de Cerén son los sitios arqueológicos mejor conservados debido a explosiones volcánicas. Pompeya pertenece a la provincia de Nápoles y sus coordenadas son cuarenta grados, cuarenta y cuatro minutos y cincuenta y ocho segundos con rumbo norte, y catorce grados, veintinueve minutos y doce segundos con rumbo este.

Esta ciudad fue enterrada por una violenta erupción del Vesubio el veinticuatro de agosto del año setenta y nueve después de Cristo, fecha de la cual da fe el romano Plinio “El Joven”. Todos sus habitantes fallecieron instantáneamente debido al flujo piroclástico; la erupción volcánica también destruyó parcialmente la ciudad de Herculano, en esta ciudad predominaron las pérdidas materiales (Laske, 2008). En Pompeya habitaban cerca de veinticinco mil personas, los presagios del catastrófico suceso consistieron en frecuentes terremotos que afectaron la ciudad, algunos pompeyanos abandonaron la ciudad por temor a los constantes terremotos y se salvaron para el momento de la erupción (Beard, 2008).

Arqueológicamente se registran dos mil víctimas de la explosión del Vesubio. La ciudad completa quedó congelada en espera de su descubrimiento, tuvieron que pasar casi diecisiete siglos para este importante hallazgo. Pompeya fue un importante centro urbano romano antes del setenta y nueve; hoy, sólo vestigios de lo que alguna vez fue una gran ciudad. Todo quedó petrificado, incluso la posición de las víctimas justo al morir (ver ilustración 2). El monte Vesubio es un volcán activo, está rodeado por un casco urbano de aproximadamente tres millones de personas, aproximadamente la mitad de la población salvadoreña. La posibilidad de que vuelva hacer erupción volcánica es muy probable y los resultados de esta nueva erupción son desorbitantes.



*Ilustración 2. Víctimas del año 72 y cráter del volcán Vesubio.*

a) Víctimas de la erupción del Vesubio, personas petrificadas en sus hogares buscando protección ante la indomable erupción del año 72 del Monte Vesubio. b) Volcán italiano Vesubio, fotografía aérea y detalle del cráter producto de las estruendosas erupciones volcánicas del pasado.

Las señales precursoras a la erupción del setenta y nueve fueron obvias y fielmente documentadas por historiadores. Diecisiete años antes ocurrió un estrepitoso terremoto con fecha cinco de febrero que destruyó generalmente la bahía de Nápoles y consecuentemente también a Pompeya (Beard, 2008). Algunos de los daños se encontraban aún en reparación cuando el Vesubio sorprendió a la ciudad. En fin, ocurrieron sismos de leve y pronunciada intensidad antes del despertar del coloso, mientras que los romanos que habitaban en sus cercanías los hacían costumbres debido a la región. La gente se acostumbró a vivir en medio de enjambres sísmicos al ver que no pasaba nada, los habitantes se confiaron y se perdió la ciudad entera, entre personas, cultivos e infraestructuras.

Según el vulcanismo histórico, los cinco volcanes que han ocasionado las peores erupciones son el Tambora de Indonesia, con ochenta mil muertes conocida como la erupción del milenio; Laki de Islandia, con cuarenta mil muertes; Krakatoa de Indonesia, con treinta y siete mil muertes que hizo explotar la isla entera con el mismo nombre; Monte Pelée de Francia, con treinta mil muertes y Nevado del Ruíz de Colombia, con veinticinco mil muertes. Es necesario buscar formas válidas y adecuadas de reducción de pérdidas materiales, las herramientas estadísticas y tecnológicas explican hallazgos que

cambian las perspectivas actuales. No es fácil, pero el camino es muy fértil para la búsqueda de modelos estadísticos que estudien el comportamiento sísmico-volcánico.

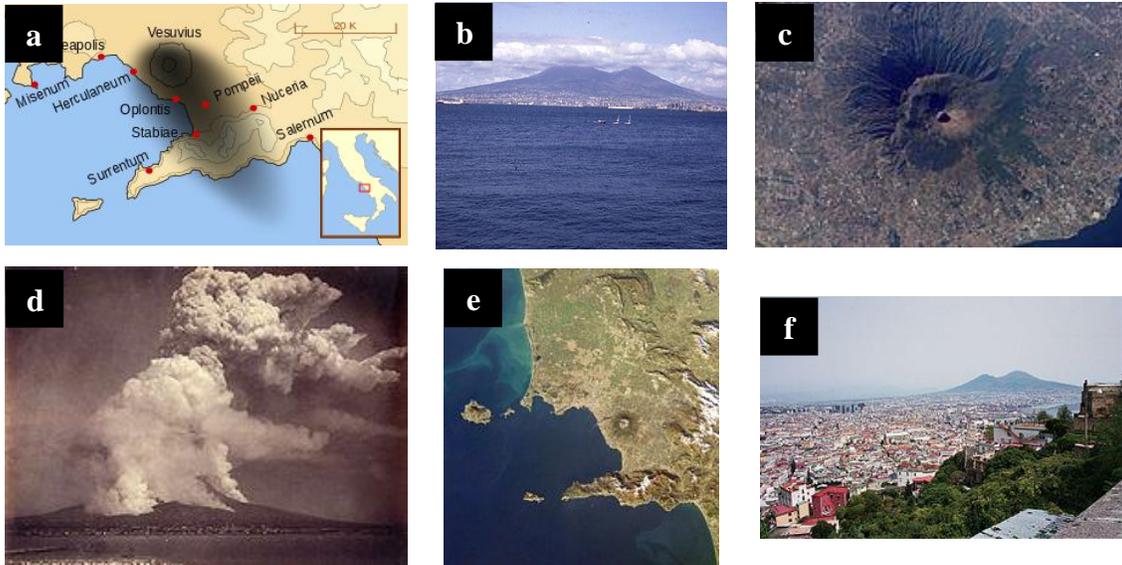


Ilustración 3. Descripción física del monte Vesubio.

a) Simulación de la caída de ceniza de la erupción del 79. b) El Vesubio desde una distancia semejante a la que tuvo que observar Plinio el Joven el día de la erupción y que remite a Tácito en una de sus cartas. c) Monte Vesubio desde el satélite. d) Erupción de 1872. e) Nápoles y el monte Vesubio desde el espacio. f) La ciudad de Nápoles

### 1.3 Planteamiento de objetivos

#### *Objetivo general*

Inferir sobre la ocurrencia de erupciones volcánicas y terremotos de origen volcánico dadas las condiciones a través del modelamiento estadístico de las variables de interés que se originan en los complejos volcánicos de El Salvador.

#### *Objetivos específicos*

- Interpretar el comportamiento de las variables sísmico-volcánicas, en función de la construcción de un modelo estadístico certero.
- Comparar la efectividad de los modelos estadísticos empleados.

### 1.4 Planteamiento de las Hipótesis estadísticas

El término hipótesis proviene del griego Hypothesis, y es una suposición de la posible ocurrencia de una cosa, de la que se advierte una consecuencia. Sinónimos de hipótesis son suposición, conjetura o presunción. Se trata de una proposición respecto a algunos elementos empíricos y otros conceptuales y sus relaciones mutuas. La proposición surge

más allá de los hechos y experiencias conocidas, con el propósito de llegar a una mejor comprensión de los mismos (Razo, 1998). Se trata de la suposición del comportamiento provisional de un fenómeno en observación, para inferir una consecuencia; este es el comienzo para una investigación científica.

La siguiente proposición es verdadera: “Una hipótesis es una explicación anticipada y provisional de alguna suposición que se trata de comprobar o desaprobar a través de los antecedentes que se recopilen sobre el problema previamente planteado.”, (Razo, 1998). La importancia de la inclusión de hipótesis en las investigaciones científicas radica en el análisis de una solución concreta al problema planteado. Para esta investigación, el planteamiento de hipótesis se diseña así:

Hipótesis nula

- $H_0 =$  “No existe diferencia alguna entre el porcentaje de señales irregulares decisivas ante erupciones volcánicas y el porcentaje de señales regulares en el RSAM de los complejos volcánicos VSA y VSM de El Salvador.”

Hipótesis alterna

- $H_1 =$  “Existe diferencia alguna entre el porcentaje de señales irregulares decisivas ante erupciones volcánicas y el porcentaje de señales regulares en el RSAM de los complejos volcánicos de VSA y VSM El Salvador.”

## 1.5 Delimitación

### 1.5.1 Delimitación espacial

El análisis estadístico se aplica a los complejos volcánicos Santa Ana y San Miguel situados en los departamentos homónimos de El Salvador. Los modelos estadísticos se aplican a la base de datos RSAM por hora de cada volcán y a las estadísticas de la naturaleza de los sismos: tipos, magnitud, profundidad y coordenadas geográficas.

### 1.5.2 Delimitación temporal

La base de datos del RSAM por hora del complejo volcánico Santa Ana comprende el intervalo temporal desde la media noche del primero de enero del dos mil cinco, hasta las siete horas del treinta de junio del dos mil quince, mientras que la base de datos del RSAM por hora del complejo volcánico San Miguel comprende el intervalo temporal

desde la media noche del primero de enero del dos mil siete, hasta las seis horas del veintinueve de junio del dos mil quince. La base de datos referente a las características de los sismos del complejo volcánico Santa Ana incluye las variables: hora y fecha del conteo sísmico, conteo sísmico, el tipo de sismo (localizado, registrado, volcánico, vulcano-tectónico), las coordenadas geográficas (latitud y longitud), la profundidad alcanzada en kilómetros (máximo quince kilómetros), y la magnitud máxima.

Comienza, aunque fragmentado, desde las tres horas del quince de enero de mil novecientos ochenta y cuatro y finaliza hasta las cuatro horas del trece de mayo del dos mil quince. La misma base de datos pero para el complejo volcánico San Miguel comienza desde las catorce horas del cinco de enero de mil novecientos ochenta y siete, finaliza a las diecinueve horas del veintinueve de mayo del dos mil quince. Las bases de datos fueron originadas por la estación sísmica San Blas para el complejo volcánico Santa Ana, y la estación sísmica homónima para el complejo volcánico San Miguel (ver tabla 1). Esta investigación durará aproximadamente dos años desde su inscripción como trabajo de graduación.

**Tabla 1.** *Especificaciones de las estaciones sísmicas*

Complejo volcánico	Código	Nombre	Latitud	Longitud	Elevación	Tipo	Comunicación	Mod.
Santa Ana	SBLS	San Blas	13.83	-89.62	1949	Periodo corto	Radio Analógico	L4-C
San Miguel	VSM	San Miguel	13.44	-88.27	1708	Periodo corto	Radio Analógico	S-13

### 1.5.3 Delimitación estadística

Los herramientas estadísticas implementadas son el suavizado a través de la regresión no paramétrica Loess, para la suavización de observaciones y búsqueda de tendencias; análisis de la estacionalidad de las series temporales a través de la Función de Autocorrelación, el espectro de frecuencias, medias móviles y el suavizado exponencial; el empleo de la regresión no paramétrica Kernel para la búsqueda longitudinal de centroides de datos reflejados por el comportamiento sísmico; la simulación Montecarlo para la comprobación y validación de los modelos aplicados en la vulcanología. Por último las Curvas de Andrews, que trasladadas al plano polar se convierten en Gotas de

Fourier, para la generación de señales estadísticas que reproduzcan los comportamientos sísmico-volcánicos de cada volcán sin el problema de la dimensionalidad de los datos.

## Capítulo 2: Abordaje teórico

### 2.1 Enfoque de la investigación vulcanológica internacional

En el transcurso de los años, el interés y preocupación de los profesionales, funcionarios e investigadores protagonistas en la prevención y atención de desastres, ha sido la respuesta de la población ante situaciones de riesgo, alertas, predicciones y pronósticos de desastres naturales. El conocimiento en cuanto al origen generador de los fenómenos ha evolucionado considerablemente; aunque en muchos casos, todavía es un proceso difícil de modelar por las ciencias físicas, debido a sus características aleatorias, incertidumbre e imprecisión. En la vulcanología, la vigilancia, monitoreo y observación de la actividad volcánica, sísmica o hidrometeorológica concluye en la inferencia de un evento peligroso.

Sin embargo, son muy pocas las ocasiones donde existe certeza de la ocurrencia del evento y su magnitud; dando lugar a una importancia arraigada de las circunstancias, sin tener en cuenta la no ocurrencia del evento tan esperado estadísticamente. Además del pronóstico vulcanológico con la mejor tecnología y los recursos suficientes para atender a una comunidad, se necesita tener en cuenta el qué y el cómo informar a la población para que reaccione tomando medidas de protección adecuadas. Han ocurrido casos en que se dispone de buen equipo vulcanológico y recursos pero la forma de informar a la población es incipiente, la erupción de noviembre de mil novecientos ochenta y cinco del volcán colombiano Nevado del Ruíz.

La vulcanología ha cambiado recientemente, para satisfacer estas dos necesidades sociales. La evolución vulcanológica en España comienza desde mil novecientos treinta y seis (ver ilustración 4), se contaba con la geología descriptiva; desde la década de los cuarenta, con el estudio de petrología-geoquímica; desde la década de los setenta, con la geofísica; desde la década de los ochenta, con la geografía física; desde mil novecientos ochenta y cinco, con los estudios de geodesia-gravimetría; a partir de la década de los noventa, con el estudio de gases, y la incorporación de la modelización como una forma más integral, tecnológica y avanzada.

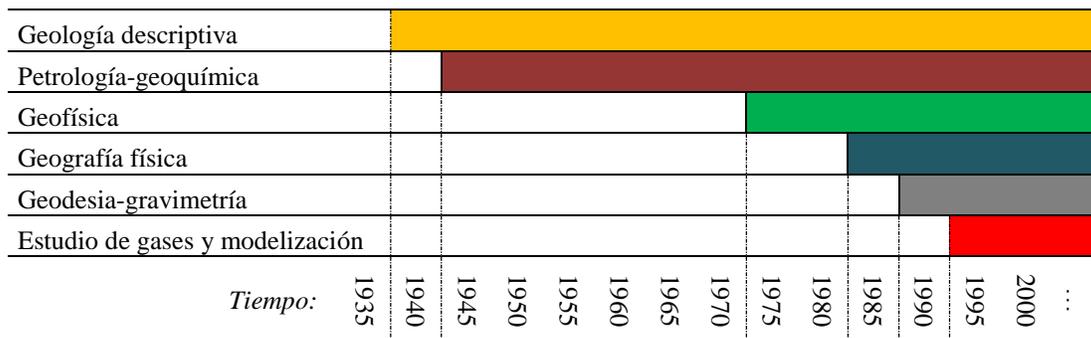
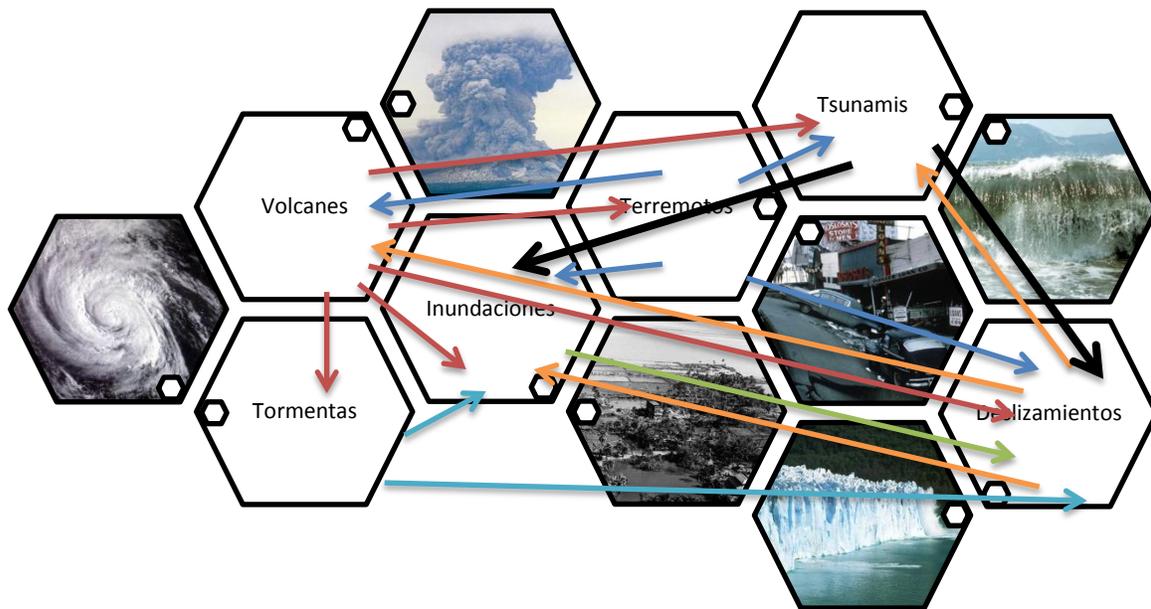


Ilustración 4. Evolución española de la vulcanología.

La vulcanología en el pasado, comprende el estudio de los volcanes, sus procesos, sus productos y formas; en la actualidad la vulcanología es revalorizada, como ciencia de la tierra que ha avanzado significativamente en las últimas décadas, convirtiéndose en una ciencia multidisciplinar dedicada al estudio de los mecanismos que generan y controlan las erupciones volcánicas y a su predicción con el fin de reducir su riesgo asociado. Para que la vulcanología logre su fin, es indispensable el abordaje integral de especialistas procedentes de otras disciplinas y de distintas coordenadas geográficas, incluida la estadística y la inferencia como ramas de la matemática. El trabajo de modelización numérica por parte del estadístico fomenta el objetivo contemporáneo de la vulcanología.

Internacionalmente, el objetivo primordial de la vulcanología es la reducción del riesgo; implica cualquier aspecto relacionado con el estudio de los volcanes y de sus productos sea directa o indirectamente enfocados al análisis de la peligrosidad, a la reducción del riesgo, o al impacto de los volcanes en el medio ambiente y el aprovechamiento económico de los volcanes. Un peligro natural no se aísla de la posibilidad de otros fenómenos como terremotos, maremotos, deslizamientos, tormentas e inundaciones (ver ilustración 5). La incidencia de muchos fenómenos en una erupción volcánica se debe al deseo de los especialistas de tener una visión más realista del problema. La evaluación de la peligrosidad, y la evaluación y gestión del riesgo también son fundamentales en la vulcanología actual.



*Ilustración 5. Dinamismo del vulcanismo.*

Interrelaciones existentes entre distintos procesos naturales peligrosos, incluido el vulcanismo.

El financiamiento en vulcanología por parte de algunos países, establece su grado de interés y preocupación por esta rama. En España, se destina un cinco por ciento del fondo de instituciones privadas; en Europa, un cuatro por ciento; en Italia, un veinticuatro por ciento proveniente de Protección Civil; Japón financia con fondos nacionales un veintisiete por ciento y Estados Unidos de América destina un catorce por ciento para el rubro del estudio vulcanológico en cada una de sus regiones, la pregunta ahora es: ¿cuánto actualmente se destina en el rubro de la protección civil, más específicamente en el área de vulcanología, en El Salvador?

Los científicos en el área de vulcanología, vigilan y reconocen indicadores de erupciones volcánicas inminentes. En una erupción volcánica, el magma provoca cambios que usualmente son detectados por varios instrumentos geofísicos y por las observaciones de rutina. También se producen enjambres de temblores, producto del rompimiento de roca debido a la emersión del magma o al incremento de la presión subterránea debido al calentamiento de fluidos volcánicos. La temperatura aumenta por el calor del magma, puede registrarse en brotes de agua caliente o en fumarolas. Pequeñas explosiones enriquecen el análisis de la composición y volumen de los gases;

estos elementos cambian cuando el magma se acerca a la superficie. La deformación superficial de un complejo volcánico es consecuencia de las inyecciones de magma.

Un sismógrafo registra grupos significativos de temblores que ocurren en los complejos volcánicos de El Salvador. En algunos volcanes ocurren enjambres significativos de sismos días o meses antes de las erupciones. Los periodos de actividad volcánica son inciertos, en las últimas décadas la investigación ha sido ardua y ha desarrollado avances sustanciales en la supervisión de volcanes y pronósticos de erupción. Existen sólo generalizaciones con respecto a la probabilidad, tipo y escala de una erupción inaplazable. La actividad precursora pasa por fases de aceleración y desaceleración, y en alguna ocasión la erupción volcánica no sucede, sino que se agota y desaparece, sin tener algún algoritmo aplicable; inversamente, no hay que descartar que la erupción pueda ocurrir sin actividad precursora sensible por la humanidad.

Japón es una nación notable por su actividad volcánica, tiene ciento ocho complejos volcánicos activos. Es afectada por la subducción de las placas del Pacífico y la de Filipinas contrapuestas a la placa de Eurasia. Japón es recordado por el más grande daño originado por la erupción volcánica del complejo volcánico Unzen, en el año de mil setecientos noventa y dos; en la cual, una avalancha causó un Tsunami que dejó sin vida cerca de quince mil personas (Fujita, 2009). La predicción de erupciones volcánicas es menos complicada de efectuar. Una erupción volcánica tiene eventos precursores, que interpretados de forma correcta se obtienen resultados esperados. Es diferente predecir el inicio de una erupción volcánica y otra predecir su culminación, o su duración. El comportamiento del magma en un volcán es un precursor ideal, en Japón se llevan predicciones de erupciones analizando el flujo magmático.

La carencia del suministro o el abastecimiento de magma de un volcán, dificulta la progresión de la predicción de la actividad volcánica. Los modelos teóricos usados son la experimentación y la simulación. El objetivo de la predicción de erupciones volcánicas es pronosticar los cinco elementos de una erupción: su duración, comprende el inicio y el final; ubicación, magnitud, tipo y progresión. La búsqueda de la predicción se asocia a tres etapas. Etapa uno, las anomalías en la actividad volcánica pueden

ser detectadas desde las observaciones directas del volcán. Etapa dos, las causas de la anomalía se estiman de observaciones del volcán y experiencias pasadas. Etapa tres, las predicciones pueden ser hechas aplicando resultados de observaciones para identificar leyes físicas, basadas en la matemática, que modelen el fenómeno volcánico.

Los esfuerzos en la predicción de erupciones volcánicas se desarrollan bajo cooperaciones internacionales, se comparte información y conocimientos de otras naciones volcánicas. Esto ha contribuido grandemente a los avances tecnológicos en la predicción de erupciones volcánicas en Japón. El trabajo predictivo también se centra en observaciones de volcanes en el Sudeste asiático y Sudamérica como parte de la contribución internacional del Japón que ha producido resultados positivos (Desastres y Sociedades, 1996).

## 2.2 Introducción al modelamiento de datos

La medida de la probabilidad de una erupción volcánica está condicionada en dos partes. Una forma es realizar análisis del historial eruptivo del volcán, la estadística se aplica a la distribución temporal de la actividad eruptiva pasada. Y también se pueden realizar análisis de información disponible del estado actual de un volcán, la estadística es determinística. La actividad o comportamiento de algunos fenómenos se advierten aleatoriamente. No basta con analizar el historial eruptivo del volcán, es enriquecedor tomar en cuenta los estudios geofísicos y las técnicas de seguimiento instrumental de la actividad volcánica. Los estudios geofísicos comprenden las ramas de la Sismología, Geodesia, Gravimetría, Geomagnetismo, Geoelectricidad, Geotermia, entre otros; aportan datos relevantes de estructura y características del volcán (ver ilustración 6).

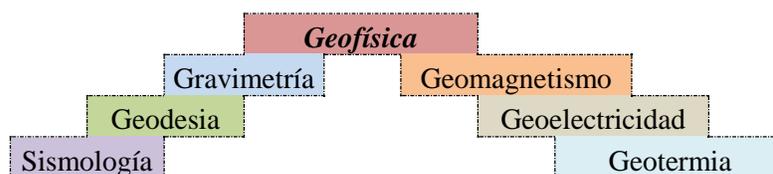


Ilustración 6. Ramas de la geofísica.

El seguimiento instrumental de la actividad volcánica establece parámetros físicos tales como la identificación del cambio de comportamiento del volcán. La observación y análisis instrumental parten del estudio de la deformación, sismicidad local, efectos

electromagnéticos y composición química de las emisiones gaseosas. Las predicciones futuras a corto plazo de la próxima erupción volcánica utilizan estudios geofísicos y el seguimiento de la actividad volcánica. El pronóstico de una erupción se basa en la detección de los fenómenos precursores de la erupción. Los precursores son variables cuantitativas o cualitativas, que pueden incorporarse adecuadamente en un modelo estadístico que prediga o pronostique la próxima erupción a corto plazo.

Hay tres casos posibles, una alarma válida si con el comportamiento del precursor en el modelo estadístico ocurre una erupción volcánica; falsa alarma, si con el comportamiento del precursor en el modelo estadístico no ocurra una erupción volcánica; o en el peor de los casos, una alarma fallida, si no se presenta el comportamiento de los precursores pero sí ocurre una erupción volcánica. El modelo estadístico debe prever estos casos, para ello se lleva a cabo la validación del modelo. Es necesario un análisis estricto para definir la eficacia de los precursores de la predicción de una erupción volcánica.

Una proporción interesante para la predicción de erupciones volcánicas es la relación entre el número de aciertos y el número de fallos para clasificar los precursores en ideales, fuertes o débiles. Naturalmente, los precursores que se desean en el modelo estadístico son de tipo ideal; o por lo menos, fuerte. La detección de un precursor ideal, se consigue con una proporción explicativa de que su eficacia sea infinita, la proporción entre los aciertos que es una constante y la suma de los fallos y falsas alarmas que corresponde a cero; forman la proporción entre una constante y cero que corresponde a una tendencia infinita. La detección de un precursor fuerte, se consigue con una proporción explicativa de que su eficacia sea mayor que la unidad, el número de aciertos es mayor que el de fallo o falsas alarmas.

Un precursor es débil cuando se activa fácilmente y provoca falsas alarmas, su eficacia es muy pequeña, significativamente menor que la unidad y con tendencia a cero. Es recomendable la combinación de varios precursores para mejorar la eficiencia del sistema de alarma, siempre y cuando esos precursores estén relacionados estadísticamente y de modo exclusivo al evento, la existencia de una serie temporal

beneficiaría en muchos aspectos. La disponibilidad de una serie de tiempo permite analizar las correlaciones cruzadas entre los distintos parámetros y determinar la existencia de causas ajenas a la actividad volcánica en la aparición, individual o conjunta, de los posibles fenómenos precursores.

En general, el pronóstico de erupciones volcánicas para la mejora del sistema de alarmas, se fundamenta en el cálculo e interpretación de una proporción que ajusta el nivel de eficacia de las variables precursoras; el objetivo general de esta investigación, trasciende más allá con la búsqueda de modelos estadísticos que mejoren la inferencia respecto a las erupciones volcánicas en El Salvador. Un análisis multivariante de Cox o un análisis de una distribución fractal ayudan a localizar futuras erupciones de complejos volcánicos. Se aplican cuando la distribución de complejos volcánicos es amplia y se desea calcular la probabilidad de que una cierta superficie contenga un futuro centro eruptivo. También se recurre a los análisis de sensibilidad producto de la investigación de operaciones; y técnicas de vigilancia instrumental, siempre y cuando se tenga claro la información que proporcionan los mejores predictores para esta ocasión.

En el desarrollo de una erupción volcánica, las técnicas más empleadas se basan en el diseño de modelos físicos que simulen el mecanismo eruptivo esperado, de tal forma que la resolución numérica de las ecuaciones que constituyen el modelo, describan la evolución de la erupción y la distribución final de los productos de la misma. En las últimas dos décadas, la modelización física de los procesos volcánicos administra una de las formas más importantes en la Vulcanología, ya que además de cuantificar las características del fenómeno, ayudan a comprender mejor la ocurrencia de los mismos (González et al., 2010). La modelización física implica un bagaje de procesos no lineales que requieren el análisis de ecuaciones de estado, transiciones de fase, mecánica litosférica, reología de las rocas y dinámica de fluidos multifase (Felpeto, 2002).

Lo complejo de la modelización física se aplica a la gestión de riesgo volcánico, ya que el papel fundamental de la modelización es el de ser capaz de simular los efectos de una erupción, su impacto en el entorno y en el tiempo. Algunos modelos físicos tienen limitaciones y dificultades en la ejecución, no retoman aspectos importantes que los

afectan intrínsecamente. En la búsqueda de modelos físicos, se aconseja descubrir los mejores precursores, no tomar en cuenta parámetros de difícil acceso con comportamientos ilimitados y que tengan la capacidad de reproducir de forma aproximada el evento de erupciones pasadas.

Los modelos físicos de procesos volcánicos, en la gestión de riesgo volcánico no se limitan a la generación y análisis de mapas de peligrosidad, sino también se aplican en ensayos de estrategias mitigadoras del riesgo volcánico. La simulación MonteCarlo calcula la efectividad de medidas de intervención y describe el desarrollo temporal de la erupción para la evaluación de planes de evacuación o sistemas de alertas. Una medida de intervención puede ser la simulación del efecto de la construcción de una barrera de contención versus el desarrollo de una colada de magma. La utilidad de un modelo estadístico en la gestión de riesgo volcánico, dispone de todos los datos que capturan a través de un entorno de Sistemas de Información Geográfica, SIG por sus siglas; las bases de datos vulcanológicas definen los parámetros de entrada del modelo, mientras que los resultados que provienen de las simulaciones numéricas permiten la evaluación cuantitativa del riesgo volcánico.

### **2.3 Introducción a la regresión de datos**

La regresión es una técnica estadística que implica ecuaciones matemáticas que expresan relaciones entre variables de interés y variables predictoras, con esto se obtiene una predicción o pronóstico de variables de interés. El término regresión fue introducido por Francis Galton en mil ochocientos ochenta y seis a sus sesenta y cuatro años. Galton trabajó arduamente en la descripción de los rasgos físicos de sus descendientes, a partir de los de sus padres. Miles de registros de grupos familiares fueron necesarios para concluir que padres muy altos tienden a engendrar hijos que heredan parte de su altura pero que también existía una tendencia de regresar a la media (Pereira, 2010). Introdujo el concepto de recta de regresión y la idea de correlación entre variables en su libro “Natural Inheritance” publicado en mil ochocientos ochenta y nueve.

Reforzar la teoría evolucionista de las especies fue la motivación de Galton al contrastar con empirismo la transmisión de los rasgos hereditarios (Peña, 2002). Galton

nombró esta tendencia como ley de regresión universal: “Cada peculiaridad en un hombre es compartida por sus descendientes, pero en medida, en un grado menor.”, (Pereira, 2010). Se dispone de datos que son las entradas y de sus correspondientes salidas en forma de un sistema lineal o no lineal. La regresión descubre la relación funcional entre la entrada y la salida del sistema, para predecir la salida cuando se presente una nueva entrada en situaciones similares. Existen fenómenos deterministas que se resuelven a través de un sistema lineal, una ecuación diferencial o un sistema no lineal.

Sin embargo, la práctica establece que la mayor parte de fenómenos naturales son no determinísticos; la aleatoriedad, la variabilidad y el comportamiento de variables están presentes en la tendencia que siguen estos fenómenos. En la ecuación regresora se justifica por medio del error aleatorio. En todo modelo de regresión es indispensable el error aleatorio. La predicción de variables de interés se diseña como la suma de un valor fijo y uno aleatorio. El valor fijo es el modelo de regresión, mientras que el aleatorio, son los errores aleatorios. El valor fijo es la parte determinista manipulada por el estadístico, el valor aleatorio es la parte del modelo que no puede controlarse debido a la naturaleza inevitable de los datos.

El objeto de la regresión es la obtención de una ecuación satisfactoria en predicción de valores que pertenecen a fenómenos naturales de interés, ayuda a predecir o pronosticar la veracidad o la falacia de una erupción volcánica. Los matemáticos clásicos establecieron una relación entre la descripción de trayectorias y una ley matemática que las explicase, surgiendo de tal modo la geometría analítica. La geometría analítica se basa en un sistema de ecuaciones que explican el comportamiento de puntos, líneas y curvas en un espacio determinado. En contraste, la estadística inferencial se sirve de ecuaciones para predicciones, pronósticos y aciertos o desaciertos futuros. La regresión utiliza sistemas de ecuaciones de la estadística inferencial actual. Formalmente, la regresión es una técnica estadística para el modelamiento de las variables de interés. El diagrama de dispersión busca relaciones entre dos variables de interés.

## 2.4 Regresión no paramétrica

Son métodos de la inferencia estadística válidos únicamente en la ausencia de hipótesis paramétricas sobre la distribución de los datos. La estimación no paramétrica de curvas es una técnica que estima la función de regresión relacionada con la distribución de probabilidad de los datos. En vulcanología, algunos modelos se basan en la regresión sesgada (suavizados), regresión Kernel (suavizados), Kernel Binarios (topografía, comportamientos por niveles de verosimilitud o regiones de confianza de dos variables) o en la regresión Loess (suavizados), (R Foundation for Statistical Computing, 2010). Los modelos no paramétricos son empíricos; mientras que la regresión paramétrica, tiene su raíz en diversas teorías que sustentan su aplicabilidad.

La sencillez y la fácil interpretación de la regresión paramétrica la convierte en preferible para los análisis, pero hay situaciones en donde ningún modelo es satisfactorio; de tal manera, ninguna transformación es óptima. Bajo estas circunstancias, la regresión no paramétrica tiene más sentido para el ajuste aceptable de los datos sin importar la complejidad relativa de la estimación. El alisador Kernel es un tipo de regresión no paramétrica que se define como un promedio ponderado de los datos. La regresión ponderada localmente, o regresión Lowess (versión anterior a la regresión Loess), es otra alternativa para la regresión no paramétrica. El alisador Kernel permite el manejo de varios valores (nivel de confianza, ponderación, ancho de banda, factor, formato) para generar distintas formas de la densidad estimada.

### 2.4.1 Introducción a la regresión Kernel

Si los datos muestran relaciones no lineales es insuficiente el sesgo introducido en la regresión sesgada, ya que es incapaz de modelar adecuadamente con un error mínimo aceptable. La regresión Kernel realiza un mapeo de los datos de entrada considerados no lineales, en un espacio de características de más alta dimensión; obteniendo de tal forma, errores de ajuste mucho menores que los conseguidos en el espacio de entrada inicial, y conservan la eficiencia del factor de regularización utilizada en la regresión sesgada. Los métodos de suavizado aplicados a un análisis exploratorio de los datos producen gráficos de la función de densidad, la función de regresión, la función de riesgo, entre otras. El

análisis multivariante tiene versiones no lineales y paramétricas de las variables predictoras principales.

En la construcción del modelo se describen fielmente los datos, produciendo modelos con ajuste óptimo. El estadístico experimenta con diferentes Kernels (Kernel de Epanechnikov, Kernel cuartico, Kernel triangular, Kernel gaussiano, o Kernel uniforme) y diferentes parámetros de regularización, para descubrir el mejor modelo explicativo y aplicarlo posteriormente en el análisis de predicciones (ver tabla 2). La validación de estos modelos se realiza a través del análisis del error cuadrado medio. Entre menor es el error cuadrado medio, mejor es el modelo, ya que tiene menos variación. La fortaleza de la regresión Kernel es la utilización de funciones Kernel que permiten construir una función de regresión lineal en un espacio de características de más alta dimensión equivalente a una regresión no lineal en el espacio de entrada. A continuación la eficiencia relativa de cada función Kernel:

**Tabla 2.** Eficiencia relativa de las funciones Kernel más importantes.

Núcleo K	Fórmula	Eficiencia= $\sigma_{K^*}R(K^*)/\sigma_KR(K)$
Epanechnikov K*	$\frac{3}{4}(1-x^2)I_{[-1,1]}(x)$	1
Biweight	$\frac{15}{16}(1-x^2)^2I_{[-1,1]}(x)$	0.994
Triweight	$\frac{35}{32}(1-x^2)^3I_{[-1,1]}(x)$	0.987
Gaussiano	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$	0.951
Triangular	$(1- x )I_{[-1,1]}(x)$	0.986
Uniforme	$\left(\frac{1}{2}\right)I_{[-1,1]}(x)$	0.930

#### 2.4.2 Estimación de la función de densidad univariante

Definición formal de función densidad: sean  $x_1, \dots, x_n$  observaciones independientes de una variable aleatoria  $X$  que tiene función de densidad  $f(x)$ . Sea  $x \in R$ . Se requiere estimar el valor de la función de densidad  $f$  en  $x$ , esto es la función  $f(x)$ . Toda función de densidad tiene las siguientes propiedades:

- $f(x) \geq 0, \forall x/x \in R, \int_{-\infty}^{\infty} f(x)dx = 1$
- $\exists f(x) \sim X \leftrightarrow \forall a, b; a \leq b/a \in R \wedge b \in R, P(a < X \leq b) = \int_a^b f(x)dx$
- Sea  $dx$  una longitud pequeña,  $f(x) \approx \frac{P(X \in [x, x+dx])}{dx}$
- Sea  $F(x)$  la función de distribución de  $X$ , entonces se tiene el siguiente resultado  

$$F(x) = \int_{-\infty}^x f(u)du, f(x) = \frac{d}{dx}F(x) = F'(x)$$

Aunque existan supuestos paramétricos para estimar la función de densidad de distribución de las observaciones, debido a la inflexibilidad en la detección de desviaciones de la hipótesis paramétrica, se aplican supuestos no paramétricos que validan la estimación de la función de densidad. Los mejores estimadores no paramétricos de la función de densidad son el histograma y el polígono de frecuencias. El histograma es muy simple en cálculo y en interpretación, el aspecto depende del ancho de los intervalos; si el ancho es pequeño el histograma tiene poco sesgo y mucha varianza, de lo contrario el histograma tiene mucho sesgo y poca varianza. No es un estimador suave de la función de densidad, es discontinuo y constante a intervalos.

Sea  $x_i$  un punto fijo y  $\hat{f}_H(x)$  estimador del valor desconocido  $f(x)$ ; se emplea el análisis del error cuadrático medio MSE para la evaluación local de la estimación, entre menor sea el valor, mejor será la estimación:

$$MSE(\hat{f}_H(x)) = E[(\hat{f}_H(x) - f(x))^2] = (\text{Sesgo}(\hat{f}_H(x)))^2 + V(\hat{f}_H(x))$$

El histograma es mejor estimador en el centro de cada intervalo que en otros puntos, entonces  $x_i$  es el punto donde se quiere estimar la densidad y  $B_x = [x - h, x + h]$  el intervalo, donde  $h$  es el parámetro de suavizado; así que el estimador de  $f(x)$  es:

$$\hat{f}_U(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} I_{[-1,1]} \left( \frac{x - x_i}{h} \right)$$

La función  $\hat{f}_U(x)$  no es suave (es discontinua y constante a trozos), la razón es que en la expresión aparece la función de densidad de variable aleatoria uniforme  $U([-1,1])$ :

$$g(u) = \frac{1}{2} I_{[-1,1]}(u)$$

Ahora bien, la localización de la estimación se calcula con la función  $\hat{f}_U(x)$ ; si se sustituye esa densidad por otra  $K(u)$  más suave (derivable varias veces), se obtiene un estimador de la densidad que hereda esas propiedades de suavidad; el estimador resultante se denomina estimador núcleo o estimador Kernel:

$$\hat{f}_K(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} K\left(\frac{x-x_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x-x_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x-x_i)$$

La función  $K$  se llama función núcleo o Kernel, es una función de densidad continua, unimodal y simétrica alrededor de cero. El último resultado de la ecuación normal anterior significa que hay  $n$  densidades (con pesos  $1/n$ ) con la misma forma que el núcleo  $K$ , reescaladas según el parámetro  $h$ , y centradas cada una en la observación  $x_i$ . El valor del parámetro de suavizado  $h$ , se llama ventana o amplitud de banda, controla la concentración del peso  $1/n$  alrededor de cada  $x_i$ . Si  $h$  es un valor pequeño, únicamente las observaciones  $x_i$  más cercanas a  $x$  serán relevantes en la estimación de  $f(x)$ , valores grandes de  $h$  permiten que observaciones más alejadas a  $x$  también intervengan en la estimación de  $f(x)$ .

La estimación final es afectada por los cambios en la elección del parámetro de suavizado; por eso, la elección de dicho parámetro es crucial en la estimación no paramétrica de la densidad. Valores grandes de  $h$  hacen que los estimadores de la densidad sean muy estables de muestra en muestra, es decir tienen poca varianza aunque las estimaciones son más sesgadas. Por el contrario, si  $h$  es pequeño el estimador varía mucho en muestras diferentes, hay mucha varianza aunque estima bien la densidad desconocida, hay poco sesgo. Una función núcleo o Kernel se define con las propiedades:

- *K es simétrica alrededor de cero:* implica que el peso  $1/n$  de cada dato observado se reparte de forma simétrica alrededor de la observación. Es deseable, pero no imprescindible.
- *K es unimodal:* tiene moda cero si es simétrica. El peso  $1/n$  de cada dato observado se reparte de forma que queda más peso en las zonas más cercanas a la observación. Es deseable, pero no imprescindible.

- *K es positiva*:  $\forall u/u \in R, K(u) \geq 0$ , no es necesaria para que el estimador núcleo tenga buenas propiedades asintóticas.
- *K integra 1*:  $\int_R K(u)du = 1$ , es una propiedad necesaria para que el sesgo asintótico del estimador sea nulo.
- *K tiene momento de orden 1 nulo*:  $\int_R uK(u)du = 0$ , se cumple si  $K$  es simétrica y tiene esperanza matemática. Si  $K$  no tiene esperanza 0 entonces el sesgo del estimador decrece más lentamente hacia 0.
- *K tiene momento de orden 2 finito*:  $\int_R u^2K(u)du = \sigma_K^2 < \infty$ , que la varianza de  $K$  sea finita es necesario para que el estimador tenga sesgo asintótico acotado. También se pueden construir núcleos no positivos con momento de orden 2 nulo que permiten reducir el sesgo asintótico, a veces son llamados núcleos de alto orden.
- *K es una función suave*: es decir tiene varias derivadas continuas. El estimador núcleo hereda las propiedades de suavidad del núcleo  $K$  a partir del que se define. Así que hay que utilizar núcleos suaves para obtener estimadores suaves.
- *K tiene soporte compacto*: es deseable computacionalmente, ya que si  $K(u)$  se anula fuera del intervalo  $[-c, c]$ , entonces para evaluar  $\hat{f}_K$  en un punto  $x$ , sólo hay que utilizar los puntos  $x_i$  situados en  $[x - ch, x + ch]$ .

Demostración del algoritmo para el cálculo de la esperanza matemática de las estimaciones del núcleo de la densidad a través de funciones Kernel. Para la demostración es necesaria la definición de convolución: Sean  $X \sim f, Y \sim g$  dos variables aleatorias independientes. La convolución  $f * g$  de las densidades  $f$  y  $g$  es la función de densidad de  $X + Y$  y vale:

$$(f * g)(x) = \int_R f(x - y)g(y)dy$$

Con la definición de convolución de dos variables aleatorias independientes con densidades conocidas, se procede a la demostración para la esperanza matemática de la estimación de la función de densidad utilizando funciones Kernel:

$$E(\hat{f}(x)) = E\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x-x_i}{h}\right)\right) = E(K_h(x-x_i))$$

$$E(\hat{f}(x)) = \int_R K_h(x-u)f(u)du = (K_h * f)(x) = \int_R \frac{1}{h} K\left(\frac{x-u}{h}\right) f(u)du$$

(por simetría de K)

$$E(\hat{f}(x)) = \int_R \frac{1}{h} K\left(\frac{u-x}{h}\right) f(u)du$$

(cambio de variable  $v = (u-x)/h$ ,  $dv = \left(\frac{1}{h}\right) du$ )

$$E(\hat{f}(x)) = \int_R K(v)f(x+hv)dv$$

(Taylor:  $f(x+hv) = f(x) + f'(x)hv + \frac{1}{2}f''(x)h^2v^2 + O(h^3)$ )

$$E(\hat{f}(x)) = \int_R K(v)\left(f(x) + f'(x)hv + \frac{1}{2}f''(x)h^2v^2 + O(h^3)\right)dv$$

$$E(\hat{f}(x)) = f(x) \int_R K(v)dv + f'(x)h \int_R vK(v)dv + \frac{1}{2}f''(x)h^2 \int_R v^2K(v)dv + O(h^3)$$

$$E(\hat{f}(x)) = f(x) + \frac{f''(x)\sigma_K^2 h^2}{2} + O(h^3)$$

Ahora, se procederá al cálculo de la varianza de las estimaciones del núcleo de la densidad a través de funciones Kernel, al igual que el cálculo anterior, siempre se necesitará de la definición de convolución, simetría y series de Taylor:

$$V(\hat{f}(x)) = V\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x-x_i}{h}\right)\right) = \frac{1}{n} V(K_h(x-x_i))$$

$$V(\hat{f}(x)) = \frac{1}{n} \left[ E\left(K_h^2(x-x_i)\right) - E(K_h(x-x_i))^2 \right]$$

$$V(\hat{f}(x)) = \frac{1}{n} \left[ \int_R K_u^2(x-u)f(u)du - \left( \int_R K_u(x-u)f(u)du \right)^2 \right]$$

$$V(\hat{f}(x)) = \frac{1}{n} \left[ (K_h^2 * f)(x) - (K_h * f)^2(x) \right]$$

(por simetría de K)

$$V(\hat{f}(x)) = \frac{1}{n} \int_R \frac{1}{h^2} K^2\left(\frac{u-x}{h}\right) f(u) du - \frac{1}{n} (f(x) + O(h^2))^2$$

(cambio de variable:  $v = (u-x)/h$ ,  $dv = (1/h)du$ )  
(el último sumando es  $O(1/n)$ )

$$V(\hat{f}(x)) = \int_R K(v) \left( f(x) + f'(x)hv + \frac{1}{2}f''(x)h^2v^2 + O(h^3) \right) dv$$

$$V(\hat{f}(x)) = \frac{1}{nh} \int_R K^2(v) f(x+ hv) dv + \left(\frac{1}{n}\right)$$

(Taylor:  $f(x+ hv) = f(x) + O(h)$ )

$$V(\hat{f}(x)) = \frac{1}{nh} \int_R K^2(v) f(x) dv + \frac{1}{nh} O(h) + O\left(\frac{1}{n}\right)$$

$$V(\hat{f}(x)) = \frac{f(x)}{nh} \int_R K^2(v) dv + O\left(\frac{1}{n}\right) = \frac{f(x)R(K)}{nh} + O\left(\frac{1}{n}\right)$$

### 2.4.3 Estimación de la función de densidad multivariante

Partiendo de lo anterior, el estimador núcleo de la densidad multivariante se plantea de la generalización natural de la estimación univariante. Donde  $K_d: R^d \rightarrow R$  es una función núcleo d-dimensional que verifica  $\int_{R^d} K_d(u) du = 1$ ;  $\int_{R^d} u K_d(u) du = 0$ ,  $0 \in R^d$ ; y  $\int_{R^d} uu^T K_d(u) du = I_d$ .

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_d(x - x_i)$$

Se toma una función densidad centrada en el  $0 \in R^d$  para que la estimación sea una mezcla de  $n$  densidades, cada una de ellas centrada en una de las observaciones  $x_i \in R^d$ . Para el cálculo de la estimación de la función de densidad utilizando el Kernel binario, se parte de que  $H = \text{Diag}(h_1, \dots, h_d)$ , donde  $h_j$  es un parámetro de suavizado adecuado para la j-ésima coordenada de  $X$ . La mayoría de paquetes estadísticos incorporan la estimación no paramétrica de densidades multivariantes, además se usa un núcleo producto, entonces el estimador del núcleo de densidad queda así:

$$\hat{f}(x) = \frac{1}{n \prod_{j=1}^d h_j} \sum_{i=1}^n \prod_{j=1}^d K\left(\frac{x_j - x_{ij}}{h_j}\right)$$

#### 2.4.4 Bandas de variabilidad y bandas de referencia normal

Las bandas muestran cómo es la variabilidad del estimador no paramétrico de la densidad  $f(x)$ . Son bandas puntuales y no uniformes de variabilidad, el propósito de estas bandas no es el mismo que el de las bandas de confianza. La siguiente fórmula genera las bandas de variabilidad puntuales para estimación de la función de densidad denotada por  $f(x)$ :

$$\left( \left( \sqrt{\hat{f}(x)} - z_{\frac{\alpha}{2}} \sqrt{\frac{R(K)}{4nh}} \right)^2, \left( \sqrt{\hat{f}(x)} + z_{\frac{\alpha}{2}} \sqrt{\frac{R(K)}{4nh}} \right)^2 \right), x \in R$$

Para el contraste de la normalidad del estimador no paramétrico de la densidad  $\hat{f}(x)$  se recurre a las bandas de referencia normal generadas por la siguiente fórmula:

$$\left( E(\widehat{f}(x)) - z_{\frac{\alpha}{2}} \sqrt{V(\widehat{f}(x))}, E(\widehat{f}(x)) + z_{\frac{\alpha}{2}} \sqrt{V(\widehat{f}(x))} \right)$$

Donde la media y la varianza muestrales son estimadores de  $\mu$  y  $\sigma^2$ , respectivamente; sus definiciones son las siguientes:

$$E(\widehat{f}(x)) = f_N(x; \mu, \sigma^2 + h^2)$$

$$V(\widehat{f}(x)) = \frac{1}{n} \left[ f_N(0; 0, 2h^2) f_N\left(x; \mu, \sigma^2 + \frac{1}{2}h^2\right) - f_N(x; \mu, \sigma^2 + h^2)^2 \right]$$

#### 2.4.5 Estimación de la función de regresión

Después de la estimación de la función de densidad, el método para la estimación de la función de regresión utilizando funciones Kernel consiste en definir el peso de  $(x_i, y_i)$  en la estimación de  $m(t)$ , es decir:

$$w_i = w(t, x_i) = \frac{K\left(\frac{x_i - t}{h}\right)}{\sum_{j=1}^n K\left(\frac{x_j - t}{h}\right)}$$

Donde  $h$  es un parámetro de escala que controla la concentración del peso total alrededor de  $t$ . Si  $h$  es pequeño, las observaciones más cercanas a  $t$  tendrán mayor ponderación, de lo contrario también se toman en cuenta las observaciones alejadas a  $t$  en la estimación de  $m(t)$ . También se le llama parámetro de suavizado o ventana y

controla el grado de localidad o de globalidad de la estimación. Sea  $m(x)$  la función de regresión:

$$m(x) = E(Y|X = x) = \int_{\mathcal{R}} y f_Y(y|X = x) dy = \int_{\mathcal{R}} y \frac{f(x, y)}{f_X(x)} dy$$

El modelo de regresión no paramétrica sugiere un estimador de  $m(x)$  obtenido a partir de los estimadores no paramétricos de las densidades  $f_X(x)$  y  $f(x, y)$ . Sean estos estimadores:

$$\hat{f}(x, y) = \frac{1}{nh_X h_Y} \sum_{i=1}^n K_X\left(\frac{x - x_i}{h_X}\right) K_Y\left(\frac{y - y_i}{h_Y}\right)$$

$$\hat{f}_X(x) = \frac{1}{nh_X} \sum_{i=1}^n K_X\left(\frac{x - x_i}{h_X}\right)$$

El estimador de  $m(x)$  se obtiene sustituyendo las densidades desconocidas por estos estimadores, al final se consigue la expresión del estimador de la función de regresión de Naradaya-Watson:

$$\hat{m}(x) = \int_{\mathcal{R}} y \frac{\hat{f}(x, y)}{\hat{f}_X(x)} dy$$

$$\hat{m}(x) = \int_{\mathcal{R}} y \frac{\frac{1}{nh_X h_Y} \sum_{i=1}^n K_X\left(\frac{x - x_i}{h_X}\right) K_Y\left(\frac{y - y_i}{h_Y}\right)}{\frac{1}{nh_X} \sum_{i=1}^n K_X\left(\frac{x - x_i}{h_X}\right)} dy$$

$$\hat{m}(x) = \frac{\sum_{i=1}^n K_X\left(\frac{x - x_i}{h_X}\right) \int_{\mathcal{R}} y \frac{1}{h_Y} K_Y\left(\frac{y - y_i}{h_Y}\right) dy}{\sum_{i=1}^n K_X\left(\frac{x - x_i}{h_X}\right)}$$

(cambio de variable  $u = (y - y_i)/h_Y$  de la ecuación  $y = y_i + h_Y u$ )

$$\hat{m}(x) = \frac{\sum_{i=1}^n K_X\left(\frac{x - x_i}{h_X}\right) \int_{\mathcal{R}} (y_i + h_Y u) K_Y(u) du}{\sum_{i=1}^n K_X\left(\frac{x - x_i}{h_X}\right)}$$

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x - x_i}{h_X}\right) y_i}{\sum_{i=1}^n K\left(\frac{x - x_i}{h_X}\right)}$$

### 2.4.6 Regresión Loess

La regresión Loess proviene de regresiones pesadas localmente lineales para suavizar los datos, se trata de suavizados de diagramas de dispersión. La regresión polinomial local es una generalización de estimación de núcleos. El objetivo de la regresión polinomial local es ajustar un polinomio de grado  $p$  alrededor de un punto utilizando los datos de un entorno. La regresión Loess ajusta una superficie polinomial determinada por uno o más predictores. La diferencia entre regresión Lowess y Loess es el límite de suavizado, la regresión Lowess suaviza en una dimensión; la regresión Loess es multidimensional. La regresión polinómica local tiene su origen en el teorema de Taylor.

La regresión Loess es flexible en comparación con las técnicas de modelamiento tradicional, ya que se usa para situaciones con forma paramétrica óptima de la superficie de regresión desconocida. También se utiliza cuando hay puntos atípicos en los datos, utiliza el método de ajuste robusto. La regresión polinomial local es una generalización de la estimación por núcleos, se ajusta un polinomio de grado cero, es decir una constante en un entorno. El objetivo de la regresión Loess es ajustar un polinomio de grado  $p$  alrededor de un punto utilizando los datos de un entorno; el ajuste puede ser de núcleo (de grado cero), lineal local (de primer grado), cuadrática local (parabólica), etcétera.

Supóngase que existe la derivada de orden  $(p + 1)$  de la función de regresión  $\varphi(\cdot)$  en un punto  $z_0 \in R$ , entonces se aproxima la función de regresión desconocida  $\varphi(z)$  localmente por un polinomio de grado  $p$  en un entorno de  $z_0$ . Aplicando el teorema de Taylor en un punto  $z$  próximo a  $z_0$  se tiene:

$$\varphi(z) = \varphi(z_0) + \varphi'(z_0)(z - z_0) + \dots + \frac{\varphi^{(p)}(z_0)}{p!} (z - z_0)^p$$

Este polinomio puede ajustarse localmente utilizando la técnica de mínimos cuadrados locales ponderados, es decir, minimizando la función:

$$\Psi(\beta) = \sum_{i=1}^n \left\{ X_i - \left( \sum_{j=0}^p \beta_j (z_i - z_0)^j \right) \right\}^2 K_{h_n}(z_i - z_0)$$

Donde  $\beta_j = \frac{\varphi^j(z_0)}{j!}, j = 0, 1, \dots, p$ ,  $h_n$  es el parámetro de suavizado y la función de densidad simétrica con soporte compacto  $K_{h_n}(u) = (1/h_n)K(u/h_n)$ . Denotando  $\hat{\beta}_j, j = 0, \dots, p$  la solución del problema de mínimos cuadrados resulta obvio por el desarrollo de Taylor que  $\hat{\varphi}_j(z_0) = j! \hat{\beta}_j$  es un estimador de  $\varphi^j(z_0), j = 0, 1, \dots, p$ . Para estimar la función  $\varphi^j(\cdot)$  se resolverá el problema de mínimos cuadrados ponderados para cada uno de los puntos  $z_0$  en el intervalo de interés:

$$\Lambda = \begin{pmatrix} 1 & (z_1 - z_0) & \dots & (z_1 - z_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (z_n - z_0) & \dots & (z_n - z_0)^p \end{pmatrix} \in \mathbb{M}_{n \times (p+1)} \quad X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n$$

$$W_n = \begin{pmatrix} K_{h_n}(z_1 - z_0) & 0 & \dots & 0 \\ 0 & K_{h_n}(z_2 - z_0) & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_{h_n}(z_n - z_0) \end{pmatrix} \in \mathbb{M}_{n \times (p+1)} \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \in \mathbb{R}^{p+1}$$

Minimizando en notación matricial se tiene  $\Psi(\beta) = (X - \Lambda\beta)^t W_n (X - \Lambda\beta)$ . Siendo  $\beta = (\beta_0, \dots, \beta_p)^t$ , utilizando la teoría de mínimos cuadrados ponderados se obtiene el estimador  $\hat{\beta}(z_0) = (\Lambda^t W_n \Lambda)^{-1} \Lambda^t W_n X = S_n^{-1} T_n$ . Obviamente, para la definición del estimador polinómico local es necesario que exista la matriz  $S_n^{-1} = (\Lambda^t W_n \Lambda)^{-1}$ , en las expresiones matriciales se ha denotado:

$$S_n = \begin{pmatrix} s_0^n & s_1^n & \dots & s_p^n \\ s_1^n & s_2^n & \dots & s_{p+1}^n \\ \vdots & \vdots & \ddots & \vdots \\ s_p^n & s_{p+1}^n & \dots & s_{2p}^n \end{pmatrix} \in \mathbb{M}_{(p+1) \times (p+1)} \quad s_j^n = \sum_{i=1}^n K_{h_n}(z_i - z_0) (z_i - z_0)^j$$

$$T_n = \begin{pmatrix} t_0^n \\ t_1^n \\ \vdots \\ t_p^n \end{pmatrix} \in \mathbb{R}^{p+1} \quad t_j^n = \sum_{i=1}^n K_{h_n}(z_i - z_0) (z_i - z_0)^j X_i$$

## 2.5 Curvas de Andrews

Este método se basa en las transformaciones de Fourier para representar datos multivariantes en dos dimensiones, grados o radianes en las abscisas y la componente de Fourier en la ordenada; la transformación de Fourier se basa en una función alternante de senos y cosenos. Cada curva recoge y resume información de todas las variables que

afectan al individuo, de manera que las diferencias entre individuos para alguna de ellas debe resultar manifiesta en alguna parte de las respectivas curvas (Andrews, 1972). El autor de esta técnica desaconseja superar las diez observaciones, ya que un conjunto numeroso de curvas resulta desconcertante. Lo mejor es tratar las curvas por separado, aunque si hay variables más importantes estas deben utilizar los valores iniciales de la serie.

Las curvas que son distintas a las demás corresponden a la presencia de datos atípicos; un subgrupo tiene un conjunto de curvas similares; el orden de las variables es muy importante en la interpretación; el orden de las variables se optimiza con un análisis de componentes principales; con un elevado número de variables se obtiene una mala señal. La distancia euclidiana se encuentra presente en la comparación de curvas próximas, ya que los valores de los elementos de ambos vectores son semejantes. Cada elemento de cada vector contribuye de forma individual a la curva, puede afectar la frecuencia, la amplitud y la periodicidad de la función, obteniéndose así una única representación bidimensional.

Sean  $f_m(t)$  y  $f_n(t)$  funciones generadoras de curvas de Andrews con  $-\pi < t < \pi$ . Si  $\forall t_0 / -\pi < t_0 < \pi$  es un punto cualquiera de las abscisas de las funciones anteriores respectivamente,  $f_m(t_0), f_n(t_0) \in [-\infty, \infty]$  los valores correspondientes en el eje de las ordenadas, entonces la distancia euclidiana entre las curvas de Andrews  $f_m(t)$  y  $f_n(t)$  se define como:  $d_{m,n} = \sqrt{[f_n(t_0) - f_m(t_0)]^2 + (t_0 - t_0)^2} = \sqrt{[f_n(t_0) - f_m(t_0)]^2}$ . Si  $d_{m,n} \cong 0 \rightarrow f_m(t_0) \cong f_n(t_0)$ . De forma general  $d_i = \sqrt{[f_n(t_i) - f_m(t_i)]^2 + (t_i - t_i)^2}$ , debido que se trata de un análisis al mismo instante se tiene  $d_i = \sqrt{[f_n(t_i) - f_m(t_i)]^2}$ ; aplicando sumatoria para el caso discreto  $\sum_i d_i = \sum_i \sqrt{[f_n(t_i) - f_m(t_i)]^2}$ , e integrando ambos miembros de la igualdad para el caso continuo resulta que:

$$\int_{-\pi}^{\pi} \sqrt{[f_n(t) - f_m(t)]^2} dt = \pi \sum_i [f_n(t_i) - f_m(t_i)]^2$$

Si  $\sum_i d_i \cong 0 \rightarrow \forall t_i, f_m(t_i) \cong f_n(t_i) \therefore f_m(t) \cong f_n(t)$ . También el significado es análogo para el caso continuo, donde:  $\int_{-\pi}^{\pi} \sqrt{[f_n(t) - f_m(t)]^2} dt \cong 0$ . Si se detectan

subconjuntos con distancias aproximadamente iguales, el resultado es una partición del total de observaciones, en grupos con características semejantes (ver tabla 3). En vulcanología, ya sea un estado sísmico-volcánico normal, relativamente normal, o la presencia de una anomalía segura, permite analizar plenamente el comportamiento propio de cada volcán, incluso pueden existir coincidencias entre volcanes. Si los datos son incorrelados, conservan varianza común  $\sigma^2$ , entonces la varianza de valores de la función para cada  $t$  es igual a:

$$\sigma_t^2 = \sigma^2 \left( \frac{1}{2} + \sin^2 t + \cos^2 t + \sin^2 2t + \cos^2 2t + \dots \right)$$

**Tabla 3.** Detalle de la existencia de siete o menos clúster.

Funciones	Distancias euclidianas $i \neq j$	Resultados posibles
$f_1(t)$	$\int_{-\pi}^{\pi} \sqrt{[f_i(t) - f_j(t)]^2} dt$	$k_l$
$f_2(t)$		$k_m$
$f_3(t)$		$k_n$
...		$k_o$
$f_i(t)$		$k_p$
...		$k_q$
$f_n(t)$		$k_r$

También es posible inferir:

$$\sigma_t^2 = \begin{cases} \frac{1}{2} p \sigma^2, & \text{si } p \text{ es impar} \\ \sigma^2 \left( \frac{p-1}{2} \right) \leq \sigma_t^2 \leq \sigma^2 \left( \frac{p+1}{2} \right), & \text{si } p \text{ es par} \end{cases}$$

La función mantiene las medias aritméticas, sea  $\bar{x}$  la media de las  $k$  observaciones del vector  $X_i$ ; para cada valor de  $t$ , el valor de la función que corresponda a  $\bar{x}$ , es la media de los  $k$  valores de la función aplicada a  $X_i$ . Las curvas de Andrews han sido aplicadas al análisis de los servicios de salud de la República de Chile para analizar el comportamiento de dotaciones, días camas disponibles, días camas ocupadas, días estada, promedio camas disponibles, número egresos y número egresos fallecidos; obteniéndose la identificación de varios conglomerados de ubicaciones geográficas en todo el territorio chileno (Schiattino, 2013). De individualidad de las observaciones se interpretan los resultados, determinando tendencias, estándares y clústeres (Schiattino, 2013).

Una nación que cuente con un número considerable de complejos volcánicos puede efectuar una investigación para la búsqueda de tendencias semejantes en el comportamiento de cada volcán y un posterior análisis de semejanza significativa, un planteamiento de hipótesis o la validación estadística del clúster. Las curvas de Andrews se aplican únicamente a datos cuantitativos de cualquier dimensión, se fundamenta en matemática sofisticada; la aplicación, resultados y análisis son interesantes. Sea  $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ik})$  el vector  $i$ -ésimo de dimensión  $1 \times k$ , representado gráficamente en dos dimensiones por las curvas de Andrews en el intervalo  $-\pi < t < \pi$ ; y  $f_i(t)$  la definición de la función generadora de las curvas de Andrews para la observación  $i$ -ésima, la curva que representa la media de la población y una variante de las curvas de Andrews se detallan a continuación respectivamente:

$$f_{ix}(t) = \frac{x_{i1}}{\sqrt{2}} + x_{i2} \sin t + x_{i3} \cos t + x_{i4} \sin 2t + x_{i5} \cos 2t + x_{i6} \sin 3t + \dots$$

$$f_{i\mu}(t) = \frac{\mu_{i1}}{\sqrt{2}} + \mu_{i2} \sin t + \mu_{i3} \cos t + \mu_{i4} \sin 2t + \mu_{i5} \cos 2t + \mu_{i6} \sin 3t + \dots$$

$$f_{ix}(t) = x_{i1} \sin 2t + x_{i2} \cos 2t + x_{i3} \sin 2^2t + x_{i4} \cos 2^2t + x_{i5} \sin 2^3t + \dots$$

Otra clase de curva estudiada y definida por Andrews incluye el concepto de potencia y elimina el término independiente de la serie, así como también a los términos que llevan  $\sin t$  y  $\cos t$ , además la sucesión aritmética de los argumentos es transformada a una sucesión geométrica. Actualmente, se desconoce la relación concreta entre las crestas de la curva y los valores de las variables, ya que los datos brutos son transformados por las curvas de Andrews. Las gotas de Fourier son transformaciones polares de las curvas de Andrews. La información contenida en ambas técnicas es exactamente la misma, la apariencia es distinta. Consiste en transformar las ondas producto de la alternancia de senos y cosenos a un sistema de coordenadas polares, las ondas se convierten en gotas, manchas o amebas. Cada vector  $X_i$  se representará por una gota, la forma de las gotas depende de los valores de los elementos de cada vector; la ventaja de las gotas de Fourier, es que las gotas no se solapan como ocurre en las curvas de Andrews.

## 2.6 Introducción a las series temporales

Una serie temporal es una secuencia de valores ordenados cronológicamente. Existen dos tipos de series temporales: las estocásticas y las deterministas. Si con los valores anteriores de una serie temporal no es posible predecir con total certeza el próximo valor de la variable, entonces se dice que se trata de una serie temporal estocástica o aleatoria (Molinero, 2004). El interés de esta investigación se centra en las series temporales no deterministas. Un proceso estocástico es la secuencia ordenada de variables aleatorias y su distribución de probabilidad asociada. Se define mediante una componente regular, una componente irregular y el ruido aleatorio.

El objetivo del análisis de series temporales es la predicción, el control de un proceso, la simulación de un proceso, y la generación de nuevas teorías físicas o biológicas. Un campo notorio en el que se aplica este análisis, es en la meteorología o en la predicción de otros fenómenos naturales (Molinero, 2004). El control de procesos sigue la evolución de una variable determinada con el fin de regular su resultado. La estacionalidad de una serie temporal se refiere a propiedades estadísticas en el tiempo. En otras palabras la estacionalidad es un comportamiento estadístico en el tiempo, el cual tiene una constante distribución de probabilidad. Usualmente se consideran los primeros dos momentos de la serie temporal, ya que definen una fuerte o débil estacionalidad (Montgomery, 2008):

- Los valores esperados de la serie no dependen del tiempo.
- La Función de Autocorrelación para cualquier periodo  $k$ , sólo es en función de los periodos y no del tiempo.

Una serie es estacionaria cuando se encuentra en equilibrio estadístico, es decir, sus propiedades no varían a lo largo del tiempo, y por lo tanto no existen tendencias. Es necesario comprender los conceptos de tendencia, estacionalidad o variación periódica, y otras fluctuaciones irregulares. La tendencia es la dirección general de la variable en el periodo de observación; la estacionalidad son fluctuaciones periódicas de la variable, en periodos relativamente cortos. Al extraer la tendencia y la variación periódica, surge una serie temporal de valores residuales, los cuales pueden o no ser aleatorios (Molinero,

2004). La tendencia es una función de tiempo que describe la evolución lenta y a largo plazo del nivel medio de la serie. Esta función depende de parámetros que deben estimarse (Gómez, 2006). El modelo completo puede ser aditivo o multiplicativo (ver tabla 4), si  $y_t$  es la definición del modelo,  $S_t$  la componente regular estacionaria,  $T_t$  la componente regular de tendencia, y  $\varepsilon_t$  la componente del error aleatorio:

**Tabla 4.** Descomposición de series temporales

<b>Modelo aditivo</b>	<b>Modelo multiplicativo</b>	<b>Implicaciones</b>
<p><i>Componente regular</i> + <i>componente irregular</i></p> $y_t = S_t + T_t + \varepsilon_t$	<p><i>Componente regular</i> × <i>componente irregular</i></p> $y_t = S_t T_t \varepsilon_t$ <p>ó</p> $y_t = S_t T_t e^{\varepsilon_t}$	$\left\{ \begin{array}{l} T_t = \beta_0 + \beta_1 t \\ T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \\ T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 \\ T_t = e^{(\beta_0 + \beta_1 t)} \\ T_t = \frac{\beta_2}{1 + \beta_1 e^{-\beta_0 t}} \\ \ln T_t = \beta_0 + \beta_1 t + \eta_t \\ y_t - T_t = S_t + \varepsilon_t \\ \frac{y_t}{T_t} = S_t + \varepsilon_t \end{array} \right.$

Las funciones de tendencias son de naturaleza lineal, cuadrática, cúbica, exponencial, logística, o logarítmica lineal. Para el caso de una tendencia cuadrática, la función será monótona creciente si los dos parámetros son mayores que cero; por el contrario, se trata de una función monótona decreciente. Si el parámetro beta uno es mayor que cero y el parámetro beta dos es menor que cero, entonces la función es cóncava hacia abajo o simplemente cóncava; inversamente, la función es cóncava hacia arriba o convexa.

Para el análisis de tendencia de la serie temporal (eliminación de la estacionalidad de la serie), se aplican diferentes filtros a los datos brutos. Un filtro es una función matemática que se aplica a la serie original para la obtención de una nueva serie con características determinadas y sin estacionalidad. Los filtros empleados en esta investigación son, las Medias Móviles, las diferencias y el Suavizado Exponencial (ver tabla 5); aunque también se utiliza el ajuste de polinomios, el alisado mediante funciones exponenciales, Kernel, Loess, Lowess, entre otros; la filtración de la tendencia constituye la componente regular. Las fórmulas matemáticas que se utilizan en los filtros son las siguientes:

**Tabla 5.** *Filtros que se aplican en la descomposición de series de tiempo.*

Filtro	Expresión matemática	Significado
	$m(x_t) = \frac{x_{t-1} + x_t + x_{t+1}}{3}$	Media móvil orden 3
Medias Móviles	$m(x_t) = \frac{x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}}{4}$ $Var(m(x_t)) = \frac{\sigma^2}{N}$	Media móvil de orden cuatro (cuando es par el orden se toma la mitad de los extremos)
Diferencias	$\nabla x_{t+1} = x_{t+1} - x_t$ $\nabla^2 x_{t+2} = x_{t+2} - \nabla x_{t+1}$ $\nabla^d x_{t+d} = x_{t+d} - \nabla^{d-1} x_{t+d-1}$	Diferencia de orden 1 Diferencia de orden 2 Diferencia de orden d
Suavizador exponencial	$\tilde{y}_1 = \lambda y_1 + (1 - \lambda) \tilde{y}_0$ $\tilde{y}_2 = \lambda y_2 + (1 - \lambda) \tilde{y}_1 = \lambda(y_2 + (1 - \lambda)y_1) + (1 - \lambda)^2 \tilde{y}_0$ $\tilde{y}_3 = \lambda(y_3 + (1 - \lambda)y_2 + (1 - \lambda)^2 y_1) + (1 - \lambda)^3 \tilde{y}_0$ $\tilde{y}_T = \lambda(y_T + (1 - \lambda)y_{T-1} + \dots + (1 - \lambda)^{T-1} y_1) + (1 - \lambda)^T \tilde{y}_0$	Simple u orden 1 Suavizado de orden 2 Suavizado de orden 3 Suavizado de orden T
Regresión Loess y Kernel	Regresión local	Diferentes suavizados

En comparación a las funciones deterministas de tendencia, los filtros convierten las señales en procesos estocásticos. En el método de suavizado exponencial, los datos se han designado como  $Y = y_0, y_1, \dots, y_T$ , mientras que para el resto como  $X = x_1, x_2, \dots, x_T$ ; sólo es cuestión de nomenclatura. Para el caso de la suavización exponencial, es imprescindible contar con el valor inicial  $y_0$ ; si el proceso comienza constante localmente, entonces se prefiere  $\tilde{y}_0 = \bar{y}$ . Si los cambios en el proceso se esperan que ocurran al instante y rápidamente, entonces es razonable  $\tilde{y}_0 = y_1$ . Si el valor del parámetro de suavizado lambda es cercano a la unidad, el suavizado es débil y hace más énfasis en la última observación; si el parámetro de suavizado lambda es cercano a cero, el suavizado es fuerte y entre más cercano a cero esté, el suavizado parecerá como una constante.

Si se asume varianza independiente y constante, entonces la varianza del suavizado exponencial es:

$$\begin{aligned}
 Var(\tilde{y}_T) &= Var\left(\lambda \sum_{t=0}^{\infty} (1 - \lambda)^t y_{T-t}\right) \\
 Var(\tilde{y}_T) &= \lambda^2 \sum_{t=0}^{\infty} (1 - \lambda)^{2t} Var(y_{T-t}) \\
 Var(\tilde{y}_T) &= \lambda^2 \sum_{t=0}^{\infty} (1 - \lambda)^{2t} Var(y_T) \\
 Var(\tilde{y}_T) &= Var(y_T) \lambda^2 \sum_{t=0}^{\infty} (1 - \lambda)^{2t} \\
 Var(\tilde{y}_T) &= \frac{\lambda}{(2 - \lambda)} Var(y_T)
 \end{aligned}$$

Para el análisis de estacionalidad de una serie temporal se utiliza la Función de Autocorrelación; esta función mide la correlación entre los valores de la serie

distanciados un lapso de tiempo  $k$ , en computación el lapso se llama “lag”. Si hay  $N$  pares  $x, y$ ; el coeficiente de correlación simple es:

$$r = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (x_i - \bar{x})^2}} = \frac{Cov(x, y)}{Var(y)Var(x)}$$

Si existe una secuencia temporal de  $N$  observaciones  $x_1, \dots, x_N$ ; entonces se forman  $N - 1$  parejas de observaciones contiguas, esto es  $(x_1, x_2), (x_2, x_3), \dots, (x_{N-1}, x_N)$ . Al calcular los coeficientes de correlación, se obtiene la Función de Autocorrelación de orden uno y se denota como  $r_1$  (ver tabla 6).

**Tabla 6.** Coeficientes de autocorrelación que forman la Función de Autocorrelación.

Definición del coeficiente	Relación de los datos	Orden
$r_0 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$	Varianza de la serie	0
$r_1 = \frac{\sum_{i=1}^{N-1} (x_{i+1} - \bar{x})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{N-1} (x_{i+1} - \bar{x})^2 \sum_{i=1}^N (x_i - \bar{x})^2}} = \frac{Cov(x_{i+1}, x)}{Var(x)}$	$(x_1, x_2), (x_2, x_3), \dots, (x_{N-1}, x_N)$	1
$r_2 = \frac{\sum_{i=1}^{N-2} (x_{i+2} - \bar{x})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{N-2} (x_{i+2} - \bar{x})^2 \sum_{i=1}^N (x_i - \bar{x})^2}} = \frac{Cov(x_{i+2}, x)}{Var(x)}$	$(x_1, x_3), (x_2, x_4), \dots, (x_{N-2}, x_N)$	2
$r_k = \frac{\sum_{i=1}^{N-k} (x_{i+k} - \bar{x})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{N-k} (x_{i+k} - \bar{x})^2 \sum_{i=1}^N (x_i - \bar{x})^2}} = \frac{Cov(x_{i+k}, x)}{Var(x)}$	$(x_1, x_k), (x_2, x_{k+1}), \dots, (x_{N-k}, x_N)$	$k < \frac{N}{2}$

Así como el error estándar del coeficiente de correlación lineal simple, también se calcula un error estándar y los intervalos de confianza para el coeficiente de autocorrelación. La Función de Autocorrelación es el conjunto de coeficientes de autocorrelación desde uno, hasta un máximo que no pueda exceder la mitad de las observaciones y su importancia se debe al estudio de la estacionalidad de la serie temporal. También existe el coeficiente de autocorrelación parcial, el cual mide la correlación entre pares de valores observados, pero además, elimina el efecto debido a la correlación producida por retardos anteriores.

## 2.7 Simulación MonteCarlo

El método MonteCarlo, es una simulación que requiere muestras aleatorias repetidas y

análisis estadístico para computar los resultados. Es descendiente del método de la caminata del borracho, y está estrechamente vinculado con experimentos aleatorios, experimentos que no se sabe el resultado adelantadamente. En las ciencias naturales se usan los modelos matemáticos para describir las interacciones de un sistema utilizando expresiones matemáticas. La diferencia entre modelo matemático y la modelización estadística radica en que, un modelo matemático depende de varios parámetros de entrada, los cuales se procesan a través de fórmulas matemáticas en el modelo y los resultados son uno o más análisis; mientras que la modelización estadística, comprende el peor, el mejor y el caso normal para ser interpretado, permite una flexibilidad al valorar resultados.

Esta simulación, es una técnica numérica para calcular probabilidades y otras cantidades relacionadas a la aleatoriedad utilizando números aleatorios. John Von Neuman, en la década de los cuarenta y con los primeros ordenadores, aplica la simulación para resolver problemas complejos nunca antes resueltos de forma analítica (Rodríguez, 2011). Montecarlo y su casino están relacionados en cuanto a simulación; la ruleta, juego poderoso de los casinos, es uno de los aparatos mecánicos más sencillos que nos permiten obtener números aleatorios para simular variables aleatorias (Rodríguez, 2011).

### 2.7.1 Obtención de números aleatorios

El origen de la simulación MonteCarlo son los números aleatorios. Hay tres fuentes de números aleatorios: las tablas de números aleatorios (ver tabla 7), los generadores de números aleatorios y los números pseudo aleatorios. Es necesario medir la aleatoriedad de los números usados para medir la efectividad de las fuentes, y darse cuenta de la influencia de la periodicidad; aunque también existan tests de rachas para la detección de patrones, la fórmula para medir la aleatoriedad de la fuente de las tablas de números aleatorios, se calcula a partir de una tabla (ver tabla 7).

**Tabla 7.** Frecuencia de las cifras de las tablas de números aleatorios.

Cifra	0	1	2	3	4	5	6	7	8	9
$f$	$v_0$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_9$

Sea  $v_i$  la frecuencia correspondiente a  $d_i$  y  $d_i \in D = \{0,1, \dots,9\}$ , y teniendo en cuenta que  $\forall d_i, P(d_i) = 0.1, i = 0,1, \dots,9$  entonces:

$$\sum_{i=0}^9 (v_i - 0.1d_i)^2$$

En el caso de los generadores de números aleatorios, la base fundamental es la utilización del ruido:

$$f(x) = \begin{cases} 0, P(0) = \frac{1}{2}; & \text{si } x \text{ es par} \\ 1, P(1) = \frac{1}{2}; & \text{si } x \text{ es impar} \end{cases} \quad \text{cuando } x \notin \Delta t$$

Está claro que no hay fórmula alguna para la obtención de números aleatorios, de lo contrario debe ser muy ingeniosa; sólo existen métodos para el cálculo de números pseudo aleatorios: el método de los centros de los cuadrados, los métodos congruenciales, el generador multiplicativo, el generador mixto. El método de los centros de los cuadrados fue desarrollado por Von Neuman, y consiste en fijar un número inicial de cuatro cifras llamado semilla  $\gamma_0$ , elevarlo al cuadrado para obtener un decimal de ocho cifras, el segundo valor se obtiene al seleccionar las cuatro cifras centrales del decimal de ocho cifras, y así sucesivamente (ver tabla 8).

**Tabla 8.** *Idea intuitiva del método de los centros del cuadrado.*

$\gamma_0 = 0.9876$	$\gamma_0^2 = 0.97535376$	$\gamma_1 = 0.5353$	$\gamma_1^2 = 0.28654609$	$\gamma_2 = 0.6546$	...
Semilla	Cuadrado de la Semilla	Valor secundario	Cuadrado del valor secundario	Valor terciario	...

El método presenta el problema de que la obtención de números pequeños tiene mayor frecuencia que los números grandes. El método congruencial utiliza la comparación de los módulos de dos números, es decir, tienen el mismo residuo al dividirlos por  $m$ :  $\forall x, x \bmod(m) \equiv y \bmod(m) \rightarrow x \equiv y \bmod(m)$ . Sean  $a$  y  $b$  dos números elegidos convenientemente,  $\gamma_0$  la semilla, la expresión para el cálculo de números aleatorios se define de la siguiente manera:  $\gamma_n = (a\gamma_{n-1}) + b \bmod(m)$ . El método multiplicativo es una variante del método congruencial, la única diferencia es el valor que tiene un número  $b = 0$ , por congruencia modular, la fórmula queda así:

$$\gamma_n = (a\gamma_{n-1}) + b \text{ mod}(m)$$

$$\gamma_n = (a\gamma_{n-1}) + 0 \text{ mod}(m)$$

$$\gamma_n = (a\gamma_{n-1}) \text{ mod}(m)$$

Al usar la definición de congruencia modular, la elección adecuada de  $a$  y  $b$  hacen que el periodo de repetición de los números aleatorios obtenidos, se incremente hasta  $m$ ; la elección adecuada es que  $a$  y  $b$  sean números primos. Lo ventaja de la simulación con números pseudo aleatorios es evidente, no necesita elevados recursos computacionales ya que se basa en operaciones aritméticas básicas; y se comprueba la calidad de la secuencia en la reproducción de números pseudo aleatorios. El punto débil de los números aleatorios es, su periodicidad.

### 2.7.2 Simulación de variables aleatorias continuas

Sean  $f(x) \sim X, F(x) \sim X$  la función de densidad en  $x$  y la función de distribución en  $x$  respectivamente, para simular los valores de una variable aleatoria continua, se supone que también  $x_i \sim X$ :

$$\int_{-\infty}^{x_i} f(x)dx = \gamma_i, \gamma_i \sim U(0,1)$$

El método para calcular la función de distribución inversa, consiste en, dada la función de densidad  $f(x)$  obtener su función de distribución  $F(x)$  y calcular  $F^{-1}(x)$ :

$$F: E \rightarrow [0,1]$$

$$x \rightarrow F(x) = P(X < x)$$

$$x = F^{-1}(\gamma) \leftarrow \gamma$$

Simulación de una variable aleatoria que siga una distribución uniforme  $U(a, b)$ , a partir de un generador números aleatorios que siga una distribución  $U(0,1)$ , (ver tabla 9).

**Tabla 9.** Funciones de densidad, de distribución y distribución inversa VAU

$f(x) = \begin{cases} \frac{1}{b-a}, x \in (a, b) \\ 0, x \notin (a, b) \end{cases}$	$F(x) = \begin{cases} 0, x < a \\ \frac{x-a}{b-a}, x \in (a, b) \\ 1, x > b \end{cases}$	$\gamma = \frac{x-a}{b-a}$ $x = F^{-1}(\gamma) = \gamma(b-a) + a$
Función de densidad	Función de distribución	Función de distribución inversa

Simulación de una variable aleatoria que sigue una distribución normal estándar  $N(0,1)$ , (ver tabla 10)

**Tabla 10.** Funciones de densidad, de distribución y distribución inversa VAN

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in (-\infty, \infty)$	$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \gamma$	No admite solución explícita
Función de densidad	Función de distribución	Función de distribución inversa

Las distribuciones continuas que siguen una ley Exponencial y Normal son las más importantes en el área de la vulcanología en la modelización del RSAM y otras variables cuantitativas relacionadas con la naturaleza de los sismos; se tomó la idea intuitiva de la distribución uniforme para la generalización de distribuciones anteriores. En esta investigación se utilizó la simulación de distribuciones normales a través del Método MonteCarlo. La distribución normal general  $N(\mu, \sigma)$  es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in (-\infty, \infty)$$

## 2.8 Validación del modelo

Los modelos de regresión se utilizan para predicción o estimación de valores de interés, generalmente la persona que utiliza el modelo es alguien distinta de quien desarrolló dicho modelo. Antes que el modelo pase a otras manos, debe hacerse la última prueba, la validación del modelo. Comprobación de la adecuación del modelo y validación del modelo son diferentes. La comprobación de la adecuación del modelo es el análisis estadístico pertinente, mientras que la validación se concentra en si funcionará correctamente con los datos de la misma naturaleza pero en tiempo real y de forma generalizada, describiendo excelentemente el comportamiento de las variables de interés en la práctica.

La validación del modelo es la garantía tanto para el desarrollador como para el usuario del modelo en cuestión. Hay tres tipos de validar un modelo, pero aquí se utilizará únicamente la simulación de datos a través del Método MonteCarlo. Se efectuará un estudio y comparación del comportamiento aleatorio de los datos reales con los pronosticados. Otra alternativa utiliza un subconjunto de datos de la base original en

calidad de placebo, o hacer una nueva recolección de datos, con características similares, para investigar el desempeño del modelo. Y la última, de carácter empírica, es la obtención de pronósticos para realizar comparaciones lógicas con la experiencia, la teoría física, otros modelos analíticos y con resultados de simulaciones ya realizadas.

## **Capítulo 3: Metodología**

### **3.1 Confiabilidad de los datos**

Los datos indispensables para realizar la búsqueda de modelos estadísticos que permitan interpretar el comportamiento de las variables sísmico-volcánicas de los principales complejos volcánicos de El Salvador y, a través de este modelo hacer inferencias (pronósticos); tienen su fuente en el Ministerio de Medio Ambiente y Recursos Naturales, es sobre esta base de datos que se efectuarán los análisis estadísticos pertinentes. El problema de investigación es la búsqueda de un modelo estadístico que pronostique erupciones volcánicas, a partir de datos de años anteriores.

### **3.2 Descripción del complejo volcánico VSA**

El volcán Ilamatepec se sitúa en el departamento de Santa Ana, región centro occidental de El Salvador y zona fronteriza Santa Ana-Sonsonate. Su altitud es de dos mil trescientos ochenta y un metros sobre el nivel del mar y su composición orgánica está conformada por basalto de olivino y piroxeno; es un estratovolcán que ha tenido numerosas explosiones a lo largo del tiempo y durante la última década su inestabilidad ha sido muy pronunciada. La actividad volcánica es freatomagmática. Las coordenadas geográficas de este complejo volcánico son: latitud norte de trece grados con cincuenta y un minutos y once segundos, y longitud oeste de ochenta y nueve grados con treinta y siete minutos y cuarenta y ocho segundos.

Es el volcán más alto de El Salvador, sus últimas erupciones fueron en mil novecientos cuatro, mil novecientos veinte y en dos mil cinco. Forma parte de la cordillera de Apaneca. En agosto del dos mil cinco se presentaron las premisas necesarias para una erupción volcánica, pero no ocurrió algo. En octubre del mismo año hubo una explosión con la que los cielos se tornaron grises por la ceniza y rocas expulsadas (ver ilustración 7). También un alud de agua caliente descendió del cráter del volcán, se registraron dos personas muertas y protección civil forzó la evacuación de San Blas, zona aledaña al volcán. La última alerta emitida por las instituciones gubernamentales data del dos mil cinco.



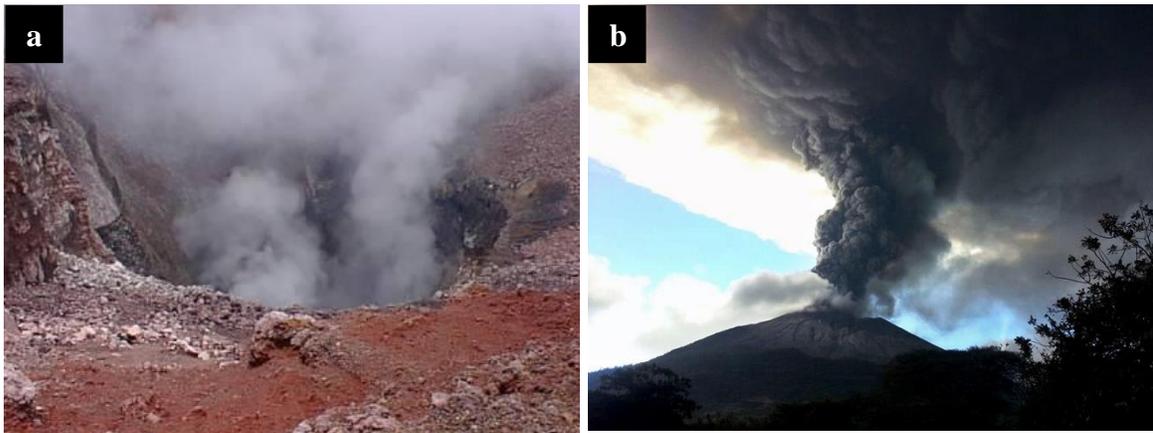
*Ilustración 7.* Laguna Verde y erupción del 01 10 05, VSA.  
 a) laguna verde debido a la concentración de azufre, del volcán Santa Ana de fecha 31 de enero del 2000. Cortesía del Ministerio de Medio Ambiente y Recursos Naturales. b) Vista del volcán Ilamatepec desde Sonsonate (15 km SO del cráter) en la que se aprecia una columna cargada de cenizas producto del monitoreo del 1 de octubre del 2005. Cortesía de La Prensa Gráfica, El Salvador.

El Salvador decreta alerta roja en un radio de cuatro kilómetros alrededor del volcán Santa Ana debido a la violenta erupción que trajo consigo ceniza y rocas. El Presidente de la República: Antonio Saca, hace un llamado nacional para guardar la calma. La erupción es anticipada a un enjambre de sismos y ruidos ensordecedores, los expertos lo catalogan como un volcán de un gran riesgo potencial debido a sus características: actividad y volumen, pero su erupción fue leve. Durante la erupción sucedieron eventos como aperturas de boquetes en el suelo de donde emergía agua hirviendo. El lanzamiento de rocas se hizo en un área mínima. El Ministro de Gobernación, René Figueroa, informó sobre el trabajo de la Fuerza Aérea de El Salvador. Entre los hallazgos de la Fuerza Aérea local se tiene la observación de una columna de humo de cincuenta mil pies de altura expulsada por el volcán (ver ilustración 7).

### **3.3 Descripción del complejo volcánico VSM**

Se conoce como volcán Chaparrastique, está ubicado en la ciudad de San Miguel del departamento homónimo, a ciento cuarenta kilómetros de la capital. Pertenece a la cordillera de Chinameca, tiene una altura de dos mil ciento veintinueve metros sobre el nivel del mar, es el tercer volcán más alto de El Salvador y presenta un cráter volcánico de unos ochocientos metros de diámetro, el cono de este volcán es el mejor formado del país. Es simétrico y está rodeado de planicies en todas direcciones, a excepción de la

región norte. Sus laderas tienen las ciudades San Miguel, Quelepa, Moncagua, Chinameca, San Jorge, San Rafael Oriente y El Tránsito; incluso hasta la carretera panamericana que atraviesa todo el país descansa sobre las laderas de este macizo, todas pertenecientes al departamento de San Miguel.

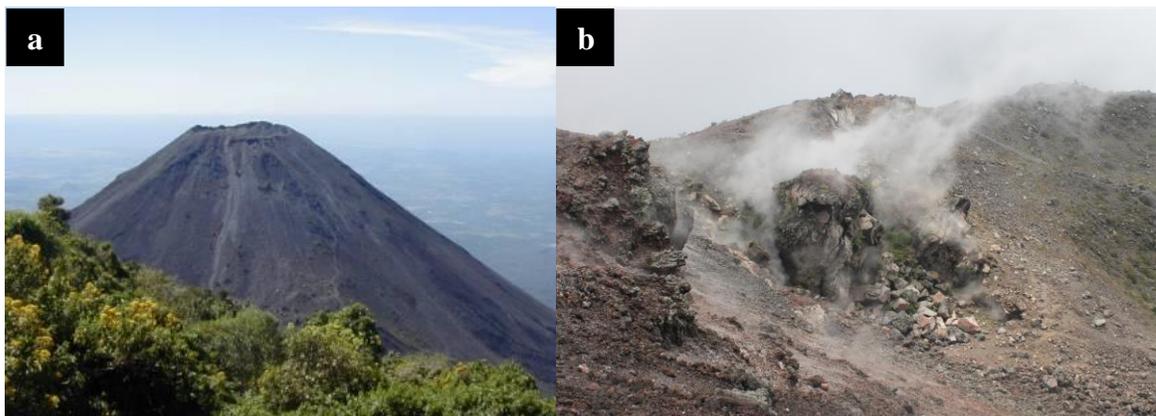


*Ilustración 8. Cráter y erupción del 29 10 13, VSM.*  
a) Cráter del volcán San Miguel en noviembre del 2000. Ministerio de Medio Ambiente y Recursos Naturales. b) mismo volcán de fecha 2013. Cortesía de Wikipedia.

Recientemente, la primera de las sucesivas erupciones fue el veintinueve de diciembre del dos mil trece, en donde a las diez y media, hora local, expulsó una columna de humo y ceniza con una altura de cinco kilómetros (ver ilustración 8), actualmente aún sigue en actividad volcánica moderada; el doce de enero del dos mil dieciséis ocurrió la última actividad volcánica: expulsión moderada de ceniza.

### **3.4 Descripción de los complejos volcánicos VIZA, VSS, VILO y VSVI**

El volcán Izalco, es un estratovolcán de la zona volcánica de Santa Ana, está ubicado al oeste de San Salvador, capital de El Salvador (ver ilustración 9). Se ha mantenido activo desde la erupción que lo formó alrededor del año mil setecientos setenta, hasta mil novecientos cincuenta y ocho; es llamado el Faro del Pacífico debido a su constante actividad volcánica. Su última erupción data del año mil novecientos sesenta y seis. La erupción más catastrófica fue en el año mil novecientos veintiséis quemando el pueblo de El Matazano y matando a cincuenta seis personas. Esta región es fértil en producción de café, cacao y caña de azúcar.



*Ilustración 9.* Volcán Izalco y altas temperaturas en el cráter.  
a) Forma cónica del volcán Izalco, 23 de noviembre del 2000. b) Altas temperaturas en el volcán Izalco, 6 de junio del 2013. Cortesía del Ministerio de Medio Ambiente y Recursos Naturales.

La altura del volcán Izalco es de mil novecientos cincuenta metros sobre el nivel del mar y su ubicación geográfica tiene latitud trece grados con cuarenta y ocho minutos y cuarenta y siete segundos, y longitud ochenta y nueve grados con treinta y siete minutos y cincuenta y nueve segundos. El volcán San Salvador, conocido por Quetzaltepec, es un estratovolcán situado al noroeste de la capital de San Salvador. Antes tenía una laguna, pero debido a su anterior actividad volcánica, desapareció. Dentro del cráter se encuentra un volcán más reciente llamado Boquerón. Dada la cercanía de la ciudad de San Salvador, alguna actividad volcánica potente puede resultar en destrucción catastrófica.

El volcán San Salvador tiene mil ochocientos noventa y tres metros de altitud (ver ilustración 10), y su última erupción fue en el año mil novecientos diecisiete. Actualmente su actividad volcánica es muy baja, muchas antenas de telecomunicaciones yacen sobre El Picacho, máxima elevación del volcán con una altura de mil novecientos sesenta metros de altitud. Tiene ubicación geográfica de trece grados con setecientos treinta y cuatro milésimas de grado latitud norte, y ochenta y nueve grados con doscientos noventa y cuatro milésimas de grado longitud oeste. La caldera de Ilopango es un cráter en forma de lago (ver ilustración 10), el segundo lago más grande del país, tiene un área de setenta y dos kilómetros cuadrados. La caldera de Ilopango está ubicada en los límites fronterizos de los departamentos de San Salvador, La Paz y Cuscatlán.

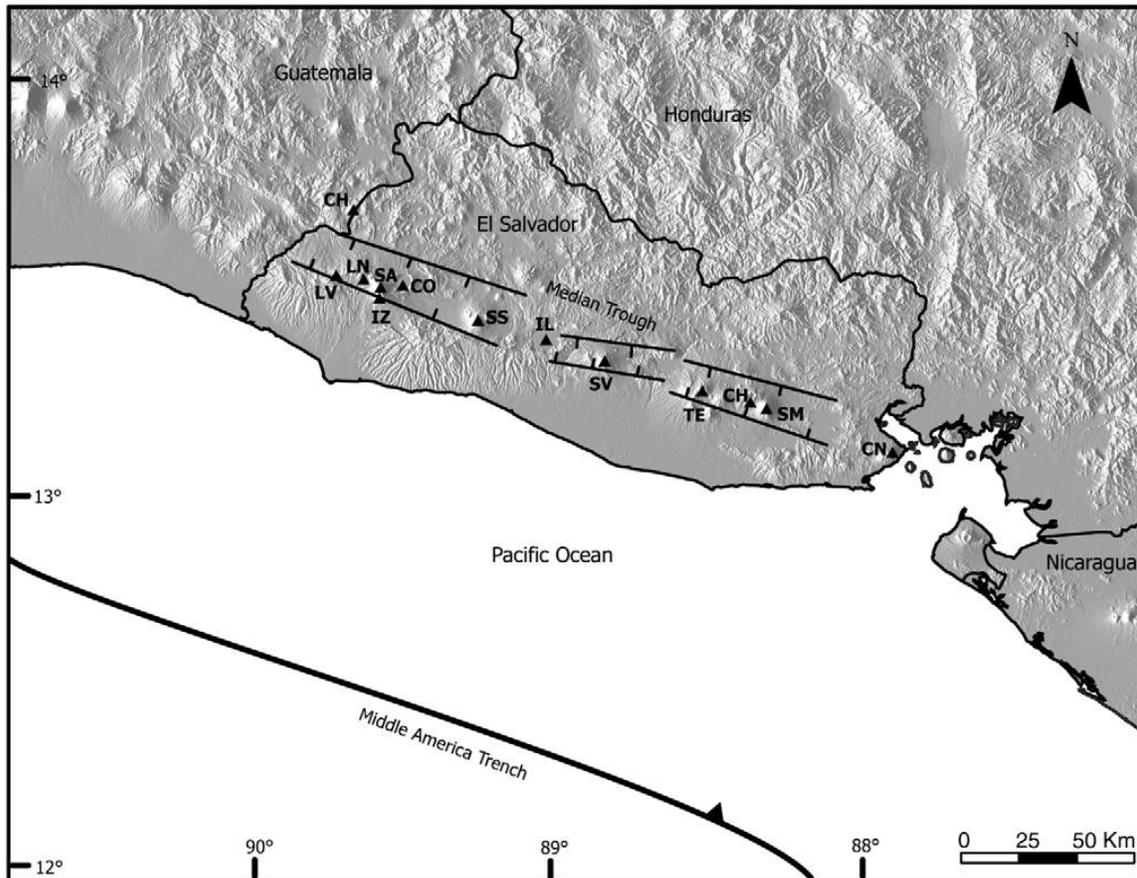


*Ilustración 10.* VSS, VILO y VSVI.

a) Boqueroncito dentro del cráter del volcán San Salvador en agosto de 1999. Cortesía del Ministerio de Medio Ambiente y Recursos Naturales. b) Caldera Ilopango en noviembre de 1998. Cortesía del Ministerio de Medio Ambiente y Recursos Naturales. c) Volcán San Vicente en 1994. Cortesía de Wikipedia.

Sus coordenadas geográficas son trece con sesenta y siete milésimas grados latitud norte, ochenta y nueve con cinco centésimas grados longitud oeste. Su altitud es de cuatrocientos cincuenta metros. Según el radiocarbono, la caldera colapsó entre el año cuatrocientos diez y quinientos treinta y cinco antes de Cristo, produciendo flujo piroclástico muy difundido en las regiones próximas. Con un volumen alrededor de veinticinco kilómetros cúbicos de tefra y un valor de seis para el índice de explosividad volcánica que devastó las ciudades mayas (Major et al., 2001). La única erupción con historiales que ocurrió desde el treinta y uno de diciembre de mil ochocientos setenta y nueve, hasta el veintiséis de marzo de mil ochocientos ochenta; produjo un domo de lava que alcanzó la superficie del lago y formó los islotes conocidos como Islas Quemadas.

El volcán San Vicente, llamado Chinchontepec o cerro de Las Chiches (ver ilustración 10), es un estratovolcán ubicado cerca de la ciudad de San Vicente, es el segundo volcán más alto en El Salvador con dos mil ciento ochenta y dos metros de altitud para la cumbre más alta. Su ubicación geográfica está determinada por las coordenadas trece grados con treinta y cinco minutos, cuarenta y dos segundos al norte; y ochenta y ocho grados con cincuenta minutos, trece segundos al oeste. El volcán tiene dos cráteres localizados en cada cumbre, aunque no están exactamente en la parte más alta; la densa vegetación cubre ambas cimas. Tiene muchas fumarolas en los flancos norte y oeste. La última erupción significativa data de más de hace mil setecientos años.



*Ilustración 11.* Ubicación geográfica de los complejos volcánicos, El Salvador. Localización espacial de los principales complejos volcánicos de El Salvador. SA = Santa Ana; IZ = Izalco; SS = San Salvador; IL = Ilopango; SV = San Vicente; SM = San Miguel. Cortesía de Scolamacchia.

### 3.5 Descripción genérica de las variables sísmicas y volcánicas

Las variables de estudio son de tipo cuantitativo continuo y cuantitativo discreto. Las variables de origen sísmico volcánico son la medición del RSAM, acrónimo de Real-time Seismic-Amplitude Measurement; la temperatura, la latitud geográfica, la longitud geográfica, la magnitud máxima, la profundidad en kilómetros lineales, y el conteo de sísmico (ver tabla 11). El RSAM es una variable cuantitativa continua y representa el tamaño de señales completas durante periodos de diez minutos. En situaciones donde el número de sismos es bastante alto de tal forma que los sismos individuales no pueden ser vistos, o los tremores volcánicos que los sismógrafos no pueden verificar en el nivel de la señal, entonces el RSAM es una excelente manera de mostrar los cambios a tiempo.

**Tabla 11.** *Detalle de las variables más importantes en la modelización.*

<b>Variab</b> les	<b>Variable Independiente</b>	<b>Variable dependiente</b>	<b>Variable interviniente</b>	<b>Variable de contexto</b>
Tiempo	o			
RSAM	o	o		
Número de sismos		o		
Magnitud máxima		o		

El tipo de sismo contabiliza el número de sismos diarios y los clasifica en sismos registrados o sismos localizados, es una variable cuantitativa discreta. La temperatura es una variable cuantitativa continua que es obtenida a través de un termómetro, termocuplas, o cámaras termográficas. Las mediciones son efectuadas en diversas áreas de un volcán, pueden hacerse en el interior, dentro del cráter, en lagos de origen volcánico o en el exterior.

**Tabla 12.** *Estaciones sísmicas de los principales complejos volcánicos ES.*

<b>Volcán</b>	<b>Código</b>	<b>Nombre</b>	<b>Latitud</b>	<b>Longitud</b>	<b>Elevación</b>	<b>Tipo</b>	<b>Comunicación</b>	<b>Modelo</b>
San Miguel	VSM	San Miguel	13.4413	-88.2725	1708	Periodo Corto	Radio Analógico	S-13
San Salvador	BOQ6	Boquerón	13.7350	-89.2800	1830	Periodo Corto	Radio Analógico	SS-1
Santa Ana-Izalco	SBLS	San Blas	13.8393	-89.6230	1949	Periodo Corto	Radio Analógico	L4-C
Ilopango	LBR3	Las Brisas	13.7383	-89.0433	770	Periodo Corto	Radio Analógico	SS-1
San Vicente	SNVI	San Vicente	13.6125	-88.8388	1312	Periodo Corto	Radio Analógico	L-22

### 3.6 Recursos

Los recursos necesarios para la correcta ejecución de esta investigación son los siguientes: entre los recursos humanos, el personal del MARN, siglas de Ministerio de Medio Ambiente y Recursos Naturales; personal del Departamento de Matemática de la Universidad de El Salvador. Los recursos técnicos utilizados son el software R-project, una laptop, la base matemática, la base estadística, impresora, internet, memoria USB, CD-ROM, correo electrónico y teléfono celular. Los recursos financieros se resumen en el gasto de los viáticos del estadístico, impresiones, fotocopias, gastos de imprenta y algunos otros gastos indirectos asociados con el trabajo de investigación, aunque definitivamente, el más valioso recurso es el tiempo. Los recursos materiales necesarios son bolígrafos, marcadores, papel, entre otros.

Como bien se sabe, una investigación científica conlleva tiempo y recursos, por lo cual se le considera una actividad productiva importante. La distribución específica del tiempo se encuentra plasmada en el cronograma de actividades añadido en la sección de anexos.

### **3.7 Muestra y población**

La precisión de esta investigación se refleja en el alcance que pueda tener; el modelo de regresión es ideal para pronóstico de observaciones de interés y valor. La teoría de regresión especifica el riesgo de la extrapolación, ya que esta se ubica fuera de la concentración o densidad de los datos. Evidentemente, el pronóstico será malo si se encuentra considerablemente lejos del centroide de los datos. La perspectiva es, que el modelo óptimo puede generalizarse para las mismas variables bajo las mismas circunstancias; es decir, el modelo se generaliza hacia la población de datos ya que se parte de una muestra de datos para describir el comportamiento de la variable con las mismas características.

Es necesario no olvidar el factor no determinístico; el factor aleatorio, también del efecto externo o contexto que puede incidir para producir cambios drásticos en el pronóstico de nuevas observaciones. En el desarrollo de esta investigación se cumplen los estándares globales de la teoría de la regresión y correlación. El muestreo es una herramienta poderosa de la estadística soportada por leyes matemáticas. En la regresión se fundamenta estimando los parámetros beta del modelo. Para llegar a un modelo de regresión óptimo debe cumplirse con todas las premisas. A diferencia de muchos casos donde la población de los datos se mantiene constante, es decir, la población es finita, y la muestra es un subconjunto directo de la población.

En este estudio la muestra es un subconjunto indirecto de la población ya que se trata de una serie originada a través del tiempo, o sea una serie temporal, cuya naturaleza es de tipo infinita. Para realizar la regresión de los datos, se tomó en cuenta una muestra de los datos definidos en un intervalo de tiempo no mayor a diez años. Estos datos se obtuvieron de las instituciones públicas a través de medios tecnológicos guardando su integridad, sin dar lugar a alteraciones o a otras corrupciones de datos. En primer lugar,

la transferencia de datos se realizó a través del correo electrónico y en segundo lugar, utilizando CD-ROM y dispositivos de almacenamiento masivo como la USB.

### **3.8 Aparatos y dispositivos de la vulcanología y sismología**

Los dispositivos usados en El Salvador en materia de vulcanismo son el termómetro FLUKE 52 II, termocuplas manuales, cámara termográfica modelo P65, estaciones DOAS, mini DOAS, cámara térmica, monitoreo hidrogeoquímico, entre otros.

### **3.9 Método cuantitativo**

La naturaleza de las variables sugiere el empleo del método cuantitativo para el análisis de los datos. Los datos se obtuvieron de dispositivos tecnológicamente avanzados, propiedad de las instituciones dedicadas al trabajo vulcanológico. El método cuantitativo recopila y analiza los datos de interés. El análisis cuantitativo puede ser descriptivo, exploratorio, inferencial univariado, inferencial multivariado, modelización y contrastación. El análisis descriptivo se orienta en las características de las observaciones; el análisis exploratorio estudia las relaciones entre variables, la dimensión de los datos y en caso necesario la omisión de variables que no contribuyan al modelo.

El análisis inferencial univariado contrasta las relaciones entre las variables; el inferencial multivariado contrasta más de dos variables, ya sean dependientes o independientes; y por último, la generación de ecuaciones estructurales para contrastar dimensiones complejas entre las variables de interés, a esto se le llama modelización. En general, es una investigación cuantitativa de tipo longitudinal o diacrónica, las observaciones se analizan durante cierto tiempo continuamente. Una investigación correlacional determina el grado de variación entre factores de interés, es una fuerza estadística calculada a través de coeficientes de regresión y correlación. Mientras que la investigación sobre hechos cumplidos o ex post facto, sirve para establecer posibles relaciones de causa-efecto y buscar en el pasado su posible origen, aunque tiene la debilidad de que a veces no es viable establecer seguramente la causa del fenómeno; es apropiada cuando prácticamente resulta imposible la experimentación, utiliza técnicas estadísticas tales como la correlación parcial y la regresión múltiple.

### **3.10 Diseño de la investigación**

El diseño de la investigación es de tipo descriptivo-inferencial. La estadística descriptiva se fundamenta propiamente en el estudio de la distribución de frecuencias de los datos recolectados que posteriormente son presentados de forma textual, tabular, o a través de gráficos, no cabe duda de que se puede implementar una combinación. El estudio de la distribución de los datos depende de la posición, dispersión, asimetría y curtosis de los datos. Los tipos de gráficos que se utilizan en este estudio son de tipo diagramáticos y cartográficos. La estadística inferencial se utiliza para estimar parámetros y probar hipótesis. Las estimaciones se llevan a cabo a partir de muestras representativas de la población y siempre están sujetas a los intervalos de confianza para mejorar dicha estimación.

### **3.11 Instrumentos de investigación**

El origen de los datos proporciona confianza y una fuente válida para iniciar un análisis de regresión, el cual estará sustentado por unas muy buenas características. De acuerdo a la necesidad del estudio, entre los instrumentos de investigación más destacados y adecuados están la recopilación documental, tanto para la descripción de las unidades de estudio como para los modelos estadísticos; una entrevista dedicada al personal del área de vulcanología en El Salvador, perteneciente al MARN; la observación, y por último, la experimentación con los datos a través de la regresión no paramétrica y la simulación MonteCarlo. La entrevista cualitativa es una recopilación de la información pertinente de forma directa.

Se sigue un orden de preguntas y se adapta de acuerdo a lo que el entrevistado va presentando. El algoritmo para efectuar una entrevista efectiva consiste de cuatro fases. La primera fase es la apertura, se logra con una breve presentación de las dos partes con la intención de explicar el objetivo y solicitar cooperación para que el entrevistado brinde la información requerida. La segunda fase es la iniciación, se despierta el interés con preguntas breves, simples y de sondeo tratando de centrar y hacer más robusta la conversación. La tercera fase es el clímax, se obtiene información medular para la investigación gracias al aumento del interés del entrevistado. Y la última, es la fase del

cierre; se agradece la participación del entrevistado y se deja un espacio prudente por si el entrevistado desea complementar la información brindada, a veces también es muy útil esta información.

La entrevista puede ser libre o dirigida. Aquí se abordó la entrevista libre, es un formato básico que sigue la opinión y criterio del entrevistador, la calidad de la información es más profunda e íntima aunque se corra el riesgo de rodeos o alejamiento del tema principal; mientras que la entrevista dirigida, las opiniones se centran en un guion preestablecido sin permitir variación, la calidad de los datos varía y la participación es menos profunda. En su mayoría, las preguntas de la entrevista son de tipo abiertas y algunas son preguntas de sondeo y otras de cierre. Las preguntas abiertas permiten libremente la divagación en la opinión del entrevistado, las de sondeo se utilizan para comprobar la veracidad de las respuestas; y las de cierre, limitan, concentran y dirigen las respuestas del entrevistado hacia el tema principal.

La técnica de la observación es muy importante en el método de investigación cuantitativo. Se define como un examen atento de los diferentes aspectos de un fenómeno a fin de estudiar sus características y comportamiento en un contexto determinado. La observación fomenta el planteamiento adecuado de la problemática a estudiar. Una ventaja de esta técnica es que permite vislumbrar oportunamente qué utilizar, cómo utilizar, dónde utilizar y cuándo utilizar planes, programas, técnicas y herramientas. Existen muchos tipos de observación pero la que se utiliza en esta investigación es la observación no participativa; el experimentador evita participar en el fenómeno a fin de no impactar su conducta, características y desenvolvimiento, al hacer esto se obtiene información más veraz.

### **3.12 Procedimientos y especificaciones**

Y en última instancia se trata de la experimentación, el investigador participa activamente, y conforme a un algoritmo, introduce cambios que modifican sistemáticamente el comportamiento del fenómeno. Las modificaciones surgidas se valoran cuantitativamente y cualitativamente para analizar las repercusiones de esos cambios en el fenómeno observado y ampliar así su conocimiento. Se clasifica en

experimentos exploratorios, experimentos confirmatorios y experimentación crucial. En la experimentación caben las siguientes palabras técnicas: constante, variable, variable dependiente, variable independiente, variable recurrente, variable ajena, variable discreta, variable continua, causalidad, variación concomitante, comparabilidad, entre otras.

La manipulación y procesamiento de los datos es sistemática y requiere del software de alto nivel y de licencia pública R-project, para el análisis de regresión. La parte teórica estadística es el fundamento científico para la investigación. La versión del software R-project es, dos punto quince punto uno, Roasted Marshmallows, dos mil doce; instalado en un ordenador Windows de treinta y dos bits. El copyright es propiedad de R Foundation for Statistical Computing. R es un proyecto colaborativo con muchos contribuyentes y puede redistribuirse bajo los términos GNU GPL, además es un software libre. Las curvas de Andrews fueron graficadas a través del software Graphmatica versión dos punto cero e.

El resguardo de la base de datos que contiene las variables sísmico-volcánicas, yace en hojas de cálculo de Excel, del paquete Microsoft Office dos mil diez; cada columna corresponde a una variable y cada fila, a una observación. Para el diseño del informe Word, siempre del mismo paquete; y para su publicación, Adobe Reader X versión diez punto uno punto siete.

## Capítulo 4: Modelamiento estadístico VSA

¿Cuál es la diferencia entre sismos registrados y sismos localizados? Los sismos registrados son de tipo volcánico o vulcanotectónico; un sismo es localizado cuando es identificado geográficamente por lo menos en tres unidades sísmicas ubicadas estratégicamente en los flancos de cada volcán; es registrado, cuando se trate de un sismo volcánico o vulcanotectónico; para la clasificación se utiliza el análisis de la Onda Primaria y Onda Secundaria (Olmos, 2016, comunicación personal). Los sismos volcánicos se originan por movimientos de fluidos en el interior del volcán, mientras que los vulcanotectónicos se refieren al sismo cuyo origen es asociado a fracturas de la corteza del volcán.

La cantidad de sismos volcánicos es elevada en comparación con la de los vulcanotectónicos, las frecuencias diarias son mayores. El ascenso del conteo diario de sismos, indistintamente del tipo, es de forma exponencial positiva; antes de la erupción volcánica, la observación correspondiente al veintiocho de agosto del dos mil cinco contabilizó cerca de cuatrocientos sismos de origen volcánico y uno de origen vulcanotectónico. El movimiento magmático es explícito a dos meses antes de la erupción. Las cifras ascendieron hasta llegar al primero de octubre del dos mil cinco, contabilizando dos sismos de origen vulcanotectónico y doscientos dos de origen volcánico; a partir de aquí comienza el descenso de sismos de origen volcánico.

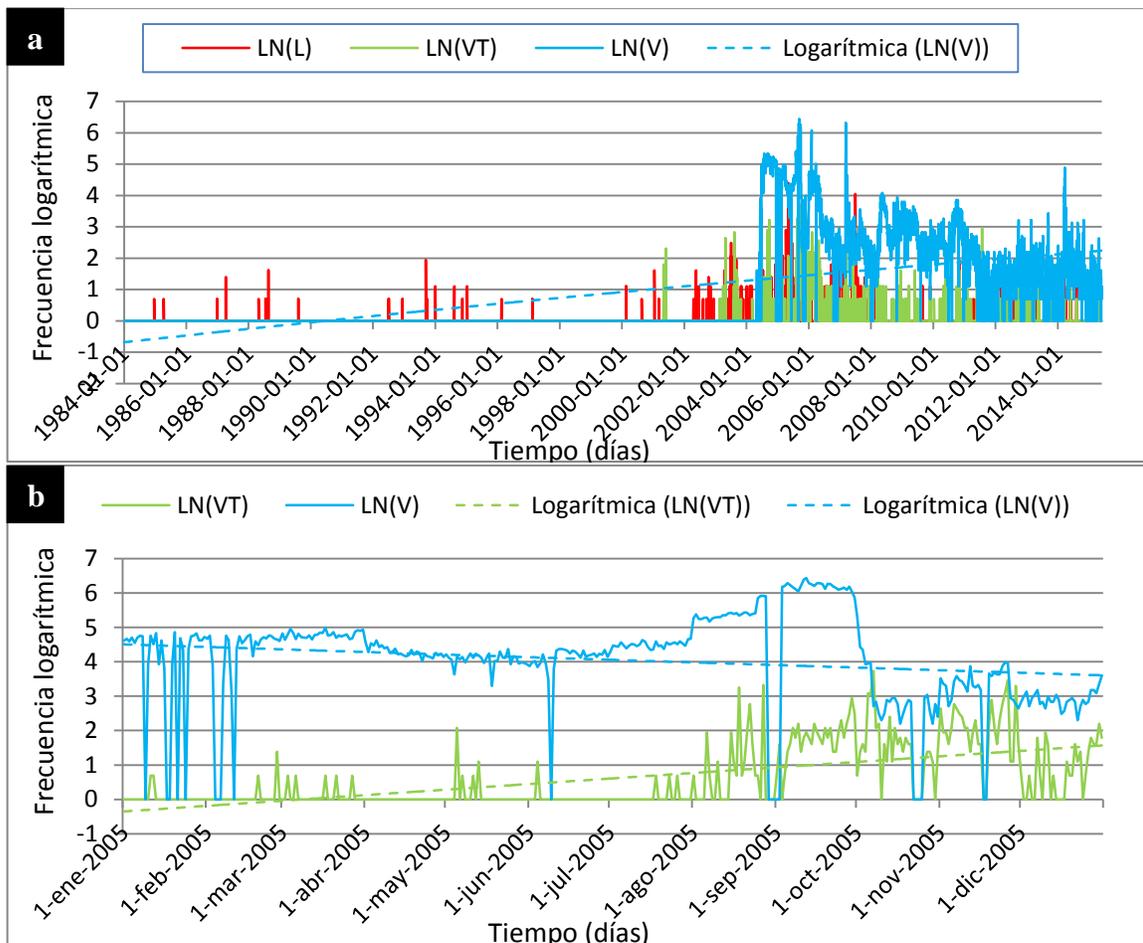
El día en que se registró el mayor número de sismos volcánicos fue el doce de septiembre del dos mil cinco alrededor de un mes antes de la erupción, el conteo fue de seiscientos veintidós sismos volcánicos y siete vulcanotectónicos. El día en que se registró el mayor número de sismos vulcanotectónicos fue el siete de octubre del dos mil cinco, una semana después de la erupción volcánica, el conteo fue de cuarenta y dos sismos vulcanotectónicos y quince de origen volcánico. No existe contradicción al afirmar que probablemente las fracturas terrestres se originaron producto de las intromisiones de magma dentro del volcán, debido al ascenso de sismos vulcanotectónicos una semana después de la erupción principal.

La media de sismos de origen volcánico es de ciento doce y sesenta y nueve

centésimas por día, mientras que la de los sismos de origen vulcanotectónico es de dos y setenta y siete centésimas por día. La media es mayor cuando el origen del sismo es de tipo volcánico, aunque presenta una amplia variación. Las distribuciones de frecuencia tienen sesgo positivo y son de tipo leptocúrticas, es decir, tienen apuntamientos hacia arriba. Un periodo normal tiene la mayoría de sismos localizados, al no tratarse de sismos registrados advierte que el origen del mismo no se debe a fracturas terrestres ni intrusiones de magma en el interior del volcán. Es frecuente la aparición de sismos volcánicos o vulcanotectónicos alrededor del momento de la erupción volcánica.

Si no se presentan indicios de una posible erupción, no se cuenta con sismos de esta naturaleza; mientras que si se toman en cuenta los periodos de tiempo cercanos a la erupción, se encontrará un significativo número de sismos volcánicos. Esto es un caso muy normal, pues la naturaleza del vulcanismo lo demuestra. Para el año dos mil cinco se observa un ascenso de los sismos volcánicos ante un ascenso con poca pendiente de sismos vulcanotectónicos. Desde comienzos del año dos mil cinco se observa elevado número de sismos volcánicos, produciéndose la mayor parte entre los meses de agosto y septiembre, a tan sólo dos y un mes respectivamente, posteriormente se ha trabajado con estos intervalos, pues se consideran de suma importancia en la búsqueda del modelo adecuado que pronostique una futura erupción.

Existe la limitante de que el tiempo previo sea demasiado corto ante una erupción volcánica. El gráfico semilogarítmico de la clasificación de los sismos ocurridos durante el año dos mil cinco (ver ilustración 12) afirma que hubieron fluctuaciones de sismos vulcanotectónicos a mediados de enero, en todo marzo, a principios de junio, finalizando con un ascenso sorprendente que comienza a mediados de julio; la erupción volcánica del primero de octubre ocurre, los sismos volcánicos disminuyen aproximadamente a la mitad de ellos, pero los vulcanotectónicos se mantienen hasta finalizar el año. Lógicamente, los sismos volcánicos producen sismos vulcanotectónicos en la mayor parte de veces, después de la erupción se mantienen muy elevados en número; aunque no es necesaria ni suficiente esta causa para que ocurran, puede haber factores endógenos o externos que los originen.



*Ilustración 12.* Secuencia gráfica de los principales tipos de sismos, VSA.

a) Gráfico semilogarítmico de los sismos localizados (L), registrados (R) y volcánicos (V) del VSA; con la línea de tendencia logarítmica de los sismos volcánicos, el intervalo de tiempo incluye información de tres décadas 1984-2015.  
 b) Gráfico semilogarítmico de los sismos vulcanotectónicos (VT) y volcánicos (V) para el año 2005 en el volcán Santa Ana; con sus respectivas líneas de tendencias logarítmicas.

La ubicación geográfica de la ocurrencia de los sismos pertenecientes al volcán Santa Ana, se mantiene con valores de longitud discreta entre trece, catorce y diecisiete grados; mientras que los valores de la latitud discreta se mantienen entre ochenta y nueve grados negativos, todas las medidas respecto al meridiano de Greenwich y al Ecuador. La magnitud sísmica es uniforme, además es baja en amplitud y alcanza la escasa unidad y media de la escala logarítmica. La profundidad de los sismos es superficial, que por lo común es normal; con el análisis del Kernel Binario (ilustración 13) se afirma que en el periodo pre-eruptivo cercano a la erupción principal, la

profundidad es superficial comparada con el mismo periodo a largo plazo.

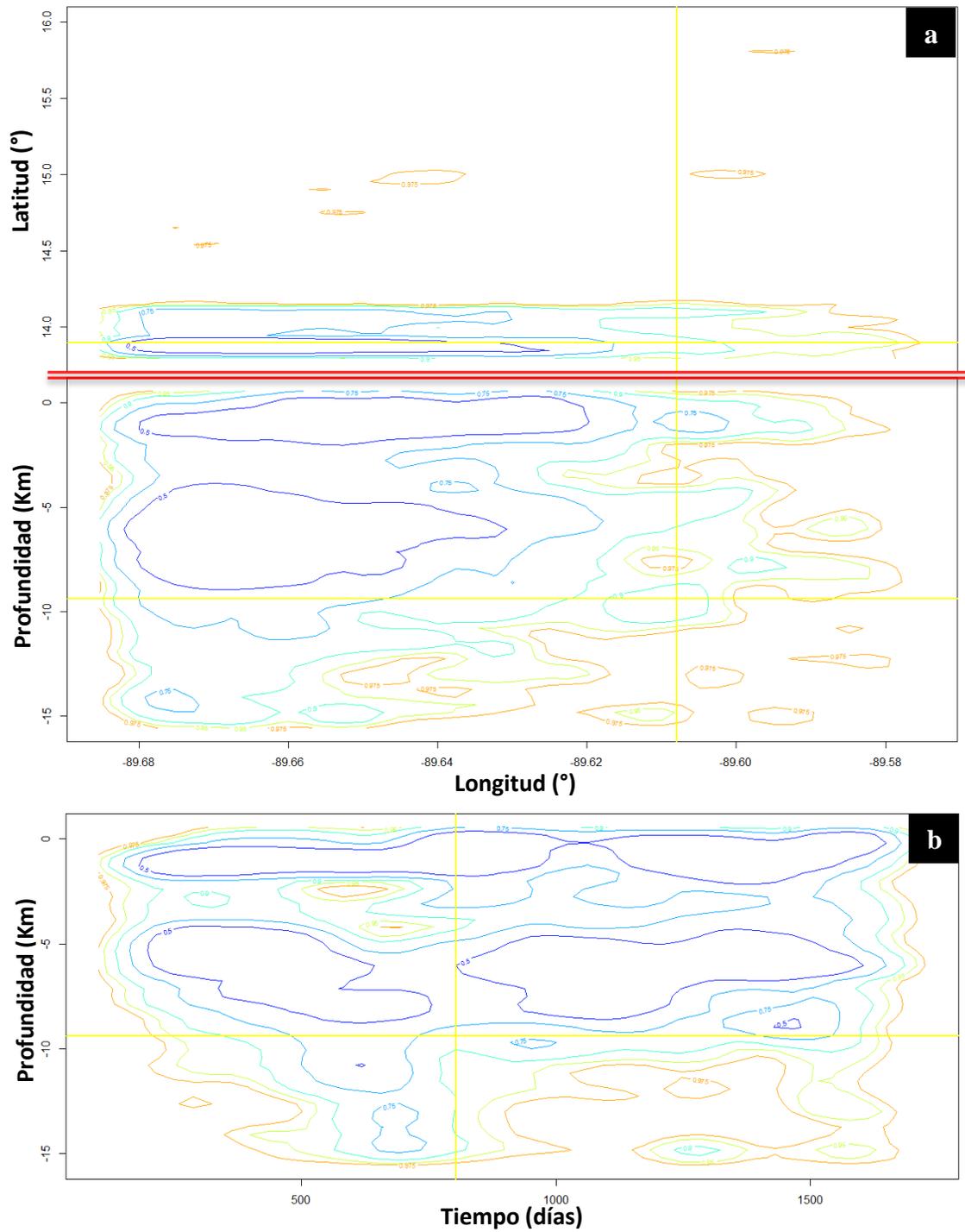


Ilustración 13. Kernels profundidad-coordenadas geográficas VSA.  
 a) Kernels Profundidad versus coordenadas geográficas. b) Kernels de la Profundidad de los sismos. Nota: el parámetro de asociación es de  $n_{bin}=45$ . La observación 804 corresponde a las 14:00 del 1 de octubre del 2005.

La magnitud es uniforme durante la erupción (ver ilustración 14), es notable que al principio de la serie temporal sea moderadamente fuerte. Las coordenadas geográficas se mantienen con poca variación (ver ilustración 14), el origen de cualquier sismo está delimitado por una zona. En los casos anteriores no se presentan patrones claros alrededor de la erupción volcánica, así que no se consideran precursores. La magnitud de la amenaza volcánica del volcán Santa Ana llega hasta la ciudad de Chalchuapa del departamento de Santa Ana, el peligro llega por la región de San Sebastián Salitrillo, al norte del volcán, debido a una singular planicie en forma de canal que los cerros rodean. La amenaza también llega al sur, hasta la ciudad del departamento homónimo.

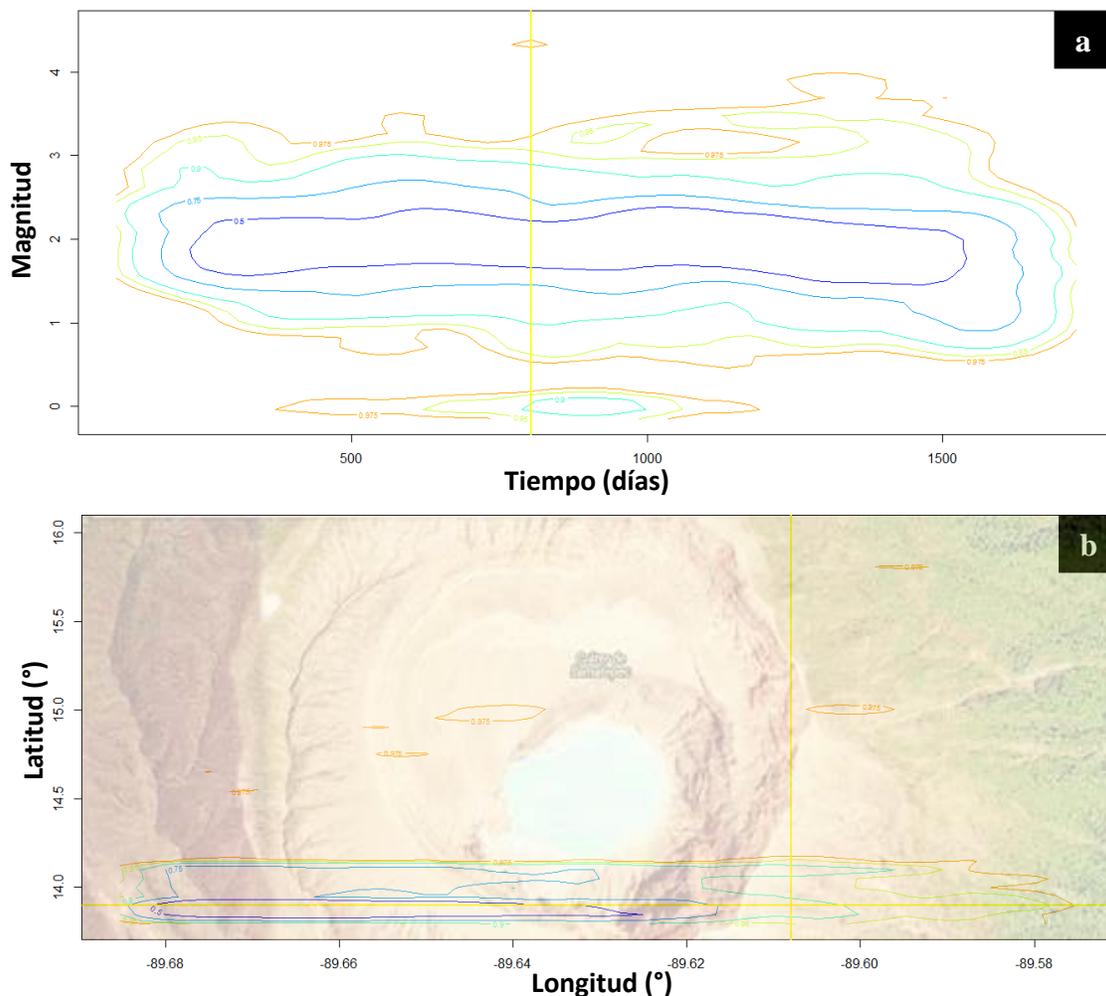


Ilustración 14. Kernels magnitud y coordenadas geográficas VSA.  
a) Kernels de la Magnitud máxima de los sismos. d) Coordenadas geográficas utilizando Google Earth. La observación 804 corresponde a las 14:00 del 1 de octubre del 2005.

El sur del volcán está definido como una densa planicie, por lo que la amplitud de un posible flujo de lava es mayor en relación a la del norte. La geomorfología y la geología afirman que esta planicie se debe a un colapso estructural que ocurrió hace miles de años (Scolamacchia, Pullinger, Caballero, Montalvo, Beramendi & Gonzalez, 2010). La magnitud sísmica máxima durante este periodo es de cuatro punto seis, y la profundidad para el mismo intervalo de tiempo es de quince kilómetros lineales. Han ocurrido treinta sismos de profundidad máxima alcanzada durante el año dos mil cinco, año de la última erupción. Es interesante que la profundidad máxima alcanzada sea de quince kilómetros lineales, esta cantidad es relativamente suficiente para examinar posibles movimientos magmáticos o acomodación de corteza volcánica por movimientos internos producidos por magma.

En el año dos mil cinco aconteció la última erupción y se tienen pocos sismos con profundidad relativamente alta. Este hecho alega dos cosas; la primera es que la mayoría de sismos asumieron una intensa profundidad lo suficiente para no ser registrados y peor aún, difícilmente localizados. Y la otra, que la mayoría de sismos asumieron una intensidad leve y poca indicadora de una probable erupción volcánica, aunque el conteo sísmico puede ser considerado como una excelente premisa ante este hecho. La magnitud sísmica máxima es débil comparada con los colosos mundiales. La magnitud registrada durante el año dos mil cinco alcanza un máximo de cuatro y seis décimas.

También hay otra vía al este del volcán que desemboca en el lago de Coatepeque, vestigio de un poderoso volcán. Esta zona lacustre está más cerca del volcán que las ciudades de Chalchuapa y Sonsonate; debido a las hermosas vistas y agradable clima, se ha convertido en un foco turístico bastante concurrido diurna y nocturnamente, en la actualidad operan muchos lugares turísticos, y en los alrededores del volcán viven al menos veinte mil personas, esa fue la cantidad de evacuados, la zona es muy reconocida por la producción de café de altura, apreciado por los importadores extranjeros. Es necesario un futuro estudio que contraste las hipótesis estadísticas sobre si es determinante la profundidad sísmica y significativamente ineludible o no al momento de una erupción volcánica; la investigación debe analizar las bases de datos de los

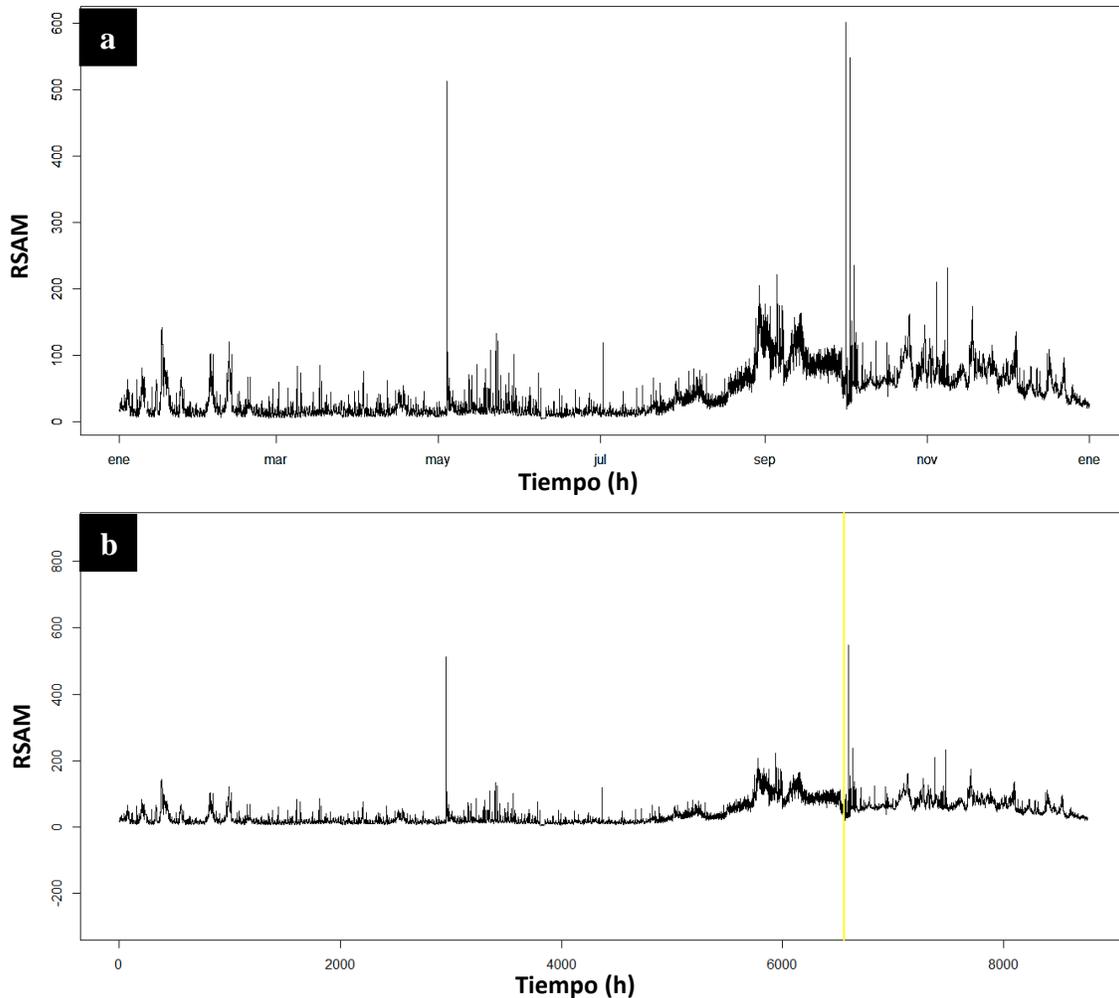
complejos volcánicos de El Salvador existentes.

La latitud sísmica oscila en catorce grados, existen algunos sismos cuya latitud es de quince, dieciséis, diecisiete y alcanzan la latitud dieciocho (ver ilustración 14), como por ejemplo la observación del catorce de septiembre del año mil novecientos noventa y tres. En el periodo pre-eruptivo la latitud estuvo variando considerablemente y es justamente en ese periodo donde las latitudes cambiaron considerablemente. La latitud varía menos que la longitud, pero la variación de la longitud es más uniforme que la de la latitud, ya que el rango de su variación es más amplio aunque posee periodos donde se estabilice. La longitud no posee periodos estables de variación pero lo hace a un ritmo suave en comparación con la latitud.

El RSAM y la medición de la temperatura diaria son premisas fundamentales de una erupción volcánica. La amplitud sísmica o RSAM, es la principal revelación volcánica convertida en sismicidad. Es imposible la ocurrencia de una erupción volcánica sin la existencia de pruebas relacionadas con las amplitudes sísmicas registradas; claro, en los siglos anteriores quizá no se haya incorporado, pero existieron otros métodos más rudimentarios que los actuales. La tecnología beneficia claramente al vulcanismo y la sismología, hoy es posible registrar mediciones de interés tan exactas sin contar con personal humano en el área restringida por el mapa de peligrosidad, sólo es suficiente la instalación de unidades sísmicas que en poco tiempo, segundos o menos, envían información actualizada a las entidades de control que supervisan estos fenómenos naturales.

Los científicos supervisan detalladamente la actividad volcánica actualizada de los complejos volcánicos del territorio nacional en sus oficinas, valiéndose de numerosas técnicas, como la observación volcánica a través de grandes pantallas y análisis que utilizan alta velocidad computacional. El año dos mil cinco (ver ilustración 15), dos mil seis y dos mil siete son los más inestables en comparación con los demás años, el RSAM deja de ser alto hasta principios del dos mil ocho (ver ilustración 16). Han ocurrido fluctuaciones aisladas entre el año dos mil diez y dos mil trece; en el año dos mil

catorce, se observa un ascenso moderado de la actividad sísmica. La efectividad en la medición del RSAM, depende del intervalo de tiempo con que se realice.

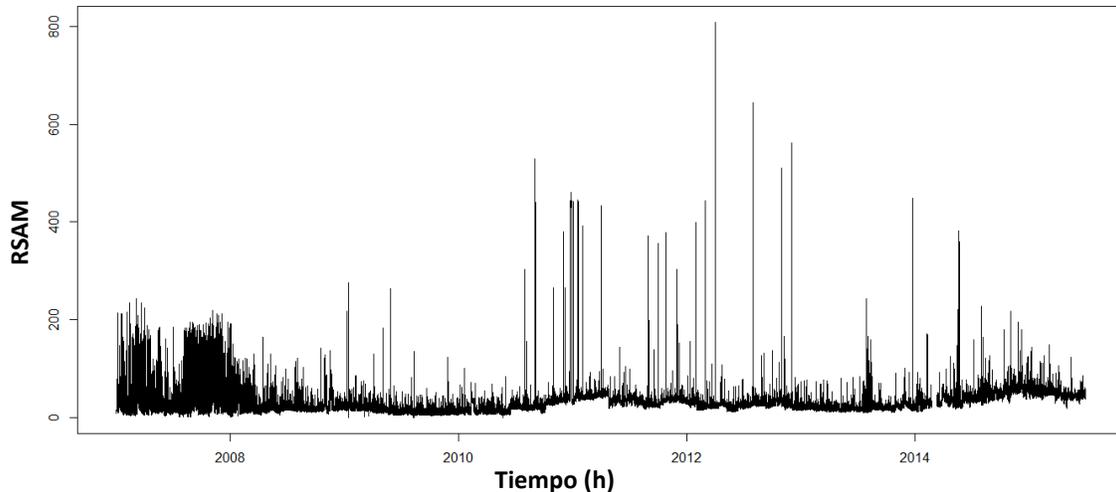


*Ilustración 15.* RSAM por hora 2005, VSA.

a) RSAM por hora VSA, año 2005. b) RSAM por hora con aumento en la escala de las ordenadas, aspecto 1, año 2005. La recta vertical se refiere al momento de la erupción volcánica.

La variación del RSAM diaria es más homogénea que la del RSAM por hora y RSAM cada diez minutos. De hecho el RSAM cada diez minutos, presenta una variación heterogénea que incluye cada por menor del comportamiento volcánico. La debilidad del empleo de promedios matemáticos consiste en lo poco probable que la representación sea significativa y apegada a los datos verdaderos. El comportamiento de cada volcán es completamente distinto, cada volcán reacciona bajo sus propias circunstancias, es idónea una búsqueda de patrones que funcionen de forma apropiada en los volcanes teniendo en

cuenta su tipo. El mínimo valor RSAM por hora es de cero unidades y un máximo, teóricamente, de casi ciento veintisiete unidades.



*Ilustración 16.* Secuencia gráfica del RSAM, periodo post-eruptivo, VSA.  
 Secuencia gráfica del RSAM por hora del periodo post-eruptivo, procedente del volcán Santa Ana, del 1 de enero del 2007 al 30 de junio del 2015.

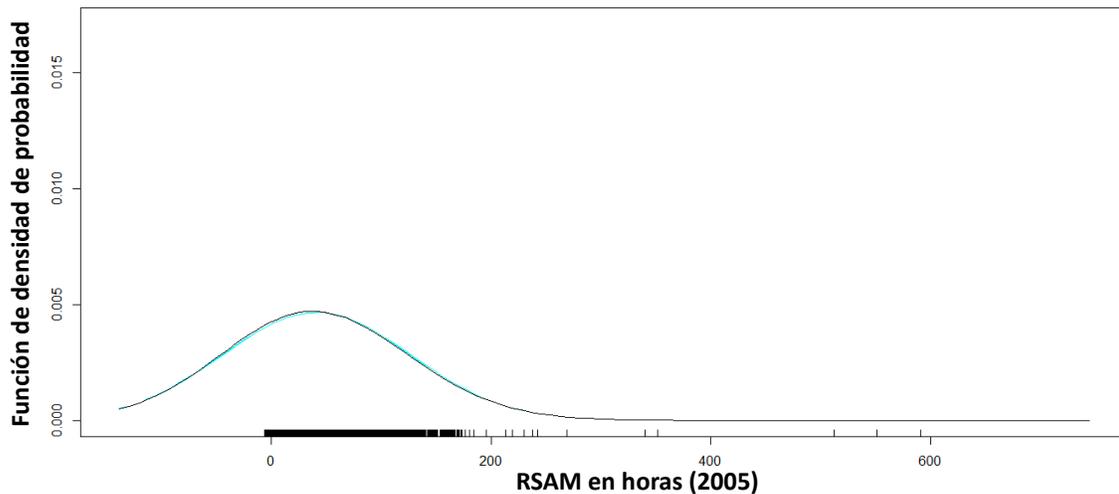
En la práctica se obtiene un mínimo igual al anterior y un máximo de aproximadamente ochocientos ocho unidades, este valor es seis veces y medio el límite superior del RSAM por hora o un incremento del seiscientos treinta y seis por ciento. La distribución de frecuencias del RSAM no es simétrica, sino sesgada positivamente o hacia la derecha; es decir, tiene la cola derecha más larga que la izquierda; además su tipo de curtosis es leptocúrtica, tienen un vertiginoso apuntamiento hacia arriba. Una variable externa que afecta directamente al RSAM por hora es la presencia de valores no válidos debido a una ausencia por erupciones o a su invalidez.

**Tabla 13.** Límites y fronteras del RSAM por hora

Forma canónica	Forma real	Interpretación
$[\bar{x} - s, \bar{x} + s]$	[0, 63.1624]	El 68.27% de los casos están comprendidos, es decir la mayoría de datos.
$[\bar{x} - 2s, \bar{x} + 2s]$	[0, 94.93681]	El 95.45% de los casos están comprendidos, es decir casi todos los datos.
$[\bar{x} - 3s, \bar{x} + 3s]$	[0, 126.71115]	El 99.73% de los casos están comprendidos, prácticamente todos los datos, fuera de este intervalo son valores anormales.

Estudio de la distribución de los valores RSAM por hora y de los límites para la interpretación estadística con un nivel de significancia del 95%, asumiendo normalidad en los datos.

Hay setecientos setenta y nueve observaciones que no se disponen para una mejor interpretación y modelización, un estudio sobre la distribución de las observaciones faltantes es fortuito. En la estimación de la función de densidad con Kernel, los valores del RSAM por cada hora se estabilizan entre cero y alrededor de ciento ochenta unidades, hay un solapamiento de valores que generan una frontera estadística lineal del comportamiento individual del volcán Santa Ana, con respecto a los demás volcanes. La máxima probabilidad para este intervalo de tiempo es de cinco milésimas que corresponde a cincuenta unidades RSAM (ver ilustración 17). Lo anterior afirma que la medida de la amplitud sísmica se considera normal cuando cubre el valor de cincuenta, además las observaciones se sitúan alrededor de este valor para un periodo normal.



*Ilustración 17.* Función de densidad RSAM, 2005, VSA.  
Función de densidad del RSAM por hora del volcán Santa Ana, año 2005.

La probabilidad de que el RSAM sea de cien unidades es de cinco milésimas; y que sea de ciento ochenta unidades, es de una milésima. Son más probables los RSAM de la frontera antes mencionada, que los RSAM mayores que doscientas unidades, algunos valores tienden a probabilidad cero; estos valores son los que han sucedido de manera esporádica. El análisis de las densidades estimadas a través del Kernel Binario con “nbins” de ciento cincuenta permite obtener información importante, el valor por defecto para este parámetro es de cincuenta y uno pero el investigador apoyado por su experiencia puede manipular este parámetro para producir mejores esquemas.

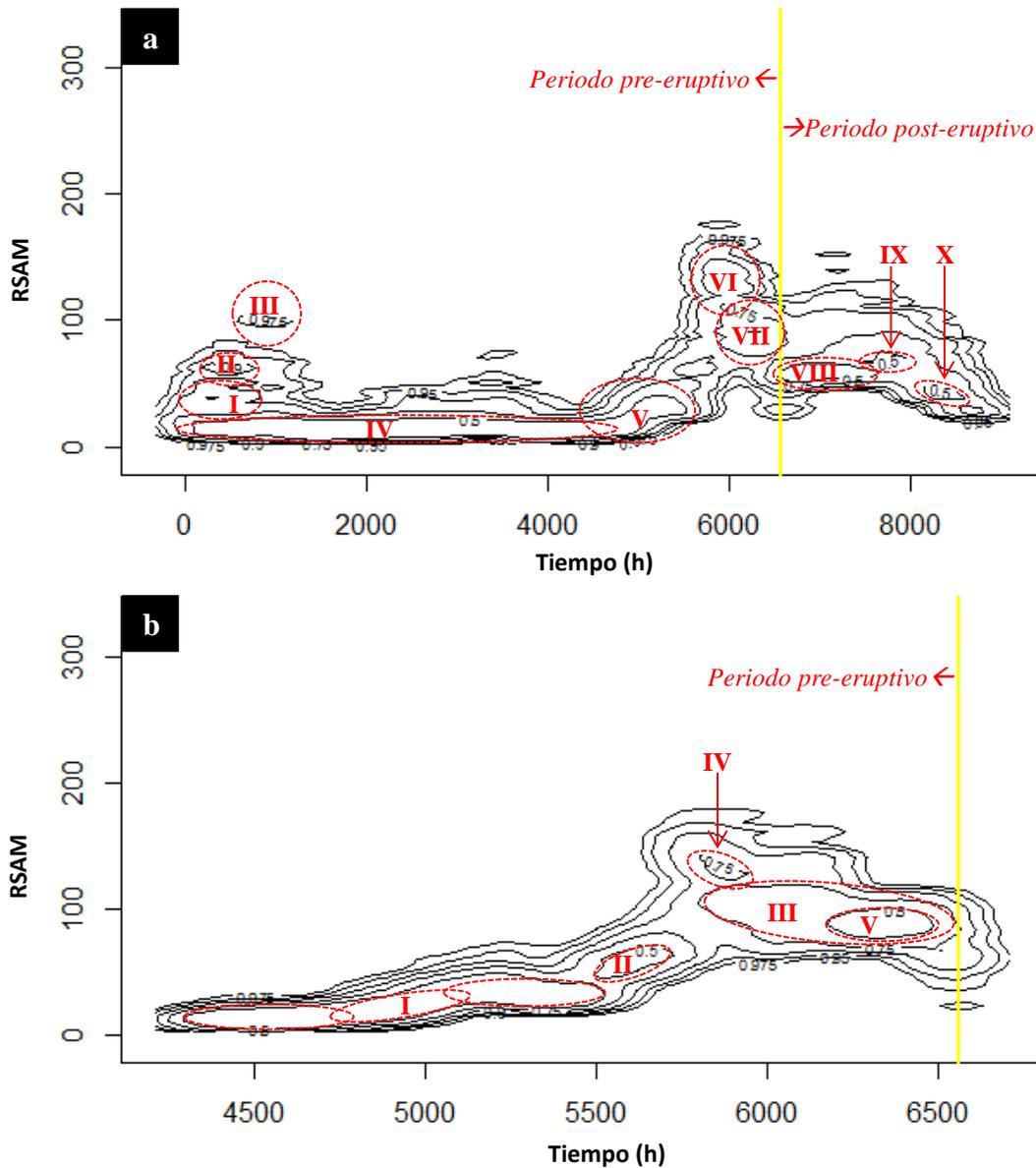


Ilustración 18. Kernels RSAM 2005 y julio del 2005, VSA.

a) Densidad con Kernel Binario del RSAM por hora del volcán Santa Ana, para el año 2005. b) Densidad con Kernel Binario del RSAM por hora del volcán Santa Ana, para julio previo a la erupción volcánica.

Los Kernels Binarios ubican las concentraciones y son más sensibles en la discriminación (análisis discriminante de diferentes niveles de actividad volcánica a partir del RSAM por hora) entre ellas. Antes de la erupción volcánica hay siete centroides bien formados (ver ilustración 18), después de la erupción hay tres concentraciones muy notables, la mayor concentración ocurre días después de dicha

erupción; para el mes de julio las concentraciones se reducen a cinco (ver ilustración 18). El Kernel IV es el más pequeño del escenario I literal b para el volcán Santa Ana. Dura ciento cincuenta horas, se anticipa veintisiete días a la erupción principal y tiene una media de ciento veintiocho unidades RSAM; este núcleo se forma con un nivel de confianza del setenta y cinco por ciento (ver tabla 14).

**Tabla 14.** Escenario de Kernels I VSA

Fig.	N	Tamaño relativo del Kernel	Límite inferior (h)	Límite superior (h)	Nivel de confianza (%)	Media RSAM	Días de anticipo
a	I	Grande (1000h-41.67d)	1	1000	90%	29.09	231.71
	II	Pequeño (500h-20.83d)	250	750	90%	27.17	242.13
	III	Pequeño (500h-20.83d)	750	1250	97.5%	27.64	221.29
	IV	Grande (5000h-208.33d)	1	5000	90%	19.85	65.04
	V	Mediano (1500h-62.5d)	4250	5750	50%	31.11	33.79
	<b>VI</b>	<b>Mediano (500h-20.83d)</b>	<b>5750</b>	<b>6250</b>	<b>90%</b>	<b>111.66</b>	<b>12.95</b>
	<b>VII</b>	<b>Mediano (500h-20.83d)</b>	<b>6000</b>	<b>6500</b>	<b>75%</b>	<b>94.94</b>	<b>2.54</b>
	VIII	Mediano (1250h-52.08d)	6750	7750	50%	73.50	49.54*
	IX	Pequeño (250h-10.42d)	7750	8000	50%	75.95	59.96*
	X	Pequeño (500h-20.83d)	8250	8750	50%	44.92	91.21*
b	I	Grande (1155h-48.13d)	4345	5500	50%	25.26	44.21
	<b>II</b>	<b>Mediano (250h-10.62d)</b>	<b>5500</b>	<b>5750</b>	<b>50%</b>	<b>63.78</b>	<b>33.79</b>
	<b>III</b>	<b>Grande (826h-34.42d)</b>	<b>5750</b>	<b>6576</b>	<b>75%</b>	<b>101.23</b>	<b>0**</b>
	<b>IV</b>	<b>Mediano (150h-6.25d)</b>	<b>5750</b>	<b>5900</b>	<b>75%</b>	<b>128.00</b>	<b>27.54</b>
	<b>V</b>	<b>Mediano (225h-9.38d)</b>	<b>6275</b>	<b>6500</b>	<b>50%</b>	<b>88.76</b>	<b>2.54</b>

\* Kernels posteriores a la erupción del 1 de octubre del 2005 del volcán Santa Ana.

\*\* Tiene 0 días anticipados debido a que incluye el momento justo de la erupción principal del volcán Santa Ana, aunque el Kernel tiene una duración de casi 34 días, un poco más de un mes previo.

En el comportamiento del RSAM a lo largo del año dos mil cinco, se presenta la concentración mayor de la serie (ver ilustración 18). Esta concentración corresponde al núcleo IV del escenario I literal a, tiene un nivel de confianza del cincuenta por ciento, la mayoría de las observaciones se concentran alrededor de veinte unidades, un valor normal para esta variable (ver tabla 14). Este núcleo tiene la característica que es usado como referencia en relación al resto. Un panorama volcánico donde no se presente erupción volcánica presenta más evidente la concentración del cincuenta por ciento; por el contrario, en una erupción volcánica las concentraciones son más heterogéneas y en cierto caso están aisladas entre sí.

La caldera Ilopango es de comportamiento sereno y no ocurren fenómenos que indiquen actividad volcánica inestable; en contraposición, el volcán San Miguel tiene un

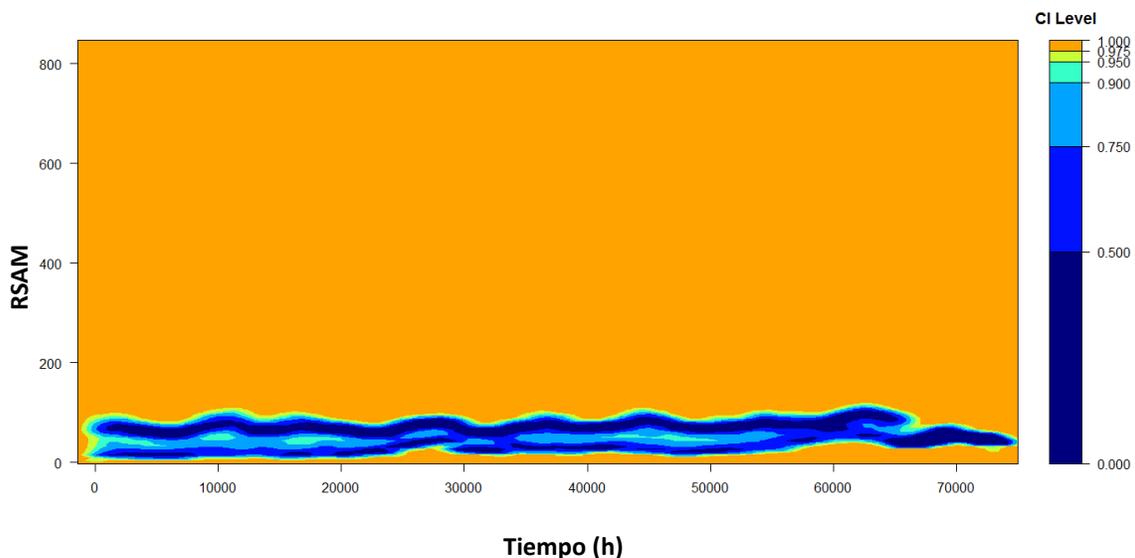
comportamiento muy inestable que se aprecia en la dispersión de los núcleos de datos. El volcán Santa presenta un esquema normal, el comportamiento volcánico es gradual y estadísticamente ideal. Para el año dos mil cinco se contabilizan diez núcleos de datos (ver ilustración 18). La concentración que presenta mayor densidad se trata del numeral cuatro, desde que inicia el año dos mil cinco hasta a mediados del mismo año. Esta concentración a lo sumo tiene veinte unidades RSAM, aunque es grande en densidad, se considera normal.

Los núcleos identificados por los numerales seis y siete (ver ilustración 18), permiten afirmar el despertar decisivo del volcán Santa Ana; para el primer caso a trece días y para el segundo, dos días y medio. El núcleo seis con un nivel de confianza del noventa por ciento (ver ilustración 18) es el de mayor intensidad aunque sea medianamente denso, posee una media aritmética de ciento once con sesenta y seis centésimas; el núcleo siete con un nivel de confianza del setenta y cinco por ciento (ver ilustración 18), posee un promedio de noventa y cuatro con noventa y cuatro centésimas. Estos dos núcleos son los más importantes de todo el año dos mil cinco, el núcleo cuatro sólo sirve de referencia del comportamiento volcánico.

En la búsqueda de Kernels a partir del mes de julio hasta el día de la erupción principal, la cantidad de núcleos son cinco (ver ilustración 18). Los núcleos encontrados no cambian su forma a pesar de que se trata de un fragmento de tiempo. El núcleo uno es el utilizado de referencia (ver ilustración 18), el resto tiene sus propias peculiaridades. En el intervalo de tiempo comprendido desde julio hasta el momento de la erupción volcánica, se halla el núcleo dos que demuestra fuerte pendiente o cambio brusco que se genera en el comportamiento del volcán Santa Ana a principios de septiembre (ver ilustración 18). El comportamiento del núcleo de referencia se separa de los demás, a partir de ese momento se presentan la diversidad de núcleos cuya intensidad va en ascenso hasta llegar al auge.

El núcleo tres y cinco son los más próximos a la erupción (ver ilustración 18), también son muy intensos. En el primero no hay días de anticipo a la erupción, en el segundo hay dos días y medio de anticipo. El núcleo cuatro es el de mayor relevancia

con densidad pequeña y promedio de ciento veintiocho unidades RSAM (ver ilustración 18); se registra entre veintiocho días y un mes de anticipo. Son notables las concentraciones antes de la erupción volcánica, graficada por una línea vertical amarilla, después son leves y con decrecimiento. A partir del mes de julio, se observa que los núcleos con un nivel de confianza del cincuenta por ciento siguen una tendencia creciente, y hay un lapso donde la pendiente se incrementa más del doble (alrededor del veintiocho de julio del dos mil cinco), luego se normaliza, sigue constante pero con valores el doble que los iniciales (ver tabla 14).



*Ilustración 19.* Kernel Binario del RSAM, VILO  
Del 1 de enero del 2007 al 25 de mayo del 2015, nbin=150.

Para el mes de agosto, se observan al menos tres concentraciones (ver ilustración 20). Estas concentraciones son de tipo moderado-fuerte y una se aproxima a las cien unidades RSAM (treinta de agosto del dos mil cinco), a dos meses de la erupción volcánica. Septiembre se antecede a la erupción volcánica con tres centroides (ver ilustración 20). El centroide dos es el más intenso y tiene un nivel de confianza del noventa y cinco por ciento con un promedio de ciento dieciséis y nueve centésimas; en este momento, el RSAM por hora es constante en cien unidades. Naturalmente, la probabilidad clásica es la que se encarga de estimar numéricamente el grado de ocurrencia o no ocurrencia de un fenómeno natural o de otra índole.

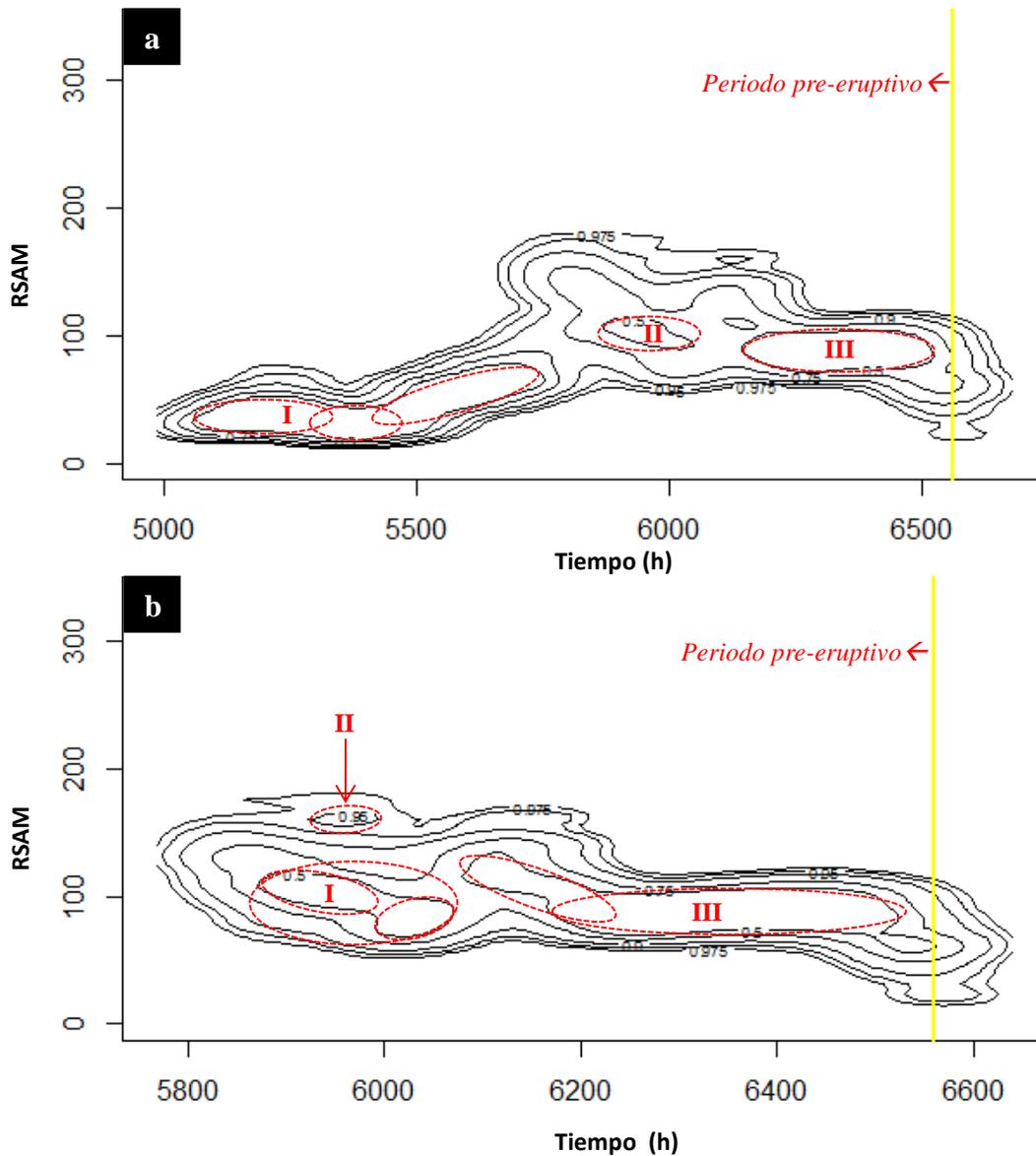


Ilustración 20. Kernels RSAM agosto 2005 y septiembre del 2005, VSA.

a) Densidad con Kernel Binario del RSAM por hora del volcán Santa Ana, agosto previo a la erupción volcánica. b) Densidad con Kernel Binario del RSAM por hora del volcán Santa Ana, para septiembre previo a la erupción volcánica.

Si se considera su definición, luego se establece que dicha estimación se define como el cociente entre el número de aciertos y el máximo de resultados posibles, sin tomar en cuenta factores externos. El número de aciertos es fácilmente calculable, pues basta con enumerar cuántos centroides de datos hay en la serie temporal para un momento determinado, y se obtiene el valor del numerador. Mientras que para la estimación de la

máxima probabilidad de ocurrencia de los centroides, es necesario aplicar el teorema de Dirichlet. Es una solución a los problemas de tipo discreto que tienen circunstancias requeridas para su debida aplicación. El cálculo de este cociente es aplicado a un intervalo de tiempo igual a un mes, manteniendo circunstancias típicas del vulcanismo, pueden existir de forma general treinta núcleos.

**Tabla 15.** Escenario de Kernels II VSA

Fig.	N	Tamaño relativo del Kernel	Límite inferior (h)	Límite superior (h)	Nivel de confianza (%)	Media RSAM	Días de anticipo
a	I	Grande (461h-19.21d)	5089	5550	50%	37.69	42.13
	<b>II</b>	<b>Mediano (430h-17.92d)</b>	<b>5570</b>	<b>6000</b>	<b>50%</b>	<b>99.23</b>	<b>23.38</b>
	<b>III</b>	<b>Grande (375h-15.63d)</b>	<b>6125</b>	<b>6500</b>	<b>50%</b>	<b>93.50</b>	<b>2.54</b>
b	I	Grande (267h-11.13d)	5833	6100	50%	107.43	19.21
	<b>II</b>	<b>Mediano (130h-5.42d)</b>	<b>5870</b>	<b>6000</b>	<b>95%</b>	<b>116.09</b>	<b>23.38</b>
	III	Grande (400h-16.67d)	6100	6500	50%	95.56	2.54

Para una semana, considerando un centroide por cada día, se disponen siete núcleos. Para un año, considerando un centroide por cada mes, de forma general se disponen de doce núcleos, y así sucesivamente. No es un pronóstico de erupción volcánica y puede ser influenciado tanto por factores externos o endógenos como por la naturaleza aleatoria que es característica en la ocurrencia de estos fenómenos. El origen de esta idea parte de consideraciones lógicas con cierto grado de flexibilidad o especificación para mayor certeza en los pronósticos; la demostración de estas proposiciones aún no existe. La erupción del primero de octubre del dos mil cinco se puede pronosticar desde el mes de julio del mismo año, el comportamiento del volcán Santa Ana una semana antes de que ocurra la erupción lo afirma explícitamente (ver ilustración 21). El núcleo uno del escenario III literal a (ver tabla 16) es grande en dimensión e intensidad, tiene un promedio de ochenta y nueve con treinta y cuatro centésimas unidades RSAM y aparece con dos días de anticipo.

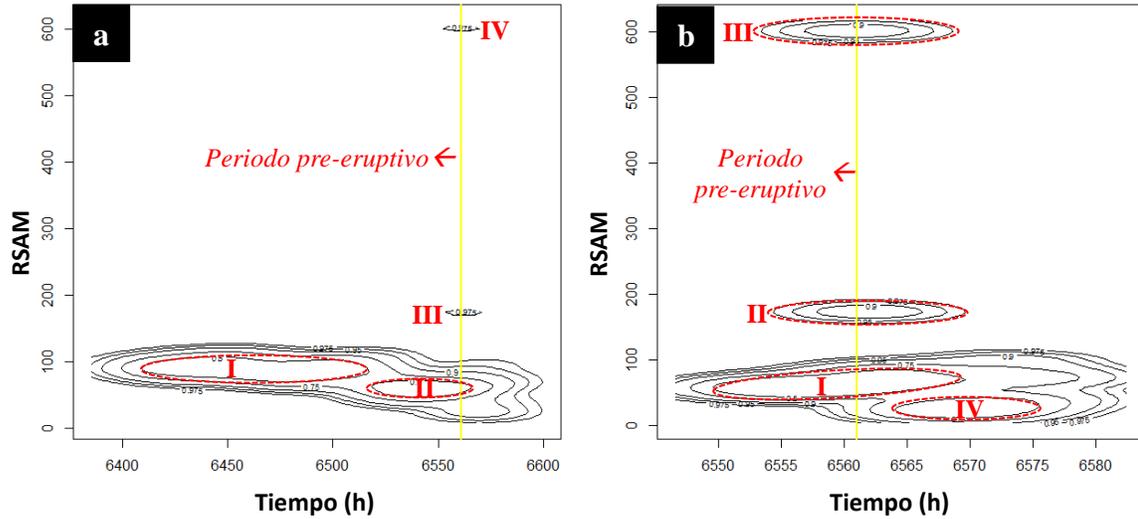


Ilustración 21. Kernels RSAM semana y horas previas a la erupción, 2005, VSA.

a) Densidad con Kernel Binario del RSAM por hora del volcán Santa Ana, una semana previa (desde el 25 de septiembre del 2015) a la erupción volcánica. b) Densidad con Kernel Binario del RSAM por hora del volcán Santa Ana, para las primeras horas del día de la erupción.

Este núcleo es el más importante de esta semana, tiene un nivel de confianza del cincuenta por ciento y es constante alrededor de las cien unidades RSAM consta de una duración de cuatro o cinco días. Para el escenario III literal a (ver tabla 16), los núcleos tres y cuatro son opuestos al núcleo uno, son de dimensión pequeña pero de alta intensidad. El comportamiento de esta pareja de núcleos refleja que aparecen traslapados en cierta forma y considerablemente separados del resto de observaciones. Ambos tienen un nivel de confianza del noventa y siete punto cinco por ciento y ofrecen alrededor de medio día de anticipo a la erupción volcánica.

Tabla 16. Escenario de Kernels III VSA

Fig.	N	Tamaño relativo del Kernel	Límite inferior (h)	Límite superior (h)	Nivel de confianza (%)	Media RSAM	Días de anticipo
a	<b>I</b>	<b>Grande (106h-4.42d)</b>	<b>6409</b>	<b>6515</b>	<b>50%</b>	<b>89.34</b>	<b>1.92</b>
	II	Mediano (45h-1.88d)	6515	6560	50%	62.13	0.04
	III	Grande (21h-0.88d)	6555	6576	97.5%	84.65	0.63*
	IV	Grande (20h-0.83d)	6560	6580	97.5%	92.29	0.79*
b	I	Grande (17h-0.71d)	6553	6570	50%	91.50	0.38*
	II	Mediano (15h-0.63d)	6555	6570	97.5%	95.78	0.38*
	III	Grande (18h-0.75d)	6552	6570	97.5%	91.50	0.38*
	<b>VI</b>	<b>Grande (11h-0.46d)</b>	<b>6564</b>	<b>6575</b>	<b>50%</b>	<b>50.14</b>	<b>0.58*</b>

\* Extrapolación de Kernels respecto a los datos originales, el anticipo de días corresponde al límite inferior del Kernel.

El mismo solapamiento ocurre en el esquema de los núcleos correspondiente a las primeras ocho horas del día de la erupción (ver ilustración 21). A diferencia del esquema de la semana del veinticinco de octubre (ver ilustración 21), los núcleos aislados del resto de observaciones no cambian de posición pero sí de magnitud. Es necesario hacer la introspección de que el resto de datos varía poco, mientras los alejados se incrementan geoméricamente. No es lo mismo observar un simple gráfico secuencial de una variable que analizar un esquema donde las observaciones se asocian a partir de un parámetro. El gráfico secuencial explica de forma individual, la aplicación del análisis Kernel se basa en una interpolación probabilística de los datos y depende del grado de asociación. El análisis discreto de los centroides se complementa con el caso continuo mediante la teoría de las secciones cónicas, especialmente la elipse y su caso trivial la circunferencia, vinculada a la geometría analítica (ver tabla 17 y tabla 18).

Hay expresiones matemáticas que inciden directamente en el caso continuo. Elementos de la sección cónica elipse como la excentricidad de las curvas de nivel, seguido por las magnitudes de su eje menor y eje mayor, sus centros, su radio, su longitud, la distancia entre los focos, la longitud del lado recto, las coordenadas de sus vértices, las coordenadas de sus focos, si se trata de una elipse con ejes rotados, las ecuaciones de sus directrices; la cercanía o solapado entre cada elipse, la simetría de cada elipse, entre otros (ver tabla 17 y tabla 18); son características indicadoras del comportamiento de las observaciones del volcán. Un análisis continuo sobre la dimensión de los centroides aportan información valiosa. La naturaleza de cada centroide brinda información estadística para la obtención de resultados más cercanos a la realidad. El estudio de estas secciones cónicas puede realizarse en sistemas informáticos inventados o que requieran su invención.

Tabla 17. La elipse en coordenadas polares

<b>Caso 1: Ecuaciones de la elipse polar</b>	
Un foco en el polo y directriz correspondiente es perpendicular al eje polar	$r = \frac{ed}{1 \pm e \cos \theta}$
Un foco en el polo y directriz correspondiente es paralela al eje polar	$r = \frac{ed}{1 \pm e \sin \theta}$
<b>Caso 2: Perímetro y área de las funciones polares</b>	
Longitud	$L = \int_{\alpha}^{\beta} \sqrt{\left(\frac{dr}{d\theta}\right)^2 + r^2} d\theta$
Área	$A = \lim_{\ \Delta\  \rightarrow 0} \sum_{i=1}^n \frac{1}{2} [f(w_i)]^2 \Delta_i \theta = \int_{\alpha}^{\beta} [f(\theta)]^2 d\theta$

Donde  $e$  es la constante de excentricidad y  $d$  la distancia no dirigida entre el foco y la directriz correspondiente de la sección cónica. Para el cálculo de la longitud y del área se debe cumplir que  $\alpha < \beta$ .

Supóngase un cono circular recto de dos mantos con sus respectivas generatrices que contienen a un mismo punto, la unión de ambos vértices. Una generatriz es simplemente una recta contenida completamente en el cono. Así pues, una elipse se obtiene a partir de una sección cónica, si el plano corta a todas las generatrices, en vez de ser paralelo a alguna generatriz. Un caso especial de la elipse es la circunferencia, la cual se forma al cortar todas las generatrices y ser perpendicular al eje del cono. “Una elipse es el conjunto de puntos de un plano tales que la suma de sus distancias de dos puntos fijos es constante. Cada punto fijo se denomina foco” (Leithold, 1998).

Tabla 18. Geometría Analítica de la Elipse y la Circunferencia

<b>Caso 1: Forma canónica o estándar de la sección cónica: circunferencia</b>	
Ecuación canónica	$(x - h)^2 + (y - k)^2 = r^2$
Coordenadas del centro y radio	$C(h, k); r$
<b>Caso 2: Forma general de la sección cónica: circunferencia</b>	
Ecuación general	$x^2 + y^2 + Dx + Ey + F = r^2$
Cuando $D^2 + E^2 - 4F > 0$	$C\left(-\frac{1}{2}D, -\frac{1}{2}E\right)$ $r = \frac{1}{2}\sqrt{D^2 + E^2 + 4F}$
Cuando $D^2 + E^2 - 4F = 0$ (punto)	$C\left(-\frac{1}{2}D, -\frac{1}{2}E\right)$
Cuando $D^2 + E^2 - 4F < 0$	$W = \{\forall x, y \in R / f_{xy} = r^2, r < 0\} = \emptyset$
Excentricidad	$e = 1$
Perímetro	$P = 2r\pi$

<b>Caso 3: Forma canónica o estándar de la sección cónica: elipse horizontal</b>	
Longitud del eje mayor	$L_+ = \overline{V'V} = 2a$
Longitud el eje menor	$L_- = 2b$
Distancia entre los focos	$\overline{F'F} = 2c$
Excentricidad	$e = \frac{c}{a} = \frac{\sqrt{a^2 - b^2}}{a} < 1$
Longitud del lado recto	$L_{recto} = \frac{2b^2}{a}$
Desigualdades	$c < a, a > b$
Teorema de Pitágoras	$c^2 = a^2 - b^2$
Ecuación canónica de la elipse	$\frac{(x - h)^2}{a^2} + \frac{(y - k)^2}{b^2} = 1$
Vértices	$V(h + a, k), V'(h - a, k)$
Focos	$F(h + c, k), F'(h - c, k)$
Centro	$C(h, k)$
Directrices	$x = h \pm \frac{a}{e}$
<b>Caso 4: Forma canónica o estándar de la sección cónica: elipse vertical</b>	
Ecuación canónica de la elipse	$\frac{(x - h)^2}{b^2} + \frac{(y - k)^2}{a^2} = 1$
Vértices	$V(h, k + a), V'(h, k - a)$
Focos	$F(h, k + c), F'(h, k - c)$
Centro	$C(h, k)$
Directrices	$y = k \pm \frac{a}{e}$
<b>Caso 5: Forma general de la elipse</b>	
Ecuación	$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, AC > 0$
Discriminante	$B^2 - 4AC < 0$

La regresión Kernel ajusta los datos a una curva tomando en cuenta la ponderación para estimar la curva de tendencia, que geoméricamente corresponde al suavizado (ver ilustración 22). Al igual que toda regresión, la regresión Kernel utiliza la interpolación geométrica para el mejor ajuste de la curva a los datos. A partir del mes de julio el comportamiento del RSAM incrementa en magnitud. Crece exponencialmente, hasta inicios y mediados del mes de septiembre decrece y nuevamente crece parabólicamente. En la segunda mitad del mes de septiembre el comportamiento es constante y alcanza niveles de cien unidades. Finalmente, la tendencia del RSAM decrece y el volcán Santa Ana hace erupción el primero de octubre del dos mil cinco.

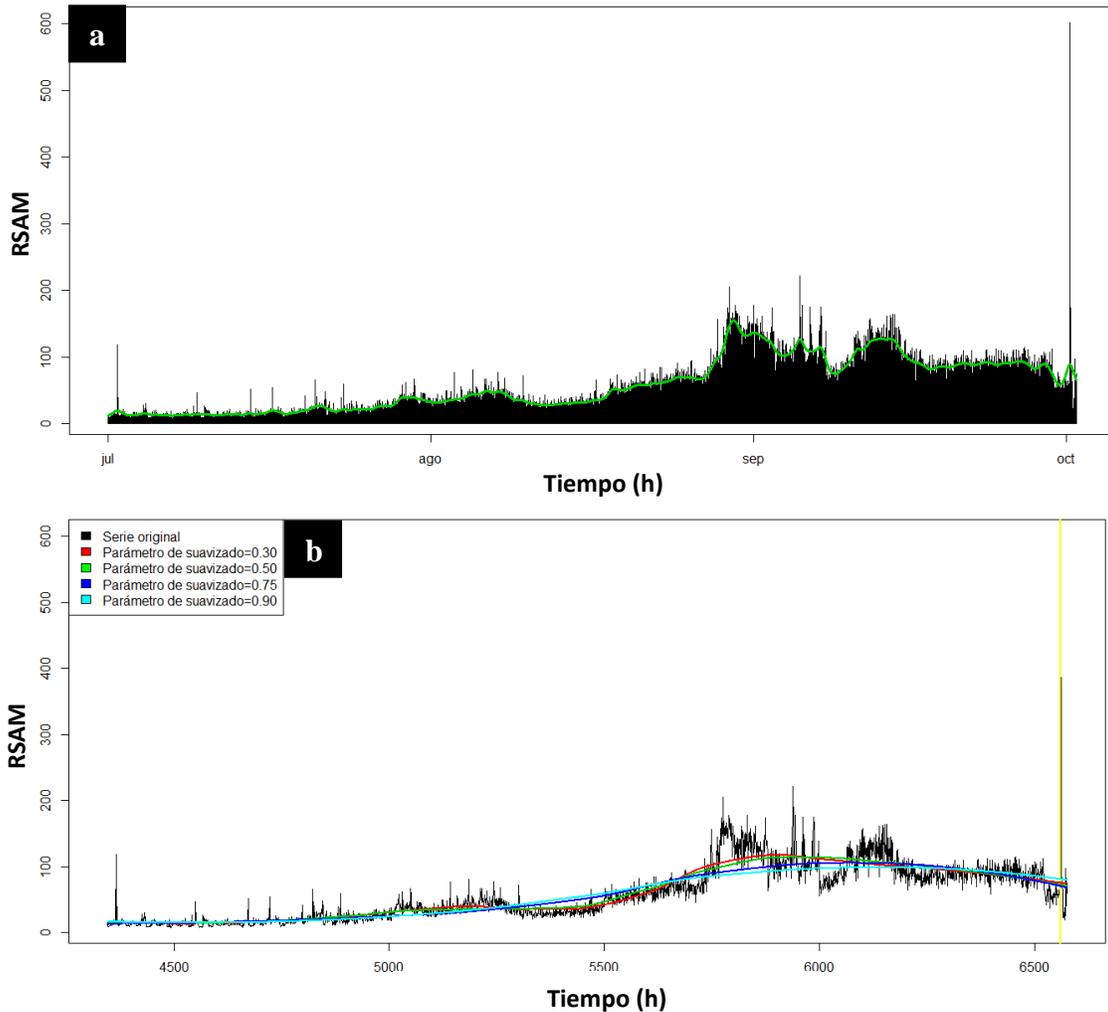


Ilustración 22. Regresión Kernel y regresión Loess, julio 2005, VSA.  
 a) Regresión Kernel del RSAM por hora VSA, julio-octubre, 2005, bandwidth=23.08. b) Regresión Loess del RSAM por hora VSA, julio-octubre, 2005.

En agosto, se incrementan las pendientes de las dos rectas de tendencias, tanto para la obtenida a través del filtro Medias Móviles, como para el Suavizado Exponencial (ver ilustración 23). Los valores de las pendientes de las tendencias anteriores, son aproximadamente igual que las correspondientes para el mes de agosto. En el análisis de tendencia se obtiene una pendiente de entre una o dos cienmilésimas. Para el caso del Suavizado Exponencial, el parámetro lambda que más ajusta a los datos es el de noventa y nueve cienmilésimas, la pendiente de la recta de tendencia es de entre una o dos cienmilésimas, exactamente es mayor que la pendiente anterior.

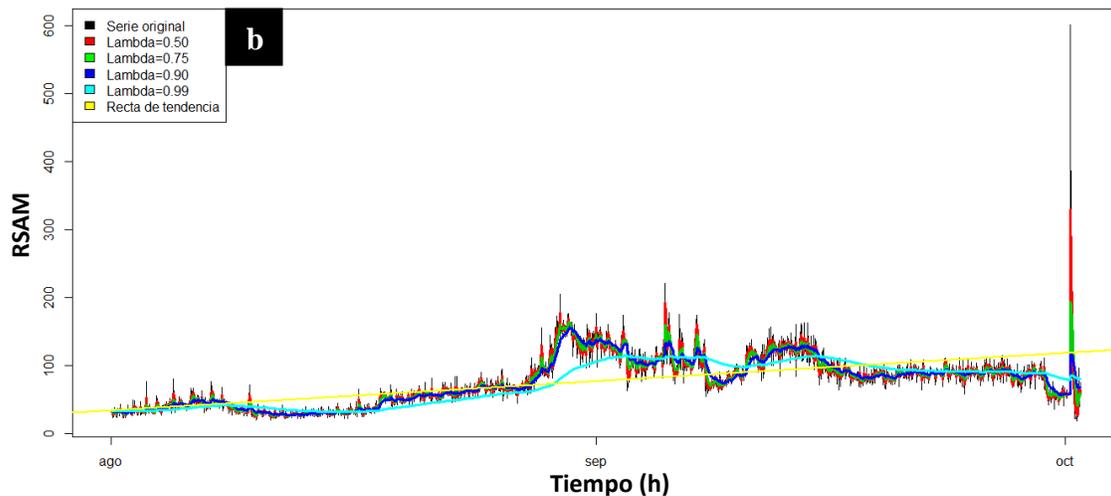
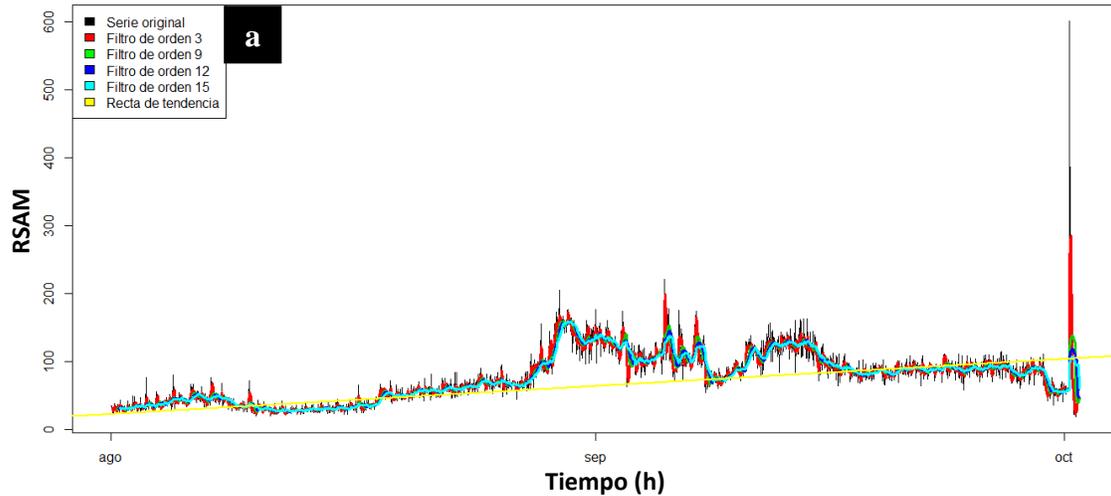


Ilustración 23. Regresión Kernel, MM y SE, Loess, RSAM, agosto 2005, VSA.

a) Análisis de tendencia con el filtro de las Medias Móviles para el RSAM, VSA, agosto-octubre, 2005. La recta de tendencia es  $RSAM = -1.710 \times 10^4 + 1.525 \times 10^{-5}Tiempo$ . b) Análisis de tendencia con el Suavizado Exponencial del RSAM, VSA, agosto-octubre, 2005. La recta de tendencia es  $RSAM = -1.81 \times 10^4 + 1.615 \times 10^{-5}Tiempo$ .

Para analizar la estacionalidad de la serie temporal del RSAM por hora se utiliza la Función de Autocorrelación y el Espectrograma (ver ilustración 24). El comportamiento de la Función de Autocorrelación para el año dos mil cinco es diferente al de las series estacionarias (ver ilustración 24). La cola derecha de la función no se aproxima a cero y decrece lentamente, es decir, la serie es muy persistente en seguir esa trayectoria, sin esperar que alcance valores negativos, incluso al cero. Al observar las líneas de la región de confianza al noventa y cinco por ciento, el descenso a valores de autocorrelación

próximos a cero parece lejano. En todo caso, se trata de una serie temporal no estacionaria. No se observan patrones ni cambios fuertes en la correlación de los periodos.

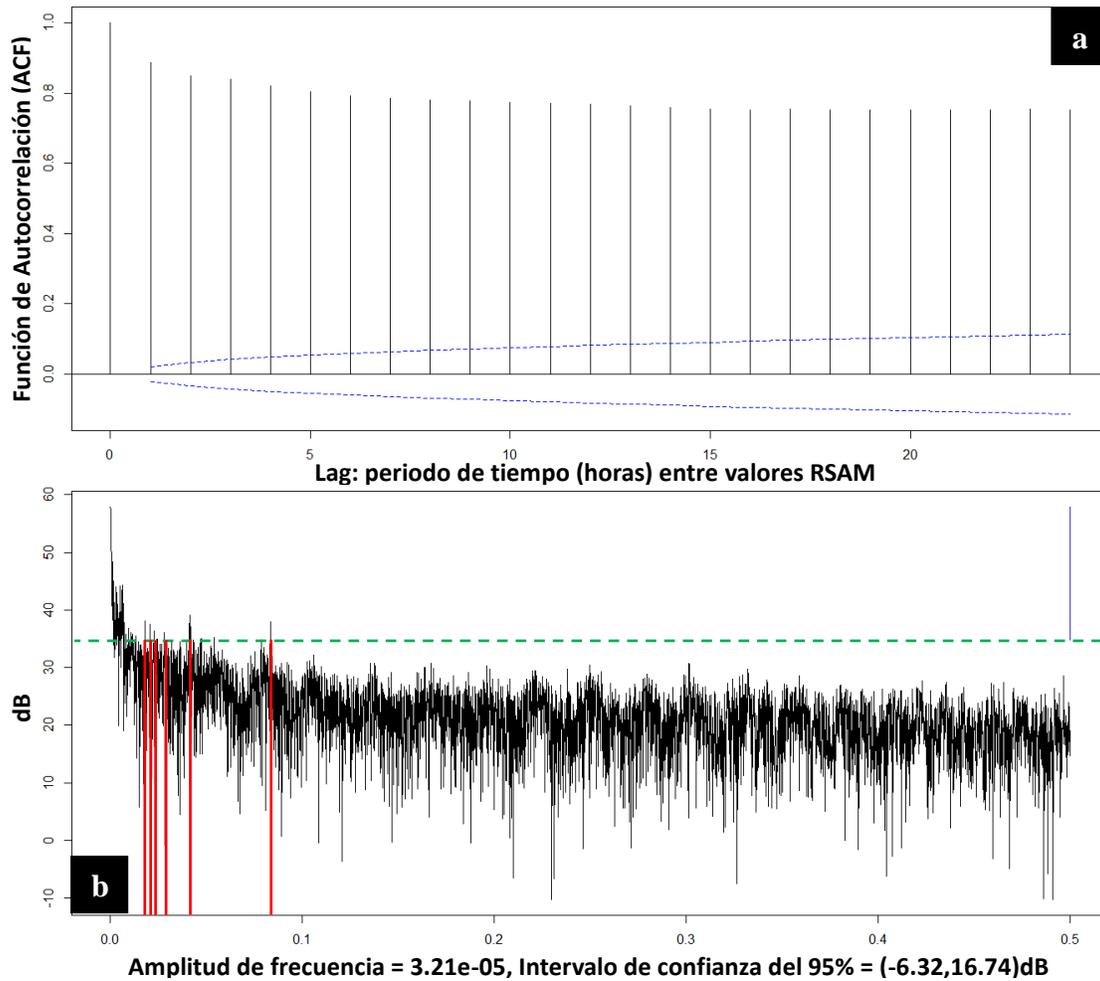


Ilustración 24. ACF con 24 lags y espectro, RSAM, 2005, VSA.

a) ACF con un nivel de confianza del 95% para el RSAM por hora VSA, año2005. b) Análisis espectral RSAM por hora con un nivel de confianza del 95%. Notas: “dB” significa decibeles.

El operador de rezago (“lag”, en inglés) de la Función de Autocorrelación que resulta conveniente analizar es el de veinticuatro horas, aunque el valor por defecto es de cuarenta horas. En la práctica es frecuente analizar la serie de modo temporal y también frecuencial. En el análisis temporal las observaciones son estudiadas tal y como son, mientras que; el análisis de frecuencias, llamado también análisis espectral, sirve para revelar la ausencia o presencia de distintas componentes, entre ellas la componente

periódica. El espectro o Función de Densidad Espectral, es una medida de la cantidad de oscilaciones en diferentes frecuencias que varían desde ciento ochenta grados negativos, hasta ciento ochenta grados positivos.

El análisis espectral revela que no hay presencia de componentes periódicas en la región de alta frecuencia ni en la de frecuencias bajas; el espectrograma también tiene un parámetro de suavizado pero no se incluyó aquí (ver ilustración 24). Al comparar el análisis de la función de autocorrelación con el análisis espectral, resulta que el primer análisis, permite obtener información indirectamente sobre las distintas frecuencias de la señal original respecto al tiempo; mientras que, el análisis espectral es una conversión de la señal original de dominio de tiempo a una señal que se basa en funciones ortogonales con dominio de las frecuencias. Este tipo de análisis fue desarrollado por Fourier entre el siglo dieciocho y diecinueve, producto de las investigaciones sobre las propiedades de la transferencia de calor en sólidos conductores.

El concepto que introdujo Fourier fue que “cualquier función arbitraria, incluso aquellas que muestran discontinuidades, puede ser aproximada mediante una expresión analítica sencilla.” Hay seis amplitudes de frecuencias importantes en el espectrograma. La frecuencia del extremo izquierdo que corresponde al recíproco de aproximadamente dos centésimas, la frecuencia del extremo derecho que corresponde al recíproco de aproximadamente nueve centésimas. La quinta frecuencia de izquierda a derecha se ubica cerca de las cuatro centésimas. Estas amplitudes son equivalentes a la aproximación de los siguientes valores, de izquierda a derecha en orden de aparición, las primeras tres tienen un ciclo cada cincuenta horas; la cuarta, un ciclo cada treinta y cuatro horas; la quinta, un ciclo cada veinte horas y la última, un ciclo cada trece horas.

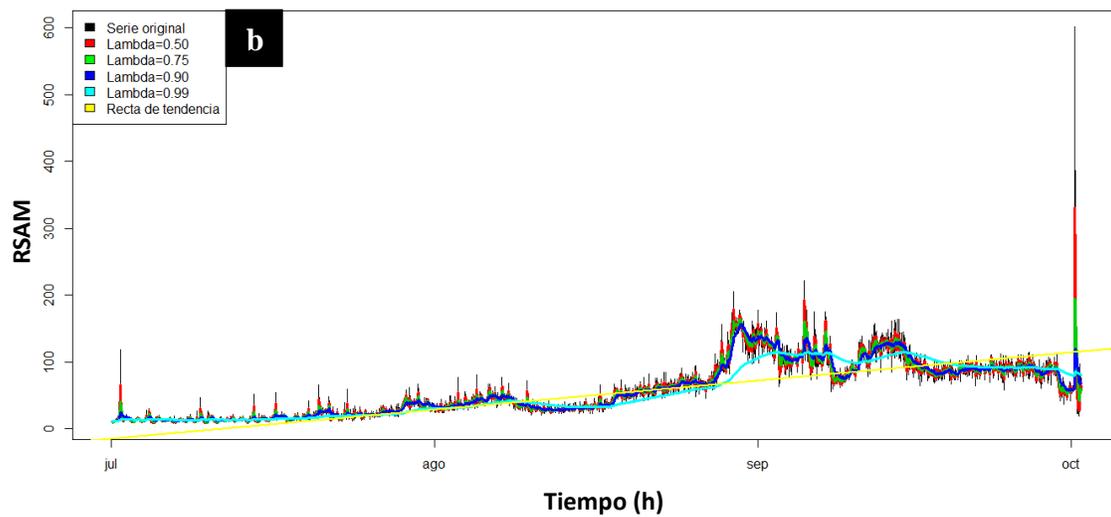
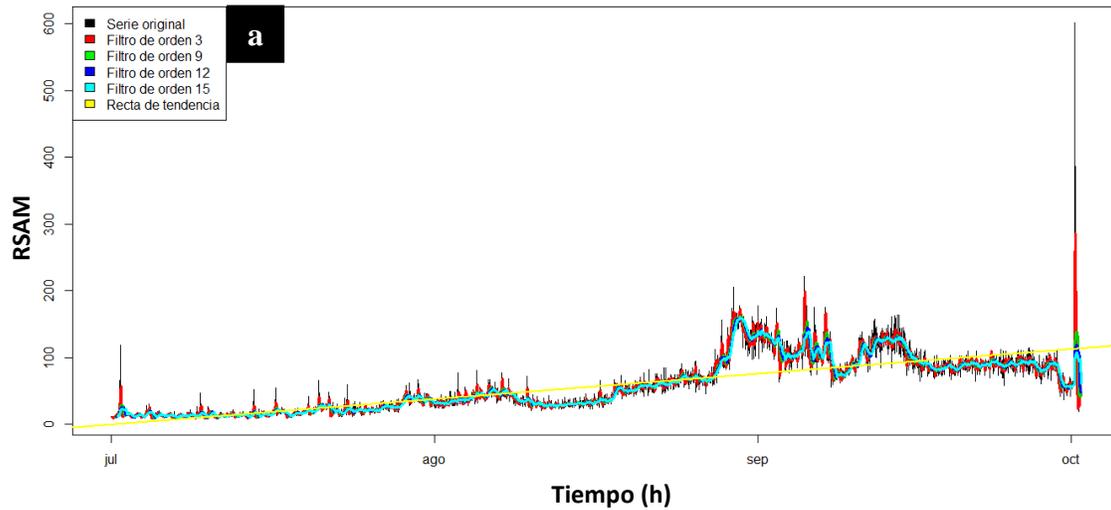


Ilustración 25. Medias Móviles y Suavizado Exponencial, RSAM, julio 2005, VSA.

a) Análisis de tendencia con el filtro de las Medias Móviles para el RSAM, VSA, julio-octubre, 2005. La recta de tendencia es  $RSAM = -1.593 \times 10^4 + 1.422 \times 10^{-5}Tiempo$ . b) Análisis de tendencia con el Suavizado Exponencial del RSAM, VSA, julio-octubre, 2005. La recta de tendencia es  $RSAM = -1.824 \times 10^4 + 1.627 \times 10^{-5}Tiempo$ .

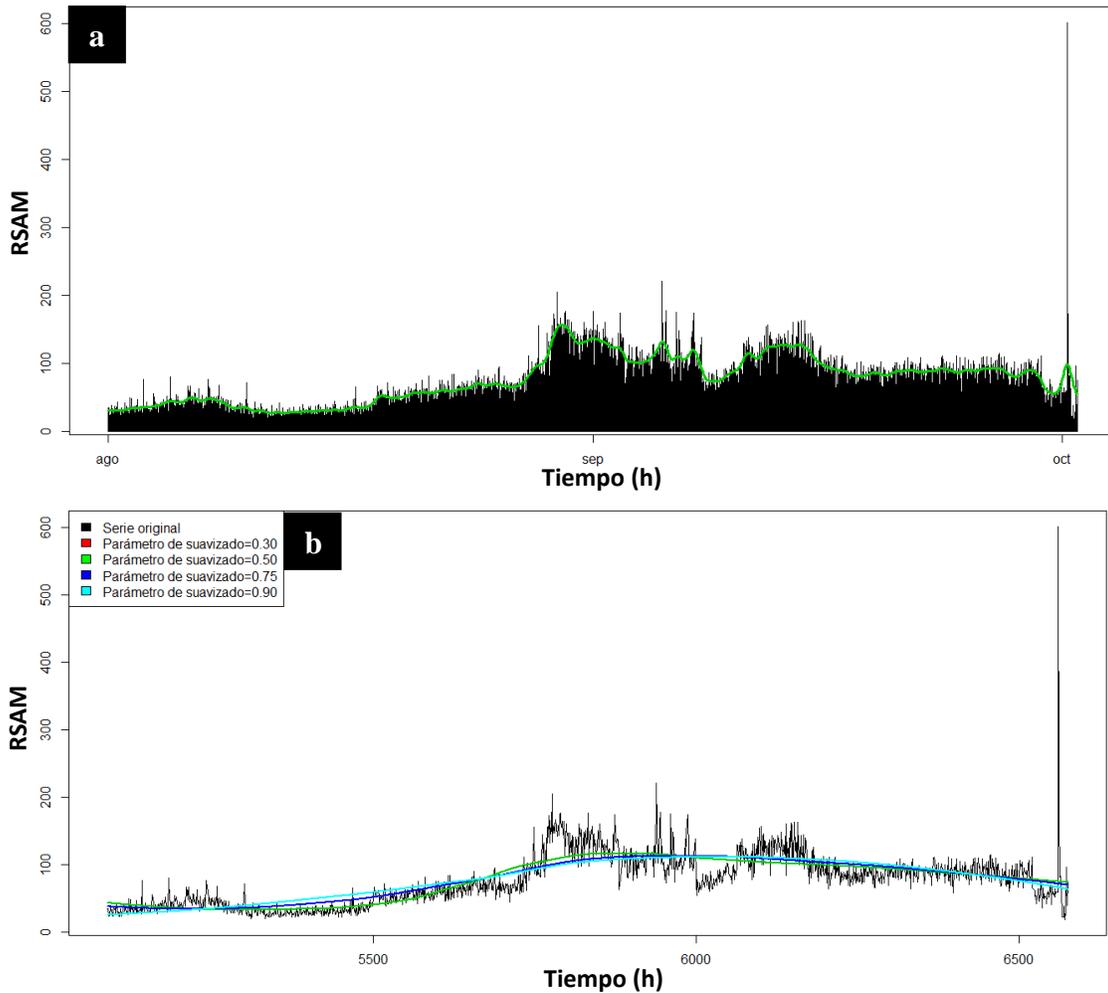
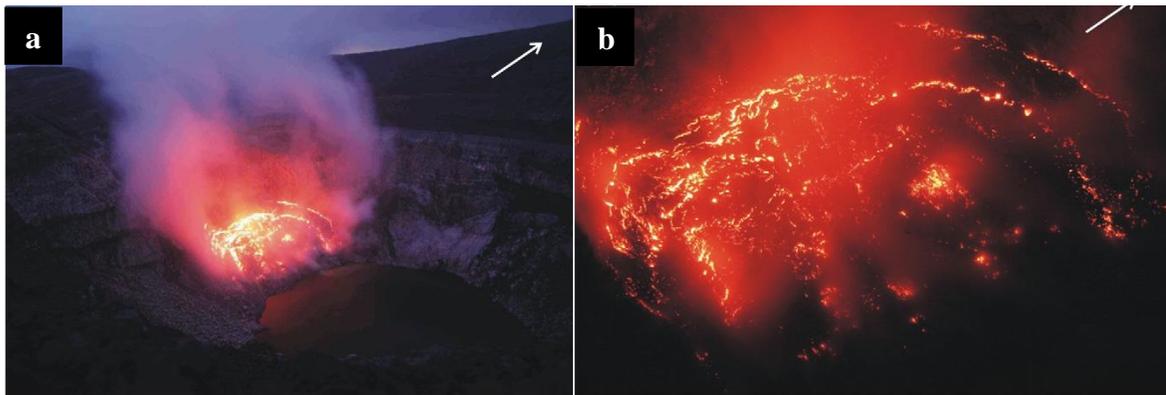


Ilustración 26. Regresión Kernel y regresión Loess, RSAM, agosto 2005, VSA.  
 a) Regresión Kernel del RSAM por hora VSA, agosto-octubre, 2005, bandwidth=17.14. b) Regresión Loess del RSAM por hora VSA, agosto-octubre, 2005.

Estas condiciones podrían afectar la naturaleza de la serie temporal original, pero debido a que estas frecuencias escasamente alcanzan la región de confianza y con auxilio de la función de autocorrelación, se puede conjeturar que no existe periodicidad en la serie temporal del RSAM por hora. Bajo este argumento se decidió la estadística se aplicaría a los datos brutos. Referente a la temperatura mensual del volcán Santa Ana, esta tiene un promedio de casi treinta y siete grados Celcius y una desviación estándar de casi cuatro grados Celcius, la distribución tiene sesgo negativo o sesgada hacia la izquierda, y su coeficiente de curtosis afirma que se trata de una distribución platycúrtica, una distribución sin apuntamientos, pero sí moderadamente achatada. Tiene un mínimo

de veintiocho grados Celcius y un máximo de cuarenta y dos grados Celcius.

A pesar de que los valores de la temperatura son de carácter continuos, la forma en que se ha realizado la medición de dicha variable es de tipo discreto y discontinuo. Esta es una barrera considerable a la hora de interpretar estadísticamente la temperatura volcánica ya que produce resultados pobres con una inmensa variación con respecto a los valores reales y con ningún sostén estadístico. En general, carecen de importancia en la búsqueda de modelos óptimos orientados al vulcanismo. No obstante, la medición de la temperatura es sustancial como premisa, pero el análisis de la temperatura anual es irrelevante debido a que no se cuenta con la información requerida por sucesos desconocidos.



*Ilustración 27.* Incandescencia del cráter en la erupción, 2005, VSA.

a) Vista del sudeste del cráter del volcán Santa Ana, el 31 agosto del 2005. Incandescencia, alta temperatura y concentración de gases. El diámetro del cráter de la laguna es de 200 metros. Las flechas apuntan al norte geográfico.  
b) Zoom del área de incandescencia. Cortesía de La Prensa Gráfica, El Salvador.

Otras variables que también son continuas pero que el ejercicio de adquirirlas es no es periódico son el dióxido de azufre, el boro, el calcio, los cloruros, la dureza, los fluoruros, el magnesio, los nitratos, el pH, el silicio y los sulfatos. Estas mediciones químicas son importantes para determinar si influyen en las erupciones volcánicas pero no son tratadas aquí por el inconveniente de la periodicidad, ya que estadísticamente son poco significantes en esa forma. Un comportamiento precursor con mucha anticipación producto de estas mediciones fue el que se observó alrededor de una década antes de la erupción volcánica del dos mil cinco. La laguna estuvo en ebullición (hervor) entre los años de mil novecientos noventa y dos y de mil novecientos noventa y tres.

En enero de mil novecientos noventa y tres se observó que la cantidad de sulfato era superior al cloro (ver tabla 19). El mismo efecto vuelve a pasar en enero y mayo del dos mil. Es evidente el cambio del índice de la relación sulfato-cloro, hay dos periodos donde el índice es mayor que uno, lo normal es que sea menor que uno. El primero con un índice promedio de uno y veinticuatro centésimas una década antes de la erupción volcánica del dos mil cinco; y el otro con un índice promedio de uno y treinta y cinco centésimas. Después de la década del noventa y tres descendió, pero después del dos mil ascendió nuevamente. El índice se mantuvo siempre anormal incluso alcanzando un máximo de uno y ochenta y un centésimas el treinta y uno de agosto del dos mil cuatro.

**Tabla 19.** Mediciones químicas durante ebulliciones de la laguna VSA

Fecha	SO <sub>4</sub>	Cl	Índice SO <sub>4</sub> / Cl	Interpretación
25-nov-92	4,300.00	6,284.90	0.68	Normal
<b>08-ene-93</b>	<b>6,300.00</b>	<b>4,947.47</b>	<b>1.27</b>	<b>Anormal</b>
<b>22-ene-93</b>	<b>6,300.00</b>	<b>5,178.39</b>	<b>1.22</b>	<b>Anormal</b>
10-mar-93	8,400.00	15,395.23	0.55	Normal
25-nov-93	4,300.00	6,284.90	0.68	Normal
<b>28-ene-00</b>	<b>9,000.00</b>	<b>6,829.00</b>	<b>1.32</b>	<b>Anormal</b>
<b>18-may-00</b>	<b>10,000.00</b>	<b>7,262.00</b>	<b>1.38</b>	<b>Anormal</b>

Detalle de la relación del sulfato y el cloro en la laguna del volcán Santa Ana en los periodos de ebullición, El Salvador. Fuente: MARN, entidad gubernamental.

Parcialmente, debido a los productos anteriores a la erupción y completamente durante la erupción del dos mil cinco, se perdió el análisis temporal de este índice muy importante y es retomado nuevamente hasta el treinta y uno de julio del dos mil cinco. El análisis químico data desde mil novecientos noventa y dos, termina hasta mil novecientos noventa y tres, es retomado en el dos mil, pero no hay datos del dos mil uno, ni del dos mil dos, ni del dos mil tres, sí lo hay del dos mil cuatro hasta el dos mil siete. Un posible modelo estadístico sería el índice precursor de una posible erupción volcánica (ver tabla 20), donde sigma es la cantidad extraída de sulfato y gamma la cantidad extraída de cloro a través de análisis químicos en la laguna del volcán Santa Ana.

Tabla 20. Índice IPPEV de cinco a diez años

Índice teórico	Discriminación del índice práctico	Interpretación
$V_{IPPEV} = \frac{\sigma}{\gamma}$	$V_{IPPEV} < 1$ $V_{IPPEV} \ll 1$	<i>Alarma correcta</i> <i>Falsa alarma</i> <i>Alarma fallida</i> Periodo normal: no hay indicios de posibles erupciones
	$V_{IPPEV} = 1$	<i>Alarma correcta</i> <i>Falsa alarma</i> <i>Alarma fallida</i> Periodo crítico: pueden ocurrir o no erupciones volcánicas
	$V_{IPPEV} > 1$ $V_{IPPEV} \gg 1$	<i>Alarma correcta</i> <i>Falsa alarma</i> <i>Alarma fallida</i> Periodo seguro: erupciones altamente probables

Interpretación de las cantidades de sulfato y cloro en las mediciones químicas en la laguna del volcán Santa Ana, El Salvador.

Si el índice precursor de una posible erupción volcánica, IPPEV, es menor que uno se espera un periodo normal sin alguna erupción, es muy seguro que se trate de una alarma correcta; si el índice tiende cero es más evidente que no pasará evento alguno. Cuando el índice es uno o alrededor de uno, podría presentarse una ambigüedad, ya que la probabilidad de que sea una alarma correcta es poco atinada, puede que se trate de una falsa alarma, o peor aún una alarma fallida. Y por último, si el índice es considerablemente mayor que uno hay tres posibles estados de la alarma, que sea correcta, falsa o fallida; entre más grande sea el valor del índice, más grande será la probabilidad de que se trate de una alarma correcta.

Lo anterior está basado en los escenarios de la década de los noventa y el año dos mil, cuando la laguna interior del volcán Santa Ana estuvo en ebullición y los niveles de sulfato y cloro cambiaron drásticamente como precursores de una posible erupción volcánica a largo plazo. Por otro lado, la búsqueda de modelos con regresión paramétrica, estimaciones a través de mínimos cuadrados; no es relevante ya que no hay una mejor línea que represente a los datos adecuadamente. El coeficiente de correlación es un valor que indica la razón entre la variación explicada respecto a la variación total;

en otras palabras, un modelo que vale la pena es aquel que tiene un coeficiente de correlación cercano al valor absoluto de la unidad, o sea que explica la mayor parte de la variación.

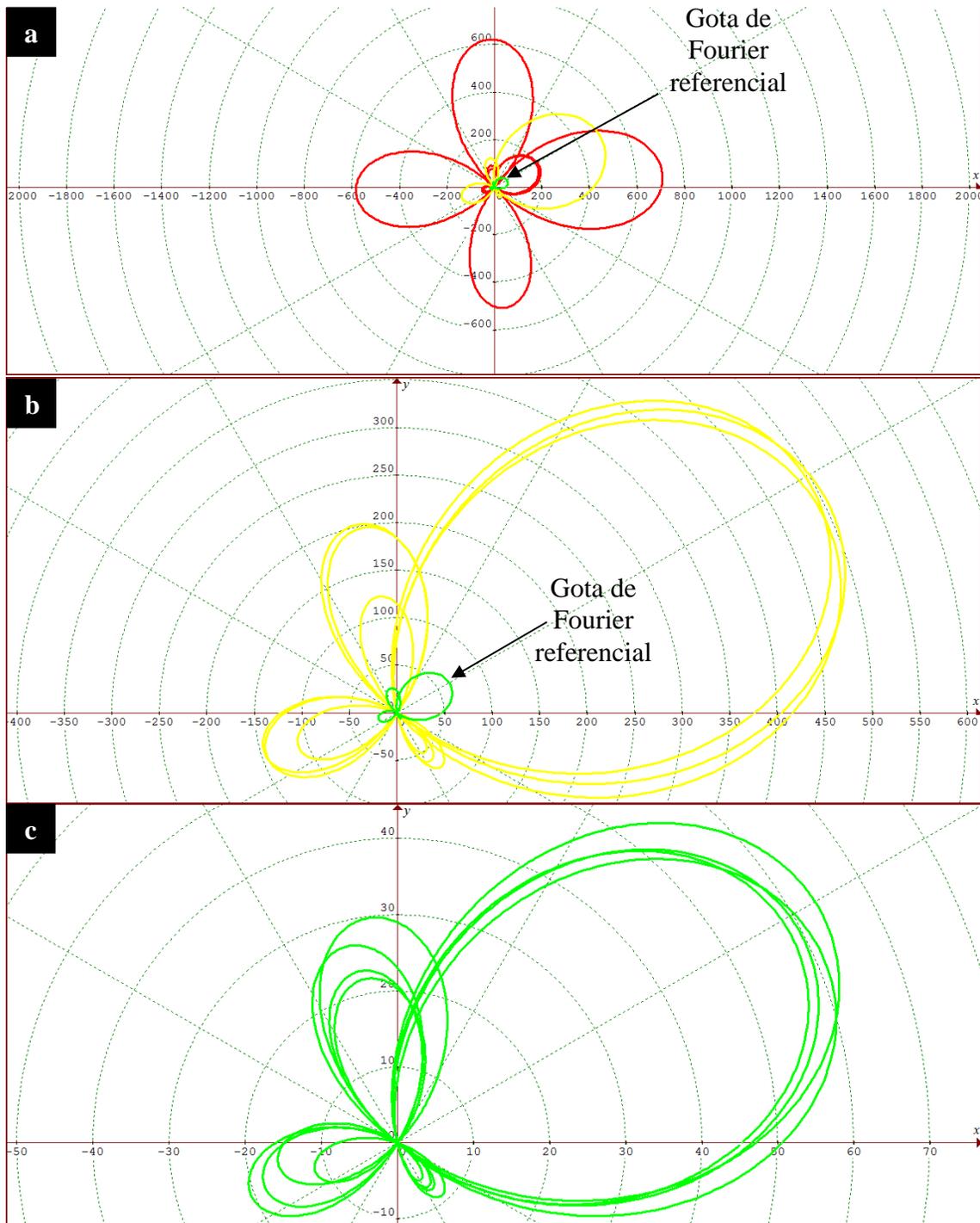
Si el modelo estadístico atraviesa esta prueba, después se le aplicarían el resto de condiciones y supuestos estadísticos, pero no es el caso. Prácticamente todos los coeficientes de correlación son bajísimos, que sea pobre un coeficiente de correlación quiere decir que su valor tiende a cero. Hay correlaciones positivas y negativas, de estas el coeficiente de mayor tendencia a cero es el de la longitud y la magnitud máxima, su valor es de tres milésimas negativas. Los mayores coeficientes son dos, pero ni siquiera alcanzan la media unidad; el mayor de todos es el correspondiente a la latitud y la magnitud máxima cuyo valor es cuatrocientas noventa y nueve milésimas positivas.

¿Para que serviría modelar estas variables? Si la latitud depende de la magnitud máxima, se estimaría la ubicación del sismo con determinada magnitud o más bien dicho, se predeciría la coordenada de la latitud dada la magnitud del sismo. Si la magnitud dependiese de la latitud, serviría para diseñar mapas donde los sismos podrían clasificarse en varios niveles debido a su magnitud alcanzada. Si la latitud depende del tipo de sismo medido, se ubicarían las regiones donde predomina algún tipo de sismo, ya sea el sismo registrado o el localizado. Y si fuese lo contrario, dada la latitud del sismo se predeciría si se trata de un sismo registrado o localizado.

La regresión lineal múltiple, a diferencia de la regresión lineal simple, incluye más de dos variables predictoras y su interpretación geométrica no es tan sencilla que se diga. También no cumple el rol de excelente modelo, ni siquiera moderado. El modelo que presenta más aceptación es en el que la latitud es la variable dependiente, y la longitud, la magnitud máxima y la profundidad en kilómetros son las variables predictoras. Finalmente, la contribución al análisis y búsqueda de modelos que permitan establecer pronósticos de erupciones volcánicas es muy variada. En este aspecto, las gotas de Fourier producto de la transformación aplicada a las curvas de Andrews, permiten monitorear todo tipo de comportamientos aleatorios, propios de cada complejo volcánico, los ejes polares están en unidades referenciales.

El primer análisis corresponde a las primeras horas del primero de octubre del dos mil cinco, año de la erupción volcánica (ver ilustración 28). Este lapso temporal inicia desde la una de la madrugada, hasta las ocho de la mañana, momento justo de la erupción. Las gotas de Fourier tienen cuatro partes en forma de pétalos, en este caso, en las primeras horas el comportamiento del RSAM no varía lo suficientemente como para darse cuenta de la cercanía de una erupción volcánica, aunque observando el eje mayor del pétalo mayor oscila entre los doscientos puntos lineales. En contraste, la observación que corresponde a la erupción volcánica es mayor que quinientos puntos RSAM.

Un mes previo el eje mayor del pétalo mayor oscila entre los cuatrocientos cincuenta y quinientos puntos y la diferencia entre cada pétalo es de aproximadamente cien puntos; esto es un ascenso de forma aritmético (ver ilustración 28). Y por último, hace un año de la erupción volcánica, el eje mayor del pétalo mayor oscila entre los sesenta puntos, hay un comportamiento que inicia de forma aritmética y finaliza de forma geométrica (ver ilustración 28). Las características en cada intervalo de tiempo son muy claras, y parecen indicar un crecimiento a veces aritmético y otras, geométrico. Los resultados de las gotas de Fourier aquí presentados sólo son intervalos diminutos, pero qué mejor si se proyectaran cada una de las observaciones registradas y mantener actualizado el esquema. Seguramente, los resultados serían innovadores en materia de vulcanismo, pues sería la medición categórica de la evolución del comportamiento de los complejos volcánicos.



*Ilustración 28. Gotas de Fourier 2005, VSA.*

a) Gotas de Fourier 8 horas previas a la erupción volcánica del 2005. b) Gotas de Fourier para las primeras 8 horas del 1 de sep. del 2005. c) Gotas de Fourier para las 8 horas del 1 de enero del 2005. Nota: los ejes de las coordenadas polares tienen valores referenciales. Verde indica un periodo muy normal; amarillo, un ascenso del RSAM; rojo, especial cuidado a los fenómenos que están ocurriendo en el interior o exterior del volcán. Nota: las flechas de las gotas detallan el movimiento creciente o decreciente. Las gotas rojas indican especial cuidado a la actividad volcánica; el amarillo, precaución; el verde, normalidad.

Para la verificación aleatoria de la aplicación de los Kernels Binarios por medio de la simulación MonteCarlo, se obtuvo una muestra de ocho mil setecientos sesenta observaciones con media aproximadamente de cuarenta y uno y desviación estándar de aproximadamente treinta y cinco (ver ilustración 29). Al graficar las observaciones correlativas, resulta una sola concentración de comportamiento normal que ondula entre la media y la desviación estándar. Las evoluciones de los esquemas ofrecen un estudio de comparación entre el comportamiento propio de cada volcán, de tal forma que resulte más fácil la aplicación de un análisis clúster y obtener información valiosa.

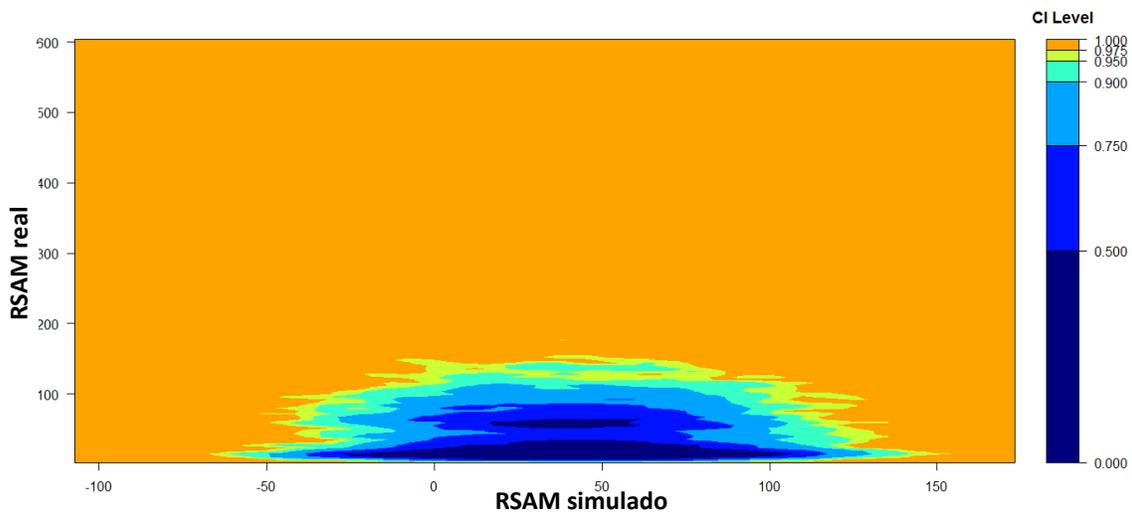


Ilustración 29. Gotas de Fourier, RSAM 2005, VSA.

c) Simulación MonteCarlo, 8760 observaciones, VSA. Kernels  $RSAM\ simulado \times RSAM\ real$  de una muestra con 8760 elementos con características normales,  $\mu = 41.1253$  y  $\sigma = 35.0097$ ; valores reales de la base de datos completa del RSAM por hora del volcán Santa Ana. Nota: los valores negativos pertenecen sólo a aspectos de forma extrapolada por el método.

## Capítulo 5: Modelamiento estadístico VSM

La discriminación de la naturaleza de los sismos previos a la erupción del veintinueve de diciembre del dos mil trece, refleja que a partir de mil novecientos noventa y seis se cuenta con el análisis de los sismos registrados (ver ilustración 30). Antes de este año, los sismos localizados son escasos y de valor mínimo. A partir de mediados de mil novecientos noventa y seis, es observable un ascenso abrupto de los tres tipos de sismos. Los sismos registrados predominan ante los sismos localizados. En un momento determinado la cantidad de sismos vulcanotectónicos y localizados son coincidentes. En el año dos mil trece es difícil analizar los tipos de sismos, ya que el comportamiento sísmico en diciembre es semejante al reflejado durante todo el año, con la única excepción de que a mediados del mismo mes, las fluctuaciones sísmicas son las más elevadas del año, y constantes en un periodo de tiempo menor a un mes.

Después de ser aproximadamente constante, desaparece estrepitosamente del análisis y no se vuelve a retomar durante el año dos mil trece. La cantidad de sismos de origen volcánicos crece de forma logarítmica a largo plazo; mientras que decrece, debido a las fluctuaciones, a corto plazo. Es contradictorio que la tendencia decrezca aun cuando se tenga presente la erupción volcánica del veintinueve de diciembre del dos mil trece, lo cual es una aseveración correcta, ya que carece de importancia debido al efecto de las grandes fluctuaciones; es decir, este fenómeno es totalmente normal cuando se aproxima una erupción volcánica (ver ilustración 30). Ahora bien, al aplicar el análisis temporal a la base de datos del volcán Santa Ana, hay periodos pre-eruptivos, periodos durante la erupción, y periodos post-eruptivos.

La ventaja de este aspecto radica en los tres diferentes tipos de análisis que se pueden realizar. En el caso del volcán San Miguel, sólo se cuenta con el periodo pre-eruptivo, ya que durante y después de la erupción volcánica se han perdido los datos probablemente resultado de la fuerte actividad volcánica. Las distribuciones de frecuencias de las variables implicadas del volcán San Miguel, en la presente investigación; tienen sesgos positivos y apuntamientos hacia arriba, es decir son leptocúrticas. El apuntamiento más fuerte lo tiene el RSAM por hora. La única base adecuada para la búsqueda de modelos

es la que está compuesta por las características fundamentales de los sismos para el intervalo de tiempo con cota inferior las catorce horas meridiano del cinco de enero de mil novecientos ochenta y siete, y cota superior las diecinueve horas meridiano del veintinueve de mayo del dos mil quince.

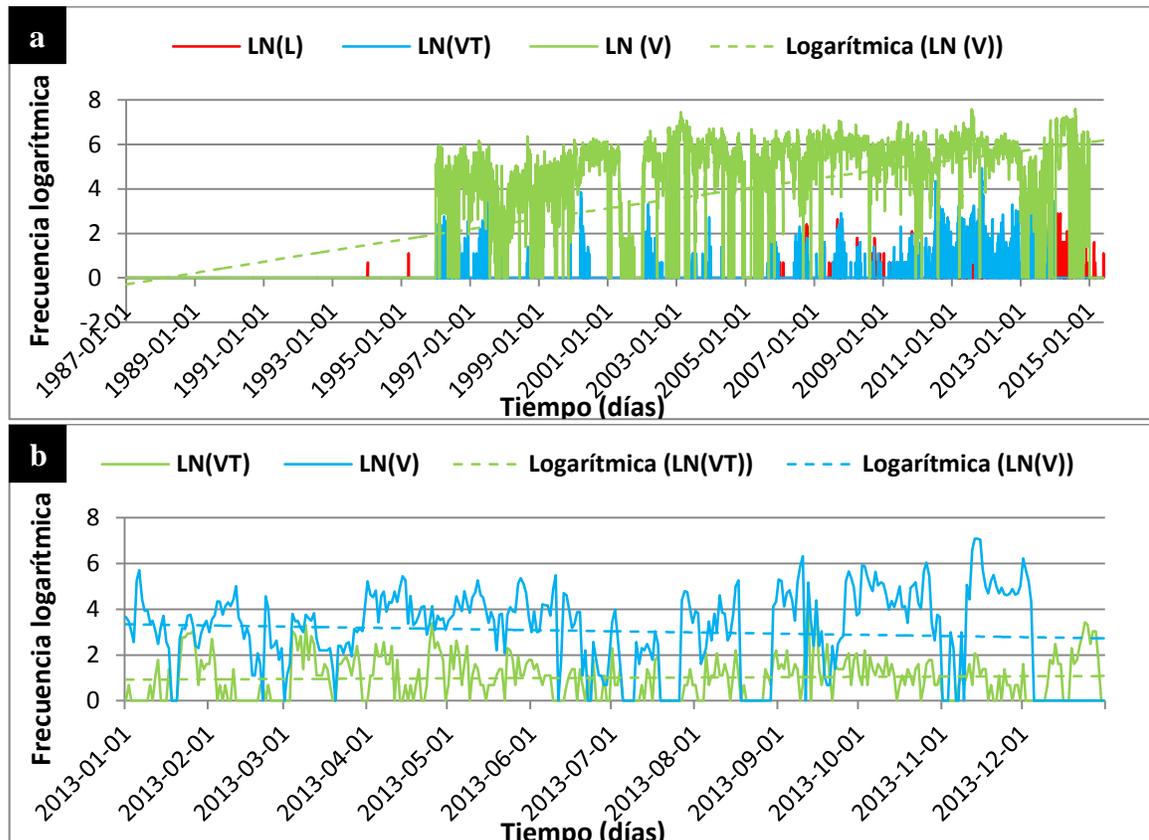


Ilustración 30. Secuencia gráfica de los principales tipos de sismos, 2013, VSM.

Gráfico semilogarítmico de los sismos vulcanotectónicos (VT), localizados (L) y registrados (R) para el volcán San Miguel; con sus respectivas líneas de tendencias logarítmicas. b) Gráfico semilogarítmico de los sismos volcánicos (V) y vulcanotectónicos (VT) para el año 2013, en el volcán San Miguel.

Las coordenadas geográficas: latitud y longitud, la magnitud máxima y la profundidad en kilómetros lineales son las únicas que se registran de manera correlativa. En el análisis de regresión paramétrica no se obtienen resultados significativos. La secuencia de las variables relacionadas con el tipo de sismo, tiene un comportamiento muy peculiar. La temperatura, aunque de forma discreta, se comporta exponencialmente; la latitud es casi una función constante a la altura de trece grados y cuarenta y cinco minutos, la longitud tiene forma de parábola negativa o abierta hacia abajo, la magnitud

máxima tiene una forma trigonométrica descrita por una función cosenoidal trasladada y la mitad de amplitud de un coseno normal.

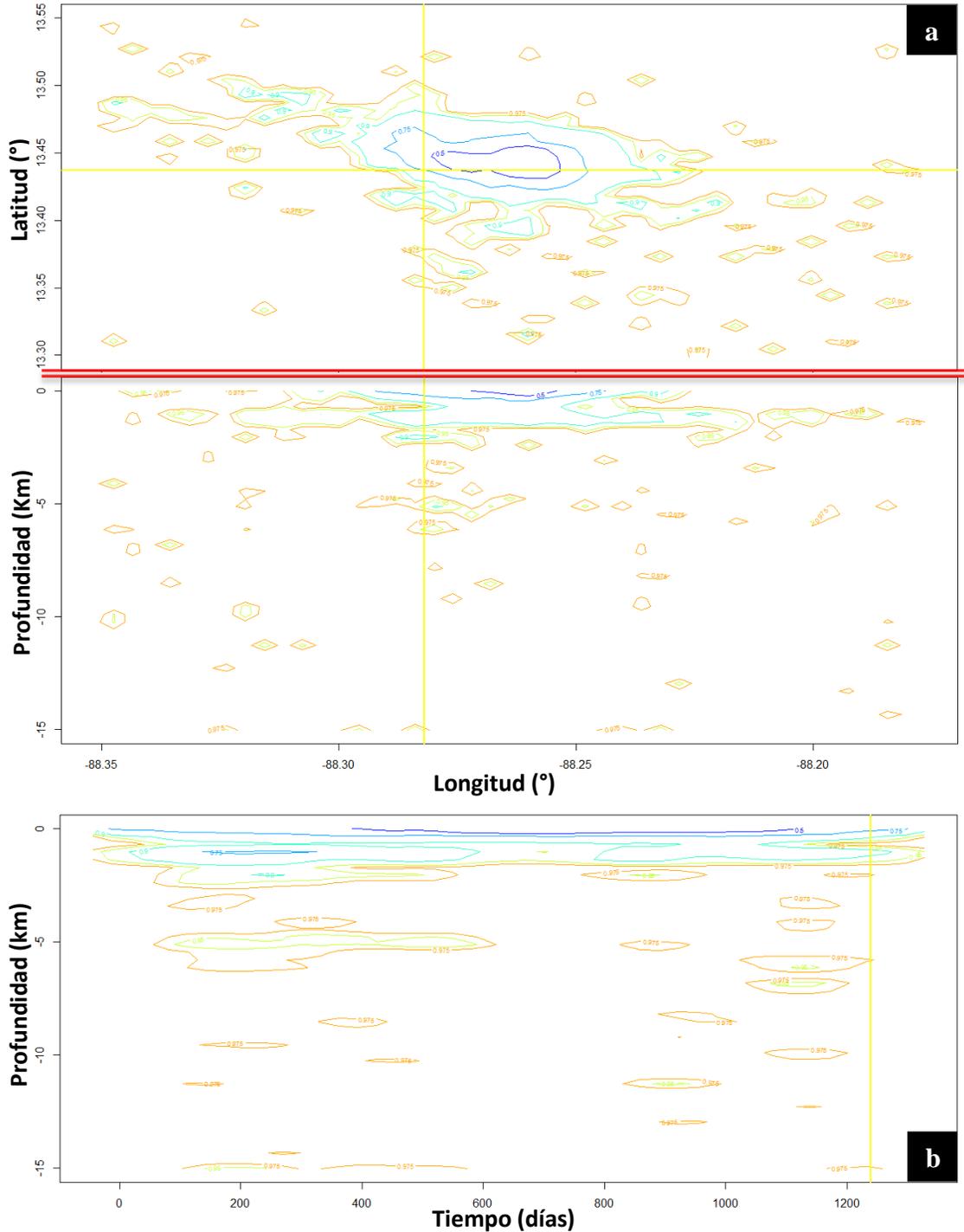
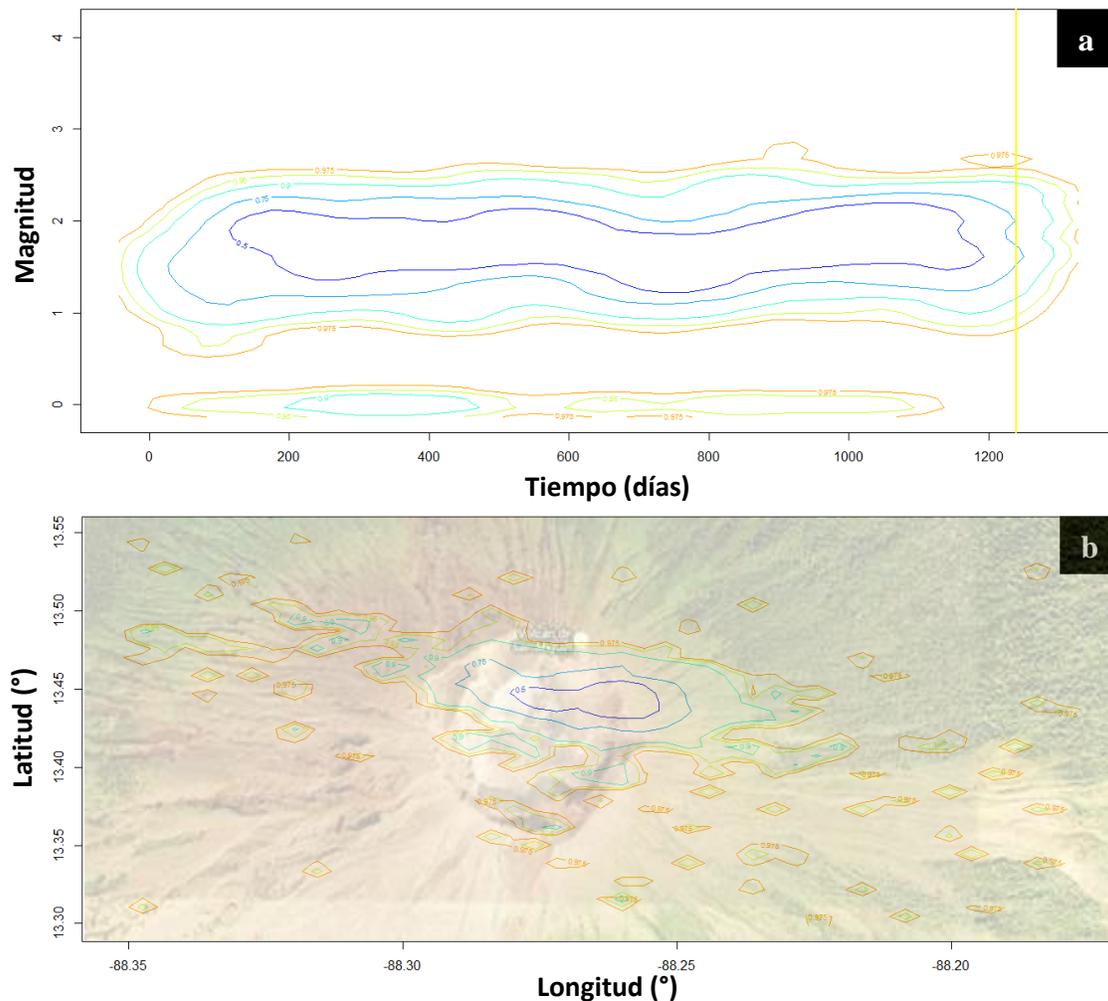


Ilustración 31. Kernels, profundidad, magnitud y coordenadas geográficas VSM.  
a) Kernels Profundidad versus coordenadas geográficas. b) Kernels de la Profundidad de los sismos.

Los Kernels de la profundidad muestran que a medida se acerca la erupción volcánica la sima de los sismos se encuentra a dos kilómetros lineales desde la superficie (ver ilustración 32). Los sismos próximos a la erupción son superficiales aunque no dejan de existir las fluctuaciones que son mayores o menores. Los Kernels aplicados a la profundidad sísmica reflejan la hipótesis del rompimiento del sistema hidrotermal del volcán (Olmos, 2016, comunicación personal). Al igual que la profundidad, la magnitud se comporta similarmente mostrando un leve descenso al acercarse al evento eruptivo (ver ilustración 32). Al comparar las observaciones anteriores al descenso, resulta que son mayores en valor absoluto.

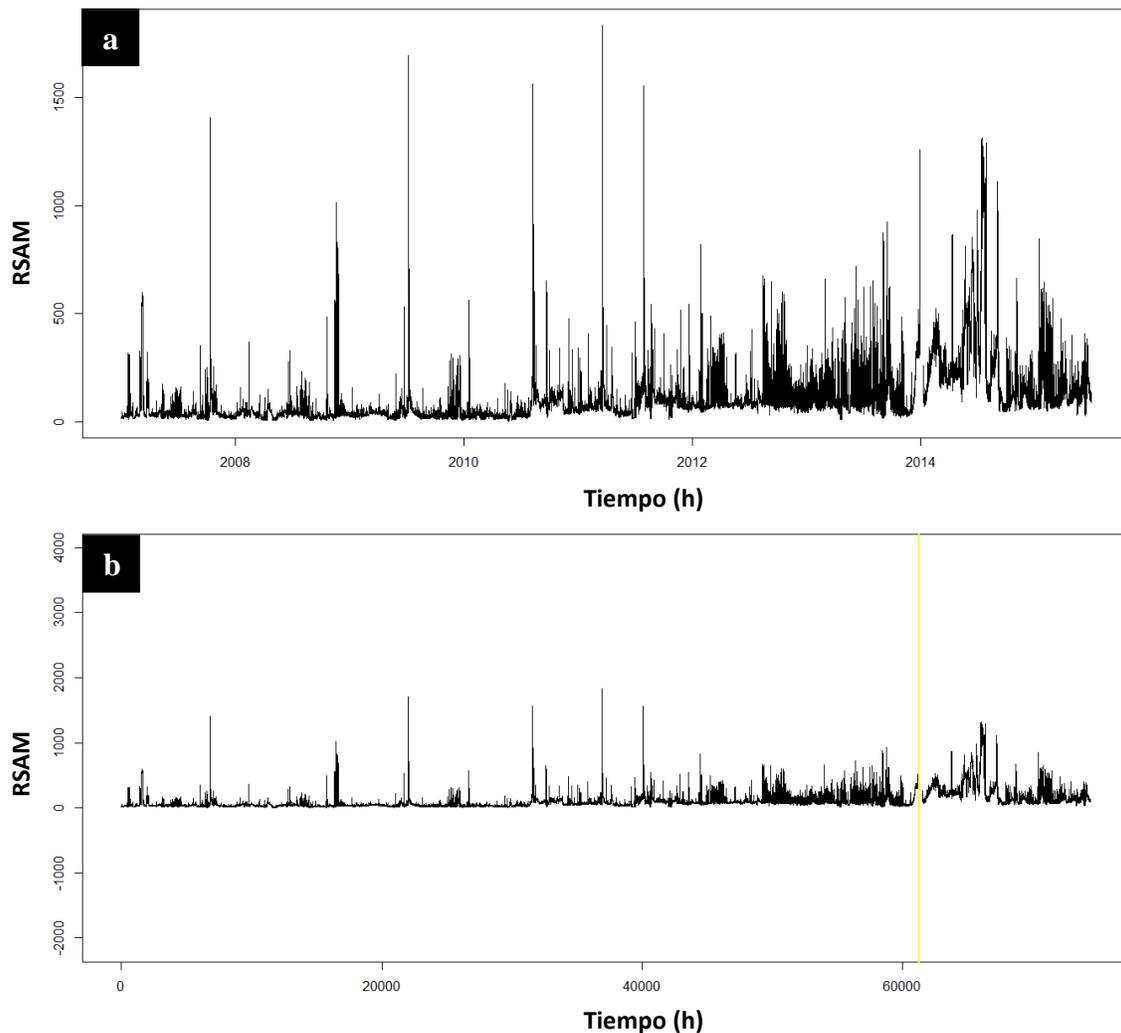


*Ilustración 32.* Kernels, profundidad, magnitud y coordenadas geográficas VSM.

a) Kernels de la magnitud de los sismos. b) Coordenadas geográficas utilizando Google Earth. Nota: el parámetro de asociación es de  $n_{bin}=45$ . Las líneas amarillas corresponden a la observación 24 de dic. del 2013 (1239), la última registrada antes de la erupción volcánica del 29 de dic. del 2013.

La concentración de coordenadas geográficas deja claro cuál es el centroide de la longitud y latitud, las líneas perpendiculares color amarillo representan el último registro de las características de los sismos, aunque no es exactamente el momento de la erupción pero sirven como una buena referencia al momento de realizar el análisis (ver ilustración 32). La ubicación real de las coordenadas geográficas admite evaluar el detalle con exactitud. Un resultado esperado y excelente es el de que bajo estas coordenadas se encuentre la cámara magmática del volcán, esta introspección es de difícil demostración con esta teoría, pues se necesita la anexión de otras pruebas y tests estadísticos. La comparación de los Kernels de la magnitud sísmica y de las coordenadas geográficas de los sismos, enuncia que son muy distintas entre sí. Los Kernels de la magnitud son más homogéneos y hay poca variación en comparación a los Kernels de las coordenadas geográficas.

Con respecto a la profundidad de los sismos se averigua que la mayoría no pasa el límite de un kilómetro, aunque hay varios sismos que alcanzan la profundidad máxima de quince kilómetros lineales. El RSAM por hora está conformado por un intervalo de tiempo cuya cota inferior es el primero de enero del dos mil siete a las cero horas meridiano, y cota superior es el veintinueve de junio del dos mil quince a las seis horas meridiano. Es un intervalo muy irregular, carece de varias observaciones faltantes debido a diversas causas tanto de la actividad sísmica como de la implementación, actualización efectividad de las herramientas del trabajo vulcanológico, lo mismo ocurre con las bases de datos del RSAM del volcán Santa Ana, la diferencia radica en que la base de datos para el volcán San Miguel se interrumpe justamente en el momento de la erupción volcánica de fecha veintinueve de diciembre del dos mil trece (ver ilustración 33).

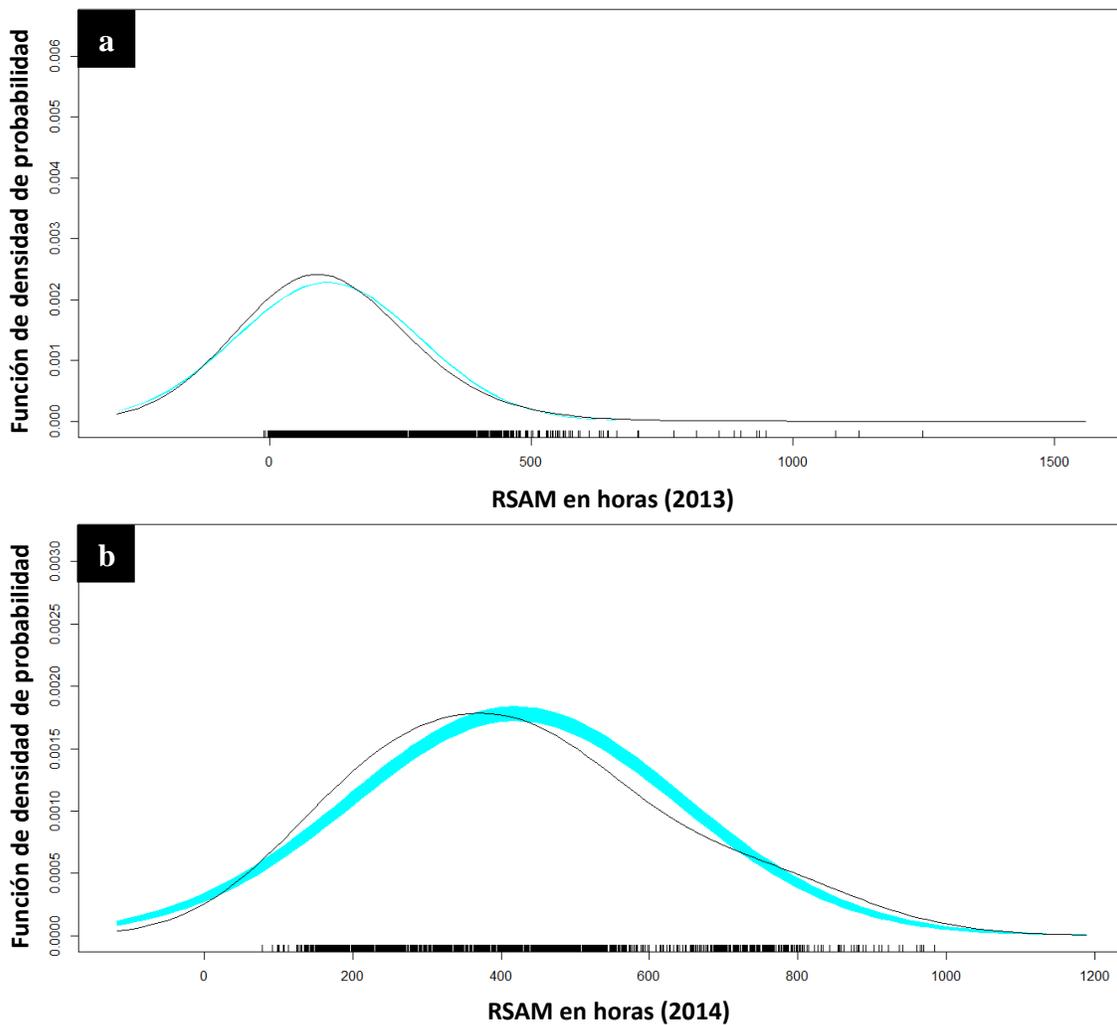


*Ilustración 33.* RSAM por hora, VSM.

a) Secuencia gráfica del RSAM por hora del volcán San Miguel 2007-2015. b) secuencia gráfica del RSAM por hora del volcán San Miguel 2007-2015, aspecto 5. El momento de la erupción se detalla por medio de una línea vertical amarilla.

Anteriormente se le denominaba el pseudónimo de última erupción volcánica del volcán San Miguel, pero otras más reciente del dos mil quince y dos mil dieciséis la han reemplazado; así que la descripción y la modelización se efectuarán a la erupción del veintinueve de diciembre del dos mil trece que relativamente se considera la primer erupción principal del volcán San Miguel. Al estimar la función de densidad existe una alta probabilidad de que ocurra un sismo, en un momento determinado, similar a los que se relacionan a las erupciones volcánicas, con una amplitud sísmica entre ciento cincuenta y doscientas unidades (ver ilustración 34). La probabilidad de que ocurra un

sismo de intensidad moderada, es de al menos tres milésimas, tres de cada mil sismos tendrán amplitud sísmica considerable. Al inspeccionar detalladamente la función de densidad del RSAM en el periodo post-eruptivo, las probabilidades de ocurrencia de un sismo disminuyen, pero las magnitudes aumentan (ver ilustración 34).



*Ilustración 34.* Función de densidad, RSAM, 2013 y 2014, VSM.

a) Función de densidad del RSAM por hora, año 2013. b) Función de densidad del RSAM por hora, año 2014.

La función de densidad para el dos mil catorce indica que la probabilidad de que ocurra un sismo de al menos cuatrocientas unidades RSAM, es de seis milésimas, esto es seis de cada mil sismos tendrán esta característica (ver ilustración 34). No hay modelo adecuado en la regresión lineal simple ni en la múltiple, el mayor coeficiente de correlación lineal es cuarenta y un centésimas negativas y pertenece a las coordenadas

geográficas. ¿Para qué serviría modelar las coordenadas geográficas? Como se afirmó anteriormente sobre la relación que guarda las coordenadas geográficas de los sismos y la ubicación real de la cámara magmática, dada la latitud como variable respuesta y la longitud como predictora, se estima la concentración de latitudes al conocer la tendencia de la longitud; esta situación es útil en el caso de que la latitud estimada coincida con la del cráter volcánico o flancos donde hay actividad volcánica (intrusiones de magma) anticipada a una erupción volcánica.

Gracias al actual apogeo de la investigación operativa, un análisis estadístico de sensibilidad reproduce regiones interesantes para la verificación de esta hipótesis. El modelo lineal simple de las coordenadas geográficas tiene un coeficiente de determinación demasiado bajo, por lo que no se puede contar con este modelo producto de la regresión lineal simple. Al aplicar la regresión lineal múltiple en las posibles combinaciones de las variables del tipo de sismo se estiman cinco modelos de alta dimensión, pero ninguno de ellos tiene un coeficiente de correlación lineal aceptable, el mayor de todos es donde la latitud es la variable respuesta. No es suficiente para empezar a manipular el modelo con las técnicas estadísticas de la regresión paramétrica.

El esquema de Kernels del periodo pre-eruptivo tiene ocho centroides en dos mil trece; para el periodo post-eruptivo existen doce centroides, el comportamiento es más heterogéneo que en el periodo pre-eruptivo. En el esquema de Kernels del periodo pre-eruptivo son notorios los centroides de datos, aunque los centroides del periodo post-eruptivo los superan en magnitud. Dada la inestabilidad sísmica con la que permanece el volcán San Miguel luego de su principal erupción, se localizan dos centroides; el primero es inmediatamente después de la erupción, el otro dista de al menos doscientos días después de la erupción volcánica.

Las curvas de nivel permiten apreciar conglomeraciones de datos. En el periodo pre-eruptivo del dos mil trece se observan al menos ocho concentraciones, el volcán Santa Ana presenta al menos siete concentraciones. La interpolación se aplica a la ciencia geofísica, especialmente a la sismología y vulcanología, se considera un problema importante resuelto por la matemática; el flujo del calor en una superficie, el potencial

eléctrico de una zona termal, las emisiones difusas de gases en los flancos de un volcán, la resistividad eléctrica de una fractura terrestre, la distribución geográfica de la pluviosidad y escorrentía superficial se expresan en campos, escalares o vectores que varían en tiempo y espacio (Lorenzo, 2004).

Hay interpolaciones determinísticas como la triangulación y el inverso de la distancia a una potencia; e interpolaciones aleatorias llamadas interpolaciones probabilísticas como el Kriging y el Kernel Binario que consideran la distribución espacial de la función de densidad de probabilidad (Lorenzo, 2004). En el análisis de la densidad Kernel del RSAM por hora del volcán San Miguel, la inestabilidad es mayor que la del volcán Santa Ana; tiene al menos cuatro concentraciones determinantes en el periodo pre-eruptivo. La mayor de las concentraciones dista de al menos trescientas horas antes de la erupción volcánica, y su magnitud es no mayor a las trescientas unidades RSAM (ver ilustración 35).

Un mes antes de la erupción del veintinueve de diciembre del dos mil trece hay tres núcleos bien definidos (ver ilustración 35). El menor en magnitud pero mayor en densidad, no supera las doscientas unidades RSAM; mientras que el núcleo menos denso está ubicado alrededor de las doscientas unidades, el próximo alcanza las trescientas unidades, el penúltimo a la erupción volcánica, ronda las cuatrocientas unidades; y el más próximo a la erupción, alcanza casi las seiscientas unidades. Todos los núcleos anteriores soportan el nivel de confianza del cincuenta por ciento, esto es muy importante porque no se tratan de observaciones aisladas o esporádicamente fuera del rango.

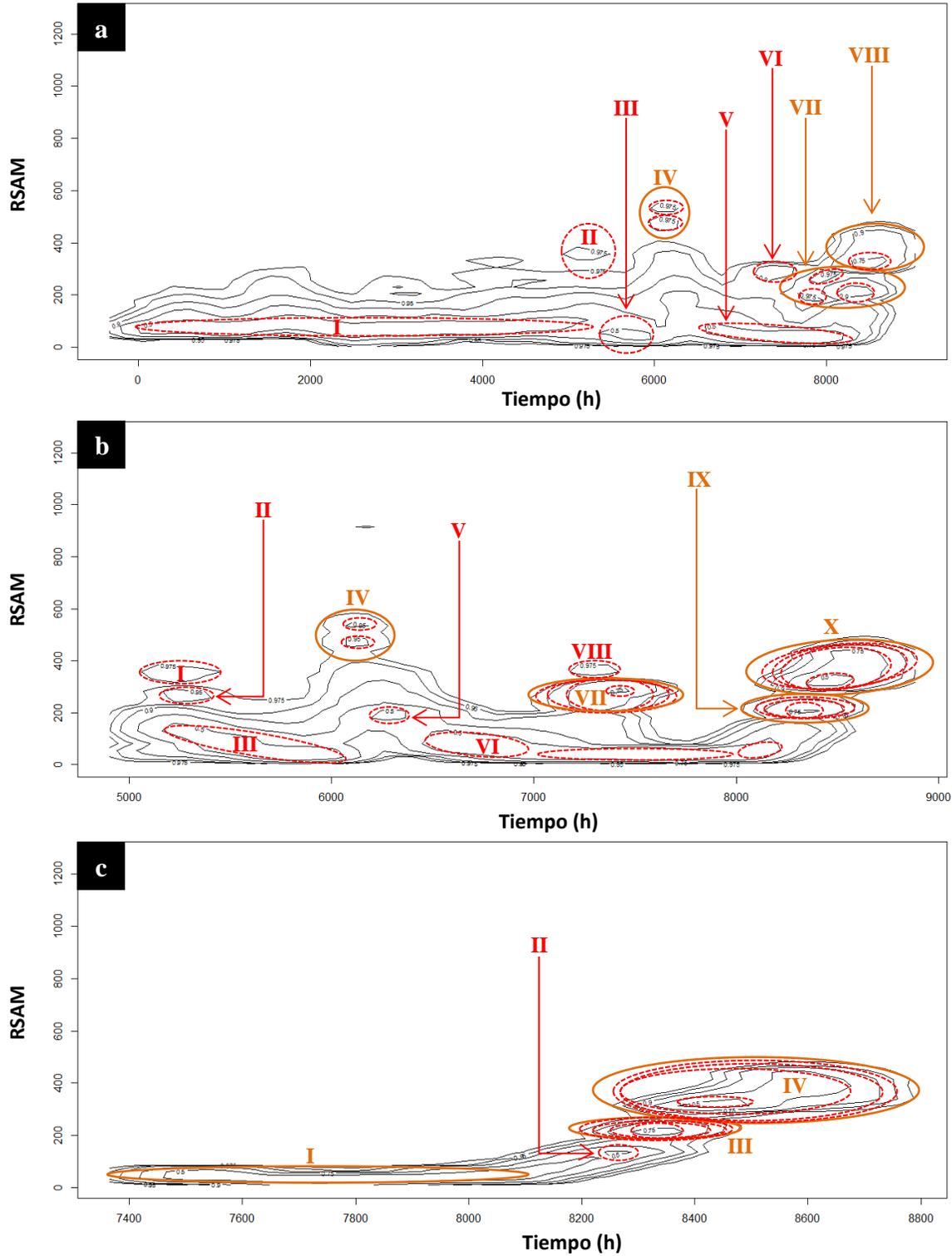


Ilustración 35. Kernels RSAM 2013, agosto 2013, noviembre 2013, VSM.  
 a) Kernel Binario del RSAM por hora VSM, 2013. b) Kernel Binario del RSAM por hora VSM, 1 de agosto 2013-29 de diciembre del 2013. c) Kernel Binario del RSAM por hora VSM, 7 de noviembre 2013-29 de diciembre del 2013).

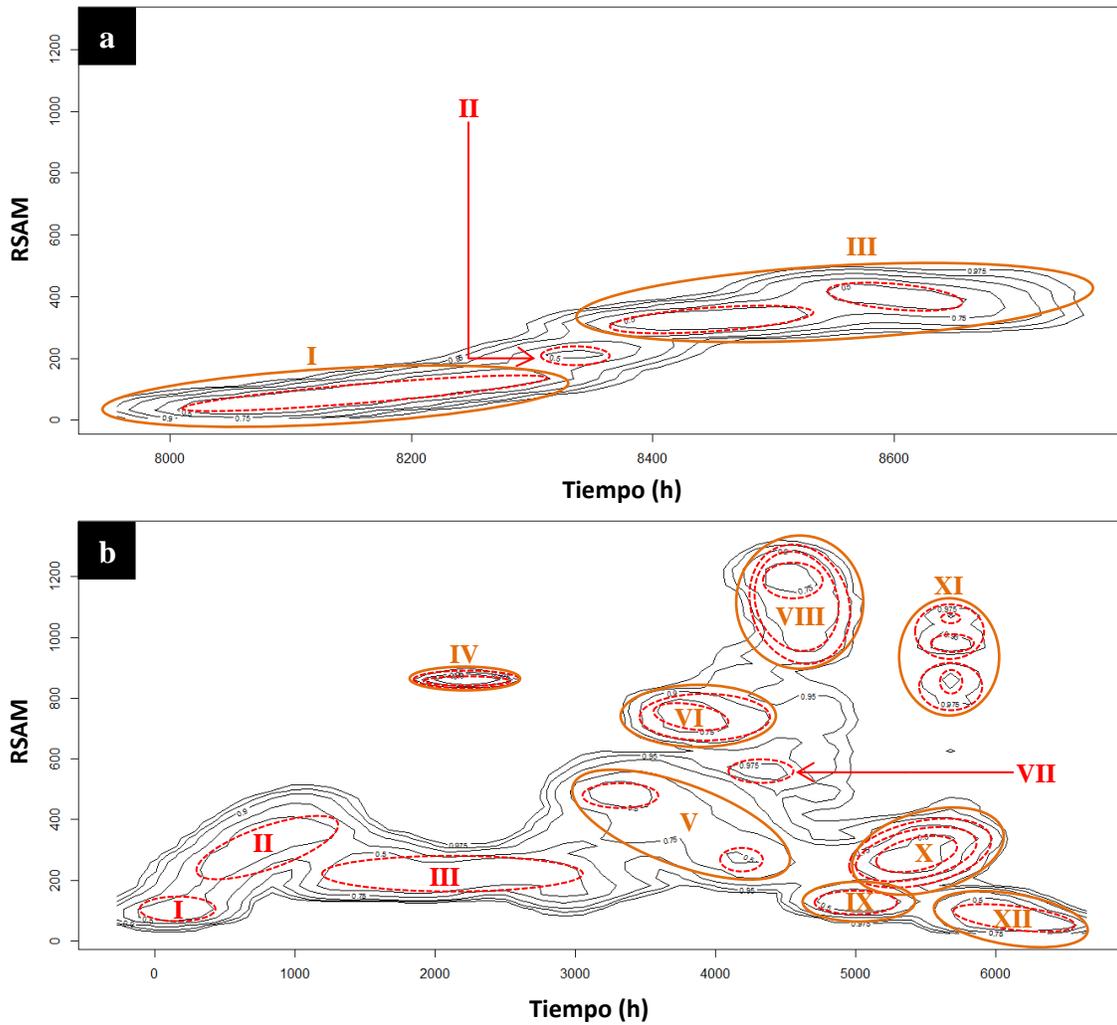
Tabla 21. Escenario de Kernels I VSM

Fig.	N	Tamaño relativo del Kernel	Límite inferior (h)	Límite superior (h)	Nivel de confianza (%)	Media RSAM	Días de anticipo
a	I	Grande (5100h-212.5d)	0	5100	50%	95.21	149.96
	<b>II</b>	<b>Pequeño (500h-20.83d)</b>	<b>5000</b>	<b>5500</b>	<b>97.5%</b>	<b>128.29</b>	<b>133.29</b>
	III	Pequeño (500h-20.83d)	5500	6000	50%	81.23	112.46
	<b>IV</b>	<b>Pequeño (300h-12.50d)</b>	<b>6000</b>	<b>6300</b>	<b>97.5%</b>	<b>232.03</b>	<b>99.96</b>
	V	Grande (1600h-66.66d)	6500	8100	50%	79.65	24.96
	<b>VI</b>	<b>Pequeño (500h-20.83d)</b>	<b>7100</b>	<b>7600</b>	<b>90%</b>	<b>113.12</b>	<b>45.79</b>
	<b>VII</b>	<b>Grande (1000h-41.66d)</b>	<b>7500</b>	<b>8500</b>	<b>97.5%</b>	<b>101.64</b>	<b>8.29</b>
	<b>VIII</b>	<b>Grande (500h-20.83d)</b>	<b>8000</b>	<b>8500</b>	<b>90%</b>	<b>164.53</b>	<b>8.29</b>
b	I	Mediano (400h-16.67d)	5100	5500	97.5%	135.37	113.29
	II	Pequeño (250h-10.42d)	5200	5450	95%	134.77	135.38
	III	Grande (900h-37.50d)	5200	6100	50%	107.97	108.29
	<b>IV</b>	<b>Grande (100h-4.17d)</b>	<b>6100</b>	<b>6200</b>	<b>95%</b>	<b>291.51</b>	<b>104.13</b>
	V	Pequeño (200h-8.33d)	6200	6400	50%	219.07	95.79
	VI	Mediano (500h-20.83d)	6500	7000	50%	92.30	70.79
	<b>VII</b>	<b>Grande (800h-33.33d)</b>	<b>7000</b>	<b>7800</b>	<b>75%</b>	<b>86.09</b>	<b>37.46</b>
	<b>VIII</b>	<b>Pequeño (300h-12.5d)</b>	<b>7200</b>	<b>7500</b>	<b>97.5%</b>	<b>82.06</b>	<b>49.96</b>
	IX	Grande (700h-29.17d)	8000	8700	95%	232.20	0
	X	Grande (800h-33.33d)	8000	8800	90%	232.20	0*
c	I	Grande (800h-33.33d)	7300	8100	97.5%	65.83	24.96
	<b>II</b>	<b>Pequeño (50h2-0.08d)</b>	<b>8250</b>	<b>8300</b>	<b>50%</b>	<b>137.24</b>	<b>16.63</b>
	<b>III</b>	<b>Mediano (300h-12.50d)</b>	<b>8200</b>	<b>8500</b>	<b>75%</b>	<b>231.61</b>	<b>8.29</b>
	IV	Grande (500h-20.83d)	8300	8800	90%	343.18	0*

\* Extrapolación de Kernels respecto a los datos originales.

Es perceptible el comportamiento sísmico originado en las horas de la tarde del veintiuno de diciembre del dos mil trece, a diez días de la erupción del veintinueve de diciembre del mismo año (ver ilustración 36). El cambio sísmico consiste en que el RSAM estaba por debajo de las cuatrocientas unidades, luego se mantiene o es mayor a esta medida. El esquema de Kernels II del periodo post-eruptivo muestra once núcleos muy cercanos (ver tabla 22). Las concentraciones ocurren luego de la erupción volcánica, indicando que se trata de un volcán sísmicamente inestable. Hay seis núcleos que superan las seiscientas unidades RSAM debido al desarrollo de posteriores erupciones volcánicas secundarias o de menor intensidad. La secuencia gráfica del RSAM del volcán San Miguel crece de forma exponencial positiva hasta llegar a la erupción del veintinueve de diciembre del dos mil trece, después continua creciendo más intensamente. Muchos días después de la erupción volcánica principal del dos mil trece,

el RSAM decrece de forma exponencial negativa y vertiginosa. Esto refleja la evidencia de otros patrones que serán estudiados más adelante en la modelización estadística.



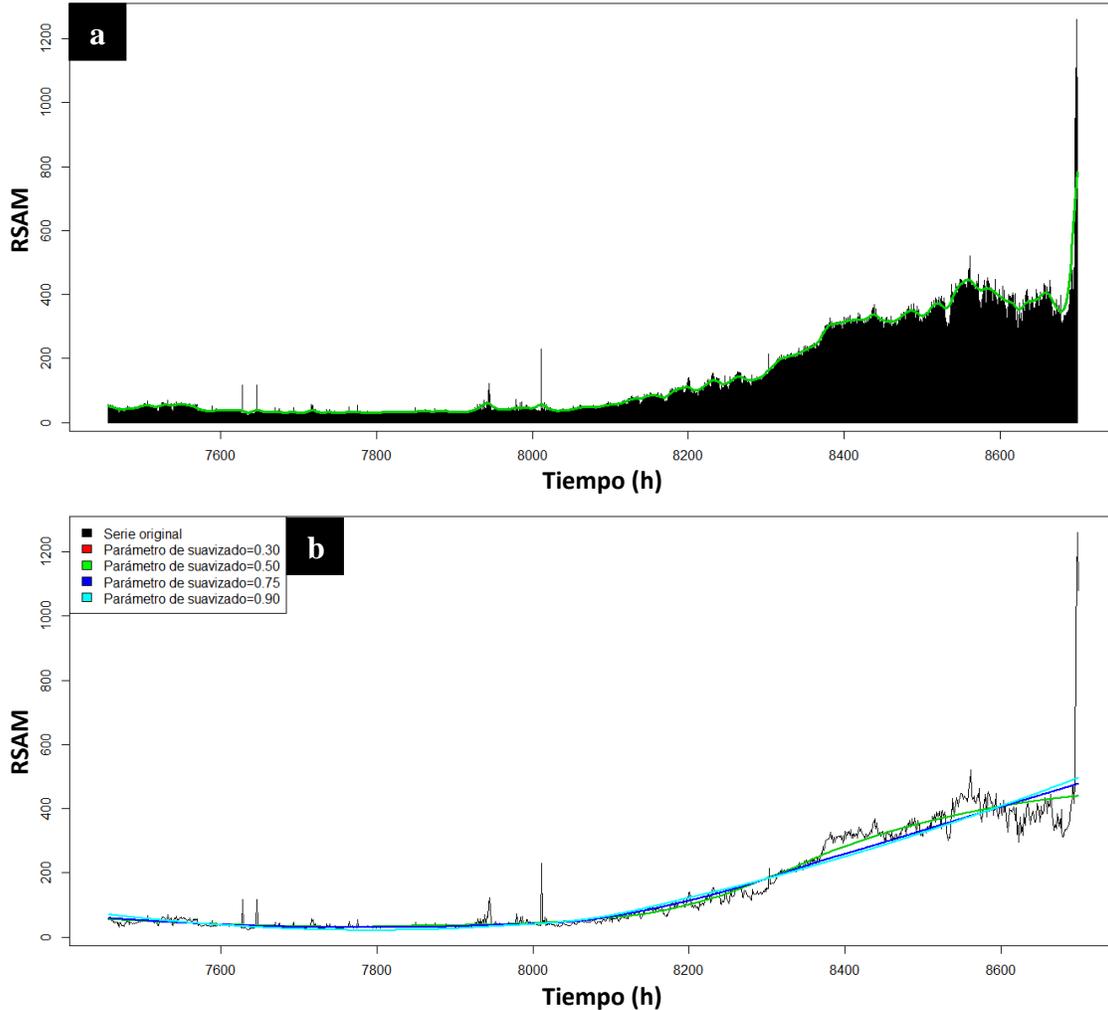
*Ilustración 36.* Kernels RSAM, diciembre 2013 y periodo post-eruptivo.  
 a) Kernel Binario del RSAM por hora del volcán San Miguel a un mes de la erupción principal (1 de diciembre 2013-29 de diciembre del 2013). b) Kernel Binario del RSAM por hora del volcán San Miguel en el periodo post-eruptivo (8 de febrero 2014-30 de septiembre del 2014).

Tabla 22. Escenario de Kernels II VSM

Fig.	N	Tamaño relativo del Kernel	Límite inferior (h)	Límite superior (h)	Nivel de confianza (%)	Media RSAM	Días de anticipo
a	I	Grande (350h-14.58d)	7950	8300	90%	78.94	16.63
	<b>II</b>	<b>Pequeño (50h-20.08d)</b>	<b>8300</b>	<b>8350</b>	<b>50%</b>	<b>200.68</b>	<b>14.54</b>
	III	Grande (550h-22.92d)	8350	8900	97.5%	363.68	0*
b	<b>I</b>	<b>Grande (225h-9.38d)</b>	<b>8500</b>	<b>8725</b>	<b>97.5%</b>	<b>403.06</b>	<b>0*</b>
	II	Mediano(50h-2.08d)	8675	8725	97.5%	496.96	0*
c	I	Pequeño (500h-20.83d)	0	500	50%	86.07	341.63**
	II	Mediano (800h-33.33d)	500	1300	50%	86.80	308.29**
	III	Grande (1900h-79.17d)	1100	3000	50%	98.42	237.46**
	<b>IV</b>	<b>Mediano (800h-33.33d)</b>	<b>1700</b>	<b>2500</b>	<b>97.5%</b>	<b>96.40</b>	<b>258.29**</b>
	V	Grande (1500h-62.50d)	3000	4500	75%	93.23	174.96**
	<b>VI</b>	<b>Mediano (1000h-41.67d)</b>	<b>3500</b>	<b>4500</b>	<b>90%</b>	<b>98.73</b>	<b>174.96**</b>
	VII	Pequeño (400h-16.67d)	4100	4500	97.5%	112.14	174.96**
	VIII	Grande (600h-25.00d)	4500	5100	90%	105.97	149.96**
	IX	Mediano (800h-33.33d)	4700	5500	50%	120.22	133.29**
	X	Grande (1000h-41.67d)	5000	6000	75%	104.80	112.46**
	<b>XI</b>	<b>Grande (500h-20.83d)</b>	<b>5500</b>	<b>6000</b>	<b>50%</b>	<b>81.23</b>	<b>112.46**</b>
	XII	Grande (800h-33.33d)	5700	6500	75%	157.56	91.63**

\* Extrapolación de Kernels respecto a los datos originales. \*\* Observaciones del periodo post-eruptivo, son los días distantes después de la erupción principal del 2013.

El promedio del RSAM por hora es aproximadamente tres veces mayor que el del volcán Santa Ana, esto se interpreta como una intensa actividad sísmica del volcán San Miguel. El RSAM por hora varía cuatro veces más que el volcán Santa Ana, el volcán San Miguel es al menos el triple de inestable que el de Santa Ana. La regresión Loess refleja el comportamiento lineal positivo que toma la variable RSAM por hora antes de la erupción volcánica principal del dos mil trece (ver ilustración 37). La tendencia sufre una leve curvatura al principio para luego adoptar la misma forma que momentos previos a la erupción, esa curvatura es decreciente (ver ilustración 37).



*Ilustración 37. Regresión Kernel y regresión Loess, RSAM, nov. 2013, VSM.*

a) Regresión Kernel con ancho de banda de 14.66 del RSAM por hora del volcán San Miguel, noviembre 2013-diciembre 2013. b) Regresión Loess del RSAM por hora, del volcán San Miguel noviembre 2013-diciembre 2013.

Los resultados anteriores demuestran que toda erupción volcánica es de carácter parabólica negativa, y los intervalos de tiempo antes y después constituyen la forma lineal creciente y decreciente respectivamente de las tendencias alrededor del evento eruptivo. Lo mismo ocurrió con el comportamiento del RSAM del volcán Santa Ana. Otra variante es la presencia de una forma de campana de Gauss, con parábola al centro y colas suaves y decrecientes. Las regresiones no paramétricas aplicadas justifican el comportamiento. Con el registro sísmico antes, durante y después de la erupción; el comportamiento que toma el RSAM es el de una parábola negativa, no se descarta que

sea una parábola rotada.

En el caso alternativo, el historial sísmico previo a la erupción, la supresión de los momentos donde se desarrolla plenamente la erupción, y la unión del momento pre-eruptivo con el post-eruptivo, muestran comportamientos similares a los descritos anteriormente. Lógicamente, el máximo de la campana de Gauss o de la parábola corresponde al valor RSAM de la erupción volcánica; no se descarta que la erupción ocurra en la parte creciente o decreciente de la forma geométrica. Antes de la erupción del veintinueve de diciembre del dos mil trece, los datos se ajustan a una línea recta con pendiente positiva (ver ilustración 38).

El comportamiento del RSAM en el periodo post-eruptivo es modelado a través de una línea de regresión o de una función exponencial positiva con suave crecimiento. La regresión Kernel confirma la tendencia lineal con pendiente positiva de los datos horas después de la erupción volcánica de la variable RSAM por hora. Primero decrece en forma exponencial, la mayor parte de los datos constituyen una línea recta; la curva decrece hasta valores normales de RSAM por hora, alrededor de treinta y cincuenta unidades para luego ascender linealmente alcanzando valores de hasta trescientas unidades, luego de la erupción principal. La interpretación se fundamenta en la inestabilidad y gran potencia sísmica debidas a la ocurrencia de erupciones secundarias.

Esta regresión también muestra que la cota superior de esa forma lineal, incluye valores de hasta mil unidades. En resumen, la regresión Loess del RSAM por hora, advierte que la forma del periodo pre-eruptivo al principio es lineal con pendiente positiva y suave; y entre más cercana a la erupción principal está, cambia de forma lineal a exponencial (ver ilustración 38). La tendencia del RSAM por hora en el periodo post-eruptivo es idéntica a la del periodo pre-eruptivo, primero comienza con un comportamiento lineal positivo y luego finaliza con un crecimiento exponencial notorio.

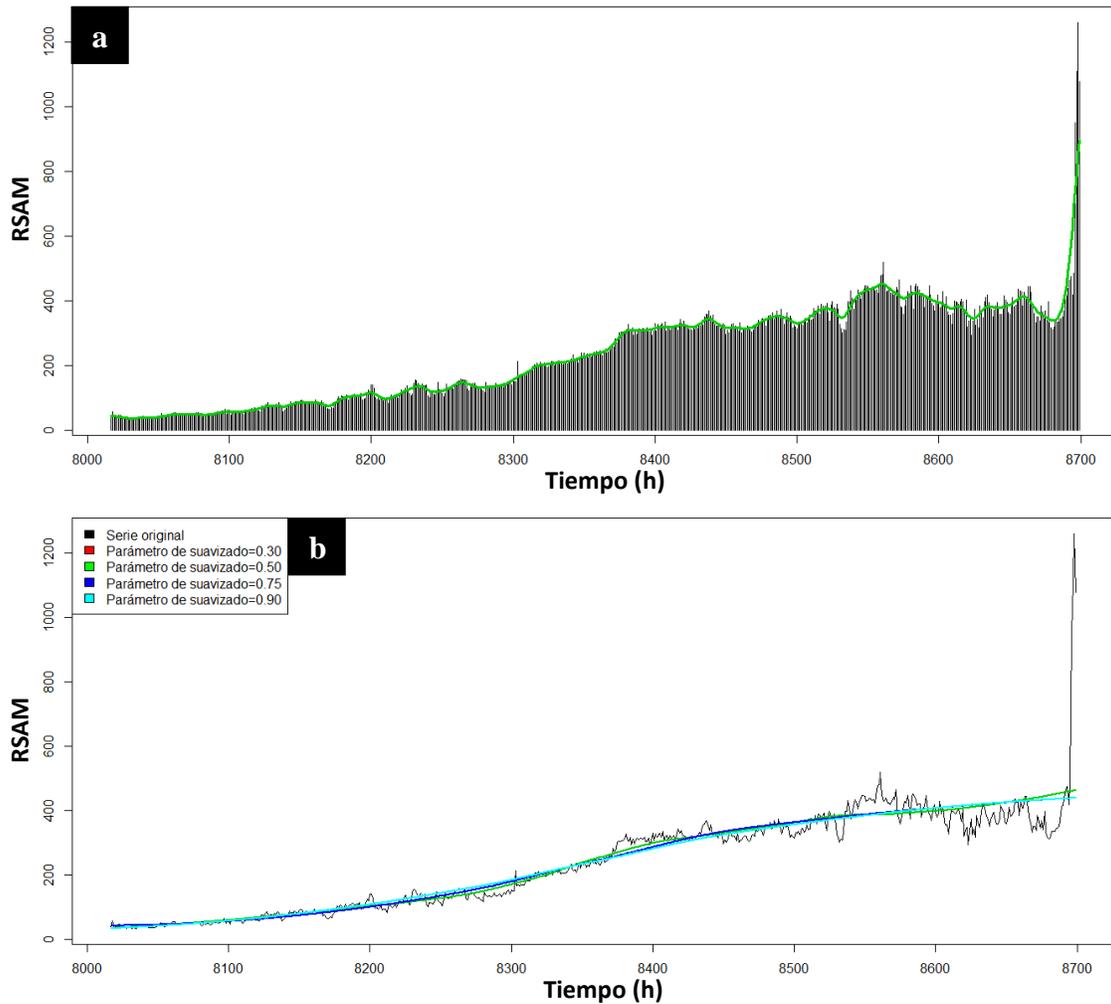


Ilustración 38. Regresión Kernel y regresión Loess, RSAM, dic. 2013, VSM.

a) Regresión Kernel con ancho de banda de 10.30 del RSAM por hora, VSM, diciembre 2013. b) Regresión Loess del RSAM por hora, VSM, diciembre 2013.

El análisis de tendencia para el mes de noviembre del año dos mil trece a través del método de las Medias Móviles, distingue los ajustes de los datos definiendo la pendiente de la recta de tendencia obtenida a través de regresión paramétrica simple (ver ilustración 39). La tendencia tiene una pendiente alrededor de una diezmilésima. En el Suavizado Exponencial, la pendiente de la recta de tendencia es menor que una diezmilésima (ver ilustración 39). Aunque las cifras tiendan a cero no deben menospreciarse; en probabilidad de predicción y pronóstico, forma ideal y real del objetivo, son cantidades importantes que contienen información requerida por los

especialistas. La tendencia para el mes de diciembre del año dos mil trece tiene pendiente de aproximadamente dos diezmilésimas, el doble que para el mes de noviembre; este resultado es válido tanto para Medias Móviles, como para el Suavizado Exponencial.

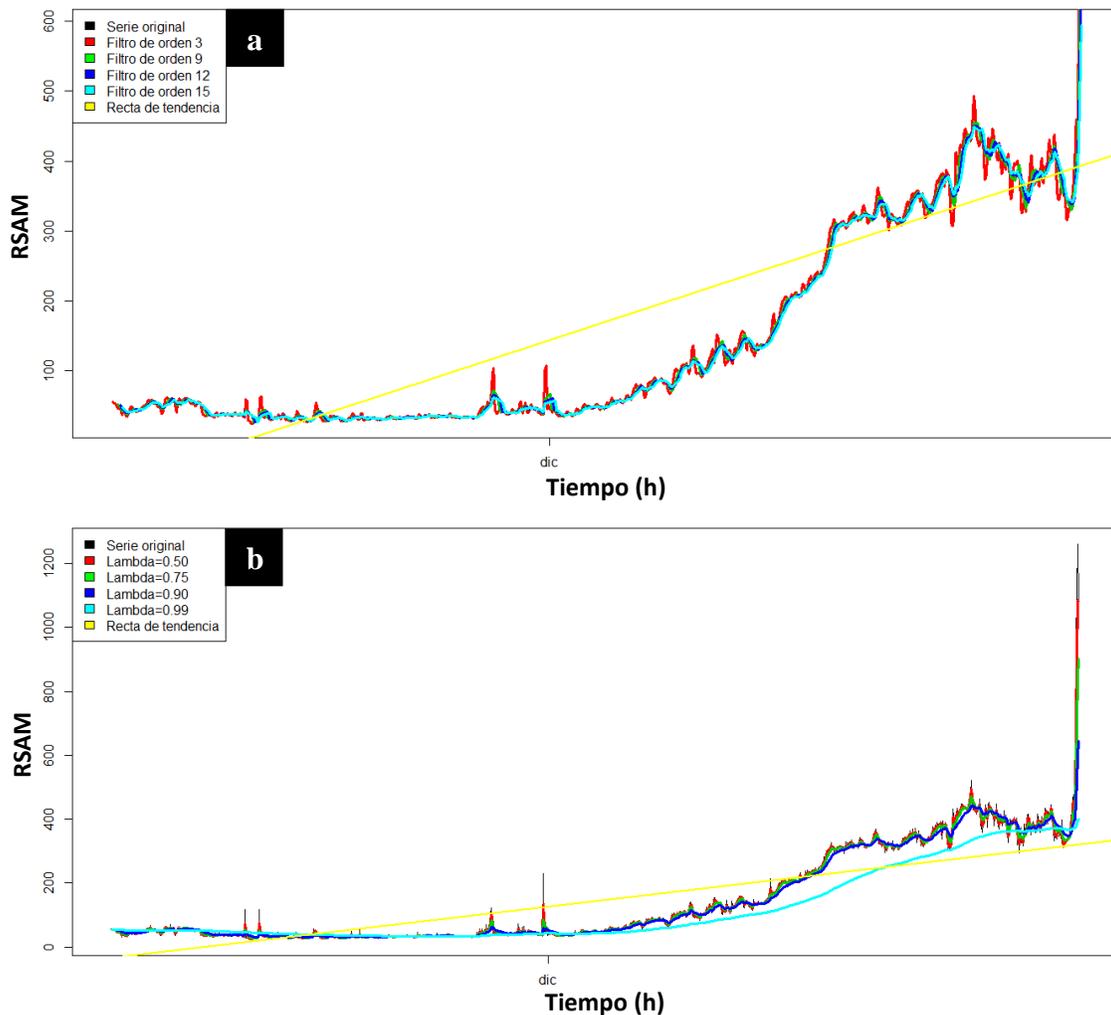
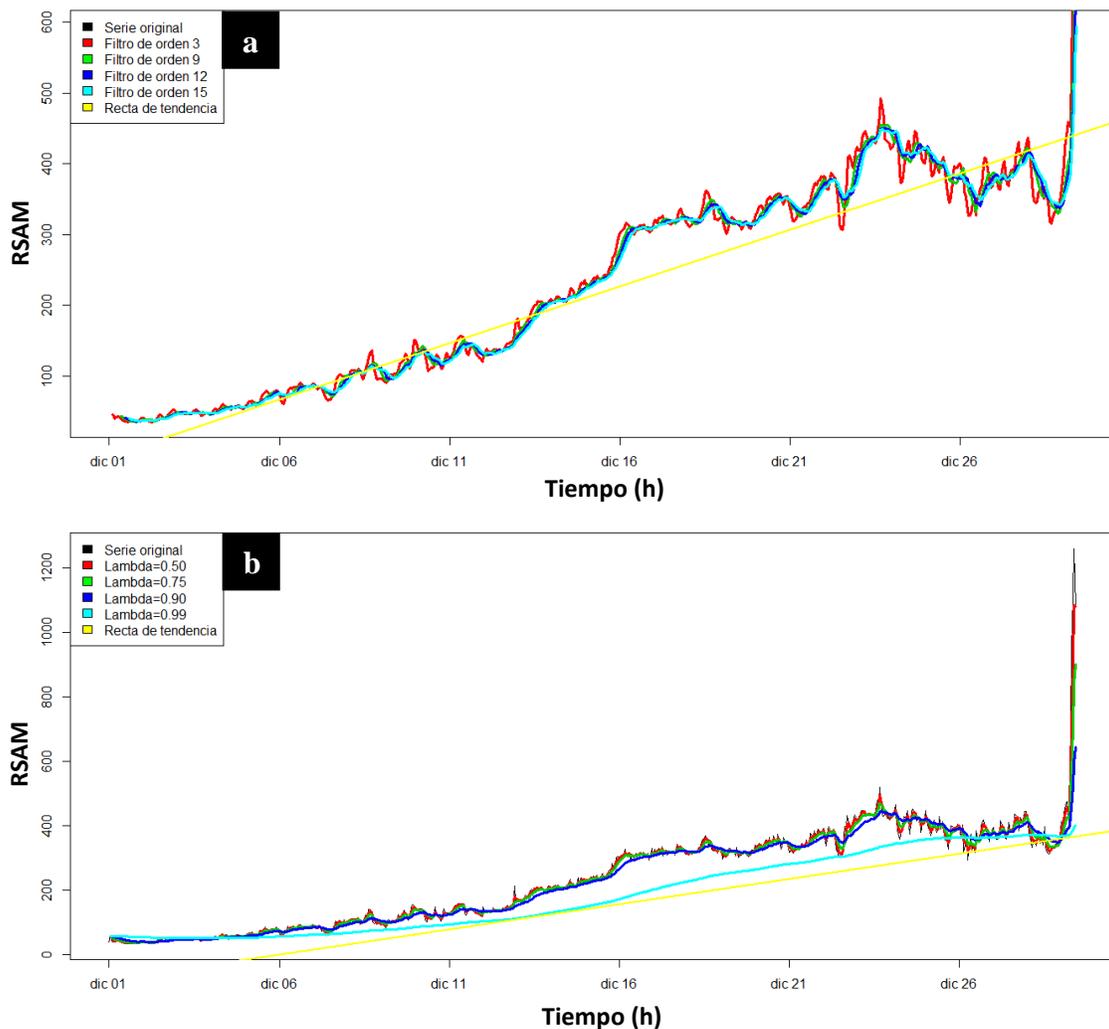


Ilustración 39. Medias Móviles y Suavizado Exponencial, RSAM, nov. 2013, VSM.

a) Desestabilización a través del filtro de las medias móviles del RSAM por hora, del volcán San Miguel noviembre 2013-diciembre 2013; con línea de tendencia  $RSAM = -1.408 \times 10^5 + 1.017 \times 10^{-4}Tiempo$ . b) Desestabilización a través del suavizado exponencial del RSAM por hora, del volcán San Miguel noviembre 2013-diciembre 2013; con línea de tendencia  $RSAM = -1.094 \times 10^5 + 7.903 \times 10^{-5}Tiempo$ .

El método estadístico que combina polinomios y funciones trigonométricas se denomina curvas de Andrews trasladadas a coordenadas polares se convierten en gotas de Fourier. Las gotas de Fourier están orientadas a la solución del problema de la

dimensión gráfica de los datos. Esta es una manera de analizar el comportamiento de las variables sísmico-volcánicas, observar cuando ocurren cambios significativos (fenómenos sísmicos o volcánicos moderados) en sus valores ya que describen un patrón único. El comportamiento se convierte en uniforme al dejar la anomalía cuando se presente cíclicamente cambios significativos en la amplitud sísmica, aunque también son aplicadas a otras variables recopiladas.

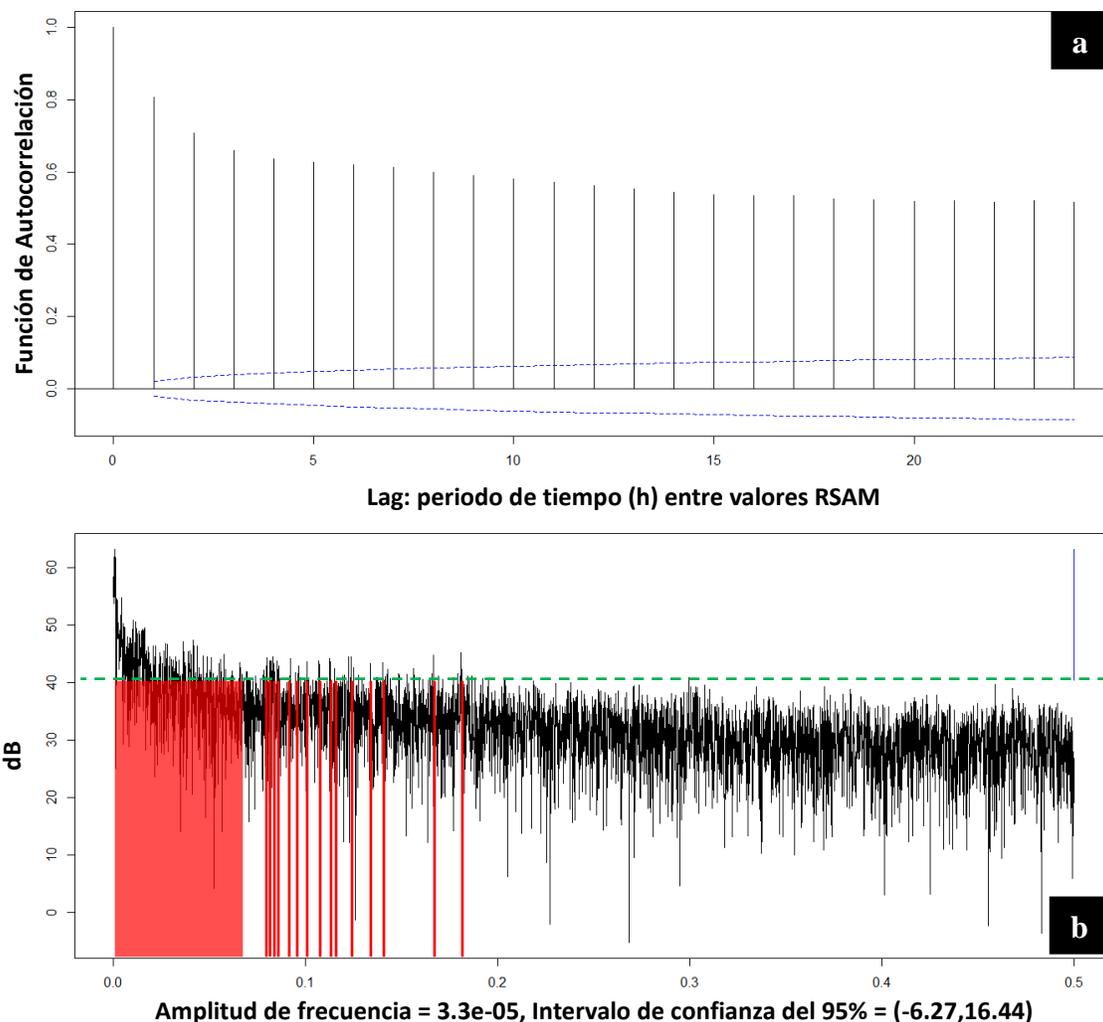


*Ilustración 40. Medias Móviles y Suavizado Exponencial, RSAM, dic. 2013, VSM.*

a) Filtro de medias móviles para el RSAM por hora, VSM, diciembre 2013; con línea de tendencia  $RSAM = -2.564 \times 10^5 + 1.85 \times 10^{-4}Tiempo$ . b) Suavizado exponencial del RSAM por hora, VSM, diciembre 2013; con línea de tendencia  $RSAM = -2.512 \times 10^5 + 1.812 \times 10^{-4}Tiempo$ .

Otra manera gráfica de contemplar la dimensión o espectro de los datos es a través del gráfico de coordenadas paralelas, se verifica la forma que toman los datos en un

intervalo de tiempo determinado, si la forma no cambia entonces no hay anomalías en los valores de interés geoestadístico. Es fácil graficar en dos dimensiones, incluso en tres, pero qué ocurre con seis variables o más. El tema dimensión no tiene sentido sin estas herramientas. Esta es la ventaja de la implementación de las coordenadas paralelas y gotas de Fourier, no importa el número de variables obtenidas de los diferentes fenómenos naturales, siempre hay una dimensión gráfica. A menudo se acostumbra mezclar las coordenadas paralelas de cada variable y verificar si existen descubrimientos, a lo largo del tiempo. Sólo se experimentó con las gotas de Fourier.



*Ilustración 41.* ACF y espectro en dB, RSAM, 2013, VSM.  
a) ACF con un nivel de confianza del 95% para el RSAM por hora del volcán San Miguel, año 2013. b) Análisis espectral RSAM por hora con un nivel de confianza del 95%. Notas: “dB” significa decibeles. Periodograma del RSAM por hora bruto.

El primer análisis corresponde a las primeras horas del veintinueve de diciembre del dos mil trece, año de la erupción volcánica (ver ilustración 42). Las gotas de Fourier tienen cuatro partes en forma de pétalos correspondientes a las primeras horas de días específicos del comportamiento del RSAM, varían lo suficiente para prever una erupción volcánica. El eje mayor del pétalo mayor oscila entre los tres mil quinientos puntos lineales, la observación que corresponde a la erupción volcánica es la mayor. Un mes previo a la erupción el eje mayor del pétalo mayor oscila las ciento sesenta unidades lineales (ver ilustración 42). Hacia un mes todavía no es perceptible el ascenso de la actividad volcánica aunque esté elevada.

Hacia un año de la erupción volcánica, el eje mayor del pétalo mayor oscila entre las ciento ochenta unidades lineales, hay un comportamiento que inicia de forma aritmética y finaliza de forma geométrica (ver ilustración 42). Las características en cada intervalo de tiempo son muy claras e indican un crecimiento a veces aritmético en su mayoría; y cada vez que se acerca al instante de la erupción, es notable un crecimiento geométrico. El crecimiento de los pétalos que forman las gotas de Fourier está estrechamente relacionado con la propiedad de cada término de la función de Andrews que genera las curvas de Andrews. Los cambios en los valores de los términos producen pétalos con nuevas dimensiones.

Cada crecimiento o decrecimiento depende a cambios en los valores de los términos de la función. Generalmente, todos los pétalos crecen o decrecen, en algunos casos el comportamiento de los pétalos es mixto; cuando algunos pétalos crecen, otros decrecen. La función trasladada a coordenadas polares producen gotas recursivas, es decir, cada vez que la función se aplica a un nuevo valor RSAM, el resto tiene movimiento decreciente, los valores antiguos desaparecen y son desplazados por los valores más recientes. La escala de los ejes en el sistema de coordenadas polares está relacionada a los efectos que generan los valores RSAM en cada término de la función de Andrews por expresiones cosenoidales y senoidales. La escala se convierte en un sistema de referencia para observar el comportamiento ascendente, decreciente o constante del RSAM, o la presencia de fenómenos originados por el volcán.

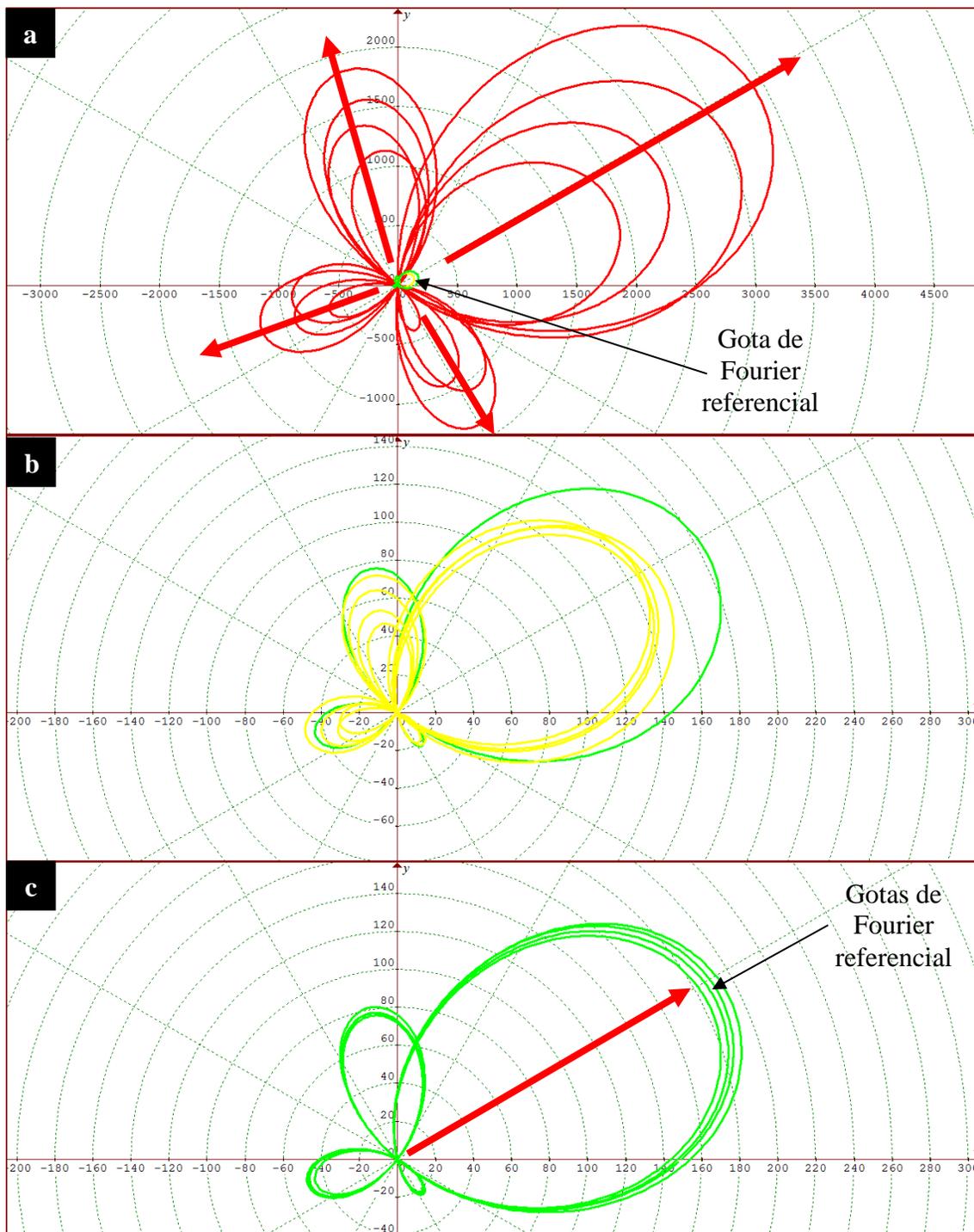


Ilustración 42. Gotas de Fourier, RSAM, 2013, VSM.

a) Gotas de Fourier para el RSAM por hora de las primeras horas del 29 de diciembre del 2013. b) Gotas de Fourier para el RSAM por hora de las primeras horas del 1 de enero del 2013. Nota: las flechas de las gotas detallan el movimiento creciente o decreciente. Las gotas rojas indican especial cuidado a la actividad volcánica; el amarillo, precaución; el verde, normalidad.

En la mayoría de casos, el número de pétalos está vinculado al número de términos que contienen senos y cosenos. Los pétalos grandes advierten presencia de fenómenos en el comportamiento sísmico del volcán, pétalos pequeños significan normalidad. Con la simulación Montecarlo se obtuvo una muestra de diez mil observaciones con media aproximadamente de ciento once y desviación estándar de aproximadamente noventa y dos (ver ilustración 43). Al graficar las observaciones correlativas, resulta una sola concentración de comportamiento normal que ondula entre la media y la desviación estándar. Al contrastar la simulación con los datos reales del RSAM por hora del mes de agosto del dos mil cinco del volcán Santa Ana, resulta una elipse de mayor magnitud, el nivel más interno pertenece al cincuenta por ciento; el intermedio, al setenta y cinco por ciento; y los más externos, al noventa y cinco por ciento y noventa siete y medio por ciento, así se distribuyen las observaciones.

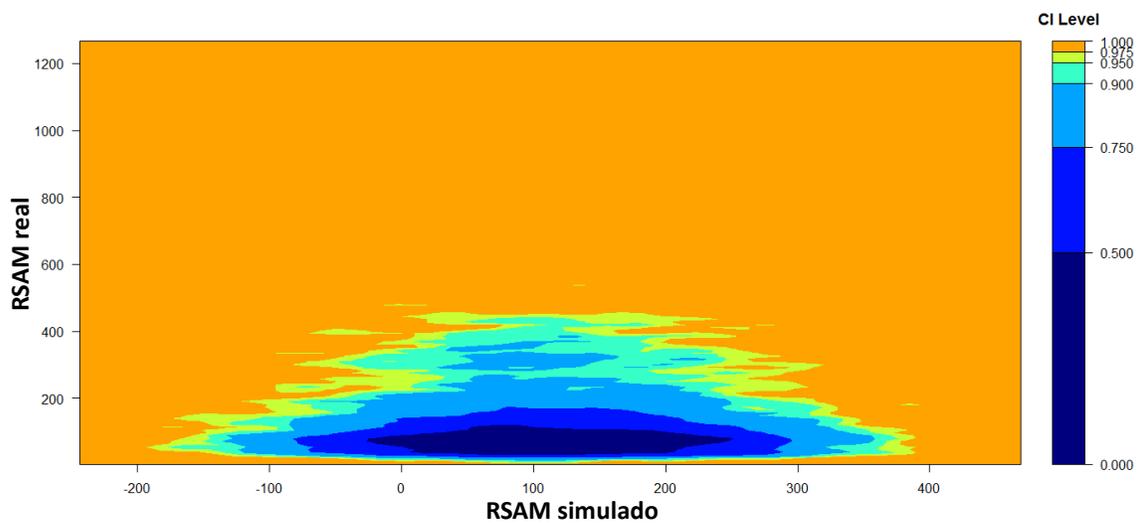
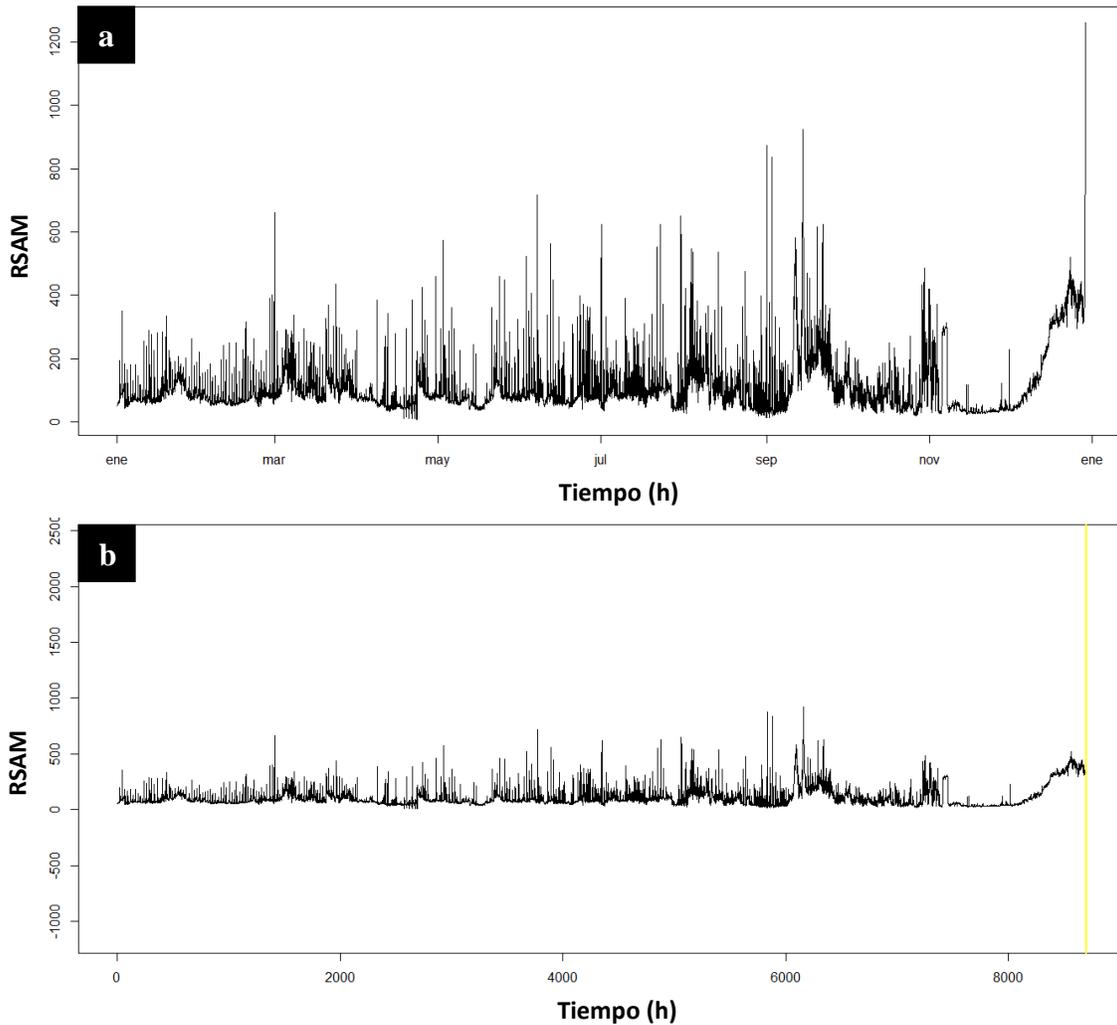


Ilustración 43. Simulación Monte Carlo, 8699 observaciones, VSM.

Kernels  $RSAM\ simulado \times RSAM\ real$  de una muestra con 8760 elementos con características normales,  $\mu = 111.4344$  y  $\sigma = 92.1817$ ; valores reales de la base de datos completa del RSAM por hora del volcán San Miguel. Nota: los valores negativos pertenecen sólo a aspectos de forma extrapolada por el método.



*Ilustración 44.* Secuencia gráfica RSAM 2013, diferentes escalas, VSM.

a) Secuencia gráfica del RSAM por hora del volcán San Miguel 2013. b) secuencia gráfica del RSAM por hora del volcán San Miguel 2013 con aumento en la escala de las ordenadas, aspecto 1. El momento de la erupción se detalla por medio de una línea vertical amarilla.

## Capítulo 6: Discusión de resultados

El índice precursor de una posible erupción volcánica, IPPEV, definido como la razón de la cantidad de sulfato respecto a la cantidad de cloro, es una aproximación al afán del pronóstico de erupciones volcánicas (ver tabla 23). Estas cantidades se obtienen de las mediciones químicas en la laguna del volcán Santa Ana, aunque generalmente puede experimentarse en otras regiones del mundo. Las regresiones locales son buenas en el ajuste de los datos, curvas de orden superior a cuatro sugieren que el modelo no es convincente, ya que es poco parsimonioso y sobreajusta las observaciones. El sentido de las concavidades es manejado por el signo del coeficiente del máximo orden relativo. El mejor ajuste es el del RSAM por hora del volcán Santa Ana, evidentemente se trata de una parábola negativa de gran amplitud.

Tabla 23. Estadística aplicada a los complejos volcánicos ES.

Modelo	Aplicación	Tipo	Interpretación
IPPEV	Laguna VSA, y posiblemente colosos con características parecidas.	Geométrico	Los cambios de cantidades de sulfato y cloro en la actividad volcánica son determinantes en la predicción.
Densidad con Kernel Binario	Todas las bases con información volcánica, puede ser aplicado también a otras áreas.	Geométrico	Importantes núcleos de datos se perciben con esta gran técnica estadística.
Curvas de Andrews	Datos multivariantes y univariantes.	Geométrico	Representación y espectro dimensional de las variables.
Médias Móviles, Suavizado Exponencial, Regresión Kernel y Regresión Loess	Filtros para el análisis de la tendencia de la serie de datos.	Geométrico	Estos filtros tienen la función de suavizar la serie, al igual que la regresión Kernel y Loess.

El modelo exponencial es determinante en el pronóstico de erupciones volcánicas. El modelo exponencial muestra el comportamiento natural y normal de muchos fenómenos. Retomando la modelización con Kernel Binarios, es útil la idea de aplicar un análisis de sensibilidad a las diferentes características del comportamiento sísmico. El índice que recoge toda la información sobre los núcleos de datos en una serie temporal, juntamente con la aplicación de la investigación operativa; producirían buenos resultados en la

inferencia estadística (ver tabla 24). Actualmente, la institución encargada de vigilar la actividad volcánica y sísmica del territorio salvadoreño es el Ministerio de Medio Ambiente y Recursos Naturales. A las diez horas y media del día seis de julio del dos mil catorce se programó una entrevista con el vulcanólogo de la institución. Se exploraron los conceptos, procedimientos y avances científicos en materia de vulcanología y sismología.

**Tabla 24.** *Índice de la modelización con Kernel Binario.*

$K_b = \text{NEZITLA}$	
N: número de núcleos	E: excentricidad
Z: magnitud del eje mayor	I: magnitud del eje menor
T: distancia euclidiana del origen al centro	L: longitud
A: área	

Los modelos encontrados al aplicar las técnicas estadísticas seleccionadas para esta investigación son de diferente naturaleza. Entre modelos polinomiales y expresiones algebraicas, se hace un pronóstico del comportamiento del RSAM por hora en el volcán Santa Ana y también en el volcán San Miguel. El primer modelo consiste en la definición de una parábola. Las restricciones pertinentes se encuentran para cada modelo estadístico. Todos los modelos son funciones, a excepción de las elipses, que son interpretaciones de núcleos de datos en diferentes niveles o intervalos de confianza. Estos niveles son el cincuenta por ciento, el setenta y cinco por ciento, el noventa por ciento, el noventa y cinco por ciento y el noventa y siete con cinco décimas, todos son ajustables.

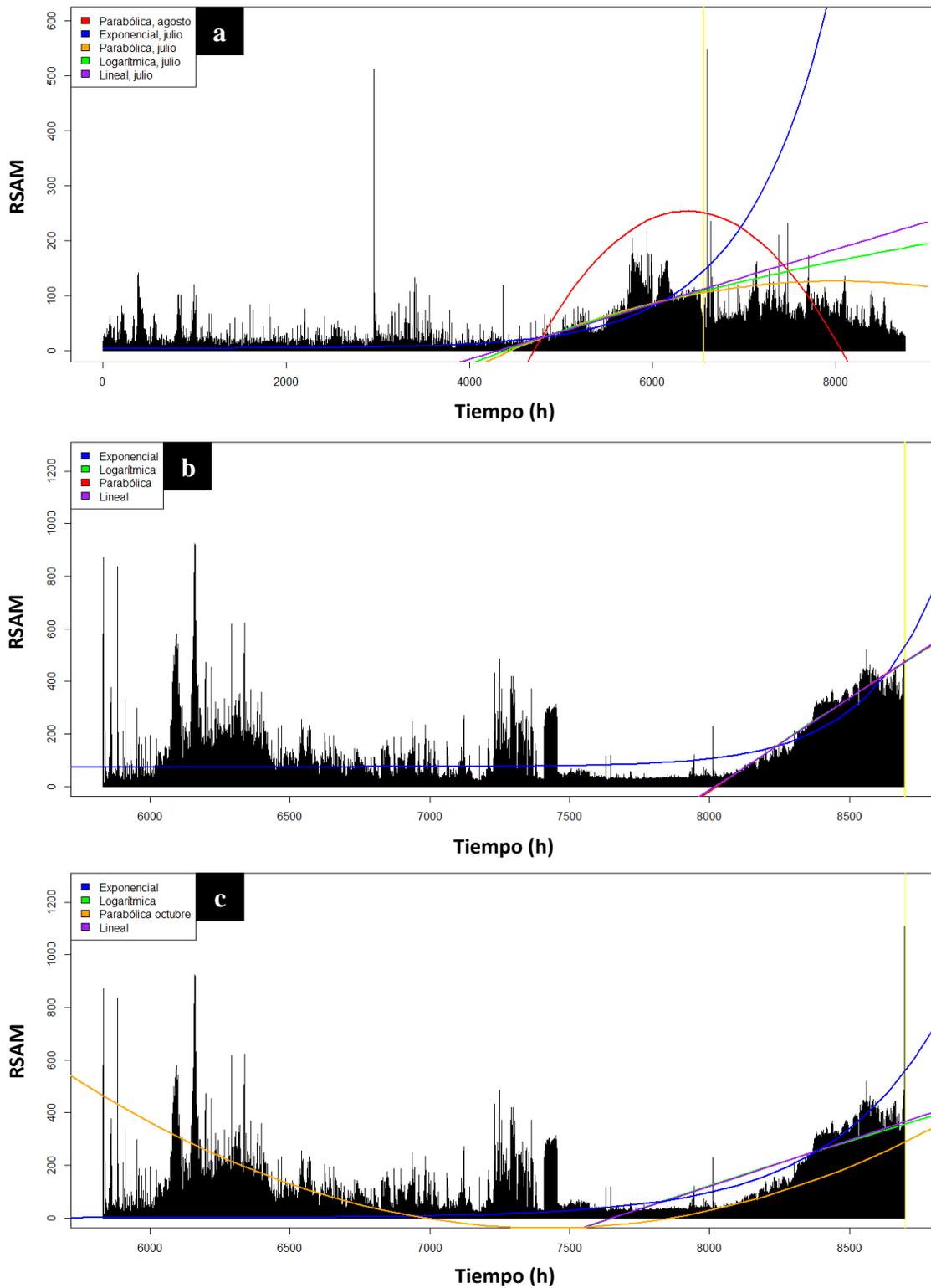
La aplicación de las elipses, ya sea en coordenadas rectangulares o en coordenadas polares, es indispensable cuando se considera la mejor representación. Las elipses, sin importar las formas que tengan, incluyendo elipses degeneradas y triviales; están presentes en prácticamente todo. Un caso particular en la medicina es el pronósticos de fecha probable de parto que consiste en realizar mediciones cuantitativas continuas de partes fundamentales del cuerpo del feto, sean estos valores la medición del cráneo, del abdomen, de la columna vertebral, del fémur, entre otros. La experiencia y las mediciones estiman la fecha en que nacerá el nuevo ser. El índice que recoge la

información de la elipse en estudio es un parámetro para medir la efectividad de pronósticos utilizando los análisis de sensibilidad de la investigación operativa.

El comportamiento parabólico de agosto cruzado con el exponencial de julio, ambos del periodo pre-eruptivo, genera dos intersecciones importantes (ver tabla 25). El primer intercepto es un cambio en el comportamiento del RSAM del volcán Santa Ana del diecinueve de julio del dos mil cinco, con una pendiente cerca de tres décimas para el comportamiento parabólico y un crecimiento ascendente con inclinación cerca de dieciséis grados. El segundo intercepto pertenece al dieciocho de octubre ubicado en el periodo post-eruptivo, el comportamiento parabólico tiene una pendiente negativa que indica que los cambios que se generan son decrecientes, la pendiente es negativa con ciento setenta y cuatro grados de inclinación.

**Tabla 25. Modelos estadísticos de pronóstico VSA**

<b>f(x)</b>	<b>Valores críticos (t,RSAM)</b>	<b>Interceptos (t,RSAM)</b>	<b>Primera derivada (pendiente)</b>	<b>Inclinación (°)</b>	
Parabólica (agosto)	Máximo (6385.5556, 254.5888)	(4786.9665, 24.5950) Exponencial (julio)	0.2877 Parabólica (agosto)	16.0505°	
			0.0221 Exponencial (julio)	1.2660°	
	Ceros (4703.6625, 0) (8067.4486, 0)	(6961.8662, 224.6967) Exponencial (julio)	-0.1037 Parabólica (agosto)	174.0796°	
			0.2423 Exponencial (julio)	13.6203°	
	Exponencial (julio)	-	(6561.0000, 251.8185) (Erupción)	0.2877 Parabólica (agosto)	16.0505°
			(4797.4052, 24.8275) Lineal (julio)	0.0224 Exponencial (julio)	1.2832°
(6013.5105, 90.1327) Lineal (julio)			0.0500 Lineal (julio)	2.8624°	
(6561.0000, 113.0072) (Erupción)			0.1559 Exponencial (julio)	8.8611°	



*Ilustración 45.* Representación de los modelos estadísticos VSA y VSM.  
 a) Modelos julio-agosto 2005, VSA. b) Modelos diciembre 2013, VSM. c) Modelos octubre-noviembre 2013, VSM.

La inclinación para el segundo intercepto es leve respecto al eje de las abscisas. Es importante el hecho que el segundo intercepto de los comportamientos parabólico-exponencial coincida con la erupción del volcán Santa Ana. La segunda combinación de modelos estadísticos definida por el comportamiento exponencial y la tendencia lineal, ambos del periodo pre-eruptivo del volcán Santa Ana, provee un mejor pronóstico. Hay dos interceptos que muestran cambios en el RSAM. El primer intercepto pertenece al diecinueve de julio; el segundo, al ocho de septiembre, ambos del año dos mil cinco. Esta combinación se basa en el criterio de preaviso de erupción volcánica, los interceptos están en el periodo pre-eruptivo y significan cambios considerables en la actividad sísmico-volcánica del volcán Santa Ana.

El análisis de estos modelos cobra importancia cuando el periodo de preaviso es distante del momento justo de la erupción volcánica. En la sección de recomendaciones se detalla una proyección del análisis que transforma el potencial del modelado de elipses que al fusionarlo con la combinación de los modelos estadísticos parabólico-exponencial o exponencial-lineal se obtienen resultados de interés para la vulcanología y la sismología. El modelo estadístico de cuarto grado en ocasiones coincide con el comportamiento parabólico, también con el de una catenaria, pero no guarda el criterio de parsimonia.

*Tabla 26. Modelos estadísticos encontrados para el RSAM VSA.*

Modelos estadísticos encontrados	R <sup>2</sup> al 95% de N. C.	Tipo	Restricciones 2005
$RSAM = 1.04 \times 10^{-1} e^{1.1 \times 10^{-3} \text{Tiempo}} + \varepsilon = 4.4632$	0.7661	Exponencial	$jul \leq RSAM \leq oct$
$RSAM = -1 \times 10^{-5} (\text{Tiempo})^2 + 1.602 \times 10^{-1} (\text{Tiempo}) - 5.1211 \times 10^2 + \varepsilon = -2.3940$ Polinomio de Taylor	0.5963	Parabólico	$jul \leq RSAM \leq oct$
$RSAM = -\frac{100000}{801} (\text{Tiempo})^2 + \frac{64313}{5000} - \frac{64313}{125}$			

Modelos estadísticos encontrados	R <sup>2</sup> al 95% de N. C.	Tipo	Restricciones 2005
$RSAM = 2.7115 \times 10^2 \ln Tiempo - 2.2737 \times 10^3 + \varepsilon = 0.2483$	0.5922	Logarítmica	$jul \leq RSAM \leq oct$
$RSAM = 5 \times 10^{-2}Tiempo - 2.1528 + \varepsilon = -2.1289 \times 10^2$	0.5884	Lineal	$jul \leq RSAM \leq oct$
$RSAM = -9 \times 10^{-5}(Tiempo)^2 + 1.1494(Tiempo) - 3.4004 \times 10^3 + \varepsilon = -1.479 \times 10$ Polinomio de Taylor	0.4606	Parabólico	$ago \leq RSAM \leq oct$
$RSAM = -\frac{9(Tiempo)^2}{100000} + \frac{5747(Tiempo)}{5000} - \frac{1819041}{100000}$ $(Tiempo - 6462)^2$ $(52)^2$ $+ \frac{(RSAM - \frac{4467}{50})^2}{(\frac{609}{25})^2} = 1$	-	Elipse horizontal*	Kernel I N. C.: 50% Escenario III $25 sep \leq Tiempo \leq 1 oct$
$RSAM = \frac{(\frac{97}{100})(\frac{104}{25})}{1 \pm (\frac{97}{100}) \cos(Tiempo)}$	-	Elipse horizontal polar**	Kernel I N. C.: 50% Escenario III $25 sep \leq Tiempo \leq 1 oct$

\* Los cálculos están basados en las características del Kernel I de la figura a) del escenario III del VSA. Es un Kernel de gran tamaño, posee sólo 106 horas, un promedio de 89.34 unidades RSAM, se anticipa a 1.92 días de la erupción principal del 1 de oct. del 2005. La abscisa del centro de la elipse es la media aritmética de las fronteras del intervalo [6409, 6515], mientras que la ordenada es la media aritmética de las observaciones pertenecientes al Kernel. El eje mayor es la distancia euclidiana de las fronteras del intervalo y el eje menor es la distancia euclidiana de las fronteras del intervalo  $[\bar{x}_{RSAM} - s_{RSAM}, \bar{x}_{RSAM} + s_{RSAM}] = [89.34 - 12.18, 89.34 + 12.18] = [\frac{1929}{25}, \frac{2538}{25}]$ , donde  $\bar{x}_{RSAM}$  y  $s_{RSAM}$  son estimaciones de la medida y desviación típica poblacionales. El intervalo anterior, con un nivel de confianza del 95%, contiene al 68.27% de las observaciones. La definición de la excentricidad de una elipse es  $e = \frac{c}{a} = \frac{\frac{1}{50}\sqrt{6651619}}{52} = \frac{1}{2600}\sqrt{6651619}$ , entonces  $e \approx 0.97 < 1$ . \*\*Las coordenadas del foco derecho son  $(\frac{162812}{25}, 89.34)$  y la directriz derecha es  $x = \frac{162916}{25}$ ; entonces la distancia no dirigida entre ambos es  $d = \frac{162916}{25} - \frac{162812}{25} = \frac{104}{25}$  Notas: el nivel de confianza para este Kernel es del 50%. N. C. es abreviatura de Nivel de confianza.

Los valores críticos de los modelos de pronóstico son preavisos de un evento eruptivo de envergadura (Olmos, 2016, comunicación personal). Las funciones se ajustan a la

serie de datos, las funciones tienen parámetros estimados por el método de mínimos cuadrados más una constante del error de estimación, las series de Taylor comprueban estos modelos. El coeficiente de correlación explica el grado de ajuste del modelo. En el caso del volcán de San Miguel, sólo hay una combinación idónea, es el comportamiento exponencial-lineal (ver tabla 27). En el periodo pre-eruptivo, la experimentación en noviembre y diciembre muestra que en ambos meses existen interceptos entre los modelos.

**Tabla 27. Modelos estadísticos de pronóstico VSM**

<b>f(x)</b>	<b>Valores críticos (t,RSAM)</b>	<b>Interceptos (t,RSAM)</b>	<b>Primera derivada (pendiente)</b>	<b>Inclinación (°)</b>
Exponencial (diciembre)	-	(8239.5403, 153.4019) Lineal (diciembre)	0.3011 Exponencial (diciembre)	16.7570°
			0.7006 Lineal (diciembre)	35.0151°
	-	(8635.3390, 430.6985) Lineal (diciembre)	1.3548 Exponencial (diciembre)	53.5684°
			0.7006 Lineal (diciembre)	35.0151°
		(8699.0000, 528.2739) (Erupción)	1.7256 Exponencial (diciembre)	59.9073°
	Exponencial (noviembre)	-	(7842.9922, 65.5316) Lineal (noviembre)	0.1638 Exponencial (julio)
			0.3501 Lineal (julio)	19.2952°
-		(8389.2757, 256.7854) Lineal (noviembre)	0.0224 Exponencial (julio)	1.2832°
			0.3501 Lineal (julio)	19.2952°
		(8699.0000, 556.9924) (Erupción)	1.3925 Exponencial (julio)	54.3166°

No se deben considerar aisladamente los modelos estadísticos ya que pertenecen a una serie temporal, y por lo tanto son influenciados por el conjunto de datos previos; una perspectiva dinámica actualizada de estos modelos permite una mejor descripción e identificación de los fenómenos del volcán. Al comparar los modelos de noviembre con los modelos de diciembre, los interceptos dependen del tiempo y se generan con más anticipación que diciembre, las pendientes de los interceptos para los modelos de

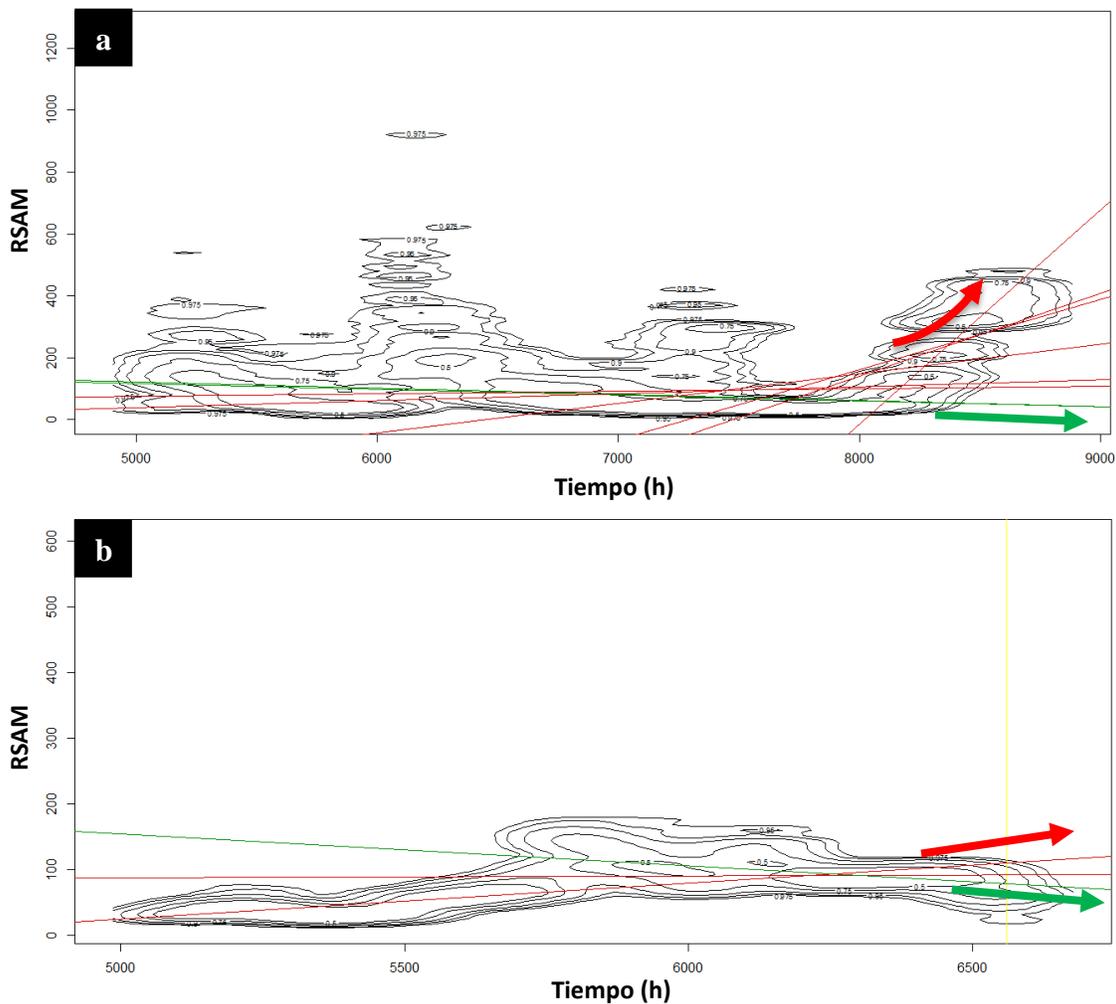
diciembre son mayores que los de noviembre, el ascenso de las unidades RSAM respecto al tiempo asciende el triple. El RSAM del volcán San Miguel también puede ser modelado de forma exponencial y parabólica como en el caso del volcán Santa Ana.

**Tabla 28.** Modelos estadísticos encontrados para el RSAM VSM.

Modelos estadísticos encontrados	R <sup>2</sup> al 95% de N. C.	Tipo	Restricciones 2013
$RSAM = 2 \times 10^{-12} e^{3.8 \times 10^{-3} \text{Tiempo}}$ $+(\varepsilon = 7.417 \times 10^{-1})$	0.8974	Exponencial	Diciembre
$RSAM = 5.8536$ $\times 10^3 \text{Ln}(\text{Tiempo}) - 5.2626 \times 10^4$ $+\varepsilon = 3.0803 \times 10^{-1}$	0.8261	Logarítmico	Diciembre
$RSAM = 7.006 \times 10^{-1} \text{Tiempo}$ $-5.6195 \times 10^3$ $+(\varepsilon = 2.8 \times 10^{-1})$	0.8261	Lineal*	Diciembre
$RSAM = -3 \times 10^{-5} (\text{Tiempo})^2$ $+1.2078(\text{Tiempo})$ $-7.7379 \times 10^3$ $+\varepsilon = -2.365 \times 10^{-1}$	0.8261	Parabólico*	Diciembre
Polinomio de Taylor $RSAM = -\frac{3(\text{Tiempo})^2}{100000}$ $+\frac{6039(\text{Tiempo})}{5000} - \frac{155231}{20}$			
$RSAM = 2 \times 10^{-7} e^{2.5 \times 10^{-3} \text{Tiempo}}$ $+(\varepsilon = -2.369 \times 10^{-1})$	0.8031	Exponencial	$nov \leq RSAM \leq dic$
$RSAM = 3.501 \times 10^{-1} \text{Tiempo}$ $-2.6801 \times 10^3$ $+(\varepsilon = -2 \times 10^{-1})$	0.7067	Lineal	$nov \leq RSAM \leq dic$
$RSAM = 2.7949$ $\times 10^3 \text{Ln}(\text{Tiempo})$ $-2.4995 \times 10^4$ $+\varepsilon = 1.4 \times 10^{-1}$	0.6924	Logarítmica	$nov \leq RSAM \leq dic$
$RSAM = 2 \times 10^{-4} (\text{Tiempo})^2$ $-2.9676(\text{Tiempo}) + 1.0969 \times 10^4$ $+\varepsilon = 8.373 \times 10^{-1}$	0.5591	Parabólico	$oct \leq RSAM \leq dic$
Polinomio de Taylor $RSAM = -\frac{(\text{Tiempo})^2}{5000}$ $-\frac{7419(\text{Tiempo})}{2500} - \frac{1105273}{100}$			

\*Los modelos lineal y cuadrático son coincidentes con el modelo logarítmico  $RSAM = 5.8536 \times 10^3 \text{Ln}(\text{Tiempo}) - 5.2626 \times 10^4$ .

El análisis exponencial se aplica a los datos previos a la erupción principal del volcán San Miguel. Distinguidos artistas dibujan su obra a partir de elipses, el cuerpo humano se descompone en elipses, la órbita que siguen los planetas y demás astros en el espacio son explicadas, estudiadas, y pronosticadas a través de leyes que rigen a las elipses formadas; la simulación de caída de ceniza y piroclastos de un volcán en erupción, la simulación de lahares, entre otros. El cráneo y el abdomen son justamente elipses. Por ello, la aplicación del tratado de las secciones cónicas, la elipse en especial, produce buenos resultados en vulcanología aunque pueda extenderse a otras áreas.



*Ilustración 46.* Modelos estadísticos VSA y VSM.

a) Modelos estadísticos para el RSAM por hora del VSM. b) Modelos estadísticos para el RSAM por hora del VSA.

La elipse deformada trivial es la que se llama comúnmente circunferencia, no es de subestimar que el invento más ingenioso que el hombre haya inventado sea la rueda, que tiene forma de circunferencia, pero realmente es una arandela. Al igual que para el elipsoide, también para la arandela existen rigurosos estudios, donde la matemática superior es aplicada sin límite alguno. Referente a las elipses deformadas, puede ser una figura que sólo tenga centro y nada más, trátase de un punto; y en el caso extremo, que los puntos que conformen la elipse estén delimitados por el conjunto vacío. Un estudio exhaustivo de la simulación de elipses de acuerdo al comportamiento del RSAM por hora, puede ser interpretado en hora buena, más cuando se trata de complejos volcánicos de alta peligrosidad, como lo son el Pinatubo, el Krakatoa, el Nevado del Ruíz, el Kilauea, el monte Fuji, el Kilimanjaro, el Tambora, el monte Etna, el monte Santa Helena, el Eyjafjallajokull por citar algunos.

*Tabla 29. Pendiente e inclinación de Kernels VSM y VSA.*

<b>Escenario Kernels II VSM</b>			
<b>Figura</b>	<b>N</b>	<b>Pendiente</b>	<b>Ángulo de inclinación</b>
a	I	-0.01946	178.8852
	II	-0.01861	178.9338
	III	-0.01861	178.9338
	<b>IV</b>	<b>0.008372</b>	<b>0.4797</b>
	V	0.02292	1.3130
	VI	0.09590	5.4779
	<b>VII</b>	<b>0.22590</b>	<b>12.7295</b>
	VIII	0.26700	14.9492
	IX	0.68960	34.5901
	X	*	*
<b>Escenario Kernels II VSA</b>			
<b>Figura</b>	<b>N</b>	<b>Pendiente</b>	<b>Ángulo de inclinación</b>
b	I	0.05432	3.1093
	<b>II</b>	0.002488	<b>0.1425</b>
	<b>III</b>	-0.04833	<b>177.2330</b>

Al trabajar con modelos estadísticos se mantiene la razón de ser de los mismos. Los modelos parsimoniosos explican lo suficiente las anomalías en la actividad volcánica de los complejos volcánicos y representan, de forma aproximada, a la totalidad de los datos cuando su orden es bajo y son efectivos. Existe una interrelación entre el vulcanismo y la sismicidad, no puede ocurrir una erupción volcánica sin la

aparición de la sismicidad, el otro caso si tiene dos opciones; hay sismicidad junto con una erupción volcánica, o simplemente hay sismicidad y ninguna erupción volcánica, pero en esta dirección siempre se relaciona debido a la dinámica de los fluidos de magma.

La relación que guarda el vulcanismo y la sismicidad se basa en que los procesos dinámicos ocurridos durante las inyecciones de magma causan sismos de origen vulcano-tectónicos, mientras los procesos debidos a movimientos de fluidos (magma, gases y vapor) en los conductos del sistema interior del volcán ocasionan la ocurrencia de sismos volcánicos, los cuales tienen características espectrales que varían según su génesis, profundidad y magnitud (Gutiérrez, 2015, comunicación personal). Una anomalía volcánica puede ser cualitativa o cuantitativa según la fuente que la origina. Entre los métodos de análisis que se aplican a los datos crudos están: límites de detección, rangos y clasificación de frecuencias, la aleatoriedad de los procesos sísmico-volcánicos obliga a la vulcanología al empleo de métodos empíricos tales como la observación, medición y experimentación.

La inclusión del método científico en el vulcanismo se forja en el planteamiento de la génesis de la fuente que origina los datos, estadísticamente hace alusión a un planteamiento de hipótesis el cual es sometido para rechazar o aceptar la hipótesis nula (Gutiérrez, 2015, comunicación personal). Pronosticar y predecir son palabras que en ocasiones se utilizan indistintamente sin saber su verdadero significado. No es lo mismo pronosticar una erupción volcánica, que predecirla; la opinión del especialista es muy atinada. El pronóstico se realiza en base a la probabilidad de ocurrencia de un futuro evento, según la información que se obtiene a través de la vigilancia volcánica. No toda intranquilidad en la actividad volcánica culmina siempre en una erupción.

En contraste, la predicción tal como lo indica su significado, no se ejerce en vulcanología. Es muy atrevido y poco ético señalar día, hora y minuto de un futuro evento eruptivo sin el debido soporte estadístico y una verosimilitud exacta, esto trae como consecuencia un sistema de alarma poco creíble (Gutiérrez, 2015, comunicación personal). Parece absurdo predecir la cantidad en toneladas de magma, la dirección del

flujo, la cantidad de ceniza. En la actualidad no hay predicción exitosa pero los avances se utilizan para tal fin. Similarmente el predecir dónde ocurrirá el próximo asesinato en el país, su género, la edad, y génesis de dicho acto; conlleva la misma limitante que el caso anterior. El empleo innovador de las gotas de Fourier y las densidades Kernel en dos dimensiones resulta potente.

Entre las técnicas estadísticas utilizadas por el MARN destacan la estadística descriptiva, los mínimos cuadrados, Pareto, regresión y muestreo. Entre las técnicas inferenciales, se tiene el análisis de las series de Fourier, los métodos numéricos de simulación. Los métodos Weibull-method, Exponential y Power Law son utilizados para determinar el volumen y masa de productos eruptivos. Actualmente no hay descubrimientos de patrones o modelos que describan el comportamiento de las variables a lo largo del tiempo, ya que la base instrumental usada corresponde únicamente a dos erupciones en volcanes distintos; a modo de reflexión, lo anterior hace alusión a que el ojo humano con suerte puede disfrutar dos erupciones volcánicas en su corto periodo de vida.

El porcentaje de las anormalidades en la variación de la medida RSAM por hora es de dos con treinta y un centésimas para el volcán Santa Ana, mientras que para el volcán de San Miguel es de dos con veintidós centésimas. Las anormalidades en el RSAM conforman el dos por ciento para los años dos mil cinco y dos mil trece respectivamente. La hipótesis alterna se rechaza ya que la diferencia entre las dos señales es significativa.

## Capítulo 7: Propuestas concretas

En todas las áreas existen bases de datos, estén o no descubiertas, a las cuales se les aplican técnicas estadísticas idóneas para la obtención de la interpretación del comportamiento de las variables en interacción. Es por eso que es imprescindible la creación e incorporación en el Ministerio de Medio Ambiente y Recursos Naturales, un SSIG, siglas de Sistema Salvadoreño de Información Geológica que contenga series de tiempo continuas de las variables más importantes para el modelamiento estadístico. La obtención de algunas magnitudes se hace automáticamente, y otras de tipo in situ. Las variables sísmico-volcánicas se almacenan en fragmentos y se recurren a ellas cuando se hace una investigación específica, teniendo las desventajas de que los intervalos de tiempo son cortos producto de la fragmentación.

Así que con la futura incorporación del SSIG, además del registro automático, existiría un sistema informático para el registro manual de las variables sísmico-volcánicas. El almacenamiento de dicho sistema es en la nube, no consume la memoria física de los ordenadores; también se trataría de un sistema creado con software de licencia libre. Serán programados análisis periódicos los días lunes de cada semana; así el modelamiento estadístico puede ser de rutina, cada siete días; o bien dependiendo de la emergencia, en cada hora o minuto, las densidades Kernel son magníficos para la ubicación instantánea de los núcleos de datos. Lo anterior estaría forjado en un cronograma de actividades vulcanológicas. La implementación se define en dos fases; la primera, cuyo interés sea el volcánico; y la segunda, orientado a la geología en general.

Con la óptima implementación de este SSIG, resultarían novedosas bases de datos y series temporales volcánicas, de dimensión apreciable, que compartir con el mundo. El trabajo del vulcanólogo no se ampliaría, sino más bien se estandarizaría, a causa de un compromiso merced a vuestro país El Salvador. ¿Por qué son tan importantes las bases de datos volcánicas? Contar con una base de datos fidedigna al comportamiento volcánico, es decir, base de datos no corrupta; permite prevenir desastres ocasionados por la actividad volcánica. La base de datos nace con el registro de observaciones continuas y equidistantes entre sí. Luego se recurre a las fases de simulación y

estimación que van de la mano con el modelamiento estadístico, anteriormente se explicaba que puede ser cada lunes o en el momento idóneo, según la necesidad.

La experimentación con materiales reales surge de las decisiones tomadas en las dos fases anteriores. La experimentación adecuada, dada la simulación y estimación, permiten la creación de modelos teóricos estadísticos que conducen a pronósticos de erupciones volcánicas exitosas. En el pronóstico de erupciones se necesitarán mapas volcánicos de peligrosidad y psicología para contrarrestar el impacto del fenómeno natural. Para darle seguimiento a esta propuesta se tiene en cuenta que la mayoría de fenómenos volcánicos dependen de las propiedades de cada volcán, pocos humanos han tenido la experiencia de presenciar múltiples erupciones volcánicas de un mismo volcán. Con la creación de la base de datos volcánicos, ahora se requiere interpretar, modelar, analizar y experimentar con la historia volcánica.

Esta última parte se divide en dos niveles interdependientes, el primero se basa en las interpretaciones de las características como premisas ante una erupción volcánica. Y segundo, en el modelado estadístico a partir de técnicas paramétricas o no paramétricas. La primera división, se trata de un árbol de eventos que muestra generalizaciones del fenómeno volcánico en términos de posibilidad, magnitud y ramificaciones, teoría de caminos de la matemática discreta. Para cada movimiento a través del árbol, la próxima rama de la progresión de la actividad volcánica es considerada y se valora desde la probabilidad con que empezó. El método para calcular la probabilidad aún está en búsqueda (ver ilustración 47), pero se dirige para ser aplicado en valoraciones cortas y largas, lo anterior realmente es una probabilidad integral, no se pierde el valor inicial de la probabilidad (Newhall & Hoblitt, 2002).

Inicio del árbol de eventos
P(1) Anormalidad
P(2 1) Causa
P(3 2) Resultado
P(4 3) Magnitud
P(5 4) Fenómeno
P(6 5) Área
P(7 6) Magnitud
P(8 7) Explosión
P(9 8) Vulnerabilidad

*Ilustración 47.* Detalle resumido del árbol de eventos de Newhall & Hoblitt, 2002.

## Capítulo 8: Conclusiones

Se observó eficiencia en el uso de técnicas estadísticas para la interpretación de las variables de tipo sísmico-volcánico por parte de la entidad gubernamental en materia de vulcanismo y sismicidad denominada Ministerio de Medio Ambiente y Recursos Naturales. El MARN está muy vinculado a la institución encargada de la protección civil, el SNET. Los especialistas afirman que se ejecutan proyectos orientados al modelamiento estadístico con finalidad del pronóstico de estos eventos de naturaleza aleatoria. Hay muchas interrelaciones entre el vulcanismo y la sismicidad, por lo que no deben tratarse aisladamente. La actividad sísmica y volcánica incluye factores que hay que tener en cuenta.

VARIABLES DE MÁXIMO INTERÉS: coordenadas geográficas del sismo, o erupción volcánica; RSAM diario y por hora, magnitud sísmica, conteo sísmico, clasificación sísmica, profundidad sísmica, deformaciones, formaciones de domos, temperatura diaria, mensual y anual; índice de explosividad, composición volcánica, análisis químicos para la obtención de mediciones de sólidos, líquidos o gases; magnitud de los desprendimientos, características de las columnas de cenizas, característica de las precipitaciones y vientos, entre otros. También hay que contar con mapas de peligrosidad y amenaza actualizados, simulaciones, excelentes historiales sísmicos y volcánicos, interrelaciones entre el vulcanismo y la sismicidad, y otros fenómenos naturales; placas tectónicas y fallas, ubicación de ciudades y centros turísticos, población y cultivos, el núcleo de la tierra, entre otros elementos.

Con el conocimiento necesario y suficiente de los vulcanólogos y los sismólogos sí es posible pronosticar erupciones volcánicas, aunque de manera empírica aún. Los pronósticos se basan en la observación y estratificación de la actividad sísmica y volcánica de los complejos volcánicos. La ciencia necesita cada vez más la aplicación de la estadística, tanto en la inferencia como en el análisis de los comportamientos que los diferentes fenómenos naturales alcanzan. Para lograr el cumplimiento de estos objetivos, en la sección discusión de resultados se han propuesto varios modelos estadísticos que ayudan al pronóstico de erupciones volcánicas; también propuestas en concreto para

alcanzarlos plenamente en un futuro próximo.

Los modelos más satisfactorios son la regresión no paramétrica Kernel, regresión Loess y las Gotas de Fourier; modelan correctamente el crecimiento natural de la anomalía de un fenómeno físico. La propuesta consiste en la creación y estandarización de un Sistema Salvadoreño de Información Geológica, el cual puede ser aplicado en dos fases. Al comparar la efectividad de los modelos desarrollados, se afirma que primero se producen transformaciones de la base de datos, para la posterior aplicación de las técnicas. Estas transformaciones son los suavizados para el análisis de tendencia. En el modelamiento se ejecutaron diagnósticos de estacionalidad a través de los filtros de Medias Móviles y el Suavizado Exponencial. No existe estacionalidad en las series temporales utilizadas, por lo que se trabajó con los datos brutos.

Por último, para responder al planteamiento de hipótesis, se afirma que sí existe diferencia alguna entre el porcentaje de señales irregulares decisivas ante erupciones volcánicas y el porcentaje de señales regulares en el RSAM por hora de los principales complejos volcánicos Santa Ana y San Miguel. El contraste consistió en calcular los tres intervalos principales que incluyen a los datos en diferentes proporciones. Se deduce que el porcentaje de la señal irregular es mucho menor que el de la señal regular. También se recurre al índice de los Kernels, un Kernel se cataloga como referencial o anormal, al comparar el Kernel de referencia con un Kernel anormal se concluye que si existe diferencia significativa.

No todo se explica paramétricamente, la mayoría de fenómenos naturales tienen comportamientos en los cuales es necesario aplicar técnicas estadísticas no paramétricas. Es bueno considerar la oportunidad de mejora y ampliación de conocimientos por parte de los especialistas en el área. La experiencia y conocimientos compartidos por las naciones sobresalientes en la vulcanología y sismología, por mencionar las más relevantes: Japón, Italia, España, y Estados Unidos; son determinantes en el intento de protección civil ante futuras catástrofes, producto de erupciones volcánicas, terremotos o tsunamis.

## Capítulo 9: Recomendaciones

El modelamiento clásico es muy bien recibido por los especialistas en distintas áreas, inclusive vulcanólogos y sismólogos. Entre los que más destacan están la regresión Loess, ya que sin ningún parámetro, únicamente el determinar cuán suave se desea la curva, describe de forma satisfactoria los datos, también la regresión Kernel y la regresión polinomial local. Toda la investigación en abordó este tipo de modelamiento estadístico, la regresión paramétrica se utilizó para definir los comportamientos volcánicos en un intervalo de tiempo previo a la erupción volcánica principal del volcán Santa Ana y volcán San Miguel. En un estudio posterior se recomienda analizar la dinámica de los Kernels en tiempo actualizado.

La tecnología provee de herramientas (scripts) que simulen los movimientos característicos (aumento, descenso, aparición o extinción) de los Kernels. El movimiento depende del tipo de pendiente y magnitud del suavizado Loess aplicado a cada Kernel y al ángulo de inclinación de la pendiente. La recomendación anterior potencia estadísticamente la observación, control y pronóstico de futuras erupciones volcánicas en los complejos volcánicos de El Salvador. La comparación de la teoría de Kernels del RSAM con las técnicas de monitoreo continuo como deformación y concentración de gases indica una posible relación importante y fuerte en pronóstico de erupciones volcánicas.

La aplicación de Redes Neuronales a los Kernels del RSAM en un determinado intervalo de tiempo es otra innovación en la búsqueda de modelos de pronóstico de eventos eruptivos. El cálculo de los elipsoides producto de la proyección de los Kernels binarios en un espacio tridimensional también es una excelente opción que involucra al análisis estadístico en pro de descubrimientos de patrones, modelos, perspectivas y pronósticos de eventos volcánicos a partir del comportamiento sísmico medido por el RSAM. El elipsoide de las coordenadas geográficas respecto a la profundidad también es interesante en el modelado estadístico.

Las Wavelets se han incorporado en la Estadística entre los años ochenta y noventa, inicialmente están vinculadas a los estudios geofísicos. Se aplican en análisis de

supervivencia, análisis estadístico de series temporales, procesamiento estadístico de imágenes, y estabilización de varianzas; consisten en pequeñas ondas que son muy potenciales. Parten de la generalización de la idea de la transformada de Fourier por lo que tienen similares aplicaciones; aunque no se expanden en series de Fourier, sino que utilizan bases ortogonales que son apropiadas para cálculos eficientes, son muy utilizadas en la compresión de datos.

Kriging es un método de interpolación que predice el valor del campo geofísico en un punto no muestreado a partir del valor observado para dicho campo geofísico en localizaciones. Para ello se emplea el variograma que permite expresar de manera directa la variación espacial observada mientras se minimiza el error de los valores predichos. Kriging recibe el seudónimo de predicción óptima, también se le conoce como BLUE (best linear unbiased estimator) que significa mejor estimador lineal insesgado. Los valores estimados son combinaciones lineales ponderadas de las variables recolectadas, y es insesgado porque la media de su error es cero.

El variograma refleja las condiciones de heterogeneidad y anisotropía de la distribución espacial de un campo geofísico, por ejemplo la permeabilidad de un conducto volcánico al paso de los gases, la distribución espacial de la emisión difusa de gas volcánico a través de los suelos. De lo anterior resulta importante el uso de la interpolación probabilística en los estudios vulcanológicos. El kriging estima las distribuciones espaciales, podría ser de cenizas, magma u otros fluidos, gases volcánicos. El kriging y el variograma permiten obtener imágenes n-dimensionales del fenómeno en estudio. Gracias a ello es posible discernir gráfica y/o numéricamente las regiones donde el campo geofísico adopta valores anómalos.

## Bibliografía

- Abdelkader, N. B. (2009). *Regresión no paramétrica en R* (tesis de maestría). Granada, España.
- Aizawa, K. *Present situation of monitoring, prediction and information for volcanic disaster mitigation in Japan*. Tokio, Japón: UJNR Panel on Earthquake Research.
- Álvarez, F. (1999). *Algunas aportaciones al análisis de datos, utilizando técnicas de representación multivariante* (tesis doctoral). Cádiz, España: Universidad de Cádiz.
- Alzola, C. & Harrell, F. (2006). *An introduction to S and the Hmisc and Design libraries*. Virginia, Estados Unidos de América: Statistical consultant.
- Aoki, Y., Takeo, M., Ohminato, T., Nagaoka, Y., & Nishida, K. (2013). *Magma pathway and its structural controls of Asama volcano, Japan*. Londres, Reino Unido: Geological Society.
- Ary, T. T. (1998). *Explorations, an introduction to Astronomy*. Massachusetts, Estados Unidos de América: WCB McGraw-Hill.
- Azofeifa, C. E. *Aplicación de la simulación MonteCarlo en el cálculo del riesgo usando Excel*. San José, Costa Rica: Tecnología en Marcha.
- Beard, Mary (2008). *Pompeya: Historia y leyenda de una ciudad romana*. Barcelona, España.
- Benito, N. (2010). *Uso de métodos gráficos en publicaciones científicas de ciencias de la documentación* (tesina). Salamanca, España: Universidad de Salamanca.
- Berrío, J. S., Paz, L. M., & Bravo, E. C. (2013). Segmentación y parametrización de líneas en datos láser 2D basado en agrupamiento por desplazamiento de media. Barranquillo, Colombia: Universidad Autónoma del Caribe.
- Biblioteca Universitaria. *Sistemas operativos*. Alicante, España: Universitat d'Alacant.
- Bidegain, M., Díaz, A., Barreiro, M. (2009). *Análisis estadístico de datos climáticos*. Montevideo, Uruguay: Universidad de la República.
- Blanco, A., & Rodríguez, J. (2007). *Intervención psicosocial*. Madrid, España: Pearson Educación.
- Bohling, G. (2005). *Kriging*. Kansas, Estados Unidos de América: Kansas Geological

Survey.

- Bonilla, G. (2005). *Estadística: elementos de estadística descriptiva y probabilidad*. San Salvador, El Salvador: UCA.
- Briones, G. (1996). *Metodología de la investigación cuantitativa en las ciencias sociales*. Recuperado de <http://www.monografias.com>.
- Canavos, G. C. (1988). *Probabilidad y estadística, aplicaciones y métodos*. México D. F., México: McGraw-Hill.
- Canizales, E. (2012). *Apuntes del curso modelos lineales*. Santa Ana, El Salvador: Universidad de El Salvador.
- Caracena, J. A. (2002). *Un procedimiento completo para la detección de estacionalidad en series económicas* (Tesina). Madrid, España: CEMFI.
- Carmona, F. (2003). *Modelos lineales*. Barcelona, España: Universitat de Barcelona.
- Carracedo, J. C., Guillou, H., Paterne, M., Scaillet, S., Rodríguez, E., Paris, R., Pérez, F. J., & Hansen, A. (2004). *Análisis del riesgo volcánico asociado al flujo de lavas en Tenerife (Islas Canarias): Escenarios previsibles para una futura erupción en la isla*. Tenerife, España.
- Centro de Escritura Javeriano. (2010). *Normas APA sexta edición*. Recuperado de <http://www.apastyle.org>.
- Chongsuvivatwong, V. (2007). *Analysis of epidemiological data using R and Epicalc*. Songkla, Tailandia: Prince of Songkla University.
- Christou, N. *Ordinary kriging using geoR and gstat*. Los Ángeles, Estados Unidos de América: University of California.
- Cohen, A. R. (1999). *An introduction to PROC LOESS for local regression*. Carolina del Norte, Estados Unidos de América: SAS Institute Inc.
- Colvin, A. S. (2008). *Crater lake evolution during volcanic unrest: case study of the 2005 phreatic eruption at Santa Ana volcano, El Salvador* (tesis de maestría). Michigan, Estados Unidos de América: Michigan Technological University.
- Comisión Nacional de Energía Eléctrica. (2011). *Informe estadístico 2011: Mercado mayorista de electricidad de la República de Guatemala correspondiente al año*

2010. Ciudad de Guatemala, Guatemala.
- Correa, J. C., & González, N. (2002). *Gráficos estadísticos con R* (tesis doctoral). Medellín, Colombia: Universidad Nacional-Sede Medellín.
- Crawley, M. J. (2007). *The R book*. West Sussex, Inglaterra: WILEY.
- Cuadras, C. M. (2012). *Nuevos métodos de análisis multivariante*. Barcelona, España: CMC.
- Delicado, P. (2008). *Curso de modelos no paramétricos*. Cataluña, España: Universitat Politècnica de Catalunya.
- Earl, D. J. & Deem, M. W. (2008). *Molecular modeling of proteins*. Springer.
- Faraway, J. J. (2002). *Practical regression and anova using R*.
- Farnsworth, G. V. (2008). *Econometrics in R*.
- Felpeo, A. (2002). *Modelización física y simulación numérica de procesos eruptivos para la generación de mapas de peligrosidad volcánica* (tesis doctoral). Madrid, España: Universidad Complutense de Madrid.
- Fernández, M. C. *Galería de curvas en el plano*. España.
- Ferrero, R. (2010). *Procesos espaciales puntuales con R, parte II*. España: IAS-CSIC. Recuperado de: <http://statisticalecology-met.blogspot.com/2010/12>
- Fujita, E. (2009). Current status and future issues of volcanic eruption prediction research. *Science & Technology Trends*, 32, 85-101.
- Fuster, J. M. *Vulcanismo y cambio climático*. Madrid, España: Real Academia de Ciencias.
- Genolini, C. (2008). *A (not so) short introduction to S4*.
- GEOS. (2005). *Sesión vulcanología*. México D. F., México.
- Gómez, F. (1996). *Desarrollo de una metodología para el análisis del riesgo volcánico en el marco de un sistema de información geográfica* (tesis doctoral). Madrid, España: Universidad Complutense de Madrid.
- Gómez, N. G. (2006). *Notas de clase: Series de tiempo en R*. Medellín, Colombia: Universidad Nacional de Colombia.
- Gómez-Cadenas, J. J. (2005). El método de MonteCarlo. *Curso de Estadística, 1*, 1-15.

- González, A. P. (2010). *Análisis predictivo de datos mediante técnicas de regresión estadística* (tesis de maestría). Madrid, España: Universidad Complutense de Madrid.
- González, E., Escobar, E., Becerra, R., Gosálvez, R. U., & Dónis, J. (2010). *Aportaciones recientes en vulcanología 2005-2008*. Castilla-La Mancha, España: Centro de Estudios Calatravos.
- González, F. *Lenguaje R: un complemento libre para las asignaturas de estadística* (tesina). Acatlán, México: Universidad Nacional Autónoma de México.
- González, G. (1997). *Series de Fourier, Transformadas de Fourier y aplicaciones*. Maracaibo, Venezuela: Universidad del Zulia.
- González, M. P. (2009). *Análisis de series temporales: modelo ARIMA*. País Vasco, España: Universidad del País Vasco.
- González, W. (1988). *Una perspectiva general con nuevos resultados de la aplicación de la estimación no paramétrica a la regresión lineal*. Santiago de Compostela, España: Estadística Española.
- Gujarati, D. N. (2004). *Basic Econometrics*. Nueva York, Estados Unidos de América: The McGraw-Hill Companies.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Pearson Prentice Hall.
- Härdle, W., & Simar, L. (2003). *Applied multivariate statistical analysis*. Berlín, Alemania: MDTECH.
- Harrell, F. E. (2002). *Statistical computing and graphics*. Virginia, Estados Unidos de América: University of Virginia.
- Hiebeler, D. (2010). *MATLAB/R reference*. Maine, Estados Unidos de América: University of Maine.
- Jerez, M. & Sotoca, S. (2010). *Econometría II, análisis de series temporales I y II*. Madrid, España: Universidad Complutense de Madrid.
- Kelmansky, D. M. (2006). *Análisis exploratorio y confirmatorio de datos de experimentos de microarrays*.
- Kitagawa, G. (2010). *Introduction to time series modeling*. Tokio, Japón: CRC Press.

- Krijnen, W. P. (2009). *Applied statistics for bioinformatics using R*. Groningen, The Netherlands: Hanze University.
- Lam, L. (2010). *An introduction to R*.
- Laske, Gabi. «The A.D. 79 Eruption at Mt. Vesuvius». *UCSD – EARTH 15: Natural Disasters*.
- Leithold, L. (1994). *El Cálculo*. México D. F., México: Oxford university press.
- Lewis, J. (2013). *GEOSTAT13 tutorial spatial interpolation with R*.
- Lorenzo, J. M. *Volcanes, terremotos y matemática*. Universidad de La Laguna.
- Major, J. J., Schilling, S. P., Sofield, Escobar, C. D., & Pullinger, C. R. (2001). *Riesgo volcánico en la región de San Salvador, El Salvador*. Washington, Estados Unidos de América: USGS.
- Mallo, P. E., Artola, M. A., Morettini, M., Galante, M. J., Pascual, M. E., Busetto, A. R., & Zanfrillo, A. I. *Conceptos preliminares de suavizado y pronóstico de series cronológicas con herramientas difusas*. Mar del Plata, Argentina: Universidad Nacional de Mar del Plata.
- Mauricio, J. A. (2002). *Funciones de Autocorrelación simple y parcial, modelos AR(1), AR(2), MA(1), MA(2) y ARMA(1, 1); algunos modelos estacionales*. Madrid, España: Universidad Complutense de Madrid.
- Mauricio, J. A. (2007). *Análisis univariante de series temporales*. Madrid, España: Universidad Complutense de Madrid.
- Mohlenkamp, M. J. *A tutorial on wavelets and their applications*. Colorado, Estados Unidos de América: University of Colorado.
- Monje, C. A. (2011). *Metodología de la investigación cuantitativa y cualitativa: guía didáctica*. Santa Fé de Bogotá, Colombia: NEIVA.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introducción al análisis de regresión lineal*. Ciudad de México, México: CECOSA.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to Time Series Analysis and forecasting*. New Jersey, Estados Unidos de América: Wiley-Interscience.

- Nason, G. P. (2008). *Wavelet methods in statistics with R*. Bristol, Reino Unido: Springer.
- Novales, A. (1993). *Econometría*. Madrid, España: McGraw-Hill.
- Olalquiaga, P., Olalquiaga, O. (2005). *El libro de las curvas*. Madrid, España: ESTEYCO.
- Olaya, J., Reina, J., Torres, L. L., Paz, M. C., Pereira, L. A. (2013). *Modelación no paramétrica de la contaminación octohoraria del aire debida al Monóxido de Carbono y al Ozono Troposférico*. Cali, Colombia: Universidad del Valle.
- Pankratz, A. (1993). *Forecasting with univariate Box-Jenkins models*. Nueva York, Estados Unidos de América: John Wiley & Sons.
- Paradis, E. (2005). *R for beginners*. Montpellier, Francia: Institut des Sciences de l'Évolution.
- Pebesma, E. (2015). *The meuse data set: a brief tutorial for the gstat R package*.
- Peña, D. (2002). *Análisis de datos multivariantes*. Barcelona, España.
- Pereira, A. (2010). *Análisis predictivo de datos mediante técnicas de regresión estadísticas* (tesis de maestría). Madrid, España: Universidad Complutense de Madrid.
- Pértega, S. (2007). *Aplicación de la regresión polinómica local al análisis discriminante y análisis clúster de series de tiempo* (tesis doctoral). Coruña, España: Universidade da Coruña.
- R Development Core Team. (2013). *Introducción a R; notas sobre R: un entorno de programación para análisis de datos y gráficos*.
- R Foundation for Statistical Computing. (2010). *R: A language and environment for statistical computing*. Recuperado de <http://www.r-project.org/>
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Massachusetts, Estados Unidos de América: Massachusetts Institute of Technology.
- Raychaudhuri, S. (2008). Introduction to MonteCarlo simulation. *Winter Simulation Conference, I*, 91-100.

- Razo, C. M. (1998). *Cómo elaborar y asesorar una investigación de tesis*. Ciudad de México, México: Pearson Educación.
- Red de Estudios Sociales en Prevención de Desastres en América Latina. (1996). *Desastres & Sociedad: predicciones, pronósticos, alertas y respuestas sociales (No. 6)*. Recuperado de <http://www.desenredando.org>
- Rencher, A. C. (2002). *Methods of multivariate analysis*. Nueva York, Estados Unidos de América: WILEY-INTERSCIENCE.
- Ricci, V. (2004). *R functions for time series analysis*.
- Ricci, V. (2005). *Fitting distributions with R*.
- Ricci, V. (2005). *R functions for regression analysis*.
- Rielo, A. F. (2002). *Modelización física y simulación numérica de procesos eruptivos para la generación de mapas de peligrosidad volcánica* (tesis doctoral). Madrid, España: Universidad Complutense de Madrid.
- Robert, C. P., & Casella, G. (2009). *Introducing MonteCarlo methods with R*. Florida, Estados Unidos de América: Springer.
- Rodríguez-Aragón, L. J. (2011). *Simulación, Método de MonteCarlo*. Aragón, España: Área de Estadística e Investigación Operativa.
- Rossiter, D. G. (2010). *Introduction to the R project for statistical computing for use at ITC*. Enschede, NL: International Institute for Geo-information Science & Earth Observation.
- Ruíz, M. (2008). *Contribución a la teoría de la estimación robusta de escala en modelos de regresión no paramétricos* (tesis doctoral). San Luis, Argentina: Universidad Nacional de San Luis.
- Sáez, A. J. (2010). *Métodos estadísticos con R y R Commander*. Barcelona, España: Vniversitas Giennensis.
- Sampieri, R. H., Collado, C. F., & Lucio, P. B. (1997). *Metodología de la investigación*. Santa Fe de Bogotá, Colombia: McGraw-Hill.
- SAS Institute Inc., SAS/STAT. (1999). *User's guide versión 8*. Carolina del Norte, Estados Unidos de América: SAS Institute Inc.

- Schiattino, I. & Silva, C. (2010). *Representación gráfica de información multivariante. Aplicación al sistema de salud de Chile*. Santiago de Chile, Chile: Universidad de Chile.
- Scolamacchia, T., Pullinger, C., Caballero, L., Montalvo, F., Beramendi, L. E., & Hernández, G. G. (2010). *The 2005 eruption of Ilamatepec (Santa Ana) volcano, El Salvador*. México D. F., México: ELSEVIER.
- Seefeld, K. & Linder, E. (2007). *Statistics using R with biological examples*. Durham, Reino Unido: University of New Hampshire.
- Sethuraman, A. (2015). *2D posterior density plots in R*.
- Sorto, M. L. (2008). *Regresión*. San Salvador, El Salvador: Universidad de El Salvador.
- Spivak, M. (1988). *Cálculo en variedades*. Barcelona, España: REVERTÈ, S. A.
- Stewart, J. (2002). *Cálculo trascendentes tempranas*. México D. F., México: Thomson Learning.
- Taha, H. A. (2004). *Investigación de operaciones*. Ciudad de México, México: Pearson Educación.
- Targett, D. (2011). *Métodos cuantitativos*. Edimburgo, Reino Unido: Edinburgh Business School.
- Tusell, F. (2011). *Análisis de regresión: Introducción teórica y práctica basada en R*. Bilbao, España.
- Tusell, F. (2012). *Análisis multivariante*. Barcelona, España.
- Urbanek, S., Theus, M. (2003). *High interaction graphics for R*. Augsburg, Alemania: University of Augsburg.
- USGS. (2008). *Understanding volcano hazards and preventing volcanic disasters*. Washington, Estados Unidos de América: United States Geological Survey.
- Ventana Systems Inc. (2007). *Análisis de sensibilidad*. España.
- Vidakovic, B. & Mueller, P. (1991). *Wavelets for kids, an tutorial introduction*. Durham, Reino Unido: Duke University.
- Vidal, J. (2008). *Interpolación de formas en imágenes usando morfología matemática* (Tesis Doctoral). Madrid, España: Universidad Politécnica de Madrid.

- Weinberger, K. Q., Tesauro, G. *Metric learning for Kernel regression*. Filadelfia, Estados Unidos de América: Universidad de Pensilvania.
- Yamasato, H., Funasaki, J., & Takagi, Y. (2013). *The Japan meteorological agency's volcanic disaster mitigation initiatives*. Tokio, Japón: National Research Institute for Earth Science and Disaster Prevention.
- Zavaleta, J. (2010). *Kriging: un método de interpolación sobre datos dispersos*. México D. F., México: Universidad Nacional Autónoma de México.

## Apéndice I: Consideraciones sobre Kernels Binarios

Contenidos importantes: Métodos de suavizado utilizando Kernels. Métodos no paramétricos. Modelamiento de la densidad. Regresión no paramétrica. Métodos paramétricos clásicos. Geometría analítica del Kernel. Densidad bivariada.

Un Kernel es simplemente una función de pesos promediados, los Kernels revelan la estructura de datos subestimada, y brindan estructuras n-dimensionales profundas.

Las densidades con histogramas tienen muchas limitantes que son cubiertas por el suavizado Kernel.

Las densidades Kernel estiman las densidades de distribución por el método Kernel (Härdle & Simar, 2003).

Las densidades Kernel se utilizan en la búsqueda de modelos, localizaciones, simetrías, asimetrías, colas de distribuciones,...

Nbin es un vector numérico de longitud uno (para ambas direcciones), o dos (para x e y separadamente) que especifica el número de puntos espaciados igualmente.

La historia del análisis de Fourier tiene más de 200 años. Sus orígenes principian unos 60 años antes del momento en que Jean Baptiste Joseph Fourier presentó la primera versión de su trabajo sobre la teoría de la conducción del calor a la Academia de París (1807). El año 1750 es un buen punto de partida: Fernando VI era rey de España, y Jorge II de Inglaterra; las colonias de América del norte estaban en medio de las guerras con los nativos y los franceses; unos años después Carlos III creaba el virreynato del Río de la Plata (1776). Voltaire, Rousseau y Kant estaban escribiendo sus libros en Europa; Bach acababa de morir, y Mozart estaba pronto a nacer; y el cálculo de Leibnitz y Newton, publicado 75 años antes, estaba permitiendo la creación de poderosas nuevas teorías sobre la mecánica celeste y la mecánica del continuo. En ese momento los esfuerzos de los físicos y matemáticos se concentraban en dos problemas principales, que sentarían las bases de lo que posteriormente se conocería como análisis de Fourier: el problema de la cuerda vibrante o la propagación del sonido en un medio elástico y la determinación de las órbitas de los planetas a partir de mediciones.

La función Kernel, función núcleo o función peso, que satisface ciertas condiciones de regularidad, generalmente es una función de densidad simétrica como por ejemplo la

de la distribución normal, con una secuencia de constantes positivas conocidas como ancho de ventana, parámetro de suavización o bandwidth. El estimador núcleo es una suma de protuberancias (del inglés bump) situadas en las observaciones. La función núcleo determina la forma de las protuberancias mientras que el parámetro determina su anchura. Una función Kernel se interpreta como una transformación continua de la función de distribución empírica de acuerdo a la función  $K(x)$ , que se encarga de redistribuir la masa de probabilidad  $\frac{1}{n}$  en la vecindad de cada punto muestral.

El estimador núcleo local permite que el bandwidth dependa de cada observación donde se realiza la estimación. Un caso particular de estimador núcleo local es el conocido como estimador núcleo de los  $k$  vecinos más cercanos (en inglés, k-nearest neighbors kernel estimator). En este estimador se usa un núcleo  $K$  con soporte en  $[-1,1]$  el parámetro de suavizado se elige como el mínimo valor que permite que en el intervalo  $[x - h, x + h]$  entren  $k$  de las observaciones (las  $k$  observaciones más cercanas al dato de referencia). El valor de  $k$  se puede elegir por validación cruzada. En ocasiones se fija la proporción de datos  $\frac{k}{n}$  que se quiere usar en la estimación de cada valor en vez de fijar  $k$ . A esa proporción se le denomina “span” en algunos paquetes estadísticos, inclusive el software estadístico R-project.

La discriminación no paramétrica (clasificación supervisada) basada en estimación de la densidad es una perspectiva interesante. Parte de la observación de  $p$  características,  $x = x_1, \dots, x_p$ , en  $n$  individuos que pertenecen a una población dividida en  $q$  subpoblaciones (clases),  $\{C_1, \dots, C_q\}$ . De cada individuo también se sabe la clase  $y_i \in \{1, \dots, q\}$  a la que pertenece. Se dispone de los datos  $(y_i; x_{i1}, \dots, x_{ip}), i = 1, \dots, n$  y se busca una regla discriminante que asigne un nuevo individuo (cuya clase se desconoce) a una de las  $q$  clases a partir de sus valores  $x_j$ . La regla óptima que minimiza la probabilidad de mala clasificación es la regla de Bayes, que consiste en asignar el individuo con observaciones  $x$  a la clase  $j$  que tiene máxima probabilidad a posteriori:

$$f(x|C_j)P(C_j) = \text{Max}_{k=1\dots q} f(x|C_k)P(C_k)$$

La regla Bayes sólo es aplicable si se conocen las probabilidades a priori de cada clase  $P(C_k)$ , y las funciones de densidad  $f(x|C_k)$  del conjunto de variables  $x$  para los individuos de cada clase. Las probabilidades a priori pueden estimarse fácilmente como las frecuencias relativas observadas de cada clase  $\hat{P}(C_k)$ . La estimación de las funciones de densidad  $f(x|C_k)$  puede llevarse a cabo mediante la estimación no paramétrica. Al estimar no paramétricamente la densidad de  $x$  en cada una de las  $q$  clases en las que está dividida la población, se usan exclusivamente las observaciones que pertenecen a la clase cuya densidad se está estimando. Finalmente, la regla discriminante consiste en asignar el individuo con observaciones  $x$  a la clase  $j$  que tiene máxima probabilidad a posteriori estimada:

$$\arg \max_{k=1 \dots q} \hat{f}(x|C_k) \hat{P}(C_k)$$

## Apéndice II: Densidad y el estimador núcleo de los $k$ vecinos más cercanos

Sea  $f(x)$  la función de densidad,  $F(x)$  la función de distribución y  $\hat{F}(x)$  su respectiva estimación,  $k(x, h) = \#\{x_i: x_i \in [x - h, x + h]\}$  para fijar que el numerador o denominador dependan de  $x$ :

$$f(x) = \frac{d}{dx}F(x)$$

$$f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$

$$f(x) \approx \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h}$$

$$\hat{f}_k(x) = \frac{\frac{k(x,h)}{n}}{2h} \text{ (numerador dependiente)}$$

$$\hat{f}_k(x) = \frac{\frac{k}{n}}{2h(x,k)} \text{ (denominador dependiente)}$$

$$\hat{f}_k(x) = \frac{1}{nh(x,k)} \sum_{i=1}^n K_U\left(\frac{1}{nh(x,k)}\right) \quad \text{(Núcleo uniforme)}$$

$$\hat{f}_k(x) = \frac{1}{nh(x,k)} \sum_{i=1}^n K\left(\frac{1}{nh(x,k)}\right) \quad \text{(Núcleo no uniforme)}$$

Si  $f(x) \approx \hat{f}_k(x) = \frac{k}{2nh(x,k)} \rightarrow h(x,k) \approx \frac{k}{2nf(x)}$ ; debido a que  $h(x) = h(x,k) \propto \frac{1}{f(x)}$ .

El parámetro de suavizado de la estimación de la densidad es  $k$ , el número de vecinos involucrados en la estimación de  $f(x)$ . Para que  $\hat{f}_k(x)$  converja a  $f(x)$  es condición suficiente que  $k$  dependa de  $n$  de tal forma:

$$\lim_n k(n) = \infty \wedge \lim_n \frac{k(n)}{n} = 0$$

### Apéndice III: Densidad y series de funciones ortogonales

Una variable aleatoria  $X$  tiene soporte en  $[-\pi, \pi]$  y su función de densidad es de cuadrado integrable en ese intervalo.  $f \in \mathcal{L}_2([-\pi, \pi])$  es un espacio vectorial euclídeo con producto escalar y una de sus bases ortonormales la constituyen las funciones trigonométricas seno y coseno. Para que la estimación sea convergente puntualmente a la función de densidad, es suficiente que el truncamiento cumpla que  $\lim_n m(n) = \infty \wedge \lim_n \frac{m}{n} = 0$ .

*Definición*  $f \in \mathcal{L}_2([-\pi, \pi]) = \left\{ \phi: [-\pi, \pi] \rightarrow R / \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi(x)^2 dx < \infty \right\}$

*Producto escalar en el espacio euclidiano*  $\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)g(x) dx$

*Base ortonormal del producto escalar*  $\{1, \cos(x), \sin(x), \cos(2x), \sin(2x), \dots\}$

*Serie de Fourier de la función de densidad con funciones ortonormales*  $f(x) = a_0 + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx))$

$$a_0 = \langle f(x), 1 \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx = \frac{1}{2\pi}$$

*Coefficientes del desarrollo en serie de Fourier*  $a_k = \langle f(x), \cos(kx) \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx = \frac{1}{2\pi} E(\cos kX)$

$$b_k = \langle f(x), \sin(kx) \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx = \frac{1}{2\pi} E(\sin kX)$$

*Estimaciones de los coeficientes*  $\hat{a}_k = \frac{1}{2\pi} \frac{1}{n} \sum_{i=1}^n \cos(kx_i) \wedge \hat{b}_k = \frac{1}{2\pi} \frac{1}{n} \sum_{i=1}^n \sin(kx_i)$

*Truncamiento en  $m$  (valor de suavizado de la estimación)*  $f_m(x) = \frac{1}{2\pi} + \sum_{k=1}^m \left( \frac{\hat{a}_k \cos(kx)}{\pi} + \frac{\hat{b}_k \sin(kx)}{\pi} \right)$

#### Apéndice IV: Principio de Dirichlet

El principio del palomar, también llamado principio de Dirichlet, establece que si  $n$  palomas se distribuyen en  $m$  palomares, y si  $n > m$ , entonces al menos habrá un palomar con más de una paloma. Otra forma de decirlo es que  $m$  huecos pueden albergar como mucho  $m$  objetos si cada uno de los objetos está en un hueco distinto, así que el hecho de añadir otro objeto fuerza a volver a utilizar alguno de los huecos. De otra manera, si se toman trece personas, al menos dos habrán nacido el mismo mes. En su versión más simple, este principio dice que no puede existir una aplicación inyectiva entre un conjunto de  $m$  elementos y otro de  $n$  elementos, si  $m > n$ . Equivalentemente, si se desean colocar  $m$  objetos en  $n$  cajas, con  $m > n$ , al menos una caja debe contener al menos 2 objetos.

Sean  $m, n$  y  $p$  tres números naturales. Si se desean colocar  $np + m$  objetos en  $n$  cajas, alguna caja debe contener al menos  $p + 1$  objetos. La demostración consiste en considerar si cada caja contiene como mucho  $p$  objetos, el número total de objetos que podemos colocar es  $np < np + 1 \leq np + m$ . Una versión generalizada de este principio dice que, si  $n$  objetos discretos deben guardarse en  $m$  cajas, al menos una caja debe contener no menos de  $\left\lceil \frac{n}{m} \right\rceil$  objetos, los símbolos denotan la función techo. Otra versión de este principio utiliza  $q_1, q_2, \dots, q_n$  números enteros positivos. Si  $q_1, q_2, \dots, q_n - n + 1$  objetos son colocados en  $n$  cajas, entonces o bien la primera caja contiene al menos  $q_1$  objetos, o la segunda caja contiene al menos  $q_2$  objetos, ..., o la  $n$ ésima caja contiene al menos  $q_n$  objetos.

Ahora supóngase que se distribuyen  $q_1, q_2, \dots, q_n - n + 1$  objetos en  $n$  cajas. Si para cada  $i = 1, 2, \dots, n$  la  $i$ ésima caja contiene menos de  $q_i$  objetos, entonces el número de objetos en todas las cajas no excede

$$(q_1 - 1) + (q_2 - 1) + \dots + (q_n - 1) = q_1 + q_2 + \dots + q_n - n$$

Puesto que el número es uno menos que el número de objetos distribuidos entre las cajas, por lo tanto para algún  $i = 1, 2, \dots, n$  la  $i$ ésima caja contiene al menos  $q_i$  objetos. Para el caso especial (forma simple o trivial de la forma fuerte del principio)

en el cual  $q_1 = q_2 = \dots = q_n = r$ ; el principio se lee “si  $n(r - 1) + 1$  objetos son colocados en  $n$  cajas, entonces al menos una de las cajas contiene al menos  $r$  objetos”, o “si  $k$  enteros  $z_1, z_2, \dots, z_k$  tienen un promedio mayor que  $r - 1$ , entonces al menos uno de los enteros es mayor o igual que  $r$ , es decir contiene  $r$  objetos.”

$$\frac{1}{k} \sum_{i=1}^k z_i > r - 1 \rightarrow \exists z_i \in Z / z_i \geq r$$

## Apéndice V: Scripts en R (correo: xxix.alexander@gmail.com)

Cambio de variable de la base de datos

```
VSM<-na.exclude(VSM) ## excluye observaciones faltantes
```

```
RSAM<-VSM$RSAM
```

```
Tiempo<-VSM$FECHA
```

Escala del plano cartesiano

```
plot(c(1,9),c(1,100), type="n")
```

Gráfico secuencial

```
plot(Tiempo, RSAM, type=c("h"), col="black", asp=3, xlab="Tiempo (h)",  
ylab="RSAM", main="Gráfico RSAM VSM")
```

Observaciones de referencia

```
abline(v=6561, col="yellow", lwd=2) ## línea vertical
```

```
abline(h=500, col="yellow", lwd=2) ## línea horizontal
```

```
abline(3, 6, col="yellow", lwd=2) ## línea oblicua intercepto-pendiente
```

```
vector<-c(-2.512*10^5,1.812*10^-4) ## vector con el intercepto y la pendiente
```

```
abline(coef=vector , col="yellow", lwd=2) ## alternativa intercepto-pendiente
```

```
lines(Tiempo,RSAM, col="yellow", lwd=2) ## agrega líneas a un gráfico actual
```

Leyenda de los gráficos

```
legend("topleft", c("Serie original", "Lambda=0.50", "Lambda=0.75", "Lambda=0.90",  
"Lambda=0.99", "Recta de tendencia"),
```

```
fill=c("black","red","green","blue","cyan","yellow"))
```

```
title("Gráfico del RSAM por hora") ## agrega título a un gráfico existente
```

Colores

```
col=rainbow(3) ##arcoiris
```

```
den<-densCols(x) ##se usa con col=
```

```
## se usa con color=Lab.palette
```

```
rgb.palette <- colorRampPalette(c("red", "orange", "blue"), space = "rgb") ## apariencia  
simétrica
```

```
Lab.palette <- colorRampPalette(c("red", "orange", "blue"), space = "Lab") ## suavizado  
uniforme
```

```
require(grDevices) ## librería esencial
```

```
color.palette=terrain.colors
```

```
## Como en matlab, se usa con color=jet.colors
```

```
require(grDevices) ## librería esencial
jet.colors <- colorRampPalette(c("#00007F", "blue", "#007FFF", "cyan", "#7FFF7F",
"yellow", "#FF7F00", "red", "#7F0000"))
```

Kernels Binarios: regiones de confianza en 2D

```
require(gplots) ## librería esencial
ci2d(Tiempo, RSAM, show="contour", col=jet.colors(8), show.points=FALSE,
nbins=150)
identify(RSAM) ## analizar observaciones
```

Estimación de la densidad con Kernel Binario 2D: el Kernel es la densidad normal bivariada estándar. (R, 2010, p. 1655-1657)

```
require(KernSmooth) ## librería esencial
```

```
estimación <- bkde(x, kernel="epanech", bandwidth=0.25, gridsize = 401L,
range.x=c(80,100)) ## "normal", "box", "epanech", "biweight", "triweight"
plot(estimación, type="l") ## Densidad de probabilidad de los datos univariantes
```

```
x <- cbind(VSA$FECHA, VSA$RSAM)
estimación <- bkde2D(x, bandwidth=c(0.7, 7))
contour(estimación$x1, estimación$x2, estimación$fhat, col=jet.colors(23)) ##
Densidad bivariada, puede usarse también filled.contour
persp(estimación$fhat, col=jet.colors(50))
image(estimación$fhat,col=jet.colors(50))
```

Fusión Densidad Kernel 2D (uso estimación\$fhat) & regiones de confianza (idea propia: observar los kernels de referencia con alta probabilidad de ocurrencia y los que tienen baja probabilidad, usualmente las anomalías)

```
ci2d(estimación$x1, estimación$fhat, show="filled.contour", col=jet.colors(10))
ci2d(estimación$x2, estimación$fhat, show="filled.contour", col=jet.colors(10))
ci2d(estimación$fhat, estimación$fhat, show="filled.contour", col=jet.colors(10)) ##
dimensión de la concentración de normalidades versus anomalías
```

Función de Densidad No Paramétrica sm: métodos de suavizado para la estimación de la función de densidad no paramétrica. Métodos Kernel de suavizado no paramétrico (Bowman & Azzalini, 1997), aplicados a datos binomiales, series temporales, análisis de supervivencia, entre otros.

```
require(sm) ## (Bowman & Azzalini, 1997)
```

```
vector<-cbind(rnorm(50)) ## 1D
```

```

vector<-cbind(rnorm(50), rnorm(50)) ## 2D
vector<-cbind(rnorm(50), rnorm(50), rnorm(50)) ## 3D

h.select(VSM) ## parámetro de suavizado óptimo para la función de densidad
sm.density(RSAM, model="Normal") ## en 1D
sm.density(RSAM, panel=TRUE) ## en 1D con panel de control
sm.density(VSM, display = "image") ## en 2D con imagen bits
sm.density(VSM, display = "contour") ## en 2D con contorno
sm.density(VSM, display = "contour", col="blue") ## en 2D con contorno y puntos
sm.density(VSM, panel = TRUE) ## en 3D (Tiempo, RSAM, probabilidad)

require(misc3d)
sm.density(VSM) ## en 3D (un vector con tres variables)

sm.ts.pdf(RSAM,lags=1:3) ## Estimación de la función de densidad de una serie
temporal, con el supuesto de estacionalidad. La densidad marginal univariante se estima
en todos los casos.

density() ## alternativa es la opción genérica, también utiliza el método Kernel

require(lattice) ## librería esencial
densityplot(VSA$RSAM)

Función de auto-regresión y regresión sm: métodos de suavizado para la estimación de la función de regresión no paramétrica. Métodos Kernel de suavizado no paramétrico (Bowman & Azzalini, 1997), aplicados a datos binomiales, series temporales, análisis de supervivencia, entre otros. Técnicas de suavizado aplicadas para el análisis de datos: la aproximación Kernel. Oxford University Press, Oxford. Págs. 29/74 sm.pdf

require(sm) ## (Bowman & Azzalini, 1997)

sm.autoregression(RSAM, h=hnorm(RSAM), maxlag=3, se=TRUE) ## Np ARF
condicional
sm.autoregression(RSAM, h=hnorm(RSAM), d=1, maxlag=3, se=TRUE) ## Np ARF
condicional

sm.regression(Tiempo, RSAM, panel=TRUE) ## Regresión no paramétrica

Regresión polinomial local (R, 2010, p. 1663)
x <- VSA$FECHA
y <- VSA$RSAM

```

```
plot(x, y)
h <- dpill(x, y) ## para el método Kernel se utiliza la función dpik(x), bandwidth.nrd(x)
para la selección del ancho de banda vía distribución de referencia normal, bcv(x,
nb=1000) para una validación cruzada parcial, nb=nbins; ucv(x, nb=1000) para una
validación cruzada imparcial, width.SJ(x,nb=1000) para el método Sheather & Jones
```

```
ajuste <- locpoly(x, y, bandwidth = h) ## graficar el ajuste con lines(ajuste)
```

#### Simulación MonteCarlo

```
y <- rnorm(50)
RSAM<-rnorm(10000,mean=96.10159,sd=129.8002) ## distribución normal
RSAM<-rexp(10000,rate=96.10159) ## distribución exponencial
```

#### Regresión Kernel

```
plot(Tiempo,RSAM,type="h")
h<-dpill(Tiempo,RSAM)
h
```

```
lines(ksmooth(Tiempo,RSAM,"normal",bandwidth=h),col="blue",lwd=3)
```

#### Regresión Loess

```
plot(Tiempo,RSAM,type="h")
vul<-loess(RSAM~Tiempo)
lines(predict(vul,data.frame(Tiempo=seq(1,10000,1))),col="blue",lwd=2)
```

#### Espectrograma

```
spectrum(RSAM,spans=100) ## suavizado
spectrum(RSAM,log="dB", ci=0.95, ci.col="red", ci.lty=1) ## dB significa decibel
```

#### Función de Autocorrelación

```
acf(RSAM,lag.max=NULL,type=c("correlation","covariance","partial"),plot=TRUE)
acf(RSAM,lag.max=24 ,plot=TRUE, ci.type="ma") ## ó white
```

#### Curvas de Andrews

```
r=57.17/(2^(1/2))+52.67*sin(t)+64.5*cos(t)+57.17*sin(2t)+59.33*cos(2t)
```

#### Gráfica de los modelos estadísticos

```
curve(-9*10^(-5)*x*x+1.1494*x-3.4004*10^3-1.479*10, 0, 9000, n = 101, add =
TRUE, type = "l", lwd=2, col="red", lty=1)
```

#### Gráfica de la pendiente Loess de los Kernels

```
vul<-loess(RSAM~Tiempo, subset=c(Tiempo>8000, Tiempo<8699))
line(vul)
```

## Apéndice VI: Constancia de Tesis UES



UNIVERSIDAD DE EL SALVADOR  
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE  
DEPARTAMENTO DE MATEMÁTICA



LA INFRASCRIPTA JEFA DEL DEPARTAMENTO DE MATEMÁTICA DE LA FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE-UES, POR ESTE MEDIO HACE CONSTAR QUE:

El Profesor **Edwin Alexander Guzmán Martínez** Carnet GM06008, es estudiante egresado de la carrera de **Licenciatura en Estadística** que sirve el departamento de Matemática, y ha inscrito en fecha 20/03/2014, su Trabajo de Grado titulado: "**Aplicación de Modelos Estadísticos para la interpretación de señales SISMICO-VOLCÁNICAS en los principales complejos volcánicos de El Salvador**", con la asesoría del Docente Director: Licdo. Juan Haroldo Linares Martínez.

Y para los usos que el interesado estime convenientes se extiende la presente constancia a los cinco días del mes de Junio de dos mil quince.

HACIA LA LIBERTAD POR LA CULTURA

f.   
Licda. Sara Vilma Rodríguez de Chica  
Jefa del Departamento de Matemática  
Coordinadora de Procesos de Graduación del  
Depto. de Matemática.



Visto bueno: f.   
Licdo. Raúl Ernesto Azuela  
Decano de la F.M.Occ.-UES



Cc: Archivo.  
SVR/Rutli

## Apéndice VII: Solicitud de información al Observatorio Ambiental



UNIVERSIDAD DE EL SALVADOR  
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE  
DEPARTAMENTO DE MATEMÁTICA



Santa Ana, 18 de Junio de 2015.

Ing. Celina Kattán  
Directora del Observatorio Ambiental  
Ministerio de Medio Ambiente y Recursos Naturales  
San Salvador

Estimada Ing. Kattán:

Reciba un cordial saludo además de éxitos en sus labores diarias y en su vida.

La infrascrita jefa del Departamento de Matemática, por este medio hace constar:

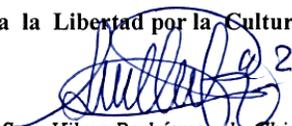
Que el alumno *Guzmán Martínez, Edwin Alexander*, carnet *GM06008* es estudiante egresado de la carrera Licenciatura en Estadística, de esta institución y que en este momento se encuentra realizando su Trabajo de Grado con una investigación científica como requisito para optar al grado de Licenciado en Estadística.

Su investigación científica se denomina “*Aplicación de modelos estadísticos para la interpretación de señales sísmico-volcánicas en los principales complejos volcánicos de El Salvador*”. Dada la naturaleza de su estudio es necesario que el alumno pueda contar con las variables de origen volcánico y sísmico.

Por lo anterior, se le solicita de la manera más atenta, proporcionar las bases de datos de las variables cuantitativas: *RSAM, frecuencia sísmica, temperatura, CO<sub>2</sub>, SO<sub>2</sub>, mediciones de las velocidades de los vientos, entre otras que se dispongan y ayuden a entender científicamente al fenómeno del vulcanismo*, que comprendan un periodo de los últimos diez años, para que el alumno no encuentre inconvenientes en el desarrollo de su investigación.

Sin nada más que agregar y esperando la colaboración de la institución que Ud. Coordina, me despido.

“Hacia la Libertad por la Cultura”

  
Licda. Sara Vilma Rodríguez de Chicas  
Jefa del Departamento de Matemática



Cc: Archivo.  
SVR/Ruth

Conmutador 2484-0800, Fax 2449-0352. Depto. de Matemática Teléfono Directo 2484-0870  
Final Ave. Fray Felipe de Jesús Moraga Sur, Santa Ana, El Salvador, Apdo. Postal 1908.  
[www.uesocc.edu.sv/departamentos/matemática](http://www.uesocc.edu.sv/departamentos/matemática)

## Apéndice VIII: Entrevista MARN 2015

Universidad de El Salvador

Facultad Multidisciplinaria de Occidente

Departamento de Matemática

\*\*\*\*\*

### Guía de entrevista para especialistas en el área de vulcanología perteneciente al Ministerio de Medio Ambiente y Recursos Naturales de El Salvador – junio 2015

**Tesis:** “Aplicación de modelos estadísticos para la interpretación de señales sísmico – volcánicas en los principales complejos volcánicos de El Salvador”

**Vulcanismo:** conjunto de fenómenos relacionados con las erupciones volcánicas y con la expansión de los magmas por la superficie terrestre

**Objetivo:** recopilar datos fidedignos concernientes a la actividad volcánica de El Salvador para generar información científica

**Indicaciones:** conteste honestamente cada pregunta perteneciente a esta entrevista dirigida al personal técnico del MARN, ya que de ello depende la validez de los resultados de esta investigación

#### 1. ¿Qué es el vulcanismo?

*Es una serie compleja de procesos geológicos que incluye el origen y el ascenso del magma a través del manto y la corteza del planeta y su manifestación (erupción) en la superficie. Donde intervienen dinámicas físicas y química de los magmas, transporte, y procesos eruptivos.*

#### 2. ¿Qué es el seísmo, o sismicidad?

*La sismicidad se refiere a la distribución geográfica e histórica de los sismos. Estudiando la propagación de las ondas elásticas en los estratos que conforman la corteza. El sismo representa los principales tipos de fuentes que generan ondas sísmicas que son oscilaciones debidas a las deformaciones elásticas que se propagan a través de la Tierra. Un sismo puede tener su origen de forma natural o creada por el hombre.*

#### 3. ¿Qué relación estrecha guardan el vulcanismo y la sismicidad?

*Los procesos dinámicos ocurridos durante las inyecciones de magma causan sismos de origen vulcano-tectónicos, mientras los procesos debido a movimiento de fluidos (magma, gases, vapor) en los conductos del sistema interior del volcán ocasionan la ocurrencia de sismos volcánicos los cuales tiene características espectrales que varían según su génesis, profundidad y magnitud.*

4. ¿Qué es una variable sísmica - volcánica?

*El estudio y análisis de los datos sísmicos aplicados a zonas volcánicas incluye idealmente la detección y registro de las señales, su identificación y clasificación para realizar un análisis cuantitativo de éstas. El objetivo de esta práctica es poder obtener información de numerosos aspectos, como los parámetros de la fuente y el desplazamiento de las ondas a través del medio de propagación con el objetivo de conocer el status de la actividad de un volcán en particular.*

5. ¿Qué variables cuantitativas y cualitativas están presentes en las prácticas de monitoreo continuo en los principales complejos volcánicos de El Salvador?

*Conteo y clasificación de sismos, medición de la amplitud sísmica (RSAM), flux de emisión de gases (SO2 toneladas/día), se tienen entre las variables cuantitativas continuas. La permanente evaluación del nivel de actividad del volcán es la variable cualitativa*

6. ¿Cómo (descripción del proceso) se obtienen estas variables?

*En la recolección de las variables sísmicas y de emisión de gases, se cuenta con una red de monitoreo permanente ubicada en las cercanías de los volcanes los cuales colectan los datos, estos son enviados a tiempo real vía telemétrica o internet y son procesados por análisis informáticos para generar el datos que luego pasa a la etapa de tabulación e interpretación*

7. ¿Qué dispositivos o máquinas se emplean para la recopilación de las variables sísmico – volcánicas?

*Estaciones sísmicas telemétricas o digitales de periodo corto (en su mayoría), de un componente o triaxiales para la adquisición de datos. Sistema informático para el procesamiento de las trazas sísmicas.*

8. ¿Alguna vez se han cometido fallas o errores de parte del equipo en el análisis volcánico? ¿Como cuáles?

*Fallas fortuitas en el sistema informático, tanto en PC, como en servidores*

9. ¿Alguna vez se han emitido falsas alarmas de posibles catástrofes, causa de desestabilización territorial?

*En el área de vulcanología no se han emitido falsas alarmas que hayan llevado a evacuaciones innecesarias*

10. ¿Cuál es el papel y el peligro del vulcanólogo en la recopilación de datos?

*La recopilación de los datos provenientes del monitoreo instrumental registrado por la red de estaciones de monitoreo es enviada en tiempo real vía telemétrica o internet al centro integrado de monitoreo donde se procesan los datos. La recopilación de los datos in-situ implica la visita a zonas volcánicas ya sean cratéricas o periféricas volcánicas con el consecuente riesgo que estas zonas representan (derrumbes, altas temperaturas, ambientes tóxicos, etc.)*

11. ¿Cuándo (factor tiempo y periodicidad) se hacen esos monitoreos?

*El monitoreo instrumental, tal como el sísmico y medición de gases se hace de forma continua. El monitoreo térmico, hidrogeoquímico y visual se realiza de forma discreta según el estado de la actividad que presente un volcán en particular.*

12. ¿Por quiénes está formado el equipo de monitoreo sísmico-volcánico?

*Por los miembros de las coordinaciones de sismología (1 coordinador, 2 sismólogos, 4 técnicos) y vulcanología (1 coordinador, 1 especialista en vulcanología, 1 vulcanólogo)*

13. ¿Según las políticas de un gobierno con transparencia, qué métodos utilizan para informar y prevenir a la población salvadoreña sobre posibles catástrofes causadas por erupciones volcánicas?

*Se siguen los protocolos de actuación que de acuerdo a la actividad volcánica pueden ser 5: Monitoreo, Pre-aviso-Aviso-Alerta y Emergencia.*

14. ¿Qué variables son las más idóneas para el análisis volcánico?

*Se puede resumir que la **actividad volcánica** es la principal variable a considerar (variable cualitativa ordenada). Esta se obtiene de la evaluación de los datos colectados a través del monitoreo diario (variables cuantitativas)*

15. ¿Cuál es el rango normal de dichas variables?

*El rango de las variables (tanto cualitativa como cuantitativa), dependerá del tipo de volcán que se está estudiando, es decir, dependiendo si las características magmáticas son básicas o ácidas tendrán comportamientos distintos, esto es debido a la cantidad de gases y viscosidad de los magmas que están presentes en el proceso volcánico. Es así como los rangos de las variables cuantitativas puedan ser distintos en cada volcán. De igual forma la configuración de las redes de monitoreo y el equipo utilizado para su colección*

*juegan un papel importante en el rango de las variables. Como ejemplo citamos, que una estación de periodo corto analógica los valores de RSAM estarían en el rango de 0 a 1800 unidades. Datos de temperatura podrían variar entre 25°C (temperatura ambiente) a aproximadamente 1200°C cuando se tiene la presencia de magma en superficie.*

16. Describa el equipo informático utilizado por la institución para analizar los datos recogidos.

*PC, con sistemas operativos Windows, UNIX y LINUX discos externos de almacenamiento de datos*

17. ¿Qué sistemas informáticos son utilizados por la institución para analizar los datos obtenidos?

*Para la ubicación de sismos se utilizan software HYP071, Hypocenter y Hypoinverse a través del programa **SEISAN ver 9.0.1. Earthworm versiones 6.3 y 7.4** en sistema Windows para computar RSAM (medición de la amplitud sísmica), SSAM (datos de espectrales de la señal sísmica) y para la creación de imágenes de sismogramas y espectrogramas. Análisis de las características se las señales sísmicas se realiza a través del software Seismic Wave Analysis Real-time Monitoring Version: 2.3.4.2 (**SWARN**). La evaluación del flux de dióxido de azufre se realiza con el sistema **NOVAC program ver 2.0**. Imágenes infrarrojas se analizan con **FLIR tolos ver 5.2.15161.1001***

18. ¿Qué métodos de análisis se aplican a los datos en crudo?

*Límites de detección, rangos, clasificación de frecuencias*

19. ¿Existe algún método empírico que se aplique a los datos?  
¿Cuál?

*La no certeza en los procesos que generan los datos recolectados, obligan en la vulcanología al establecimiento de métodos empíricos tales como la observación, medición y experimentación, entre otros.*

20. ¿Existe algún método científico (experimental) que se aplique a los datos? ¿Cuáles?

*El análisis de los datos permite plantear la génesis de la fuente que lo está originando, planteándose de esta manera algunas hipótesis de su comportamiento, las cuales deben ser probadas.*

21. ¿En el área de vulcanología, qué significa “pronosticar”?

*El pronóstico, se realiza en base a la probabilidad de ocurrencia de un futuro eventos, según la información que se obtiene a través de la vigilancia volcánica.*

*Cabe mencionar que no toda intranquilidad en el sistema volcánico culmina siempre en una erupción.*

22. ¿En el área de vulcanología, qué significa “predecir”?

*La predicción, tal como lo indica su significado, no se hace en vulcanología. Sería muy atrevido y poco ético señalar día, hora y minuto de un futuro evento eruptivo*

23. ¿Cuál es la importancia de los procesos estocásticos en la vulcanología?

*Un proceso estocástico es muy importante en la vulcanología ya que logra representar en detalle las operaciones que conforman el monitoreo, la simulación probabilística de zonas afectadas por depósitos volcánicos. En muchos casos la creación de un árbol probabilístico de eventos es de suma importancia en la toma de decisiones.*

24. ¿Los especialistas aplican técnicas de inferencia? ¿Cuáles?

25. ¿Los especialistas aplican técnicas estadísticas? ¿Cuáles?  
*Estadística descriptiva, Mínimos Cuadrados, Pareteo, regresión, muestreo*

26. ¿Los especialistas aplican técnicas matemáticas? ¿Cuáles?  
*Analiza de series de Fourier, métodos numéricos de simulación*

27. ¿Qué tipos de modelos (interferenciales, estadísticos, matemáticos) aplican para generar predicciones y pronósticos?

*Método para determinar volumen y masa de productos eruptivos tales como Weibull-method, Exponential y Power Law*

28. ¿Han ocurrido descubrimientos de patrones o modelos que describan el comportamiento de las variables a lo largo del tiempo?

*Hasta el momento no se han determinado patrones de comportamiento, recordemos que la base instrumental utilizada únicamente se ha tenido dos erupciones en volcanes distintos*

29. ¿Qué bibliografía recomienda para la aplicación de modelos o patrones?

## Apéndice IX: Biografía de Jean Baptiste Fourier



Nació en Auxerre, Francia, el 21 de marzo de 1768 y huérfano a la edad de ocho años, Jean Baptiste Joseph Fourier ingresó en 1789 en la Ecole Royale Militaire de Auxerre, donde mostró talento para la literatura, pero las matemáticas fueron su interés principal. A los 14 años había completado el estudio de los seis volúmenes del Cours de mathematique de Bézout. En 1787 solicitó su ingreso en la abadía Benedictina para consagrarse a la religión, aunque su interés por las matemáticas permanecía intacto. Su correspondencia sugiere que su propósito era destacarse en esta ciencia: es famosa una carta suya a Bonard donde anota: “Ayer fue mi vigésimo primer cumpleaños. A esta edad Newton y Pascal habían hecho méritos suficientes para alcanzar la inmortalidad”.

Dejó el monasterio benedictino en 1789 y se dirigió a París, donde presentó un trabajo sobre ecuaciones algebraicas en la Academia Real de Ciencias. En 1790 tomó un puesto de profesor en el colegio Benedictino en la Escuela Real Militar de Auxerre, donde había estudiado. Siempre había dudado entre dedicarse a la religión o a las matemáticas. Sin embargo, este conflicto se resuelve en 1793, cuando se involucra en política, donde se unió al Comité Revolucionario local. Fourier abrazó los ideales de la Revolución, aunque no apoyó el Terror.

Fourier defendió a una facción de Orleáns, y a consecuencia de ello estuvo en prisión un tiempo, con riesgo de terminar en la guillotina. Finalmente, los cambios políticos acaecidos después de la muerte de Robespierre resultaron en su liberación. En 1795 fue admitido en la Ecole Normale de París, creada para instruir a maestros, y modelo de otras escuelas normales. Allí fue alumno de Lagrange, quien, según Fourier, era “el primero entre los hombres de ciencia de Europa” y de Laplace, a quien no admiraba tanto. Luego de egresar, fue maestro en la Ecole Polytechnique, recién creada. A consecuencia de su arresto previo nuevamente fue a prisión, de donde pudo ser liberado gracias a las súplicas de Lagrange, Laplace, Monge, sus alumnos, y un cambio de clima político.

En 1798 se une a la Armada de Napoleón en la invasión de Egipto, como un consejero científico. Luego de una serie de rápidas victorias fue tomado el delta del

Nilo, pero la destrucción completa de la flota napoleónica por la armada de Nelson en la batalla del Nilo confinó a la expedición francesa a las tierras que estaba ocupando. Fourier actuó como un administrador de instituciones políticas similares a las que existían en Francia, y ayudo a establecer establecimientos educacionales en Egipto, además de efectuar diferentes exploraciones arqueológicas. Creó el Instituto de El Cairo, donde fue uno de los doce miembros de la división matemática. Fue elegido secretario del Instituto, cargo que ejerció durante todo el período de la ocupación francesa.

Napoleón volvió a Francia en 1799, y Fourier volvió a París en 1801 con el resto de la fuerza expedicionaria, retomando su puesto de profesor de análisis en la Ecole Polytechnique. Sin embargo, Napoleón tenía otros planes para él, y lo nombró Prefecto del Departamento de Isère, cargo que Fourier no deseaba, pero que no podía rechazar. Se mudó a Grenoble, y sus principales trabajos en esa posición administrativa fueron el drenado de los esteros de Bourgoin, y la supervisión de la construcción de una carretera entre Grenoble y Turín. También invirtió gran parte de su tiempo en la escritura de su Descripción de Egipto, que no estuvo concluida hasta 1810 debido a que Napoleón introdujo muchos cambios, reescribiendo algunas partes.

Para la época de una segunda edición, toda referencia a Napoleón había sido eliminada. Durante su estancia en Grenoble Fourier realizó un importante trabajo matemático sobre la teoría del calor. Comenzó a trabajar en este tema alrededor de 1804, y para 1807 había completado su importante memoria Sobre la propagación del calor en los sólidos. Esta memoria fue leída en el instituto de París el 21 de diciembre de 1807, y aunque ahora es muy estimada, en su momento causó controversia, siendo comentada por Lagrange, Laplace, Monge y Lacroix. Este comité no estaba de acuerdo con las expansiones de funciones en base a series trigonométricas (lo que hoy conocemos como series de Fourier).

Las aclaraciones de Fourier aún no lograron convencer al comité. Según autores modernos, la claridad lógica de las explicaciones de Fourier y la inhabilidad de Laplace y Lagrange para comprenderlas da una buena idea de la originalidad de su trabajo. Otro tipo de objeciones fue elevado por Boit, ya que Fourier no hacía mención al trabajo del

primero, publicado en 1804. Laplace y posteriormente Poisson tenían objeciones similares. Para el Premio Matemático de 1811, el Instituto propuso el tema de la propagación del calor en los sólidos, y Fourier envió su memoria de 1807, junto con material adicional sobre el enfriamiento de los sólidos infinitos. Sólo se presentó un manuscrito adicional, y el comité de evaluación se decidió finalmente por el trabajo de

Fourier, aunque su juicio no fue enteramente favorable: “la manera en que el autor arriba a estas ecuaciones no está a salvo de dificultades, y su análisis para integrarlas deja aún algo de desear, sea en su generalidad o en su rigor”. Ante este juicio dudoso no hubo interés en publicar el trabajo de Fourier. Durante la época de prefecto en Grenoble tuvo una serie de encuentros y desencuentros con Napoleón; finalmente, cuando Napoleón cayó, Fourier volvió a París, donde fue electo miembro de la Academia de Ciencias en 1817, y Secretario de la sección matemática en 1822. En ese año se publica finalmente su ensayo *Théory analytique de la chaleur*, cuya impresión había sido aprobada por el Secretario anterior.

Durante sus últimos años en París Fourier retomó sus investigaciones matemáticas, y publicó una serie de trabajos tanto en matemáticas puras como aplicada. Su vida no transcurrió sin problemas ya que su teoría del calor todavía provocaba controversias. Biot reclamaba prioridad sobre Fourier, una queja que Fourier pudo mostrar fácilmente que era falsa. Sin embargo, Poisson atacó las técnicas matemáticas y propuso una teoría alternativa. Fourier escribió *Historical Précis* como respuesta a estas críticas, pero aunque el trabajo fue mostrado a varios matemáticos, nunca fue publicado. Sin embargo, la obra de Fourier proveyó el ímpetu necesario para los trabajos posteriores en series trigonométricas y la teoría de funciones de una variable real.

Fourier vivió obsesionado por el calor, hasta el punto de mantener sus habitaciones insoportablemente calurosas para sus visitantes, mientras que él vestía un pesado sobretodo. Algunos autores rastrean esta excentricidad hasta sus tres años en Egipto. Falleció por un aneurisma el 16 de mayo de 1830.

## Apéndice X: Biografía de Johann Peter Gustav Lejeune Dirichlet



Nació en Düren, actual Alemania, el 13 de febrero de 1805. Su familia era natural del pueblo de Richelet en Bélgica, donde vivió su abuelo, y de donde derivó su apellido “Lejeune Dirichlet” (“le jeune de Richelet” = “el joven de Richelet”). Su padre era el jefe de la oficina de correos de Duren. El joven Johann fue educado en Alemania, y después en Francia, donde aprendió de muchos de los más renombrados matemáticos del tiempo, relacionándose con algunos como Fourier. Tras graduarse, fue profesor en las universidades de Breslau (1826-1828), Berlín (1828-1855) y Göttingen, en donde ocupó la cátedra dejada por Gauss tras su muerte. Sus aportaciones más relevantes se centraron en el campo de la teoría de números, prestando especial atención al estudio de las series, y desarrolló la teoría de las series de Fourier.

Su primera publicación comprendió una demostración particular del teorema de Fermat, para el caso  $n = 5$ , que también fue completada por Adrien-Marie Legendre, uno de sus revisores. Dirichlet completó su propia prueba casi al mismo tiempo; más adelante completó también la prueba para  $n = 14$ . Aplicó las funciones analíticas al cálculo de problemas aritméticos y estableció criterios de convergencia para las series. En el campo del análisis matemático perfeccionó la definición y concepto de función, y en mecánica teórica se centró en el estudio del equilibrio de sistemas y en el concepto de potencial newtoniano. Se casó con Rebecka Mendelssohn, que provenía de una distinguida familia de judíos conversos. Era la nieta del filósofo Moses Mendelssohn, hija de Abraham Mendelssohn Bartholdy y hermanadel compositor Felix Mendelssohn Bartholdy. Dirichlet falleció en Göttingen, actual Alemania, el 5 de mayo de 1859.

## Apéndice XI: Biografía de Johann Carl Friedrich Gauss



Nació el 30 de abril en Brunswick, Alemania. Comenzó la escuela elemental a la edad de 7 años, donde demostró su potencial casi de manera inmediata. Su maestro Büttner y su asistente Bartels quedaron sorprendidos cuando Gauss sumó los enteros de 1 al 100 casi instantáneamente ordenando la suma en 50 pares de números que sumaban 101. En 1788 comenzó su educación en el Gymnasium, donde aprendió Latin. En 1792 ingresó al Colegio Carolingio en 1792, donde descubrió el teorema del binomio, las medias geométricas y aritméticas, la ley de reciprocidad cuadrática, y el teorema de los números primos.

En 1795 ingresó a la Universidad de Göttingen, la que abandonó sin haber conseguido un título, pero habiendo hecho uno de sus descubrimientos más importantes: la construcción de un polígono regular de 17 lados utilizando regla y compás. Este fue el avance más importante en este campo desde la época de los matemáticos griegos. Retornó a Brunswick donde recibió un grado en 1799. Su tesis doctoral fue una discusión sobre el teorema fundamental del álgebra. Publicó su primer libro, *Disquisitiones Arithmeticae* en el verano de 1801. En 1807 se hizo cargo de la dirección del observatorio de Göttingen.

En 1809 publicó su segundo libro, *Theoria motus corporum coelestium in sectionibus conicis Solem ambientium*, un tratado en dos volúmenes sobre el movimiento de los cuerpos celestes. Sus contribuciones a la astronomía teórica concluyeron después de 1817, aunque siguió efectuando observaciones hasta sus 70 años de edad. Gauss dedicó mucho tiempo a la construcción del nuevo observatorio, finalizado en 1816, pero fue capaz de trabajar en otros temas. Sus publicaciones durante este período incluyen *Disquisitiones generales circa seriem infinitam*, un tratamiento riguroso de las series y una introducción a las funciones hipergeométricas, *Methodus nova integralium valores per approximationem inveniendi*, un ensayo práctico sobre integración aproximada, *Bestimmung der Genauigkeit der Beobachtungen*, una discusión de estimadores estadísticos, y *Theoria attractionis corporum sphaeroidicorum ellipticorum homogeneorum methodus nova tractata*.

Este último trabajo estuvo inspirado por su trabajo en problemas geodésicos y trataba sobre teoría potencial. En 1818 Gauss fue designado para efectuar un reconocimiento geodésico del estado de Hannover, tarea que desarrolló personalmente, haciendo mediciones durante el día y reduciéndolas a la noche, haciendo uso de su extraordinaria capacidad mental para los cálculos. Para facilitar las mediciones inventó el heliotropo, que refleja los rayos del sol utilizando un diseño de espejos y un telescopio pequeño. A pesar de este demandante trabajo publicó 70 trabajos sobre éstos y otros temas entre 1820 y 1830. En 1822 ganó el premio de la Universidad de Copenhagen por *Theoria attractionis...* junto con la idea de mapear una superficie en otra de manera que fuesen similares en sus mínimos detalles.

El artículo fue publicado en 1825, y condujo a la publicación muchos años después de *Untersuchungen über Gegenstände der Höheren Geodäsie* (1843 y 1846). El artículo *Theoria combinationis observationum erroribus minimis obnoxiae* (1823), junto con u suplemento (1828) estuvo dedicado a la estadística matemática, en particular al método de cuadrados mínimos. Desde 1800 Gauss investigó la posible existencia de geometrías no euclídeas, discutiendo ampliamente este tema con Farkas Bolyai. En 1816 presentó pruebas que deducían el axioma de las paralelas a partir de otros axiomas de Euler, sugiriendo que creía en la existencia de geometrías no euclídeas, aunque no fue claro al respecto; posteriormente se sinceró con Schumacher diciéndole que creía que su reputación sufriría si admitía en público la existencia de tal geometría.

Tuvo gran interés en geometría diferencial y publicó muchos trabajos sobre el tema. *Disquisitiones generales circa superficies curva* (1828) fue su trabajo más reconocido en este campo. En 1831 Wilhelm Weber llegó a Göttingen donde ejerció como profesor de física. Gauss, quien lo conocía desde 1828 y apoyó su contratación, había trabajado en física antes de 1831 publicando *Über ein neues allgemeines Grundgesetz der Mechanik*, y *Principia generalia theoriae figurae fluidorum in statu aequilibrii* donde discutía las fuerzas de atracción. Estos artículos estaban basados en la teoría de potencial, que probó ser de gran importancia para su trabajo en física. Creyó que su teoría y el método de cuadrados mínimos proporcionaban vínculos vitales entre la ciencia y la naturaleza.

En 1832 Gauss y Weber comenzaron a investigar la teoría del magnetismo terrestre después que Alexander von Humboldt intentó obtener la ayuda de Gauss para realizar una grilla de puntos de observación magnética sobre todo el mundo. Gauss se entusiasmó con este proyecto, y para 1840 había escrito tres trabajos importantes sobre el tema, demostrando, entre otras cosas, que sólo podía haber un par de polos magnéticos, y brindando una estimación de la ubicación del polo sur magnético. Humboldt había diseñado un procedimiento para la observación de la declinación magnética, pero luego que su nuevo observatorio magnético fuese construido en 1833, libre de cualquier tipo de material magnético, Gauss modificó muchos de los procedimientos de Humboldt, lo que permitió obtener resultados más precisos con menos esfuerzo. Entre los logros de Gauss y Weber durante los seis años que trabajaron juntos, merecen citarse el descubrimiento de las leyes de Kirchoff, y un telégrafo primitivo que podía enviar mensajes a distancias de 1,5 km. Sin embargo esto era apenas un pasatiempo para Gauss, y su interés mayor era establecer una red mundial de puntos de observación magnética.

Esta ocupación produjo muchos resultados concretos, incluyendo la publicación de un atlas de geomagnetismo, y una revista donde publicaron sus trabajos desde 1836 a 1841. Luego de la partida de Weber en 1837 la actividad de Gauss fue disminuyendo. Aún producía pequeñas notas en respuesta a los descubrimientos de otros científicos, usualmente resaltando que conocía esos métodos por años pero no que no había sentido la necesidad de publicar. En 1845-1851 se dedicó a consolidar los fondos de la Fundación de las viudas de Göttingen, lo que le dio experiencia práctica en aspectos financieros, y le permitió consolidar una importante fortuna personal invirtiendo en acciones emitidas por compañías privadas. Desde 1850 en adelante su trabajo fue de naturaleza práctica, aunque aprobó la tesis doctoral de Riemann; en 1854 discutió una modificación al péndulo de Foucault. Su salud decayó paulatinamente, y murió mientras dormía en la mañana del 23 de febrero de 1855.

## Apéndice XII: Biografía de James Cooley



Nació en EE.UU. en 1926. Se graduó en 1949, y recibió su doctorado en 1961 en la Universidad de Columbia, Nueva York, en matemática aplicada. Fue uno de los programadores de la computadora de von Neumann en el Instituto de Estudios Avanzados de Princeton, desde 1953 hasta 1956. Trabajó en cálculos de mecánica cuántica desde 1956 hasta 1962 en el Instituto Courant, de Nueva York, cuando se unió al grupo de investigación del IBM Watson Research Center, Yorktown Heights, NY. Su contribución más importante al mundo de las matemáticas y el procesamiento digital de señales fue el desarrollo de la FFT, realizado en conjunto con John Tukey mientras trabajaba para la división de investigaciones de IBM en 1965.

Este trabajo fue impulsado por el Dr. Richard L. Garwin de IBM Watson Research quien trataba de averiguar el cumplimiento del tratado de armas nucleares con los soviéticos (acuerdos SALT). La idea era instalar sensores sísmicos en los países vecinos a la Unión Soviética y detectar las vibraciones producidas por los ensayos nucleares, supuestamente abolidos. Estas vibraciones se analizaban con técnicas de Fourier, siendo necesario transformar lo más rápidamente posible un gran volumen de datos. Sugirió la idea de cómo se podían programar un algoritmo semejante a Cooley y Tukey, los que hicieron el trabajo. Se colocaron los sensores, se procesaron las señales, y se pudo detectar una explosión nuclear con una precisión de 15 kilómetros del lugar exacto. El trabajo que publicaron Cooley y Tukey en 1965 describiendo el algoritmo de la FFT disparó las aplicaciones del procesamiento digital de señales. Después de la publicación Cooley se preocupó de ayudar a otros a entender el algoritmo y su uso. Luego de retirarse de IBM en 1991 se vinculó con el Departamento de Ingeniería Eléctrica de la Universidad de Rhode Island, Kingston, donde sigue participando de proyectos de investigación en detección de señales.