

APLICACIÓN SOFTWARE DE APOYO EN LA IDENTIFICACIÓN DE
AGLOMERACIONES DE EVENTOS PARA LA TOMA DE DECISIONES. CASO
DE ESTUDIO EL DENGUE EN EL MUNICIPIO DE FLORIDABLANCA

ESTUDIANTES:
JONATHAN CONTRERAS VARGAS
OMAR YESID IBAÑEZ ORTIZ

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA
INGENIERÍA DE SISTEMAS
SANTANDER
BUCARAMANGA
2019

APLICACIÓN SOFTWARE DE APOYO EN LA IDENTIFICACIÓN DE
AGLOMERACIONES DE EVENTOS PARA LA TOMA DE DECISIONES. CASO
DE ESTUDIO EL DENGUE EN EL MUNICIPIO DE FLORIDABLANCA

ESTUDIANTES:

JONATHAN CONTRERAS VARGAS

OMAR YESID IBAÑEZ ORTIZ

TRABAJO DE GRADO PRESENTADO PARA OPTAR AL TÍTULO DE
INGENIERO EN SISTEMAS

DIRECTOR:

MIng., Esp. Ing. FEISAR ENRIQUE MORENO CORZO

UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA

INGENIERÍA DE SISTEMAS

SANTANDER

BUCARAMANGA

2019

TABLA DE CONTENIDO

TABLA DE CONTENIDO	3
TABLA DE ILUSTRACIONES.....	5
RESUMEN.....	7
ABSTRACT	8
INTRODUCCIÓN.....	9
1. PLANTEAMIENTO DEL PROBLEMA.....	10
JUSTIFICACIÓN.....	13
PREGUNTA(S) DE INVESTACIÓN	18
HIPOTESIS	19
OBJETIVOS	20
OBJETIVO GENERAL.....	20
OBJETIVOS ESPECIFICOS	20
2. MARCO REFERENCIAL	21
2.1. MARCO CONCEPTUAL Y TEÓRICO.....	21
2.1.1. SISTEMAS DE COORDENADAS.....	21
2.1.2. COORDENADAS CARTESIANAS.....	24
2.1.3. EL SISTEMA DE COORDENADAS UNIVERSAL TRANSVERSAL DE MERCATOR (UTM) .	30
2.1.4. LA FUNCIÓN K DE RYPLEY	34
2.1.5. EL MÉTODO KDE	37
2.1.6. ARCHIVOS KML (KEYHOLE MARKUP LANGUAGE).....	38
2.1.7. COLORES.....	40
2.2. ESTADO DEL ARTE	41
2.2.1. HERRAMIENTAS COMPUTACIONALES:.....	52
3. MATERIALES, INSTRUMENTOS, Y/O METODOLOGÍA	55
3.1. METODOLOGÍA DE DESARROLLO	55
3.2. DEFINICIÓN DE LA ARQUITECTURA DEL SOFTWARE	59
4. DISEÑO DE LA APLICACIÓN SOFTWARE	61
4.1. DIAGRAMA DE CASO DE USO.	61
4.2. DIAGRAMAS DE ACTIVIDAD	62
4.3. DIAGRAMAS DE SECUENCIA.....	68

4.4.	MOCKUPS.....	77
5.	RESULTADOS	83
5.1.	ALISTAMIENTO DE LOS DATOS	83
5.2.	ANÁLISIS DE RESULTADOS.....	84
5.2.1.	HALLAZGOS EN EL DESARROLLO E IMPLEMENTACIÓN DE LA APLICACIÓN SOFTWARE 85	
5.3.	ANÁLISIS DE RESULTADOS DE LA APLICACIÓN SOFTWARE DE LOS MÉTODOS IMPLEMENTADOS	93
5.3.1.	FUNCIÓN K DE RIPLEY.....	93
5.3.2.	MÉTODO KDE.....	97
5.4.	CONFRONTACIÓN DE RESULTADOS.....	100
6.	CONCLUSIONES.....	106
7.	TRABAJO FUTURO	107
8.	BIBLIOGRAFÍA.....	108
	ANEXO A. MANUAL DE USUARIO	112
A.1.	INTERFAZ CARGAR ARCHIVOS.....	112
A.2.	INTERFAZ FILTRO DE INFORMACIÓN	113
A.3.	INTERFAZ SELECCIÓN DE MÉTODO	115
A.4.	FORMULARIO INGRESO DE PARÁMETROS K DE RIPLEY	116
A.5.	INTERFAZ DE VISUALIZACIÓN RESULTADOS K DE RIPLEY	117
A.6.	FORMULARIO INGRESO DE PARÁMETROS KDE.....	118
A.7.	INTERFAZ VISUALIZACIÓN RESULTADOS KDE	119

TABLA DE ILUSTRACIONES

<i>Ilustración 1: Árbol de problemas.</i>	12
<i>Ilustración 2: Regiones más afectadas por el dengue en Colombia.</i>	14
<i>Ilustración 3: Reporte de casos de dengue clásico y dengue hemorrágico en Colombia entre el 2004 -2013. Realizado con base a la información de la base de datos SIVIGILA y del INS de Colombia.</i>	15
<i>Ilustración 4: Reporte de casos de muerte por dengue en Colombia entre 2004-2013, según la base de datos de SIVIGILA y del INS.</i>	16
<i>Ilustración 5: Valores de longitud y latitud de un punto sobre la esfera.</i>	23
<i>Ilustración 6: Elementos del sistema de coordenadas geográficas: paralelos (a), meridianos (b) y origen de coordenadas (c).</i>	23
<i>Ilustración 7: Ejes X, Y, Z del sistema de coordenadas cartesianas.</i>	25
<i>Ilustración 8: Correspondencia entre coordenadas geográficas y geocéntricas por un punto P.</i>	25
<i>Ilustración 9: Distribución de signos de las coordenadas X, Y en una malla del sistema de coordenadas proyectadas.</i>	27
<i>Ilustración 10: Proyección Mercator, utiliza un cilindro situado de forma tangente al elipsoide en el ecuador.</i>	31
<i>Ilustración 11: Proyección UTM, utiliza un cilindro situado de forma transversal respecto del eje de la tierra.</i>	32
<i>Ilustración 12: Cuadrícula UTM.</i>	33
<i>Ilustración 13: Grilla del sistema de coordenadas UTM empleada para cada zona.</i>	34
<i>Ilustración 14: Función K.</i>	36
<i>Ilustración 15: Kernel Density Estimator – KDE.</i>	38
<i>Ilustración 16: Representación gráfica común del modelo de color HSV.</i>	41
<i>Ilustración 17: Metodología primera parte.</i>	58
<i>Ilustración 18: Metodología segunda parte.</i>	59
<i>Ilustración 19: Diagrama de actividad, cargar archivos.</i>	63
<i>Ilustración 20: Diagrama de actividad: Filtrar información.</i>	64
<i>Ilustración 21: Diagrama de actividad, ingresar parámetros KDE.</i>	65
<i>Ilustración 22: Diagrama de actividad: ingresar parámetros K de Ripley.</i>	66
<i>Ilustración 23: Diagrama de actividad: ver resultado KDE.</i>	67
<i>Ilustración 24: Diagrama de actividad: exportar resultados.</i>	68
<i>Ilustración 25: Diagramas de secuencia, cargar archivos.</i>	69
<i>Ilustración 26: Diagrama de secuencia, filtrar información.</i>	70
<i>Ilustración 27: Diagrama de secuencia, ingresar parámetros K-Ripley.</i>	72
<i>Ilustración 28: Diagrama de secuencia, Ver resultados K de Ripley.</i>	73
<i>Ilustración 29: Diagrama de secuencia, ingreso de parámetros KDE.</i>	74
<i>Ilustración 30: Diagrama de secuencia, ver resultados KDE.</i>	75
<i>Ilustración 31: Diagrama de secuencia, exportar resultados.</i>	76
<i>Ilustración 32: Mockups, cargar archivos.</i>	77
<i>Ilustración 33: Mockups, Filtro de información.</i>	78
<i>Ilustración 34: Mockup, ingresar parámetros K de Ripley.</i>	79
<i>Ilustración 35: Mockup, ver resultado K de Ripley.</i>	80
<i>Ilustración 36: Mockup: Ingresar parámetros KDE.</i>	81
<i>Ilustración 37: Mockup, ver resultado KDE.</i>	82
<i>Ilustración 38: Ejemplo de resultado de la Función K de Ripley.</i>	90
<i>Ilustración 39: Ejemplo de visualización de resultados KDE.</i>	91
<i>Ilustración 40: Ejemplo de filtro de información.</i>	92
<i>Ilustración 41: Grupo de resultados de la función K de Ripley – dengue Floridablanca 2015.</i>	95
<i>Ilustración 42: Grupo de resultados del método KDE, dengue Floridablanca 2016.</i>	98
<i>Ilustración 43: Incidencia acumulada de dengue en.</i>	101
<i>Ilustración 44: Distribución del dengue en el mes.</i>	103

<i>Ilustración 45: Distribución del dengue en el mes de.....</i>	103
<i>Ilustración 46: Distribución del dengue en el municipio de Floridablanca, ancho de banda 100 metros.....</i>	104
<i>Ilustración 47: Distribución del dengue en el municipio de Floridablanca, ancho de banda 200 metros.....</i>	104
<i>Ilustración 48: Distribución del dengue en el municipio de Floridablanca, ancho de banda 500 metros.....</i>	105
<i>Ilustración 49: Distribución del dengue en el municipio de Floridablanca, ancho de banda 1000 metros.....</i>	105
<i>Ilustración 50: Interfaz de carga de archivos KML.....</i>	112
<i>Ilustración 51: Interfaz de filtro de información.....</i>	113
<i>Ilustración 52: Interfaz de filtro de información, seleccionando parámetros.....</i>	114
<i>Ilustración 53: Interfaz selección de función.....</i>	115
<i>Ilustración 54: Formulario de ingreso de parámetros KDE.....</i>	116
<i>Ilustración 55: Interfaz de visualización de resultados K de Ripley.....</i>	117
<i>Ilustración 56: Formulario ingreso de parámetros KDE.....</i>	118
<i>Ilustración 57: Interfaz de visualización de resultados KDE.....</i>	119

RESUMEN

El documento presentado muestra el desarrollo del proyecto, aplicación software de apoyo en la identificación de aglomeraciones de eventos para la toma de decisiones, elaborado por JONATHAN CONTRERAS VARGAS y OMAR YESID IBAÑEZ ORTIZ, para lo cual se usaron métodos estadísticos como el método *Kernel Density Estimator* y la función *K de Ripley*. El documento se divide en siete capítulos, que van desde la revisión de la teoría, hasta la aplicación de los métodos a los datos reales, terminando con las respectivas conclusiones y trabajo futuro.

En el primer capítulo, se encuentra todo lo relacionado con el planteamiento del problema, justificación, pregunta de investigación, también, se plantean los objetivos del proyecto. En el segundo capítulo, está el marco conceptual, todo lo relacionado con la respectiva búsqueda y documentación de la teoría relacionada con el proyecto en el que se trabajó, además, se halla la revisión de la literatura y herramientas computacionales.

En el tercer capítulo, se encuentra todo lo que tiene que ver con materiales, instrumentos y metodología que utilizamos para desarrollar el proyecto presentado, asimismo la arquitectura de la aplicación software; En el capítulo cuatro, se halla el primer resultado obtenido, el cual es el diseño de la aplicación software, este capítulo va desde los diagramas de caso de uso, hasta los mockups, pasando por los diagramas de actividad y secuencia.

En el quinto capítulo, se encuentran los resultados, este capítulo está el alistamiento de los respectivos datos que se usaron para cargar en la aplicación software, además, los hallazgos que se obtuvieron en el momento de desarrollar la aplicación, y, por último, los resultados obtenidos a medida que se iba terminando cada módulo. Luego se probó el funcionamiento de la aplicación software terminada e implementando los métodos a los datos de dengue reales.

En el sexto capítulo, está lo que corresponde a las conclusiones encontradas luego de todo el proceso de desarrollo que llevó la construcción de la aplicación software, por último, en el capítulo siete, se hallan las recomendaciones y el posible trabajo a futuro.

Palabras clave: Análisis espacial, Geocodificación, Georreferenciación
Información geoespacial, Sistemas de información geográfica.

ABSTRACT

The document presented shows the development of the project, support software application in the identification of agglomerations of events for decision making, prepared by JONATHAN CONTRERAS VARGAS and OMAR YESID IBAÑEZ ORTIZ, for which statistical methods such as the Kernel Density Estimator method were used. and the Ripley's K Function. The document is divided into seven chapters, ranging from the revision of the theory, to the application of the methods to the real data, ending with the respective conclusions and future work.

In the first chapter, you will find everything related to the approach of the problem, justification, research question, also, the objectives of the project are stated. In the second chapter, there is the conceptual framework, everything related to the respective search and documentation of the theory related to the project in which we worked, in addition, is the literature review and computational tools.

In the third chapter, we find everything that has to do with materials, tools and methodology that we use to develop the presented project, in addition to the architecture of the software application; In chapter four, we find the first result obtained, which is the design of the software application, this chapter goes from the use case diagrams, to the mockups, going through the activity and sequence diagrams.

In the fifth chapter, the results are found, this chapter is the enlistment of the respective data that were used to load in the software application, in addition, the findings that were obtained at the time of developing the application, and, finally, the results obtained as each module was finished. Then, the operation of the finished software application was tested and the methods were applied to the real dengue data.

In the sixth chapter, is what corresponds to the conclusions found after the entire development process that led to the construction of the software application, finally, in chapter seven, are the recommendations and possible work in the future.

Keywords: Spatial analysis, Geocoding, Georeference Geospatial information, Geographic information systems.

INTRODUCCIÓN

El presente proyecto o tesis de grado ha tenido como finalidad el desarrollo de una aplicación software de apoyo en la identificación de aglomeraciones de eventos con el objetivo de tener ideas que ayuden a identificar el comportamiento y la distribución en la que ocurren los eventos. Para este proyecto se tomó como caso de estudio el dengue en el municipio de Floridablanca, Santander, siendo los eventos los casos de dengue en dicha zona, ocurridos en el año 2015, los datos fueron proporcionados por el Sistema de Vigilancia en Salud Pública (SIVIGILA), adscrito al Ministerio de Salud y Protección Social. La aplicación software carga los datos y hace uso de dos métodos, estos métodos existen separadamente, pero se justifican el uno con el otro.

El primero, la función **K de Ripley**, permite determinar si la distribución espacial de los eventos es de forma aleatoria, uniforme o tienden a agruparse en el área de análisis. Si dichos eventos forman grupos, es porque existen focos lo cual indica que hay actores u objetos en el entorno que generan el evento. Luego, si por medio de esta función podemos determinar que los eventos tienden a agruparse, se procederá a la aplicación del segundo método llamado **KDE**.

El método **Kernel Density Estimator (KDE)**, permite realizar los mapas de calor de los puntos de aglomeración por medio de su localización y representación cartográfica en un mapa digital con el fin de inferir en la toma de decisiones o en la solución del problema de dicho evento. Este software también permite aplicar filtros a los datos de los eventos cargados con el propósito de reducir el tamaño de búsqueda a un conjunto menor o igual con información de mayor importancia a lo que necesitamos saber y tener mejor idea a la hora de encontrar patrones entre dichos eventos.

Este software de apoyo puede ser de ayuda a personas u organizaciones a comprender mejor los eventos en una determinada zona mediante patrones espaciales y las relaciones entre ellas para apoyar en la toma de decisiones con respecto al caso de estudio.

1. PLANTEAMIENTO DEL PROBLEMA

Todos los eventos que ocurren cotidianamente generan en el hombre incertidumbre acerca de su organización espacial y como se encuentran relacionados entre sí, lo que puede hacer necesario determinar su distribución, organización y aglomeración, con el fin de entender porque se encuentra en dicha zona, o de ser posible identificar las causas que hacen que el evento se presente allí. Dichos eventos pueden ocurrir aleatoriamente o presentarse de manera concurrente generando patrones, si se determina que no son aleatorios pueden ser analizados para identificar la manera en cómo se comportan los eventos bajo ciertas condiciones. Esto es posible mediante un sistema de información geográfico de análisis espacial, el cual se centra en el estudio de manera separada de los eventos por medio de la utilización de técnicas, procedimiento o conjunto de procedimientos basados en la localización y la información de cada uno de los casos con el fin de observar, percibir, comparar, inferir y agrupar para analizar y tener ideas que permitan describir la composición, funcionalidad y distribución de los eventos y su participación dentro de la globalidad en una región.

Debido a esto, se hace necesario una aplicación software de apoyo en la identificación de aglomeraciones de eventos, ya que por medio de esta se puede plantear e identificar ideas que ayuden a resolver o a reducir el problema. Cuando un evento es negativo en la sociedad es necesario encontrar una solución para que este disminuya o si es posible deje de existir, para ello es importante comprender mejor la forma en que estos se distribuyen. Apoyándonos en los medios computacionales podemos integrar métodos estadísticos, matemática compleja y métodos gráficos que muestre resultados por medio de mapas de calor con el fin de encontrar aglomeraciones en dichos eventos que nos ayuden a encontrar un patrón, y así, podamos resolver algunas de las preguntas presentadas a continuación en un artículo realizado por (Nyerges, 1991):

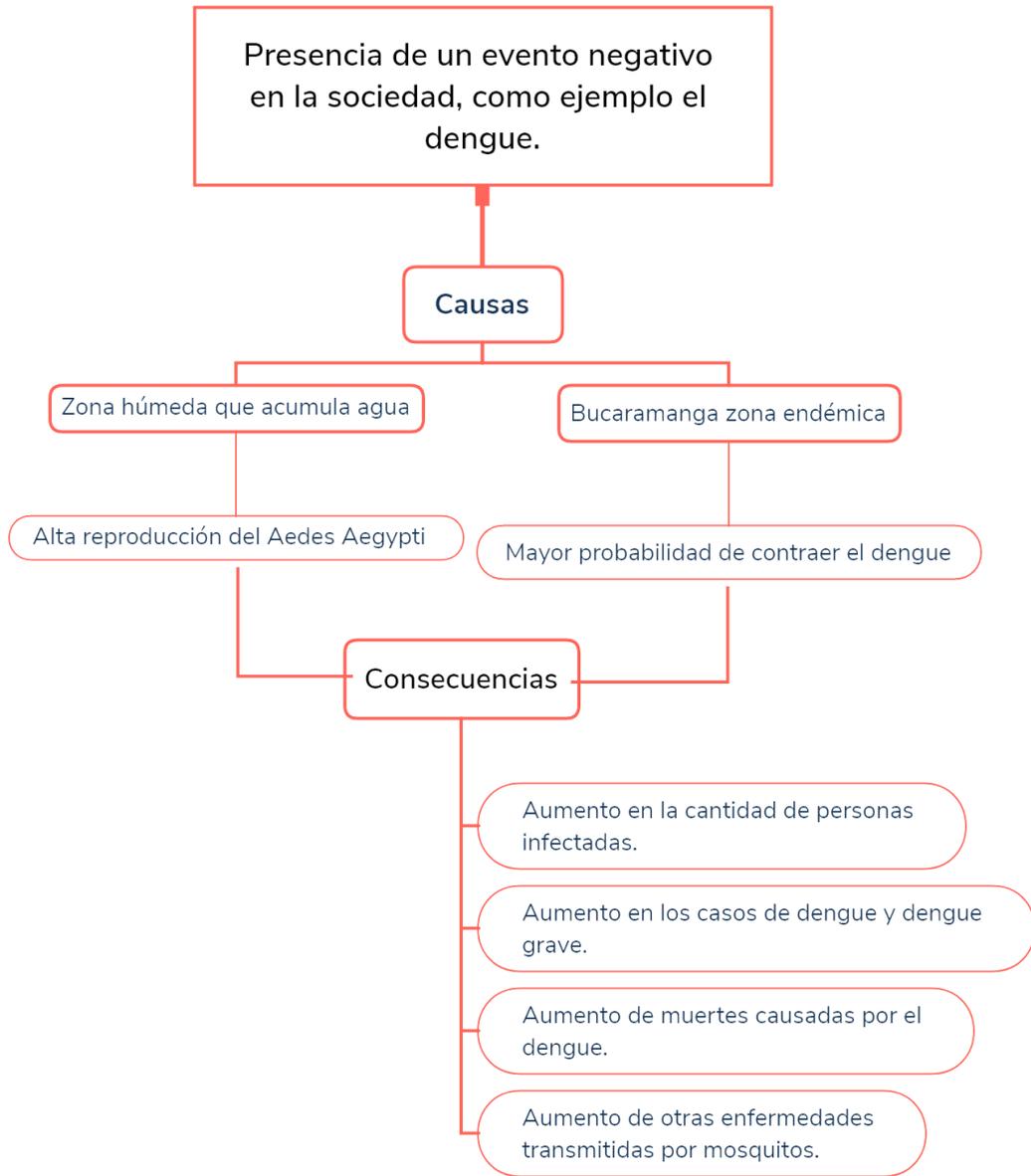
- ¿Dónde se encuentra?
- ¿Dónde tiene lugar?
- ¿Qué hay ahí?
- ¿Por qué está ahí?
- ¿Por qué no está en otro lugar?
- ¿Qué podría haber ahí?
- ¿Podría estar en otro lugar?
- ¿Cuánto hay en ese lugar?
- ¿Hasta dónde se extiende?
- ¿Por qué tiene esa estructura particular?

- ¿Existe regularidad en esa estructura?
- ¿De qué naturaleza es esa regularidad?
- ¿Por qué el patrón de distribución espacial tiene que mostrar regularidad?
- ¿Dónde se encuentra en relación con otros de su misma clase?
- ¿Qué clase de distribución conforma?
- ¿Se encuentra en todo el mundo?
- ¿Es universal?
- ¿Dónde están sus límites?
- ¿Cuál es la naturaleza de esos límites?
- ¿Por qué esos límites acotan su distribución?
- ¿Qué otras cosas encontramos asociadas espacialmente a ese fenómeno?
- ¿Aparecen esas cosas asociadas espacialmente en otros lugares?
- ¿Por qué deben estar asociadas espacialmente?
- ¿Ha estado siempre allí?
- ¿Cuándo apareció por primera vez?
- ¿Cómo ha variado espacialmente a través del tiempo?
- ¿Qué factores han influido en su dispersión?
- ¿Por qué se ha dispersado de esa manera?
- ¿Qué factores geográficos han limitado su dispersión?

Un ejemplo de estos eventos puede ser: los accidentes de tránsito en una determinada zona, casos de hurtos, casos de epidemias tales como el dengue o el VIH. Por lo tanto, el desarrollo de este software de apoyo ayuda a personas u organizaciones a descubrir las particularidades de los patrones espaciales y las relaciones entre ellas, apoyando en la toma de decisiones con respecto al caso de estudio que está siendo evaluado, y así intervenir en la posible solución del problema, ya sea erradicando por completo la ocurrencia de este evento o en su defecto disminuyéndolo.

Como se ha descrito anteriormente, pueden existir muchos eventos dentro de una localidad. Como caso de estudio y para efectos de prueba del software se trabajó como evento el dengue en el municipio de Floridablanca ubicado en el departamento de Santander, con eventos ocurridos en el año 2015 suministrados por el Sistema de Vigilancia en Salud Pública (SIVIGILA), adscrito al Ministerio de Salud y Protección Social.

Ilustración 1: Árbol de problemas.¹



¹ Fuente los autores

JUSTIFICACIÓN

Uno de los eventos que llevaron a realizar este proyecto, es el dengue, debido a que, en la actualidad, es considerado como una de las enfermedades más importantes transmitidas por mosquitos y uno de los principales problemas de salud pública en el mundo (Halstead, 2007). Según Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL en los últimos cincuenta años, el número de casos de dengue, tanto en la forma leve como grave y la incidencia de la infección por el DENV, ha incrementado de manera alarmante como se cita (Organización Mundial de la Salud., 2018). El número real de casos de dengue está insuficientemente notificado y muchos casos están mal clasificados. Según una estimación reciente, se producen 390 millones de infecciones por dengue cada año (intervalo creíble del 95%: 284 a 528 millones), de los cuales 96 millones (67 a 136 millones) se manifiestan clínicamente (cualquiera que sea la gravedad de la enfermedad).

El número anual de casos que se notificaron consecuentemente por los estados miembros de tres regiones de la OMS pasó de 2,2 millones en 2010 a 3,2 millones en 2015. En fecha reciente el número de casos notificados ha seguido aumentando. Fueron notificados 2,35 millones de casos en el año 2015 tan solo en la Región de las Américas, de los cuales más de 10.200 casos fueron diagnosticados como dengue grave provocando 1.181 muertes.

El año 2016 se caracterizó por grandes brotes de dengue en todo el mundo. La Región de las Américas notificó más de 2.380.000 casos ese año, y solo en Brasil hubo poco menos de 1.500.000 casos, es decir, cerca de tres veces más que en 2014. Se notificaron asimismo 1.032 muertes por dengue. Cada año, unas 500.000 personas que padecen dengue grave necesitan hospitalización, y aproximadamente un 2,5% fallecen (Organización Mundial de la Salud., 2018).

En Colombia en particular, un análisis hecho por la revista chilena de infectología muestra que cada año se presenta un incremento en el número de casos y las regiones más afectadas son Santander, Norte de Santander, Huila, Tolima y Valle del Cauca. El 2010 fue el año con mayor registro de eventos pues de 9.745 casos reportados en 1992, pasó a 151.774. Los resultados indican que la infección por DENV presenta un comportamiento cíclico, que muy posiblemente se repite cada tres o cuatro años y dicha ocurrencia de casos puede ser atribuida a cambios sociales y climáticos (Castrillón, Castaño, & Urcuqui, 2015).

Ilustración 2: Regiones más afectadas por el dengue en Colombia².



² Tomado de (Castrillón et al., 2015), figura 3, página 146.

Ilustración 3: Reporte de casos de dengue clásico y dengue hemorrágico en Colombia entre el 2004 -2013. Realizado con base a la información de la base de datos SIVIGILA y del INS de Colombia.

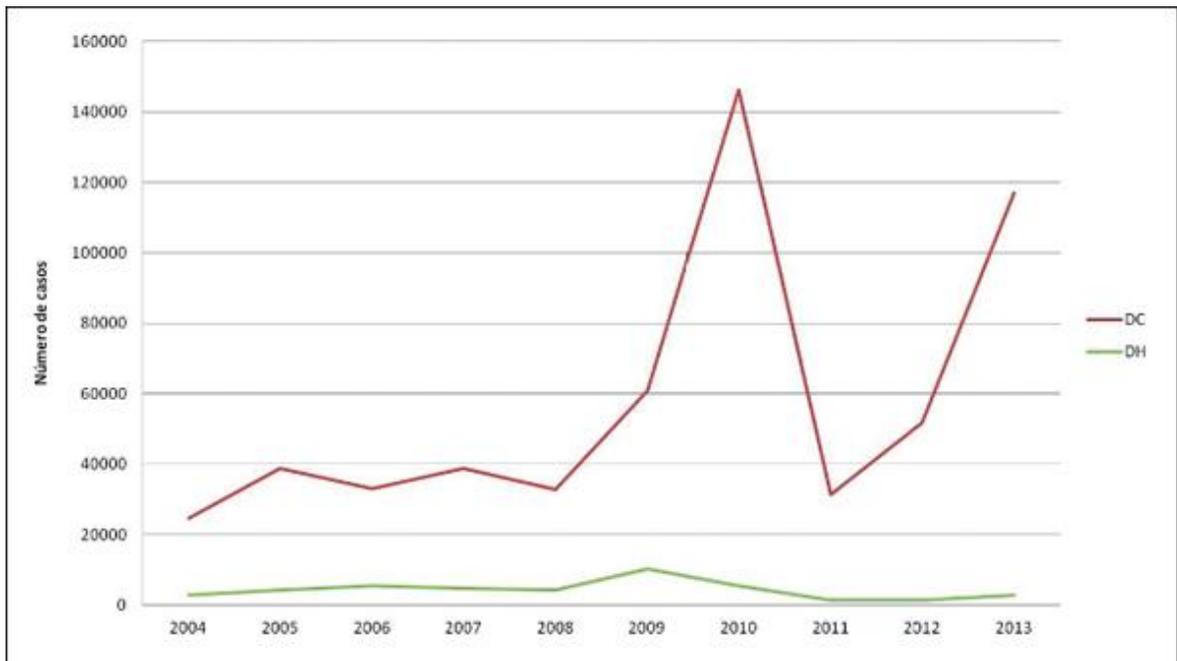
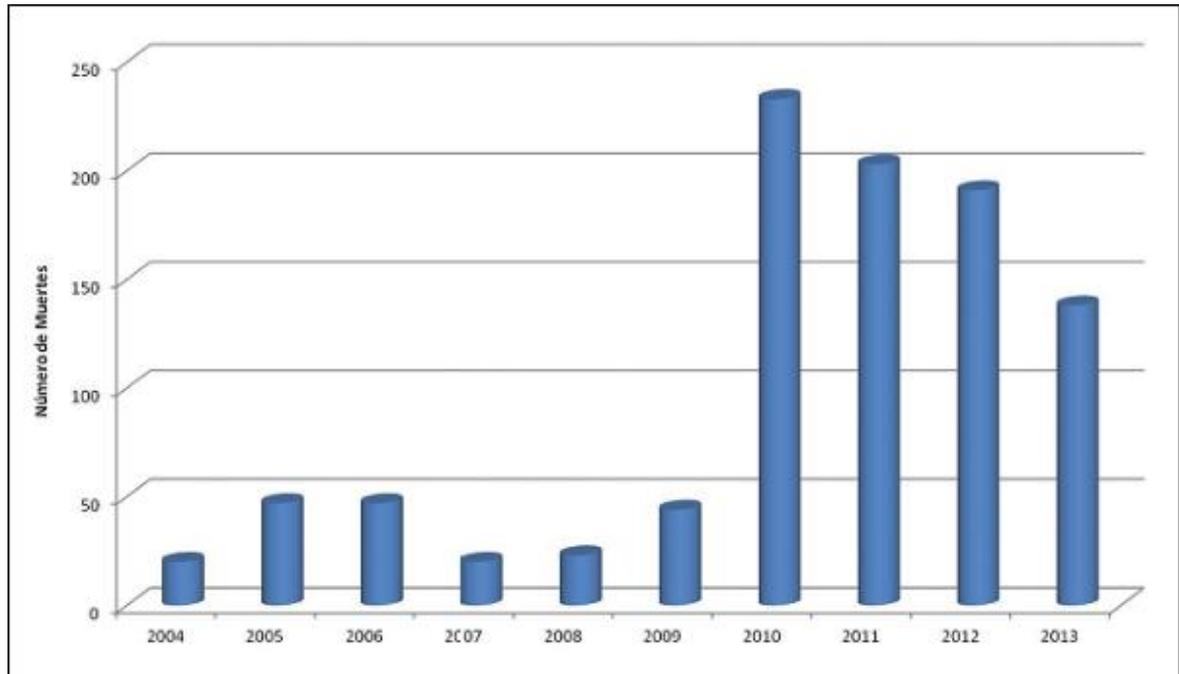


Ilustración 4: Reporte de casos de muerte por dengue en Colombia entre 2004-2013, según la base de datos de SIVIGILA y del INS.



Por otra parte, en un ámbito local encontramos a el departamento de Santander el cual, está ubicado a 1.230 metros sobre el nivel del mar con un clima cálido y húmedo, lo hace un lugar óptimo para esta enfermedad endemo-epidémica, por lo tanto, representa un problema de salud pública en el departamento. En el intervalo de 2010 hasta el año 2016 se habían infectado un total de 71.936 personas y cobró la vida de 59 de ellos (Vanguardia, 2016). Actualmente, Santander se encuentra entre sexto departamento en reportes de casos de dengue y hasta septiembre del año 2018, se registraron 1.535 eventos, 260 en las últimas semanas. Del total, 12 casos corresponden a dengue grave. El INS reporta que en Santander el 61% de los casos de dengue grave requirieron hospitalización (Vanguardia, 2018b).

No obstante, desde el año 1999 hasta el 2010 se registraron en Santander 81.947 casos de dengue de los cuales el 20,2% se registraron en Floridablanca (Padilla, Rojas, & Sáenz, 2012), colocándolo como el municipio más hiperendémico con una incidencia de casos de 139,9 para cada 100000 habitantes. Para las primeras 12 semanas del año 2018, 366 personas fueron diagnosticadas con dengue, es decir, aproximadamente 30 personas eran contagiados cada semana, la principal causa del aumento de los contagiados se debe al incremento de las precipitaciones en el departamento (Vanguardia, 2018a), lo que provoca estancamiento de agua en botellas, llantas o diferentes objetos en los que se

pueda almacenar agua, permitiendo la proliferación del mosquito *Aedes aegypti*, mosquito causante no solo del dengue, si no de muchas otras enfermedades tales como la fiebre amarilla, el Chikunguña y Zica, sin embargo, la mayor proporción de casos graves sucedieron en los años comprendidos entre el 2005 y 2008 y el 90% de los contagios ocurrieron en la zona metropolitana de Santander (Ocazonez-Jiménez, 2013).

PREGUNTA(S) DE INVESTACIÓN

¿Cómo se consigue rastrear el comportamiento de fenómenos o epidemias como el dengue por medio del análisis espacial y mediante la utilización del método KDE y función K de Ripley, con el objetivo de ayudar en la toma de decisiones?

¿Qué se puede conseguir al utilizar en asociación el método KDE y la función K de Ripley en el campo del análisis espacial de eventos?

HIPOTESIS

Cuando un evento no se distribuye de forma aleatoria en el espacio, si no que tiende a generar patrones de agrupación, es posible detectar su distribución específica, por medio de métodos estadísticos geográficos como KDE, que se implementan para estimar la densidad de probabilidad de un evento en una región o población, por otra parte, la función K de Ripley, se utiliza para determinar el comportamiento de distribución de una serie de eventos, estos dos métodos trabajando en conjunto ayudan no a entender cómo o cuales son las causas que hacen que se genere un fenómeno, pero sí, como este se distribuye y se propaga. También si existe aleatoriedad o no en estos eventos, de no ser así, indicará que existe algún foco que hace que este evento ocurra en determinadas zonas, las cuales deberán ser analizada posteriormente por los expertos en el tema.

OBJETIVOS

OBJETIVO GENERAL

Desarrollar una aplicación software de análisis espacial que implemente los métodos función **K de Ripley** y Kernel **Density Estimator (KDE)** para el apoyo en la identificación de aglomeraciones de eventos para la toma de decisiones.

OBJETIVOS ESPECIFICOS

1. Diseñar los módulos preliminares de la interfaz de la aplicación software.
2. Desarrollar un módulo de carga desde archivos KML de puntos para eventos y de polígonos para delimitaciones de áreas.
3. Implementar el método KDE con base en los archivos KML de puntos de eventos cargados y que permita aplicar filtros de información específica.
4. Implementar la función K de Ripley con base en los archivos KML de puntos de eventos y polígonos de delimitación y que permita aplicar filtros de información específica.
5. Desarrollar el módulo visor de resultados de KDE, por medio de mapas de aglomeración de eventos.
6. Desarrollar el módulo de exportación de resultados por medio de archivos KML, PNG y Excel.
7. Realizar pruebas funcionales a cada módulo del software para verificar su correcto funcionamiento.
8. Realizar un análisis del dengue en el municipio de Floridablanca en el 2015 basado en los métodos implementados.

2. MARCO REFERENCIAL

2.1. MARCO CONCEPTUAL Y TEÓRICO

Para este proyecto es necesario la utilización de un sistema de referencia terrestre y un sistema de coordenadas para la determinación de la ubicación de cada evento cargado en nuestro sistema, para ello necesitamos conocer que es la geodesia. Como se expone en el siguiente libro (Pérez Navarro, Botella Plana, & Muñoz Bolas, 2011) la geodesia es una ciencia básica con principios matemáticos y físicos. Es aplicada en diferentes ramas del conocimiento, como la topografía, la cartografía, la fotogrametría o la navegación. Su principal objetivo es el estudio y determinación de la forma y dimensiones de la Tierra, ya que está ligado a la georreferenciación, lo cual es el proceso de determinación de posiciones de puntos sobre la superficie terrestre, son importantes las implicaciones prácticas de esta ciencia. Es importante explicar los fundamentos de la georreferenciación con el fin de entender los mecanismos con los que los sistemas de información geográfica (SIG) analizan y manejan la información.

Para este software necesitamos determinar la posición en la que se encuentra cada uno de los eventos analizados, para esto es necesario la descripción del contexto matemático básico con el que expresaremos un punto sobre la esfera que es la tierra, más específico el municipio de Floridablanca, Santander, como son los sistemas de coordenadas y las proyecciones cartográficas.

2.1.1. SISTEMAS DE COORDENADAS

Según (Pérez Navarro et al., 2011) un sistema de coordenadas es una creación artificial que permite la definición analítica de la posición de un objeto o un fenómeno. Existen diferentes opciones para definir analíticamente la situación geométrica de un elemento, lo cual hace posible elegir entre diferentes sistemas de coordenadas. Matemáticamente, todos los sistemas de coordenadas son admisibles, y el único motivo para seleccionar uno u otro es la conveniencia o el hecho de que una determinada cuestión aparezca en su forma más simple. Desde lo práctico, se escogen los sistemas de coordenadas que permitan representar la cuestión objeto de estudio de una forma física y geoméricamente interpretable y susceptible de ser medida.

Los sistemas de coordenadas más utilizados son: el sistema de coordenadas geográficas, el sistema de coordenadas cartesiano y el sistema de coordenadas proyectadas.

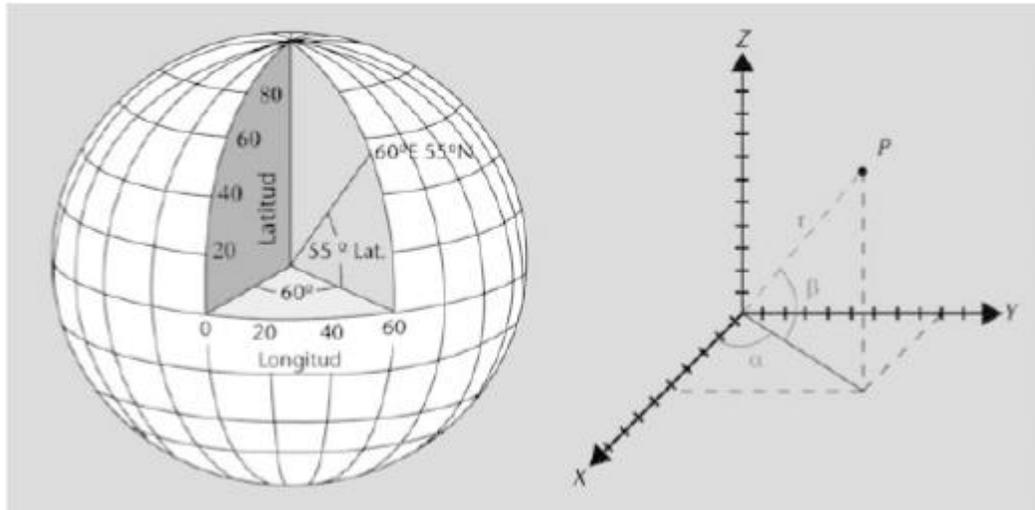
2.1.1.1. COORDENADAS GEOGRÁFICAS

Como se encuentra en el libro Introducción a los sistemas de información geográfica y geo telemática (Pérez Navarro et al., 2011), p.29) este tipo de sistemas también se denomina geodésicas, utiliza una superficie esférica tridimensional para definir las localizaciones sobre la superficie terrestre. En la ilustración 4, se muestra cómo cualquier punto sobre la superficie terrestre se determina con dos ángulos medidos desde el centro de la Tierra, que se llaman **latitud** y **longitud**.

La longitud (α) de un punto es el ángulo medido a lo largo del ecuador desde cualquier punto de la Tierra. Es medida con respecto a un meridiano principal, que normalmente es el meridiano de Greenwich (longitud cero), aunque algunos países han utilizado otras líneas de longitud que pasan por Berna, Bogotá o París. Los meridianos son las líneas verticales de igual longitud que pasan por los polos.

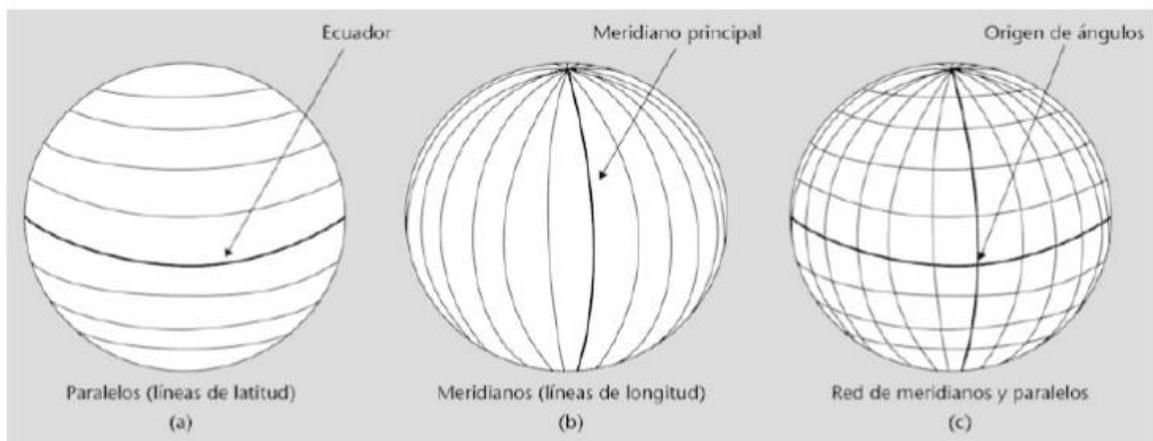
La latitud (β) de un punto es el ángulo medido desde el centro de la tierra hacia el norte, entre el ecuador y la posición de un punto sobre la superficie terrestre. Las líneas horizontales de igual latitud se llaman paralelos. Se toma el ecuador terrestre como la línea de latitud cero.

Ilustración 5: Valores de longitud y latitud de un punto sobre la esfera³.



En la ilustración 5a y 5b se puede ver un ejemplo de cómo son los paralelos y meridianos. En la parte 3c de la figura se ve cómo estos meridianos y paralelos forman una red sobre la esfera. Esta red tiene el origen (0, 0) en el punto donde se cruzan el ecuador y el meridiano principal.

Ilustración 6: Elementos del sistema de coordenadas geográficas: paralelos (a), meridianos (b) y origen de coordenadas (c)⁴.



³ Tomado de (Pérez Navarro et al., 2011), figura 2, Página 29.

⁴ Tomado de (Pérez Navarro et al., 2011), figura 3, Página 30.

La esfera terrestre es dividida en dos hemisferios por el ecuador, norte y sur, la latitud al norte del ecuador será positiva y al sur será negativa. Los valores de latitud van desde el -90° en el polo Sur hasta los $+90^\circ$ en el polo Norte. La esfera también es dividida en dos mitades por el meridiano principal, este y oeste. La longitud al este del meridiano principal será positiva y al oeste será negativa con valores entre -180° oeste y $+180^\circ$ este.

En conclusión, este sistema de meridianos y paralelos permite definir la localización exacta de cualquier punto sobre la Tierra, sin embargo, con este sistema, sólo a lo largo del ecuador la distancia representada por un grado de longitud es equivalente a un grado de latitud. Eso es debido al hecho de que el ecuador es el único paralelo de igual dimensión que un meridiano: es lo que se conoce con el nombre de **geodésicas** (círculos máximos) lo cual es una circunferencia con el mismo radio que la tierra. En particular, es el círculo que se obtiene como intersección de cualquier plano que contiene el centro de la Tierra con la superficie terrestre. El ecuador y todos los meridianos son geodésicas. Las geodésicas indican la distancia más corta entre dos puntos de la Tierra. Esta propiedad se utiliza mucho en navegación.

2.1.2. COORDENADAS CARTESIANAS

En este sistema también llamado coordenadas geocéntricas, la posición es definida en un espacio tridimensional por las coordenadas (X, Y, Z). Los ejes de coordenadas se detallan de la siguiente manera:

- El eje Z pasa por el centro de la Tierra y por los polos.
- El eje X pasa por el centro de la Tierra y por el meridiano principal de Greenwich.
- El eje Y forma un ángulo de 90° con los otros dos ejes.

Ilustración 7: Ejes X, Y, Z del sistema de coordenadas cartesianas⁵.

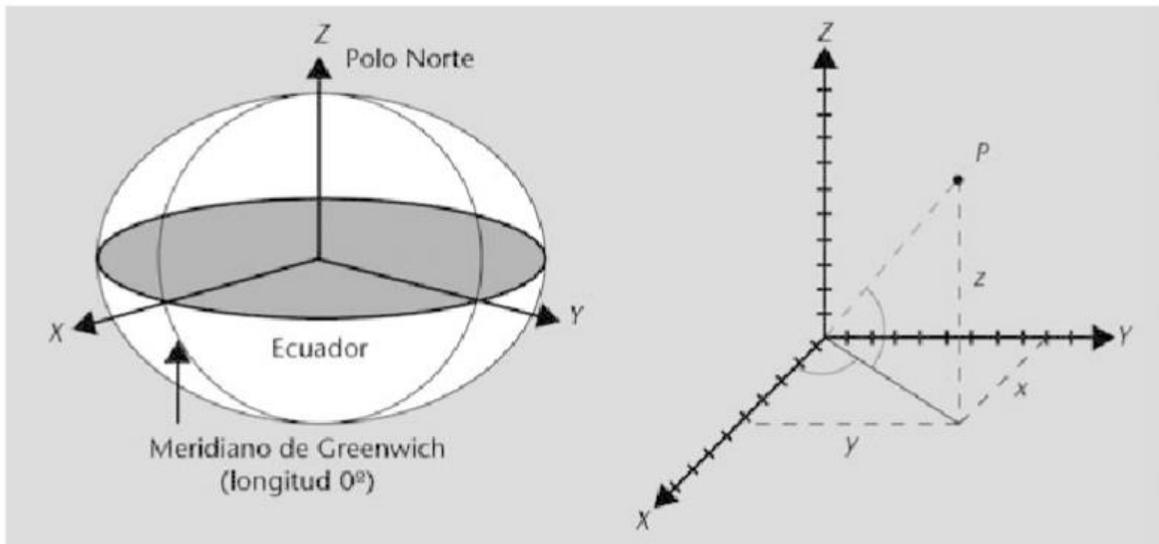
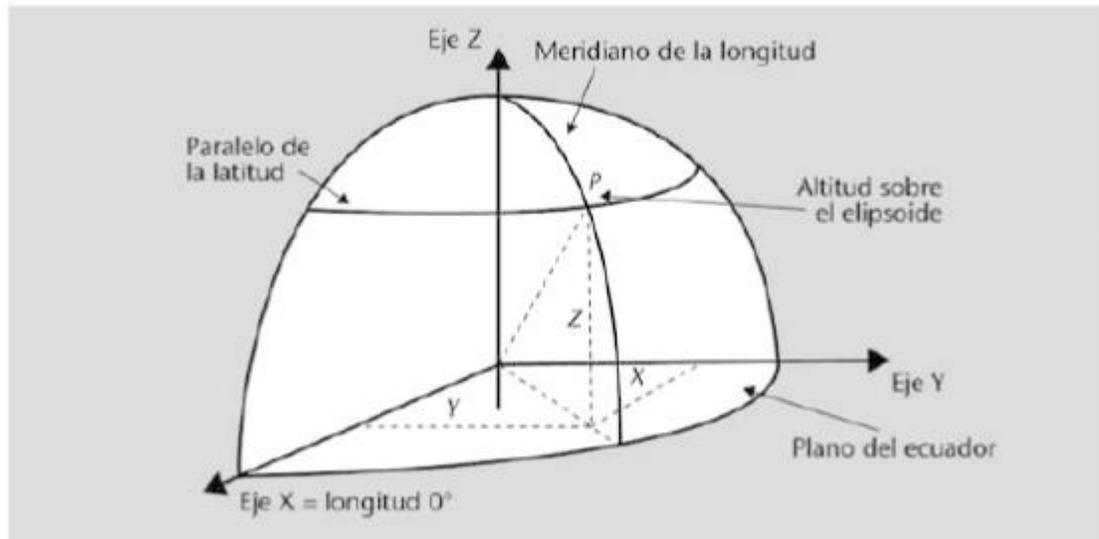


Ilustración 8: Correspondencia entre coordenadas geográficas y geocéntricas por un punto P⁶.



En la ilustración 7 se muestra cómo se relacionan entre sí los sistemas de coordenadas geográficas y cartesianas XYZ (coordenadas geocéntricas) si establecemos como origen de X el plano que pasa por la longitud 0° (hacia el

⁵ Tomado de (Pérez Navarro et al., 2011), figura 5, página 31.

⁶ Tomado de (Pérez Navarro et al., 2011), figura 6, página 32.

este), como origen de Y el plano de longitud 90° , y como origen de Z el plano que pasa por la latitud 0° (ecuador) (Pérez Navarro et al., 2011).

Por ejemplo, un punto P con coordenadas geodésicas:

- longitud = $-3^\circ 35' 57,7336''$
- latitud = $39^\circ 32' 46,8909''$

Se expresa en coordenadas geocéntricas como:

- $X = 4915215,575$
- $Y = -309184,931$
- $Z = 4039273,734$

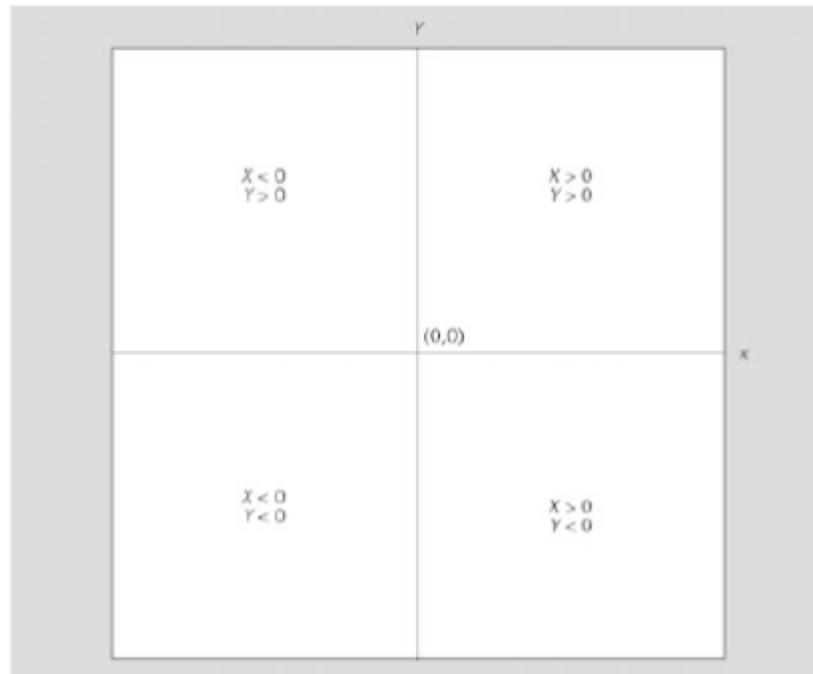
2.1.2.1. COORDENADAS PROYECTADAS

Para (Pérez Navarro et al., 2011) este sistema es definido sobre una superficie plana, donde la localización de las coordenadas se realiza con respecto a una malla (grid) donde se ha definido el origen en su centro.

La posición tiene dos valores referidos al punto central. Uno especificará la posición horizontal y es llamado coordenada X, y el otro especificará la posición vertical dentro de la malla y lo llamaremos coordenada Y. La coordenada del origen es $(X, Y) = (0, 0)$.

En el grid, la línea central horizontal es lo que se llama eje X y la línea central vertical es lo que se llama eje Y. Estos dos ejes dividen la malla en cuatro cuadrantes. Los signos de las coordenadas se distribuyen tal como se indica en la ilustración 8.

Ilustración 9: Distribución de signos de las coordenadas X, Y en una malla del sistema de coordenadas proyectadas⁷



2.1.2.2. PROYECCIONES CARTOGRÁFICAS

Según (Pérez Navarro et al., 2011) las proyecciones cartográficas se utilizan para expresar un punto sobre una superficie esférica proyectado en una superficie plana. La transformación matemática que permite la transformación de las coordenadas esféricas del globo en un sistema de coordenadas planas es lo que llamamos proyección cartográfica. Las proyecciones cartográficas son sistemas que reproducen la esfera terrestre en una superficie plana. Ya que una superficie plana no puede ajustarse a una esfera sin estirarse o encogerse, es imposible representar en un mapa los atributos de un globo terráqueo (por ejemplo, meridianos, paralelos, límites entre países, etc.) sin producir distorsiones.

Las proyecciones tratan de minimizar, en la medida de lo posible, las deformaciones que sufren en este proceso los datos o relaciones geométricas de los objetos representados. Todas las proyecciones tienen algún error de deformación, lo cual se llama distorsión. Como no es posible conservar todas las

⁷ Tomado de (Pérez Navarro et al., 2011) , figura 7, página 33.

propiedades al mismo tiempo, hay que concentrarse en qué tipo de mapa se quiere realizar, ya que cada proyección puede conservar alguna de sus propiedades geométricas: la forma, el área, la distancia o la dirección. A pesar de la distorsión, todas las proyecciones mantienen una característica importante: **la exactitud del posicionamiento**.

2.1.2.3. TIPOS DE PROYECCIONES

Para (Pérez Navarro et al., 2011) la clasificación de las proyecciones según las propiedades geométricas mencionadas anteriormente, así como la clasificación según la figura geométrica de la que derivan son las siguientes:

PROYECCIONES SEGÚN LAS PROPIEDADES GEOMÉTRICAS

Según sus propiedades geométricas básicas, podemos clasificar las proyecciones en **conformes**, equivalentes, equidistantes y acimutales. Las proyecciones que no se encuentran en ninguna de estas clases suelen llamarse de compromiso.

- **Proyecciones conformes:** Las proyecciones conformes son caracterizadas por mantener la forma de la superficie que se muestra en el mapa. No distorsiona las relaciones angulares; por ende, las superficies conservan en el mapa la forma que tienen en la superficie terrestre. La condición para que un mapa sea conforme es que meridianos y paralelos se corten en ángulo recto tal como pasa en el globo terráqueo. Sin embargo, estas proyecciones distorsionan mucho el tamaño de las superficies cartografiadas y, como consecuencia, la escala no es constante entre regiones del mapa. Por ejemplo, en un mapamundi las superficies en latitudes elevadas se muestran más grandes de lo que realmente son. (un ejemplo de proyección conforme es la proyección de Mercator de la cual se hablará más adelante y fue la escogida para utilizar en este proyecto gracias a unas bases mencionadas posteriormente).
- **Proyecciones equivalentes o de igual área:** Son caracterizadas por que las áreas proyectadas mantienen las mismas proporciones que las áreas de la Tierra que representan. Por lo tanto, 1 cm^2 representa la misma área en Estados Unidos, Argentina o Siberia. Para conseguir la equivalencia, hay que deformar los ángulos originales; por lo tanto, una proyección no puede ser al mismo tiempo equivalente y conforme. Las proyecciones equivalentes suelen utilizarse en mapas temáticos que muestran la distribución de variables como la densidad de población, la superficie agrícola, las áreas forestales, etc.
- **Proyecciones equidistantes:** Estos tipos de proyecciones conservan las distancias, o bien sólo desde el centro de la proyección o bien a lo largo de

los círculos máximos (meridianos), pasando por un punto. (Snyder, 1987, p.4) citado por (Pérez Navarro et al., 2011, p.38). Por el contrario, no es posible representar todas las distancias de manera correcta. pero, la mayoría de las proyecciones equidistantes tienen una o más líneas para las que se conserva la distancia. Se dice que estas distancias son verdaderas.

- **Proyecciones acimutales, cenitales o de dirección verdadera:** Según (Snyder, 1987, p.4) citado por (Pérez Navarro et al., 2011, p.39) este tipo de proyecciones conservan las direcciones (y, por lo tanto, los ángulos o acimuts, aunque no necesariamente las distancias) de todos los puntos del mapa con respecto a un punto de referencia: el centro del mapa. Como en el caso de las distancias, no todas las relaciones angulares pueden representarse correctamente en un mismo mapa, pero sí que es posible representar correctamente todas las relaciones angulares con respecto a un punto.
- **Proyecciones de compromiso:** Esta proyección no intenta optimizar alguna de las cuatro propiedades geométricas, intenta conseguir un balance entre propiedades diferentes. Así, por ejemplo, mientras que una proyección puede no ser ni conforme ni equivalente, puede mantener una distorsión mínima de las formas y las áreas en una región concreta. Estas proyecciones de compromiso suelen utilizarse como la base de mapas temáticos.

PROYECCIONES SEGÚN LA SUPERFICIE DE LA QUE DERIVAN

Este tipo de proyecciones según (Pérez Navarro et al., 2011) se clasifican en: Cónicas, cilíndricas o planas.

- **Cónicas:** Son aquellas en las que podemos visualizar la Tierra proyectada en un cono tangente o secante, el cual se corta longitudinalmente para visualizarlo de forma plana. Los paralelos (líneas de latitud) se representan mediante arcos circulares concéntricos que aumentan a medida que se acercan al ecuador, y los meridianos (líneas de longitud) mediante líneas rectas en forma radial.
- **Cilíndricas:** Permiten visualizar la Tierra proyectada en un cilindro tangente o secante, el cual se corta longitudinalmente para visualizarlo de forma plana. La mayoría de las proyecciones cilíndricas se derivan de manera tal que el cilindro toca el globo terráqueo en el ecuador (punto de tangencia, donde no hay distorsión). En este tipo de proyección, los paralelos se muestran como líneas rectas horizontales y los meridianos como líneas rectas verticales. Paralelos y meridianos se distribuyen uniformemente y se cruzan ortogonalmente, es decir, en ángulo recto. En esta proyección, una línea recta entre dos puntos cualesquiera sigue una única dirección llamada

rumbo. Las líneas de rumbo o loxodromias son líneas de dirección constante o, dicho de otra manera, son líneas que cortan cada meridiano en un ángulo constante.

- **Planas:** En esta proyección, una porción de la superficie de la Tierra se transforma desde la perspectiva de un punto en una superficie plana. hipotéticamente, el punto de tangencia puede ser cualquier punto del planeta, aunque normalmente se utiliza el polo Norte, el polo Sur o algún punto del ecuador. Esta proyección conserva sus propiedades geométricas en torno al punto de tangencia y aumentan las distorsiones a medida que nos alejamos de su punto de origen.

Para la selección del tipo de proyección adecuada para este proyecto nos basamos en una regla descrita por (Maling, 1992) la cual nos dice que para proyecciones en zonas ecuatoriales y tropicales se usan proyecciones cilíndricas, esto debido a que las cilíndricas no tienen distorsión en el ecuador ya que ésta incrementa a medida que se acerca a los polos. Como se ha descrito anteriormente para este proyecto se utiliza la proyección UTM (sistema de coordenadas transversal de Mercator) la cual es una proyección cilíndrica, transversa y conforme.

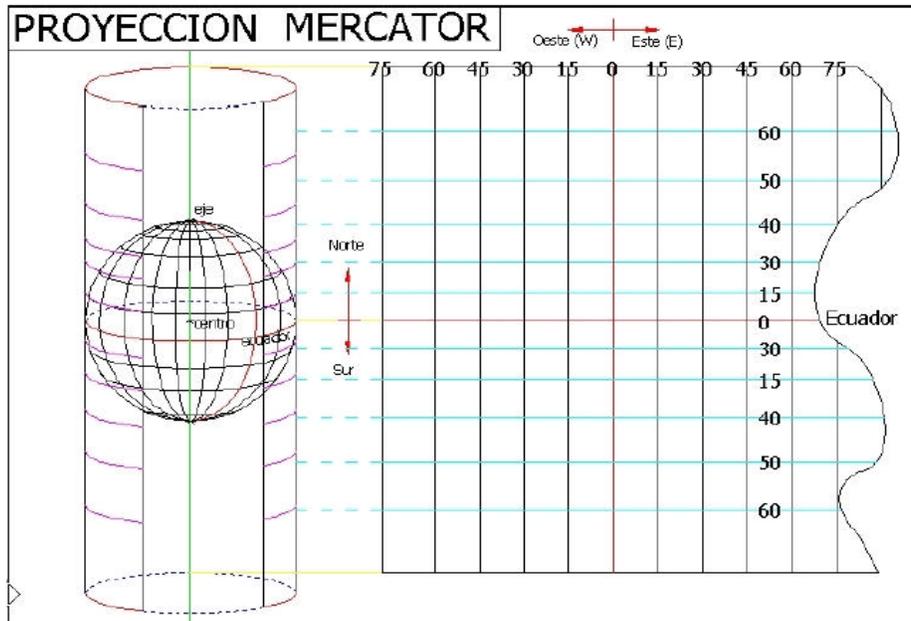
2.1.3. EL SISTEMA DE COORDENADAS UNIVERSAL TRANSVERSAL DE MERCATOR (UTM)

Este sistema de coordenadas es el más usado y conveniente para los SIG, desarrollado en los años 40 por US Army, el cual es ahora un estándar del mundo para mapeo topográfico e intercambio de datos digitales (Peña Llopis, 2005). Como se describe en la publicación hecha por (Fernández-Coppel, 2001) está basado en la proyección **MERCATOR** el cual utiliza un cilindro situado de forma tangente al elipsoide en el ecuador como se muestra en la ilustración 9. No obstante en la proyección transversal de Mercator la posición del cilindro de proyección es transversal respecto del eje de la tierra como se muestra en la ilustración 10.

El sistema **UTM** emplea husos de 6° de longitud los cuales son las posiciones geográficas que ocupan todos los puntos comprendidos entre dos meridianos generándose en cada huso un meridiano central equidistante 3° de longitud de los extremos de cada huso. La Tierra se ha dividido en 60 husos, de 6° cada uno, que se numeran del 1 al 60, de W a E, a partir del antimeridiano de Greenwich. Cada huso se divide horizontalmente en 20 bandas entre paralelos, quedando delimitado cada huso en áreas de 6° de longitud y 8° de latitud, que se denominan zonas, cada zona tiene su propio sistema de coordenadas, y constituyen la

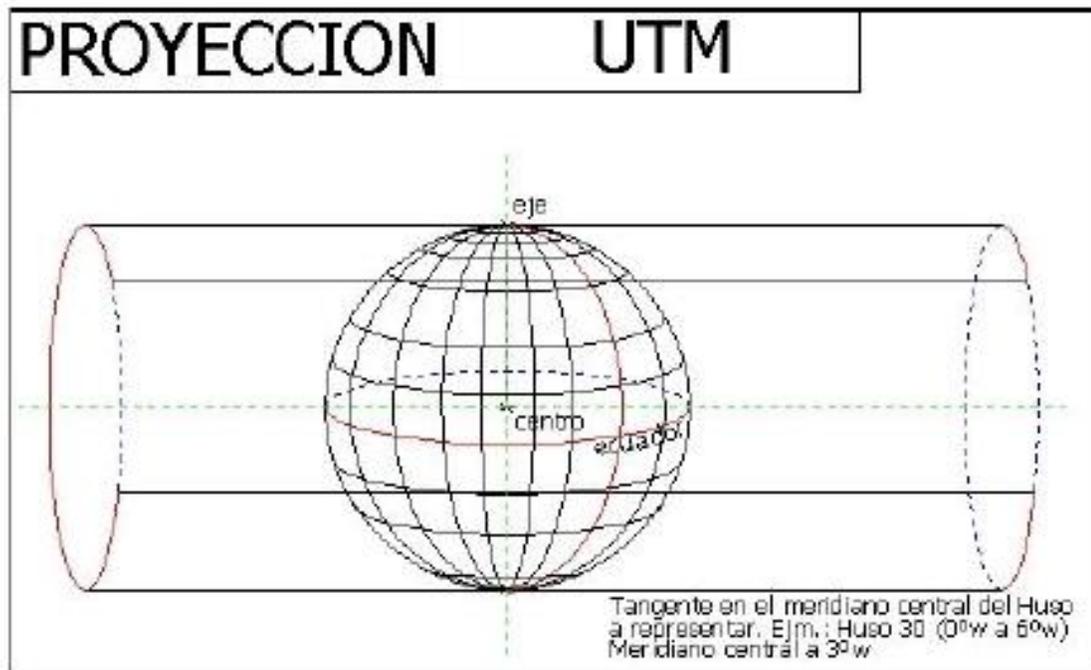
cuadrícula UTM. Estas bandas se nombran de Sur a Norte con las letras de la C a la X (se excluyen las letras (CH, I, LL, Ñ y O). Las bandas C a M se encuentran en el hemisferio Sur y de N a X en el hemisferio Norte, como se muestra en la ilustración 11 (Gallego Salguero & Sánchez Marco, 2013).

Ilustración 10: Proyección Mercator, utiliza un cilindro situado de forma tangente al elipsoide en el ecuador.⁸



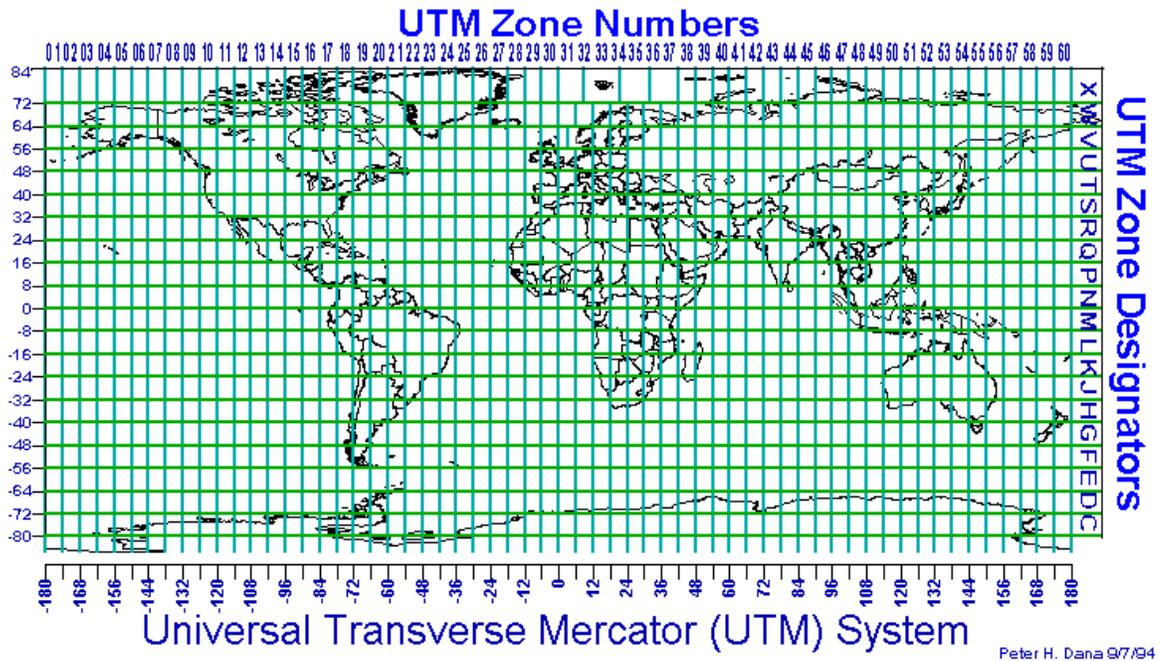
⁸ Tomado de (Fernández-Coppel, 2001) , página 23.

Ilustración 11: Proyección UTM, utiliza un cilindro situado de forma transversal respecto del eje de la tierra⁹.



⁹ Tomado de (Fernández-Coppel, 2001), página 24.

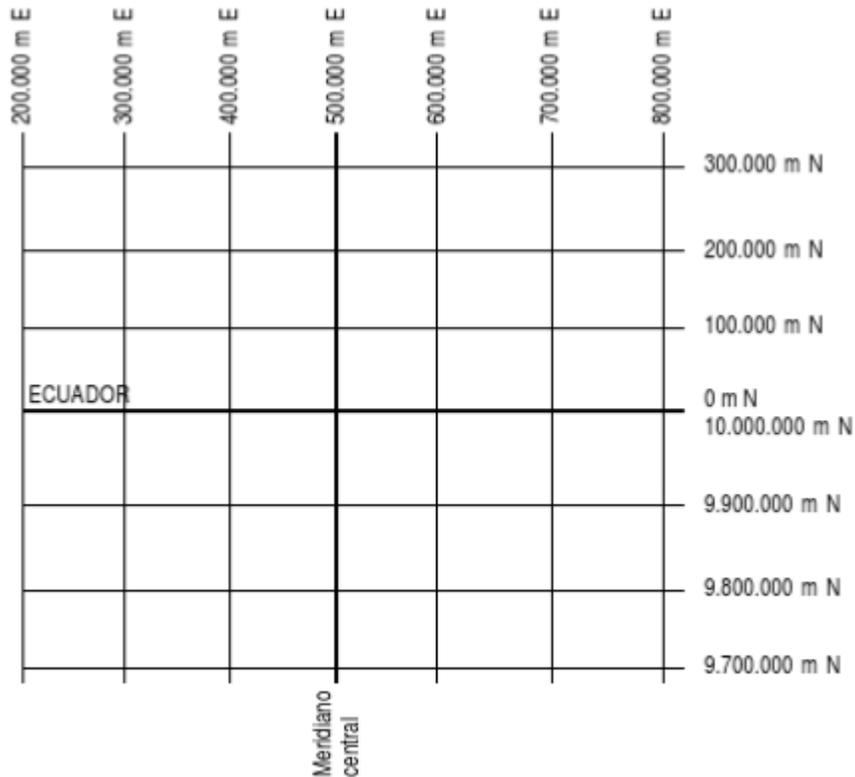
Ilustración 12: Cuadrícula UTM.¹⁰



Las cuadrículas de referencia, o grilla formadas por los husos, se componen de líneas verticales, denominadas "estes" e identificadas con la letra E (mayúscula), y líneas horizontales, denominadas "nortes" y designadas con una letra mayúscula N. En los dos tipos de líneas se anota su valor en unidades enteras, acompañado de la letra m (minúscula), para indicar que se mide en metros. Ejemplo: 230.000 m E, 456.000 m N. Las líneas verticales, estes (E), tienen su origen en cada zona del meridiano central y su valor es de 500.000 metros. Las líneas nortes (N) tienen una secuencia especial en su origen. Para el hemisferio sur se le asigna al Ecuador un valor de 10'000.000 de metros y para el hemisferio norte el valor en esta misma línea de referencia comienza con 0 metros, como se ve en la ilustración 13 (Serrato Álvarez, 2009, p.90).

¹⁰ Tomado de ("Utm Zones," 2015).

Ilustración 13: Grilla del sistema de coordenadas UTM empleada para cada zona¹¹.



Este sistema es comúnmente utilizado entre los 0° y los 84° de latitud norte y los 80° de latitud sur ya que a partir de ahí se produce una gran distorsión lo cual la hace perfecta para ser empleada tanto en el hemisferio norte como en el hemisferio Sur. La localización de un punto en UTM tiene la siguiente estructura: Un X y una Y con su medida en metros (m), un Huso (Colombia se encuentra entre las zonas 17,18 y 19 y también tiene en los dos hemisferios) y una zona y su Datum correspondiente (Fernández-Coppel, 2001).

2.1.4. LA FUNCIÓN K DE RYPLEY

La función K, fue desarrollada por Brian Ripley en el año 1976, es una herramienta utilizada para analizar datos sobre ubicaciones geográficas de eventos en un área de estudio predefinida, con el objetivo de hallar patrones de aglomeración de puntos con aleatoriedad de ocurrencia, por otra parte, realizar hipótesis de pruebas sobre el patrón, estimar y ajustar modelos de bivariado y multivariado

¹¹ Tomado de (Serrato Álvarez, 2009) , figura 19, página 91.

para describir relaciones entre 2 o más puntos (Dixon, El-shaarawi, & Piegorsch, 2002), una forma de probar los patrones de aleatoriedad se puede hacer por medio de simulaciones de Monte Carlo, lo cual consiste en generar una cantidad alta de conjuntos aleatorios de puntos, en el mismo espacio o área de estudio, y se puede aplicar a diferentes rangos de h , con respecto a cada uno de los valores originales de puntos y los generados aleatoriamente, entonces, si el valor hallado en la función K con los datos originales o valores reales, es superior a el intervalo de confianza definido, se puede afirmar que, los datos del fenómeno estudiado tienden a formar aglomeraciones y no se comporta de forma aleatoria en el espacio (Dixon et al., 2002, p.10).

La función K de Ripley general está dada por:

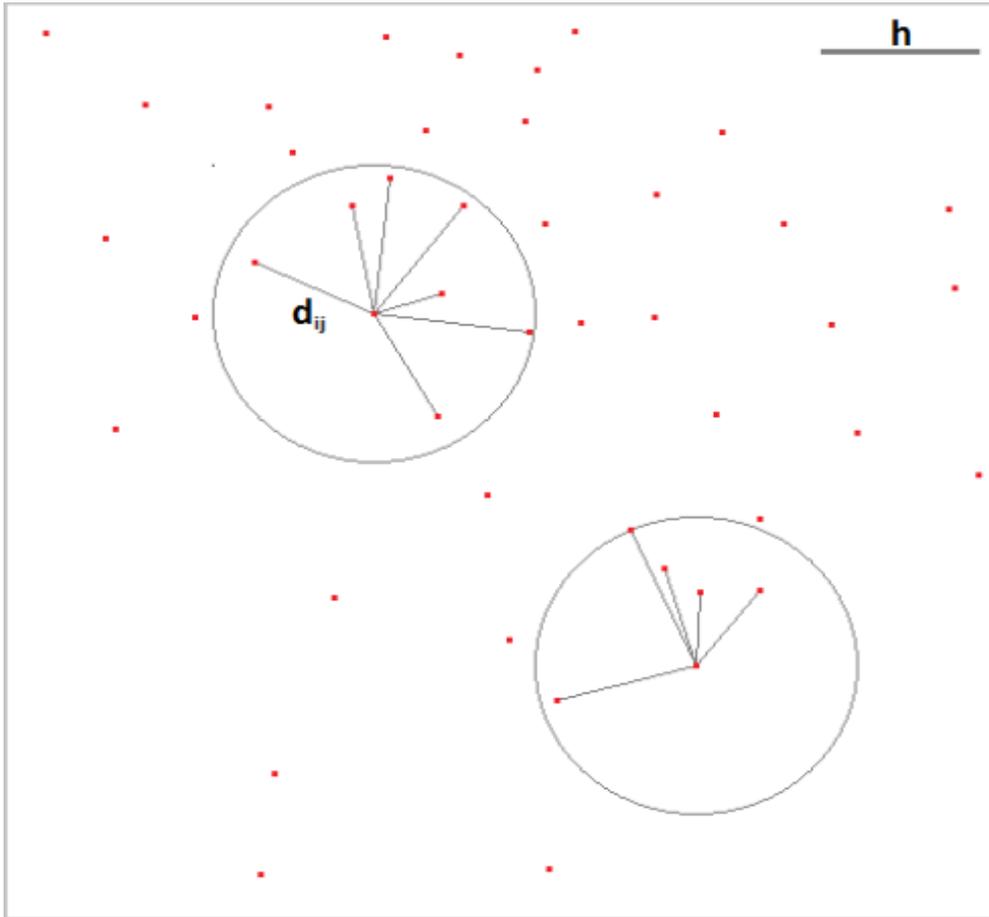
$K(h) = \lambda^{-1} E$ (1) [Número de eventos en una distancia h de un dato tomado al azar] (Dixon et al., 2002, p.1).

Donde λ , es la densidad de los eventos por unidad de área, $K(h)$ describe las características de los procesos puntuales en muchas escalas de distancia, $K(h)$ considerando que se toma un evento al azar, más comúnmente usado, para un proceso homogéneo de Poisson y distribución aleatoria espacial que es igual a πh^2 , para la estimación de $K(h)$, teniendo las ubicaciones de los eventos en un área determinada, “ $K(h)$, es una relación de un numerador y la densidad de los eventos λ , la densidad se puede estimar como $\hat{\lambda} = \frac{N}{A}$, (2) donde N , es el número observado de puntos y A , es el área de la región de estudio” (Dixon et al., 2002, p.2). Si se ignora los efectos del borde del polígono del área de estudio, es decir, que los puntos fuera del límite no se toman en cuenta dentro del numerador, así se encuentren dentro de la distancia h del punto que se tomó como aleatorio, entonces el numerador es igual a:

$$\hat{E}(h) = \frac{1}{N} \sum_{i \neq j} \sum I_t(d_{ij} < h) \quad (3)$$

Donde d_{ij} es la distancia entre los puntos i y j , por otra parte, $I(x)$ es el valor con el indicador 1 sí x es verdadero y 0 si no lo es, como se muestra en la siguiente figura (Dixon et al., 2002, p.2).

Ilustración 14: Función K^{12}



Una variedad de estimadores corregidos por el borde es:

$$\hat{K}_{ij}(h) = (\hat{\lambda}_i \hat{\lambda}_j A)^{-1} \sum_k \sum_l w(ik, jl) I(d_{ik, jl} < h) \quad (4)$$

Como la ecuación de arriba d_{ij} es la distancia entre los puntos i y j la función de peso $w(ik, jl)$ permite hacer la corrección del borde del polígono, las correcciones de bordes son de mayor relevancia cuando los h son puntos de mayor tamaño, ya que existe mayor probabilidad de que se encuentren fuera del área de estudio.

¹² Tomado de (F. Moreno, 2016), Figura 2, página

2.1.5. EL MÉTODO KDE

El método KDE (Kernel Density Estimator), es un método para el análisis de patrones de puntos muy conocido, permite calcular la densidad de los eventos en el espacio, para la implementación del método el área de estudio se divide en una cuadrícula del mismo tamaño y cada punto se va a encontrar en alguno de estos cuadrados. Además, se define un radio de búsqueda o ancho de banda que se puede hallar siguiendo el siguiente algoritmo: “calcular el centro medio de los puntos de entrada, calcular la distancia desde el centro medio (ponderado) para todos los puntos, calcular la distancia media de esas distancias, D_m , calcular el valor de la distancia estándar SD ” (ArcGIS, 2017). Por lo tanto, la ecuación para hallar el ancho de banda (R), es:

$$R = 0.9 * \min \left(SD, \sqrt{\frac{1}{\ln(2)}} * D_m \right) * n^{-0.2} \quad (5)$$

En donde n representa el número de puntos, y las demás variables fueron mencionadas anteriormente, luego, la ecuación para hallar la función de densidad para cada una de las cuadrículas generadas es:

$$\lambda(s) = \sum_{i=1}^n \frac{1}{\pi r^2} k \left(\frac{d_{is}}{r} \right) \quad (6)$$

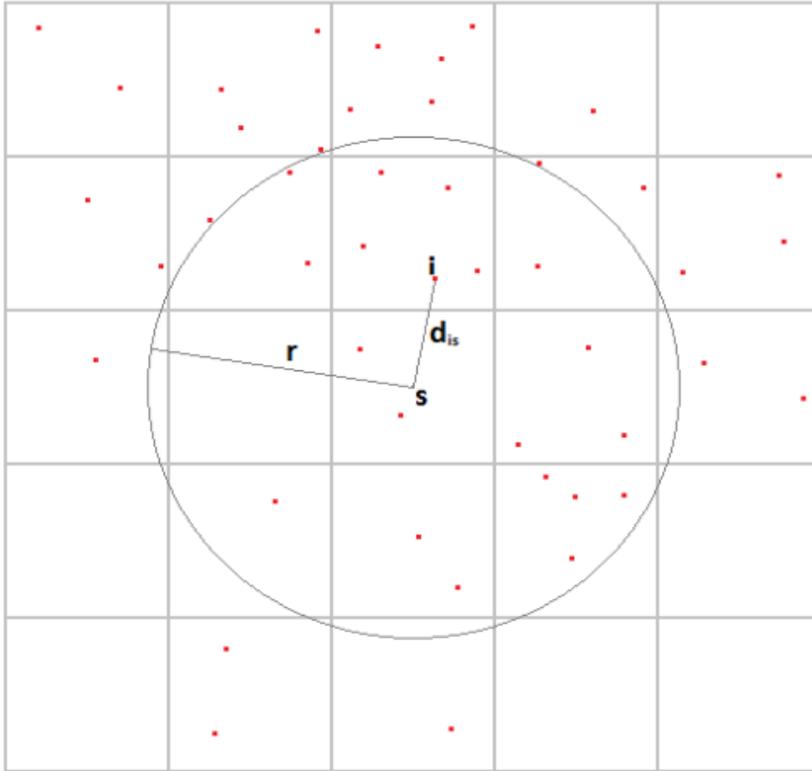
Donde $\lambda(s)$, representa la función densidad del punto s y r , representa el ancho de banda, k es la función de peso relacionada con el ancho de banda y la distancia desde la más cercana con un peso de d_{is} hasta la más lejana con 0, cuando se supere el ancho de banda (F. Moreno, 2016, p.8). Además, para hallar el valor del peso, se puede realizar por medio de la función Gaussiana la cual es:

$$k \left(\frac{d_{is}}{r} \right) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{d_{is}^2}{2r^2} \right), \text{ para } 0 < d_{is} \leq r$$

$$k \left(\frac{d_{is}}{r} \right) = 0, \text{ para } d_{is} > r \quad (7)$$

En la que K es un valor que puede ser $\frac{3}{\pi} 0 \frac{3}{4}$

Ilustración 15: Kernel Density Estimator – KDE¹³.



2.1.6. ARCHIVOS KML (KEYHOLE MARKUP LANGUAGE)

Este tipo de archivo es el utilizado para agregar la información de la ubicación de los eventos a nuestro sistema. Según la página de desarrolladores de Google (Google, 2014) los archivos KML son utilizados para mostrar datos geográficos en un navegador terrestre, como Google Earth, Google Maps y Google Maps para móviles. KML utiliza una estructura basada en etiquetas con atributos y elementos anidados y está basado en el estándar XML. Gracias a las marcas de posición (placemark) podemos marcar una posición en nuestro polígono. El archivo KML para determinar la marca del evento viene con la siguiente estructura:

```
<?XML version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2"> <Placemark>
  <name>Marca de posición simple</name>
  <description>Pegada al suelo. Se coloca de forma
```

¹³ Tomado de (F. Moreno, 2016), figura 1, página 9.

```

inteligente a la altura del relieve subyacente.
</description>
  <Point>
    <coordinates>-
122.0822035425683,37.42228990140251,0</coordinates>
  </Point>
</Placemark> </kml>

```

- Un encabezado XML. Es la línea número 1 de todos los archivos KML. Antes de esta línea no puede haber caracteres ni espacios.
- Una declaración de espacio de nombres de KML. Es la línea número 2 de todos los archivos KML 2.2.
- Un objeto de marca de posición (Placemark) que contiene los siguientes elementos:
 - Un nombre (name) que se utiliza como etiqueta para la marca de posición,
 - Una descripción (description) que aparece en una "viñeta" junto a la marca de posición,
 - Un punto (Point) que especifica la posición de la marca de posición en la superficie de la Tierra (la longitud, la latitud y, opcionalmente, la altitud).

Para crear el perímetro que encierre nuestros puntos de eventos se usa un objeto denominado (*Polygon*) con el podemos crear edificios simples y otras formas dibujando sus limitantes por medio de sus coordenadas. Su estructura en KML es la siguiente y representa un pentágono sobre la tierra.

```

<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2"> <Placemark>
  <name>El pentágono</name>
  <Polygon>
    <extrude>1</extrude>
    <altitudeMode>relativeToGround</altitudeMode>
    <outerBoundaryIs>
      <LinearRing>
        <coordinates>
-77.05788457660967,38.87253259892824,100
-77.05465973756702,38.87291016281703,100
-77.05315536854791,38.87053267794386,100

```

```

-77.05552622493516,38.868757801256,100
-77.05844056290393,38.86996206506943,100
-77.05788457660967,38.87253259892824,100
</coordinates>
</LinearRing>
</outerBoundaryIs>
<innerBoundaryIs>
<LinearRing>
<coordinates>
-77.05668055019126,38.87154239798456,100
-77.05542625960818,38.87167890344077,100
-77.05485125901024,38.87076535397792,100
-77.05577677433152,38.87008686581446,100
-77.05691162017543,38.87054446963351,100
-77.05668055019126,38.87154239798456,100
</coordinates>
</LinearRing>
</innerBoundaryIs>
</Polygon>
</Placemark> </kml>

```

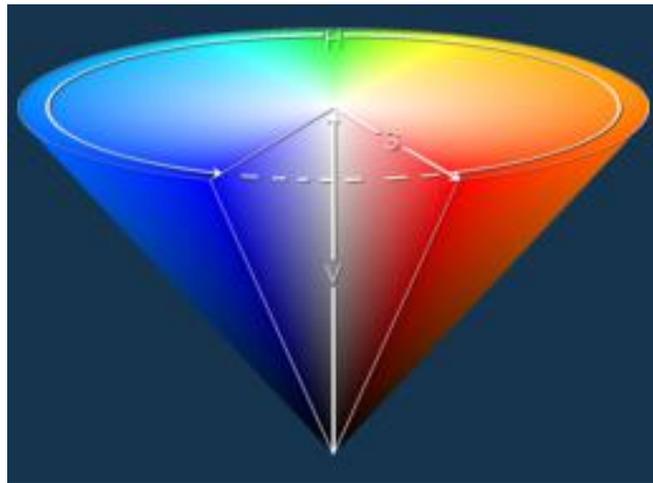
2.1.7. COLORES

Es un símbolo fundamental para representar la aglomeración de los eventos mediante el mapa de calor. Según (ArcGis, 2014) se puede clasificar en 6 modelos:

- **RGB:** un modelo de color aditivo formado por rojo, verde y azul. Esta es la opción predeterminada.
- **HSV:** un modelo de color aditivo formado por tono, saturación y valor (brillo).
- **HSL:** un modelo de color aditivo formado por tono, saturación y brillo (luminosidad).
- **Lab:** un modelo de color independiente del dispositivo formado por valores de color absoluto.
- **Escala de grises:** un modelo de color monocromático formado por tan solo un canal que varía de negro a blanco.
- **CMYK:** un modelo de color de sustracción formado por canales cian, magenta, amarillo y negro.

Para este proyecto se utiliza **HSV**. Este sistema de color se define en función de los valores de tres atributos: matiz, saturación y brillo. El matiz (*Hue*) hace referencia al tipo de color, va desde 0 a 360° y cada valor corresponde a un color; La saturación o intensidad indica la concentración de color en el objeto, su valor va entre 0 y 100% indicando la intensidad del color; El brillo (*Value*) denota la cantidad de claridad que tiene el color (tonalidad más o menos oscura). Cuando hablamos de brillo hacemos referencia al proceso mediante el cual se añade o se quita blanco a un color, su valor oscila entre 0 y 100% (L. Moreno, 2004).

Ilustración 16: Representación gráfica común del modelo de color HSV.¹⁴



2.2. ESTADO DEL ARTE

El siguiente estado del arte, se realizó por medio de la búsqueda de diferentes documentos y artículos en las bases de datos brindadas por la UNAB, especialmente la IEEE, además, de fuentes brindadas por el director de proyecto, dicha búsqueda se hizo por medio de palabras claves como: análisis espacial, método KDE, función K-Ripley, SIG, distribución espacial del dengue, las cuales nos permitieron encontrar y seleccionar los documentos más relevantes con respecto a la ubicación en la que se desarrolló el artículo, del mismo modo, se trató de escoger los documentos más recientes con el objetivo de obtener información más actualizada, en dichos documentos encontramos las diferentes estrategias de análisis espacial que se utilizaron de acuerdo a los problemas que se presentaban y que de alguna manera afectaban a una comunidad, por ejemplo, el dengue, Chikunguña, casos de hurto, actos ilegales, minas antipersonales, robos residenciales, entre otros. Dichas estrategias de análisis espacial las

¹⁴ Tomado de (Colorizer, 2014).

realizaron por medio, de la implementación de métodos como K-Ripley, KDE o utilización de algunos softwares como ArcGIS y crimestate, con el objetivo de identificar las zonas en las cuales se encontraba mayor aglomeración del evento, para así, enviar los diferentes recursos para combatir dicho evento de una manera óptima, reduciendo considerablemente la presencia del evento en dicha región, ya que se evidenciaba de una manera gráfica las zonas más críticas.

Tabla 1: Documentos, proyectos y revistas relacionadas con el proyecto Aplicación software de apoyo en la identificación de aglomeraciones de eventos

Referencia	Geographical Information Systems for Dengue Surveillance (Duncombe et al., 2012).	
Problema	Solución	Aporte
<p>El incremento del dengue en la región del Asia pacífico en países como: Filipinas, Indonesia, Vietnam y Tailandia. En donde aumentó en 30 veces llegando así a 50 millones de infectados anualmente. Por lo tanto, la prevención de los vectores causantes del dengue por medio de la fumigación química, control biológico, eliminación física de los sitios de reproducción y mejora de la infraestructura, además, la educación comunitaria. pero el factor más importante es para prevenir el dengue es por medio de la vigilancia debido a que permite identificar el número y la identificación de los casos y la gravedad de la enfermedad en una población o región.</p>	<p>Un SIG (Sistema de Información Geográfico) o GIS (por sus siglas en inglés) para la vigilancia del dengue, ya que este sistema permite identificar la distribución de la enfermedad y los cambios que va teniendo en el tiempo. Y es que el mapeo de la distribución del dengue les ha permitido la identificación visual de manera instantánea de las zonas con mayor riesgo y por lo tanto, hace que los recursos sean enviados a dichas áreas de forma más rápida.</p>	<p>El uso de un SIG para la vigilancia del dengue en una región con el motivo de identificar patrones de aglomeración de casos de dengues en una región en específico. Estos mapas pueden ser utilizados por el personal encargado de la salud pública para dirigirse a las regiones más afectadas, ya que, en muchas ocasiones los responsables de la formulación de políticas responden de manera más positiva a los mapas que a los números brutos o gráficos estadísticos.</p>
Referencia	<p>Web mapping GIS: GPS under the GIS umbrella for Aedes species dengue and chikungunya vector</p>	

mosquito surveillance and control (Palaniyandi, 2014).		
Problema	Solución	Aporte
<p>La necesidad para vigilar enfermedades como dengue y chikungunya y cómo esta epidemia ha aumentado año tras año en India y que ha causado miles de afectados y cientos de muertos, un país con grandes problemas de salud pública debido a las grandes aglomeraciones de personas, además, de otros factores que ayudan a crear un ambiente de propagación del mosquito, por ejemplo, las personas del sur de India, en épocas secas almacenan agua en estanques plásticos o de cemento. Se necesitaba conocer las especificaciones de las casas y datos de la densidad de los mosquitos en las diferentes zonas para delimitar vulnerables a las epidemias como el dengue o chikungunya en India específicamente en una ciudad llamada Pondicherry.</p>	<p>Un GIS de cartografía web para la vigilancia del dengue y chikungunya con el fin reconocimiento sistemático con un intervalo de 500 metros para evaluar la evolución de la población del mosquito en una zona delimitada y de igual forma el mapeo de los aspectos que ayudan a la reproducción del mosquito como botellas, neumáticos, llantas, cáscaras de coco, huecos de árboles etc. Con el fin de identificar las áreas vulnerables o con alta probabilidad de ser afectadas por el dengue para planificar medidas de control con el objetivo de impedir la cría, proliferación y transmisión del mosquito vector en Pondicherry.</p>	<p>La implementación de un SIG que tiene la capacidad de mapear la densidad de casos de dengue por medio de la visualización rápida y clara a través de colores graduados y mediante la implementación de métodos cuantiles y el uso de las coordenadas de cada uno de los casos ocurridos en un rango de tiempo y en una región específica. Con el fin de ser útil para la rápida toma de decisiones de las instituciones encargadas de las operaciones de salud pública.</p> <p>Otro aspecto clave fue de cómo este software ha impactado de manera positiva en India en la prevención y control de brotes de dengue ya que ha permitido que las personas tomen conciencia de la importancia de la reducción de las fuentes clave para la reproducción del mosquito vector.</p>
Referencia	<p><i>Using GIS and KDE analysis spatial distribution on public housing households: A case study</i> (Zuo Zhang, Yanzhong Liu, Biao Chen, & Kunlun Chen, 2013).</p>	

Problema	Solución	Aporte
<p>En Wuhan una región de China el gobierno proporciona viviendas públicas por medio de diferentes fuentes como nuevas casas, casas abandonadas o propiedades bajo alquiler a las personas de más escasos recursos. Actualmente, por el aumento de las ciudades, mayor población y el período de rápida urbanización, para dicha política, el gobierno de China tiene una enorme tarea de proporcionar vivienda digna a todo su pueblo a un precio asequible, por lo tanto, puede ser un problema encontrar una zona determinada y apta para la construcción de nuevas viviendas, además de la lucha por encontrar un hogar, una vivienda, la comodidad y la independencia que genera una casa para cualquier persona hace que esta tarea no sea tan fácil.</p>	<p>Uso de un Sistema de Información Geográfico y el método KDE, En este trabajo se presenta una metodología para la identidad densidad variable de familias de bajos ingresos que viven en viviendas públicas, a su vez crear una técnica de agrupación que determina los indicadores casuales más probabilidades de ser presentado en ciertos grupos, por lo tanto, ser capaz de comparar puntos de agrupación similares a través del tiempo y el espacio por medio del método KDE que les permitió estimar la densidad de los datos espaciales se hicieron diferentes clasificaciones de los tipos de viviendas con pesos probabilísticos, otra clasificación que se usó para localizar las áreas de viviendas fue la clasificación de las diferentes zonas dependiendo de su densidad poblacional, facilidad de acceso a transporte y servicios públicos, entonces en el proceso KDE calculan las familias de bajos ingresos que viven en viviendas públicas en la región de Wuhan y herramientas SIG se emplean para</p>	<p>Implementación del método KDE que nos permitirá hallar la densidad de la probabilidad de que ocurran los eventos en una región delimitada, entonces el método realiza una división del área en cuadrículas o celdas y se grafica el dato plenamente identificado por otra parte se hace un coloreado al punto a una distancia circular con un radio alrededor de cada punto y la suavización de color a medida que se aleja del evento graficado, dicho punto puede ser un caso de dengue , caso de hurto o cualquier eventualidad de la cual se esté trabajando dicha representación gráfica permitirá al usuario hacer las diferentes tomas de decisiones de una manera más fácil.</p>

	hacer el análisis de la distribución espacial.	
Referencia	<i>Application of density analysis for landmine risk mapping</i> (Alegria, Sahli, & Zimanyi, 2011).	
Problema	Solución	Aporte
La evidencia de problemas que presentan las minas terrestres para la seguridad pública nacional en muchos países debido a que no se conoce la distribución de dichas minas y que su localización requiere mucho trabajo debido a que se realiza en muchos casos de manera manual, además, generan cientos de víctimas al año en un país como Colombia y por supuesto la falencia de un software o herramienta de investigación para el análisis espacial, la vigilancia y evaluación de los riesgos que permite el apoyo a las diferentes decisiones por medio del procesamiento de distintas fuentes de datos ya sean cualitativos o cuantitativos, por ejemplo datos históricos de antecedentes.	El desarrollo de un sistema de apoyo en la toma de decisiones SIG, con el propósito de detectar zonas con alta vulnerabilidad de poseer minas en las diferentes regiones de Colombia, por medio de la utilización de diferentes métodos como el K de Ripley y el KDE (Kernel Density Estimator), El primero fue utilizado con el objetivo de estudiar o evaluar el área de agrupación o dispersión espacial de los diferentes casos de minas terrestres en un rango de distancias y el método KDE para la transformación de los eventos, casos de punto a una superficie plana o ubicación plana para lograr la localización, extensión o distribución y la intensidad de aglomeración de los casos en puntos calientes por otra parte se toma en cuenta como la población, la vecindad, red de transporte influyen en la distribución espacial de los incidentes.	La implementación de los métodos K de Ripley y KDE en nuestro sistema de información geográfica. El primero, con el objetivo de la distribución de los diferentes eventos en un rango de distancia en una región, como por ejemplo los casos de dengue en el municipio de Floridablanca y el KDE para convertir las aglomeraciones de los eventos en puntos localizados y por supuesto permite la clasificación y la representación gráfica de cada uno de los eventos por medio de puntos de calor y los diferentes esquemas de color a integrar en el software.
Referencia	Análisis espacial de transmisión del dengue en	

	una ciudad de tamaño medio (JC, 2005).	
Problema	Solución	Aporte
<p>La tendencia de aumento de la transmisión de enfermedades como el dengue, en los años 1990 al 2002 en Brasil, presentándose en los 25 de los 27 estados del país ya que por cada 100.000 habitantes; en el 33% entre 1000 y 1999; y en el 26% entre 2.200 y 4.999. En el 5% de los tractos, la transmisión alcanzó 5.000 o más casos por 100.000 habitantes y en al menos una zona, se confirmaron 18.200 casos por 100.000 habitantes, que asciende a aproximadamente el 20% de los residentes. El aumento de la enfermedad se da debido a la variedad de factores de riesgo que están asociados con la presencia de la enfermedad y el vector, los factores más relevantes fueron; el crecimiento demográfico, la urbanización inadecuada, la migración, el transporte aéreo y el deterioro de los sistemas sanitarios. Lo que hace que el control de la enfermedad sea difícil en dichas áreas, además como sucede en muchos países latinoamericanos los recursos son</p>	<p>El uso de una herramienta de software para el análisis espacial llamado Mapinfo Professional que permite a los usuarios visualizar, analizar, editar, interpretar, comprender y generar datos para revelar relaciones, patrones y tendencias en una región. Y el mapa base de São José do Rio Preto en el Universal Transversa de Mercator.</p> <p>El objetivo del trabajo fue analizar la transmisión del dengue en un municipio de tamaño y población medio con la función de mejorar, adaptar y optimizar el uso de los instrumentos de vigilancia y control disponibles en el país.</p> <p>De un total de 14.431 casos de dengue que se produjeron en el área urbana del municipio entre los años 1994 y 2002 se lograron georreferenciar un total de 13.998, por su parte, los 433 casos que no pudieron georreferenciar debido a que su dirección no coincidía o no se encontraba en el mapa.</p>	<p>El uso de un banco de datos notificados y confirmados por diferentes entidades del estado o privadas y exista la necesidad de estudiar o analizar el comportamiento de los eventos por medio del sistema de análisis espacial con el objetivo de encontrar patrones de distribución, transmisión y relaciones de ocurrencia en la zona que está siendo evaluada, además de detectar condiciones favorables para que un evento ocurra dentro de dicha región con el objetivo de optimizar los recursos, adoptar una única estrategia de vigilancia, control y focalizar el trabajo, una vez que se han identificados estas áreas prioritarias con el mayor porcentaje de puntos calientes para controlar que dicho evento siga sucediendo.</p>

limitados.		
Referencia	Distribución espacial del dengue basado en herramientas del Sistema de Información Geográfica, Valle de Aburrá, Colombia (Londoño C., Restrepo E., & Marulanda O., 2014).	
Problema	Solución	Aporte
El aumento de las enfermedades tropicales como la malaria, el dengue, la enfermedad de Chagas, la leishmaniasis, la fiebre amarilla en Colombia y las cuales se habían convertido en un verdadero problema de salud pública en el país en los años de los 80s a los 90s mostrando así un patrón endemo-epidémico, debido a que, diferentes enfermedades aumentaron en un 65% en localidades como la ciudad de Medellín, y por lo tanto la necesidad de hallar estrategias de prevención para los brotes de dichas enfermedades.	El uso de un SIG para el estudio de los patrones de la distribución espacial de enfermedades infecciosas. como caso de estudio ellos tomaron los datos de dengue reportados por la Secretaría de Salud del Municipio de Medellín para los años 2008, 2009 y 2011; para lograr georreferenciar los casos de dengue, por lo tanto hicieron el diseño e implementación de un modelo gravimétrico que le permitió estudiar la distribución espacial del fenómeno en el área estudiada, por otra parte, se utilizó con el fin de observar la variación espacial del fenómeno con respecto a un punto de ocurrencia del evento para conocer el comportamiento del dengue en dicha región.	La cultura del uso de los SIG en el sector salud, como en otros sectores con el fin de monitorear diferentes eventos que afectan a la sociedad de la actualidad y de la cual se desea conocer su comportamiento, esta necesidad de vigilancia permitirá el incremento de los SIG y diferentes herramientas de análisis espacial en la medida en que puedan ser reconocidas sus ventajas y en especial sus favorables efectos en la vigilancia, control y protección de la salud y el bienestar de la población.
Referencia	<i>A method based on spatial analyst to detect hot spot of urban component management events</i> (Dai, Jing, Du, & Zhou, 2015).	
Problema	Solución	Aporte
El rápido crecimiento de las instalaciones públicas en zonas urbanas permite	Se propusieron métodos y flujos de análisis espacial con el objetivo de hallar	Uso de métodos de análisis espacial como por ejemplo la función K

<p>el incremento de diferentes eventos que pueden afectar a una sociedad, dichos eventos pueden ser la acumulación de desechos, anuncios ilegales entre otros. Otro problema que ellos encontraron es que no existen muchas investigaciones sobre eventos de gestión urbana y ésta sigue haciéndose de manera tradicional lo que hace que el trabajo de hallar patrones de correlación del evento sea más tedioso y lleva más tiempo analizar los datos, además ignora las características de la distribución espacial, por lo tanto, se le es más difícil para las diferentes entidades encargadas de la administración urbana la toma de decisiones y permita resolver los eventos a tiempo.</p>	<p>los diferentes puntos calientes de un determinado evento en una región y de esta forma apoyar en la toma de decisiones.</p> <p>Lo que ellos hicieron fue un conteo de los datos ocurridos en el mismo lugar, además agrupar datos con las mismas características con el objetivo de medir la autocorrelación espacial de los eventos agrupados con la ayuda de la I de Moran Global, por otra parte, ellos implementaron la función K de Ripley que les permitió hallar el patrón de distribución espacial de los eventos, luego para identificar el borde espacial de los puntos calientes, se empleó el análisis de puntos calientes espaciales.</p> <p>El resultado fue la de detección del hotspot, la mayor densidad de agrupamiento de hotspots en este caso de ocurrió dentro del área de negocios y lugares donde la población fluye con mayor frecuencia.</p>	<p>de Ripley con el fin de hallar patrones de distribución espacial, ya sea de agrupamiento o dispersión de los eventos estudiados, con el objetivo de encontrar zonas de puntos calientes y luego graficar dichos puntos en la zona en la que se está analizando el comportamiento de los eventos, para que sea más fácil la toma de decisiones por parte del usuario de nuestro software.</p>
<p>Referencia</p>	<p><i>Crime Hotspots: An Evaluation of the KDE Spatial Mapping Technique</i> (Johansson, Gåhlin, & Borg, 2015).</p>	
<p>Problema</p>	<p>Solución</p>	<p>Aporte</p>
<p>La delincuencia común</p>	<p>El mapeo de puntos</p>	<p>La implementación del</p>

<p>como por ejemplo robos residenciales y las afectaciones que esto trae a la comunidad, por ejemplo en el año 2004, el 78% de los aproximadamente 5,9 millones de delitos reportados en algunas regiones de Inglaterra y Gales fueron delitos como los que se mencionaba anteriormente además se dice que la menor parte de los delincuentes cometen la mayoría de los robos residenciales, los cuales se dificulta a las autoridades el detectar manualmente patrones y series entre los delitos. Por lo tanto, la necesidad de predecir o pronosticar con precisión los casos de hurtos cometidos es clave para ayudar a reducir dicho delito, se tiene entendido que los casos de hurto no se distribuyen uniformemente si no que tienden agruparse, es decir, que están presentes en algunas partes del área estudiada mientras que en otras zonas no se presentan casos.</p>	<p>calientes como método analítico y el método KDE para mapear y predecir patrones de distribución espacial en los datos estudiados, lo datos que ellos tomaron fueron proporcionados por el Departamento de Policía de Suecia y contienen información sobre 5, 681 robos residenciales reportados en la parte sur de Suecia. Los datos consisten en información espacial (longitud y latitud, así como código postal) y fecha.</p> <p>Se demostró que el uso de puntos calientes para el mapeo del crimen junto con el método KDE pueden ayudar a la reducción de los delitos en las áreas que se detectó alta densidad de delincuencia, usaron el método KDE ya que se encontró mucho más eficiente y eficaz en términos de predicción debido a los datos que ellos tenían y la forma en la que debían presentar los datos, para el estudio correspondiente ellos utilizaron la biblioteca de Scipy stats.gaussian.kde.</p>	<p>método KDE estimación de densidad de Kernel ya que, Además, de ser una de las técnicas más populares en el mapeo, debido a que utiliza todos los puntos de datos de dengue en el área geográfica investigada por ejemplo Floridablanca y nos permitirá producir una estimación de densidad del fenómeno o evento y de igual forma producir el mapa con tonos de colores suaves, degradados y precisos.</p>
<p>Referencia</p>	<p><i>GIS mapping and spatial analysis of cybersecurity attacks on a Florida university</i> (Hu, Baynard, Hu, & Fazio, 2015).</p>	
<p>Problema</p>	<p>Solución</p>	<p>Aporte</p>
<p>La necesidad de hallar</p>	<p>Mapeo GIS para saber</p>	<p>El uso de funciones de</p>

<p>patrones y relaciones en el origen de los ataques cibernéticos que sufren a diario las diferentes entidades gubernamentales, pero en específico, la UNF, (Universidad del Norte de Florida), en los Estados Unidos, que, de enero a marzo de 2014, hubo 1,300,000 ataques contra esta entidad, entre los cuales hubo 72 categorías filtradas. Los ataques principales son exploits, spyware, virus, vulnerabilidades, DDoS y reconocimiento, dichos ataques son perpetrados por piratas informáticos, especialmente los financiados por un gobierno extranjero, son capaces de romper los cortafuegos y explotar vulnerabilidades en el software que atacan. También son expertos en engañar a personas confiadas o sencillas mediante estafas o diferentes técnicas de hackeo con el objetivo de buscar y capturar información que es restringida por ejemplo investigaciones, descubrimientos valiosos y novedosos, dicha información llega a ser muy lucrativa para los que la poseen.</p>	<p>dónde se originan estas intrusiones lo que permite una defensa más específica, por lo tanto, parten de dos aspectos fundamentales contra la ciberdelincuencia, primero que todo la detección de intrusos y luego, la prevención de intrusos. La detección se encarga de registrar, clasificar y notificar al personal de seguridad de un evento y la prevención corresponde realizar acciones apropiadas, como bloquear el tráfico, poner en cuarentena o enviar paquetes de restablecimiento, pero el enfoque que ellos le dieron fue el de sólo detectar los eventos.</p> <p>La UTF usa Tipping Point, es un software para el monitoreo y prevención de ingresos no deseados en la red, este software tiene la capacidad de detectar a qué dispositivo de la red de la universidad está siendo dirigido el ataque, por ejemplo, impresoras, estaciones de trabajo, computadoras de escritorio, servidores web, tabletas y teléfonos y de cada eventualidad de intrusión proporciona la hora del día y filtra los ataques por métodos de ataque.</p> <p>Para la geo codificación</p>	<p>estadística y análisis espaciales en los datos con el objetivo de determinar si el patrón del evento es aleatorio o agrupado y tratar de establecer el comportamiento del fenómeno con el pasar de los años. Por otra parte, refinar y automatizar la preparación de datos para el análisis espacial continuo para monitorear y mapear conjuntos de datos cada vez más grandes que incorporan períodos de tiempo más largos, lo que permitirá que la predicción del comportamiento sea más exacta y, por lo tanto, les sea más fácil controlar dicho fenómeno.</p>
--	--	---

	de los datos lo hicieron mediante la herramienta GEO2 Lite, este les permite mapear patrones entre los eventos, para ellos saber de dónde se originan los eventos permite una defensa más específica.	
Referencia	<i>A Spatial Analysis of Selected Categories of Offences in Krakow Based on Data from the National Safety Risk Map</i> (Polonczyk & Lesniak, 2018).	
Problema	Solución	Aporte
Conocer la distribución espacial de dos tipos de delitos, actos de vandalismo y el consumo de alcohol en lugares prohibidos, cometidos en la ciudad de Cracovia entre los años 2016 y 2017. Aunque existía ya una plataforma llamada Mapa Nacional de Riesgos de Seguridad, la cual que fue lanzada por la policía de Polonia con el objetivo de identificar y visualizar diferentes tipos de delitos y delitos cometidos, por otra parte, dicha plataforma muestra un mapa interactivo en la cual un usuario puede por medio del uso de una ubicación de red determinada, enviar, no más de una vez al día, un informe anónimo que indique la infracción particular cometida, así como su ubicación exacta	<p>Análisis y representación gráfica de datos por medio del uso de software ArcGIS y CrimeStat, ellos mapean los dos tipos de delitos en mapas y luego hicieron un análisis para determinar la relación y/o dependencias existentes entre los eventos por medio de la implementación de dos métodos, análisis espacial de regresión y una estimación de la densidad del núcleo dual con un análisis de "punto caliente".</p> <p>Los datos que se estudiaron fueron los que habían sido verificados por la policía, mientras que los que habían sido desestimados o todavía estaban en espera de verificación fueron ignorados, además sólo se analizaron los delitos dentro del área de la</p>	Presentación visual de los resultados obtenidos en mapas de calor, por medio del análisis a los datos de un evento, con la implementación de la función K de Ripley y el método KDE, ya que se demuestra que dicho resultado gráfico que muestra la distribución de un fenómeno, permite que las diferentes autoridades o entidades encargadas de hacer control sobre algún evento que ha venido afectando a una comunidad, lo haga de manera más eficiente debido a que se puede llegar a identificar las zonas de una región en las que se halla el mayor número de eventos, lo que permitirá que sea más rápido tomar medidas para tratar reducir el fenómeno.

<p>en el mapa, El gran desafío al que se enfrentaron fue la muestra de datos relativamente pequeña, porque la plataforma de la policía llevaba pocos meses que estuvo en servicio (aproximadamente 15 meses), por otra parte, a la distribución heterogénea de los datos.</p>	<p>ciudad de Cracovia, entonces, se realizaron el mapa con una dispersión de puntos (casos de delitos), que a su vez mostraba la distribución espacial de ambas categorías de delitos que sucedieron en la ciudad.</p> <p>Los resultados del estudio pueden llegar a constituir una fuente importante de conocimiento para las diferentes agencias locales encargadas de dar cumplimiento de la ley en sus esfuerzos por garantizar una mayor vigilancia y observación de las zonas en las que se halló mayor número de datos reportados.</p>	
---	---	--

2.2.1. HERRAMIENTAS COMPUTACIONALES:

A continuación, se presentan una serie de herramientas computacionales más relevantes que implementan métodos de análisis espacial y patrones de puntos.

- **CrimeStat:** Es un programa de estadísticas espaciales para el análisis de ubicaciones de incidentes de delitos que incluye el método KDE, desarrollado por Ned Levine & Associates, este programa actualmente está siendo usado por diferentes entidades encargadas de hacer cumplir la ley, ya que reduce los esfuerzos en el mapeo del crimen, lo que les permite tomar decisiones de manera más rápida, los datos (ubicaciones de robos), son ingresados a el sistema por medio de archivos en formato dbf, shp, ASCII u ODBC que trabajan con coordenadas de proyección o esféricas. El programa está dividido en 5 secciones o módulos, configuración de datos, descripción espacial, modelado espacial, crimen modelado de demanda de viajes y Opciones (CrimeStat, 2005).
- **PASSaGE:** Es un software gratuito usado para análisis estadísticos espaciales y de patrones de punto, sobre datos biológicos y otro tipo de datos, los datos pueden ser importados a el software desde archivos de texto, hojas de cálculo, shapefiles, entre otros, la exportación de datos

también es muy flexible con muchos formatos de archivos permitidos, Por otra parte, el software posee una interfaz gráfica muy amigable para el usuario, actualmente tiene miles de descargas en cerca de 60 países y es usado en más de 170 instituciones de los Estados Unidos, el Software fue desarrollado por Michael S. Rosenberg y Corey D. Anderson, la sección de análisis del software se divide en: Análisis de patrones de puntos, análisis de datos dispersos y datos de las líneas o polígonos (Rosenberg & Anderson, 2011).

- **SANET:** Es una caja de herramientas para analizar eventos que suceden en redes, algunos ejemplos típicos pueden ser, accidentes automovilísticos en carreteras y salones de belleza en calles céntricas, robos en calles peatonales, tiendas minoristas a lo largo de calles en un área urbanizada, refugios de castores en arroyos respectivamente. Dicha herramienta se permite hacer diferentes análisis, por ejemplo, diagramas de Voronoi, estimación de la densidad del núcleo, método de distancia global vecino más cercano automático, método de la función Auto K global, método de la función de cruz global K, interpolación, Método de agrupamiento de puntos, entre otras herramientas que constituyen a este robusto software que fue desarrollado por investigadores de la Universidad de Tokio, Japón (SANET, 2009).
- **GeoDa:** Es una herramienta de código abierto y gratuita que se usa para el análisis de datos espaciales y temporales, el software proporciona una interfaz gráfica, intuitiva y fácil de usar para los métodos de análisis espacial de datos espaciales, como patrones de autocorrelación espacial para los datos agregados, detección y comparación de mapas y gráficos generados. Actualmente, es usado por laboratorios en universidades como Harvard, MIT y Cornell. El programa permite archivos de entrada shapefiles, geodatabases, GeoJSON, MapInfo, GML, KML y otros formatos de archivos. Así pues, permite conectarse con la base de datos CartoDB para el almacenamiento de resultados (GeoDa, 2016).
- **ArcGIS:** Es el conjunto de herramientas de software se usa para la captura, mapeo, edición, análisis de información geográfica, este robusto software posee herramientas como ArcReader, ArcMap, ArcCatalog, ArcToolbox, ArcScene y ArcGlobe, además de diversas extensiones, herramientas de desarrollo, servidores y soluciones, que lo hacen el sistema más completo y líder en lo que comprende el análisis espacial de patrones de puntos (Esri, 2018).
- **ArcGIS Spatial Analyst:** Es una extensión del robusto sistema de información geográfico ArcGIS, implementa el método KDE, permitiendo estudiar distribuciones espaciales de los datos, con el objetivo de mostrar aglomeraciones entre los puntos de eventos y hallar zonas con puntos caliente, para que sea más fácil predecir el comportamiento de un fenómeno, por otra parte, permite hacer un análisis de distancia y costo de

viaje en línea recta de un punto y diferentes operadores matemáticos (Esri, 2004).

- **Kosmo Desktop:** Es una plataforma SIG de código abierto, desarrollado en el lenguaje de programación Java, y la implementación de diferentes bibliotecas como Geotools, Castor, Ermapper, GDAL, Log4J y de la plataforma JUMP. El diseño y arquitectura de la plataforma se basan en el análisis y gestión de la información territorial a través de bases de datos espaciales, la primera versión estable fue entregada en 2006, de ahí en adelante, se han entregado 7 versiones (Kosmo, 2013).
- **QGIS:** Es un robusto Sistema de Información Geográfica, de código abierto, desarrollado por QGIS Development Team, lanzado en el año 2002, QGIS admite archivos vectoriales como Shapefile, Arcinfo, Mapinfo, GRASS GIS, DXF, DWG, asimismo, de diferentes tipos de archivos ráster, implementa diferentes métodos para hacer análisis espacial que permiten conocer la distribución de un fenómeno y distancia a el centro más cercano, permite visualizar los eventos mediante mapas de calor y puntos, su versión más reciente fue presentada en junio de 2018 (QGis, 2013).
- **gvSIG Desktop:** Es un potente Sistema de Información Geográfica con precisión cartográfica, de código abierto, fue desarrollado en el lenguaje de programación Java e implementación de librerías como Geotools, permite formato de archivos vectoriales con extensión SHP, GML, KML, DXF, DWG, DGN, además de múltiples archivos ráster, acceso a bases de datos, construcción de mapas, vista 3D, importación y exportación de scripts y muchos otros Geo procesos típicos de un SIG (gvSIG, 2015).

3. MATERIALES, INSTRUMENTOS, Y/O METODOLOGÍA

3.1. METODOLOGÍA DE DESARROLLO

El presente proyecto fue desarrollado por dos personas, por lo tanto, fue necesario establecer una dinámica de desarrollo de trabajo en grupo, debido a esto, se implementó un proceso de desarrollo por etapas e iterativo, el cual fue la metodología ágil de desarrollo de software Crystal clear, desarrollada por Alistair Cockburn.

Utilizamos esta metodología debido a que minimiza el riesgo de error en el proceso de desarrollo de software, además está diseñada para equipos de trabajo pequeños, ya que esta metodología se centra 3 prioridades que son: “entrega frecuente, comunicación osmótica y mejora reflexiva, por consiguiente, aprovecha el tamaño pequeño del equipo y la proximidad para fortalecer la comunicación dentro del equipo de trabajo” (Cockburn, 2004, p.34)., debido a esto, hace que la metodología tenga una capacidad adaptiva, es decir, hace que sea más fácil realizar cambios en el proyecto en el momento que se está desarrollando, a partir de nuevas necesidades que se den.

- **Entrega Frecuente:** Una de las propiedades más relevantes de un proyecto, es la de entregar código funcional y en ejecución para usuarios reales, lo que puede traer numerosas ventajas para el equipo de trabajo, ya que permite que los usuarios den una realimentación en cada iteración y la respectiva mejora reflexiva, aunque no existe un período establecido para cada una de las entregas, se dice que no debe exceder los cuatro meses.
- **Mejora reflexiva:** La reunión del equipo de trabajo una vez finalizada cada iteración es importante y no conlleva mucho tiempo con respecto a los beneficios que tiene dialogar los hábitos de desarrollo, conocer las dificultades que se presentaron durante la iteración anterior e identificar las prioridades para la siguiente iteración, lo que permitirá reflexionar y encontrar nuevas estrategias de trabajo, por ejemplo, “los equipos pueden probar, en varias formas: programación de pares, pruebas de unidad, desarrollo guiado por pruebas, sala simple versus múltiple asientos en la sala, diversos niveles de participación del cliente e incluso diferentes longitudes de iteraciones, estas son todas las variaciones apropiadas dentro de Crystal Clear” (Cockburn, 2004, p.39). Si es necesario, con el objetivo de mejorar el funcionamiento del equipo de desarrollo.

- **Comunicación osmótica:** Es la versión de la comunicación más poderosa que pueden realizar los pequeños proyectos, ya que permite que los integrantes del equipo de trabajo mantengan una comunicación cercana, generando que el conocimiento se difunda de manera eficiente y rápido, y exista la proporción de realimentación deseada con el objetivo de identificar y corregir pequeños errores antes de que estos se conviertan en errores más grandes. Para que la comunicación osmótica sea la deseada, se necesita que los integrantes del equipo estén muy cerca el uno del otro.

Aunque, existen un total de 7 propiedades pertenecientes a la metodología Crystal, son necesarias, para el correcto funcionamiento y finalización de un proyecto, las tres desglosadas anteriormente, por ser las de mayor relevancia. No obstante, la seguridad personal, el enfoque, fácil acceso a usuarios expertos y un entorno técnico con pruebas automatizadas, permiten que el equipo se adentre en la zona de seguridad, por lo tanto, se genere un mejor rendimiento del equipo.

En resumidas cuentas, la metodología de desarrollo de software Crystal Clear consiste en la seguridad personal que da a las personas, el coraje para compartir lo que descubren, la comunicación osmótica aporta a una mejor oportunidad de descubrir información importante entre ellos, por otra parte, la mejora reflexiva les da un canal para aplicar retroalimentación a su proceso de trabajo, el fácil acceso a los usuarios expertos les permite descubrir rápidamente información relevante de los usuarios, la frecuente la entrega genera comentarios sobre los requisitos del sistema y el proceso de desarrollo, el entorno de desarrollo técnico incluye pruebas automatizadas, configuración, la gestión y la integración frecuente permiten a las personas realizar cambios seguros en el sistema, el enfoque permite al equipo gastar su energía bien en las cosas más importantes (Cockburn, 2004, p.60).

Pese a que, Crystal Clear no requiere ninguna estrategia, si es bueno tener herramientas a la mano para poder guiar el proyecto en los primeros meses de su desarrollo y lograr sus respectivas entregas, las estrategias seleccionadas por Cockburn son:

- **Estudio 360°:** Al comienzo del proyecto se debe establecer el equipo que se necesita, valor del proyecto, plan de proyecto, requerimientos, tecnología a utilizar y objetivos a cumplir.
- **Victoria temprana:** Busca lograr la primera pieza de código probado, por lo general consiste en una pequeña función del software a desarrollar, por lo general, no es mucho más que la capacidad de agregar un elemento a la base de datos del sistema, por otra parte, Cockburn nos recomienda no empezar por

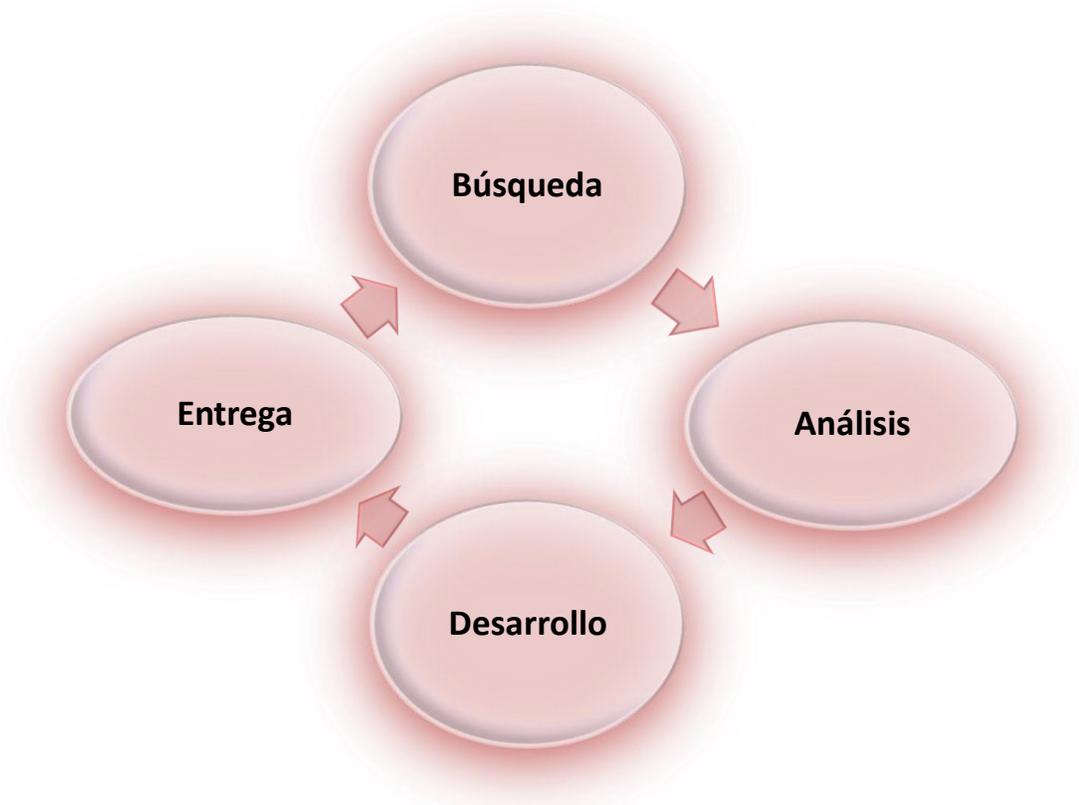
lo más difícil del proyecto, ya que se puede llegar a generar problemas de frustración en el momento en el que no se pueda lograr el objetivo.

- **Esqueleto caminante:** Es una pequeña implementación del sistema que realiza una pequeña función en específico, no es completo ni robusto, sólo corre, pero no es un código que se desecha, todo lo contrario, está destinado a crecer con el sistema a medida que este se va desarrollando.
- **Cambio de arquitectura incremental:** La arquitectura del sistema debe ir evolucionando a través de cada iteración, manteniendo un avance incremental y el sistema funcionando, hasta llegar a la funcionalidad final del sistema o deseada por el cliente.
- **Radiadores de información:** Generalmente suele ser un gran cartel de papel, ubicada en una zona, en la que pueda ser vista por personas externas a el equipo del proyecto, por ejemplo, la sala del equipo de desarrollo o el pasillo, su objetivo es mostrar información actualizada y relevante como, por ejemplo, “el conjunto de trabajo de la iteración actual (casos de uso o historias), las asignaciones de trabajo actuales, el número de pruebas escritas (o aprobadas), el número de casos de uso (o historias) entregados, el estado de los servidores clave (arriba, abajo, en mantenimiento), el núcleo del modelo de dominio, los resultados del último taller de reflexión” (Cockburn, 2004, p.37).

Por otra parte, la documentación de la aplicación software que representa su funcionamiento por medio de los diferentes diagramas, se realizó por medio de lenguaje unificado de modelado, UML. Se uso este lenguaje de modelado debido a que tiene un mayor soporte a la planeación y al control de proyectos, ayuda a minimizar costos en el caso en que se realice un buen diseño, además, que es el más conocido y utilizado en la actualidad, lo cual lo hace un lenguaje de modelado idóneo para visualizar, especificar, construir y documentar la aplicación software de nuestro proyecto.

Como metodología de desarrollo del proyecto general se dividió el trabajo en dos partes, en la primera parte iniciamos con la fase de búsqueda de la información y la documentación necesaria para el proyecto, en la segunda fase tenemos el análisis de los mismos, en la fase posterior tenemos el inicio del desarrollo del software iniciando por los módulos de carga y lectura de datos y como fase final de esta primera parte tenemos la entrega tanto del documento como del software hasta el momento.

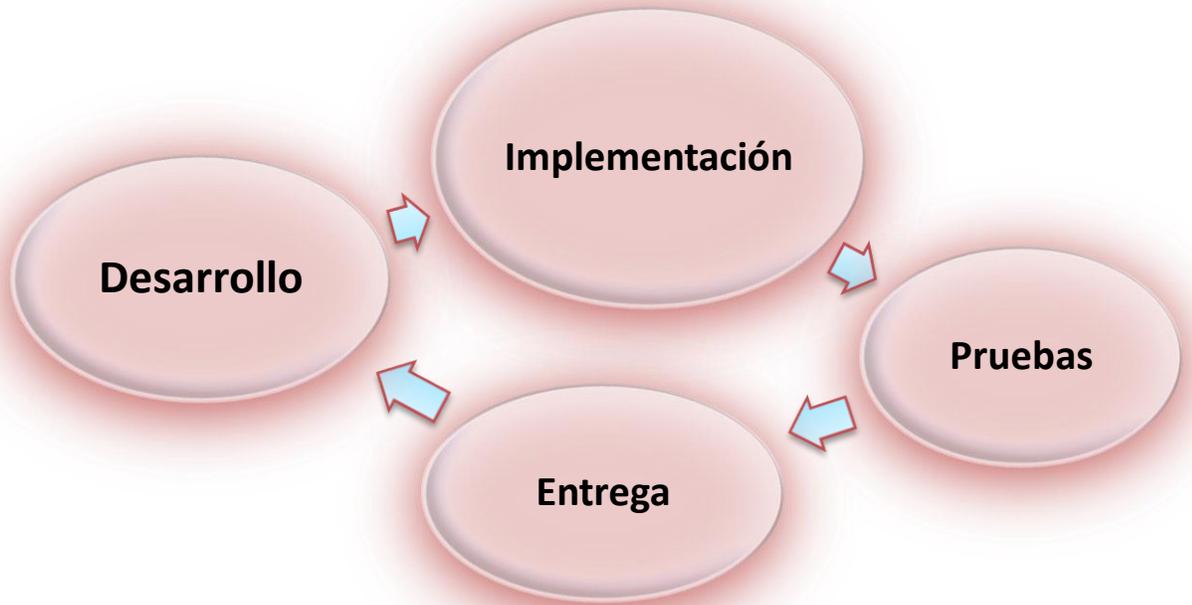
Ilustración 17: Metodología primera parte¹⁵.



En la segunda parte, iniciamos con la fase del desarrollo de todo el software, continuamos con la fase de la implementación donde se utilizó el software con los datos proporcionados del dengue en el año 2015, para la fase posterior tenemos las pruebas del software, donde se validará que cada componente del software trabaje correctamente y los datos arrojados por el mismo sean congruentes, de no ser así, se reanudaba el ciclo hasta la fase de desarrollo para la corrección de los respectivos errores, para proceder a su fase final la cual sería la entrega del software.

¹⁵ Fuente los autores

Ilustración 18: Metodología segunda parte.¹⁶



3.2. DEFINICIÓN DE LA ARQUITECTURA DEL SOFTWARE

La aplicación software construida maneja métodos de estimación de densidad como el KDE y comportamiento de puntos K de Ripley, asimismo, maneja tamaño de archivos relativamente grandes, debido a lo anterior se determinó que los tiempos de cálculo y respuesta podrían llegar a ser extensos, por lo tanto, se decidió desarrollar una aplicación software de escritorio y no una aplicación web o de otro tipo, el software fue desarrollado en el lenguaje de programación Java, ya que, además, de destacarse como uno de los lenguajes de programación más sencillos de utilizar y de los más fáciles de dominar, es un lenguaje multiplataforma, es decir correrá en cualquier sistema operativo, por otra parte, hace un manejo automático de memoria, lo que mejora el procesamiento de los cálculos necesarios, utilizamos el IDE NetBeans porque es un entorno de desarrollo muy completo, seguro y profesional, ya que implementa muchas funcionalidades que permite un desarrollo más ágil de distintos tipos de aplicaciones.

¹⁶ Fuente los autores

Utilizamos la tecnología de software JavaFX, que, implementada junto con Java, permite crear y desplegar aplicaciones con un aspecto vanguardista y contenidos avanzados, por otra parte, es gratuita, además, mejora la potencia de Java haciendo que los desarrolladores puedan implementar cualquier biblioteca de Java en aplicaciones JavaFX. Por lo tanto, se puedan crear aplicaciones con un mejor atractivo visual.

En el caso de los datos de entrada como lo son puntos de fenómeno y del polígono del área, que utiliza la aplicación software son archivos necesariamente con extensión KML (*Keyhole Markup Language*), el cual es un lenguaje de marcado basado en XML, que es implementado por Google Earth, es usado para almacenar datos geográficos, como entidades ráster, que incluyen puntos, polígonos, líneas, e imágenes.

4. DISEÑO DE LA APLICACIÓN SOFTWARE

4.1. DIAGRAMA DE CASO DE USO.

Ilustración 19: Diagrama de caso de uso¹⁷

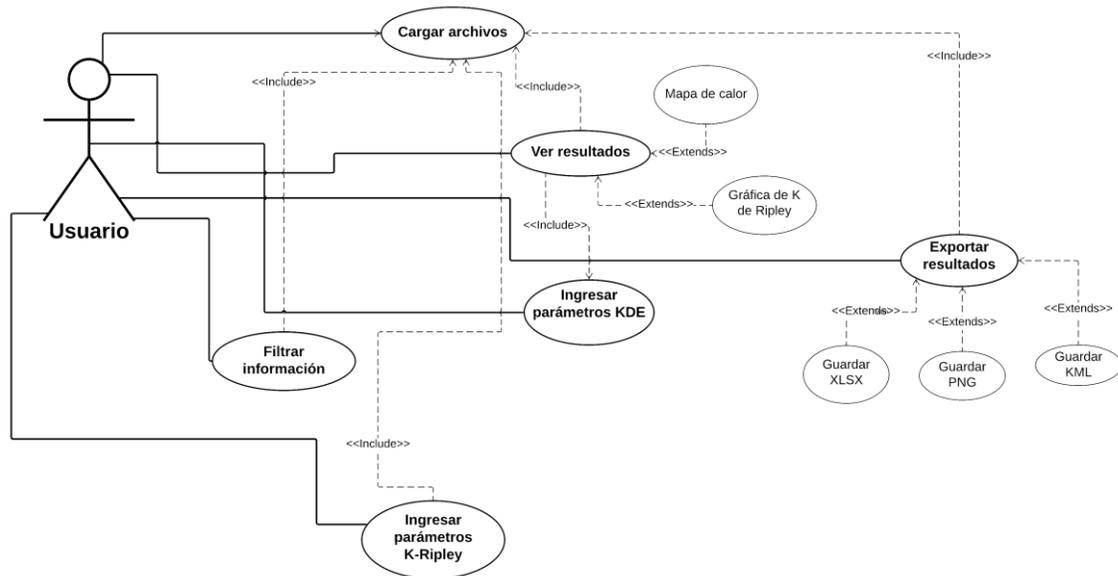


Tabla 2: Diagrama de caso de uso

No	Caso de uso	Incluye	Extiende	Autor(es)
1	Cargar archivos			Usuario
2	Ingresar parámetros KDE	1		Usuario
3	Filtrar información	1		Usuario
4	Ingresar parámetros K-Ripley	1		Usuario
5	Ver resultados	1,4	Ver mapa, ver gráfica K de	Usuario

¹⁷ Fuente los autores

			Ripley	
6	Exportar resultados	1	Exportar en KML, PNG o XLSX	Usuario

A continuación, describimos todos los casos de uso identificados por medio de flujos básicos y alternativos, a través de diagramas de actividades UML, lo cual permite la identificación de objetos y métodos de programación, y los diferentes mensajes entre objetos, gracias al diseño de los diagramas de secuencia UML.

4.2. DIAGRAMAS DE ACTIVIDAD

1. CARGAR ARCHIVOS

Este caso de uso permite la carga de los archivos KML de los puntos del fenómeno que se quiere evaluar y del polígono de la región sobre la cual se va a hacer el análisis espacial por medio del método KDE y la función K de Ripley.

Actores

Esta acción la hará el usuario directamente tan pronto ingrese a la aplicación software, será el primer paso, para poder continuar con los diferentes procesos.

Precondiciones:

No existe ninguna precondición para este caso de uso.

Flujo básico

- 1) El usuario selecciona el botón cargar archivos
- 2) Selecciona cada uno de los archivos KML respectivamente, en el orden que desee.
- 3) Los archivos son procesados correctamente por el software
- 4) Los archivos se cargan o guardan en memoria

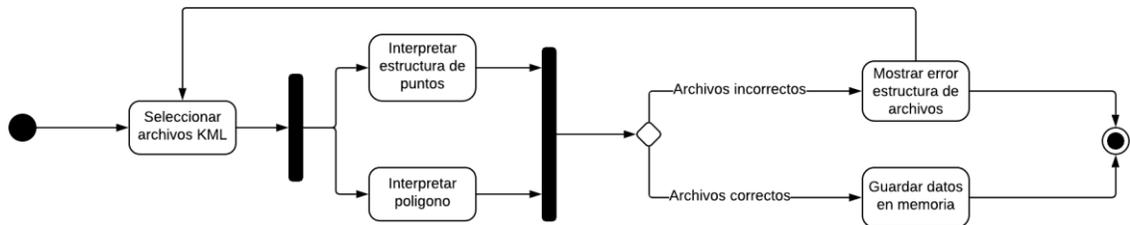
Flujo alternativo 1

En el paso 3

- 5) Los archivos son procesados correctamente por el software, debido a un error de estructura de los archivos
- 6) Se muestra un error

- 7) El usuario regresará a la ventana de carga de archivos y tendrá que volver a cargar los archivos.

Ilustración 19: Diagrama de actividad, cargar archivos¹⁸



2. FILTRAR INFORMACIÓN

Este caso de uso permite el filtro de información por parte del usuario, por medio de parámetros de filtro como lo pueden ser (sexo, edad, localización, fecha, hora, estrato social).

Actores

El usuario será el encargado de realizar esta acción

Precondiciones:

El usuario debe haber cargado los archivos KML y el software debe haberlos procesado y cargado en memoria correctamente

Flujo básico

- 1) El usuario selecciona filtrar información
- 2) El usuario selecciona los parámetros o atributos de la opción de filtro.
- 3) El usuario acepta una vez seleccionado algún parámetro
- 4) El en caso de que no haya ningún error se guarda y se finaliza el proceso

Flujo alternativo 1

En el paso 3:

- 5) En el caso de que ocurra algún error se muestra un mensaje de error
- 6) Se finaliza el proceso

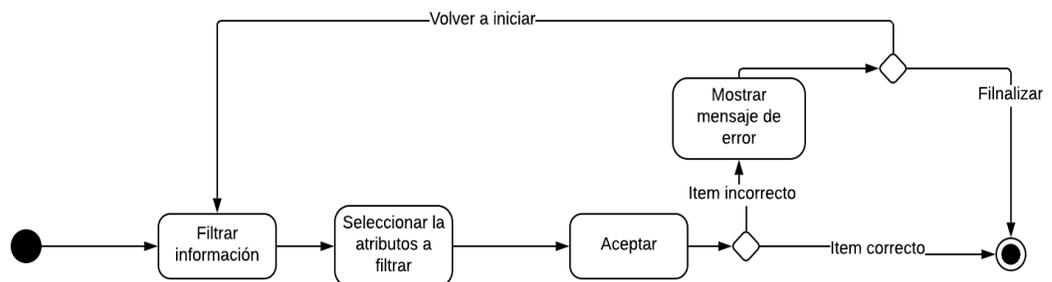
¹⁸ Fuente el autor

Flujo alternativo 2

En el paso 5:

- 7) Volver a la interfaz de filtrar información

Ilustración 20: Diagrama de actividad: Filtrar información¹⁹



3. INGRESAR PARÁMETROS KDE

Este caso de uso permitirá al usuario ingresar los parámetros necesarios (ancho de banda, resolución y método), para que el método KDE pueda calcular la densidad de núcleo.

Actores

Este caso de uso es realizado por el usuario

Precondiciones

El usuario debe haber cargado los archivos KML de puntos y del polígono previamente para que se habilite la opción de ingreso de parámetros.

Flujo básico:

- 1) El usuario ingresa los parámetros correspondientes en el orden que desee
- 2) El usuario selecciona el botón aceptar o siguiente
- 3) Si los datos son correctos se termina esta actividad

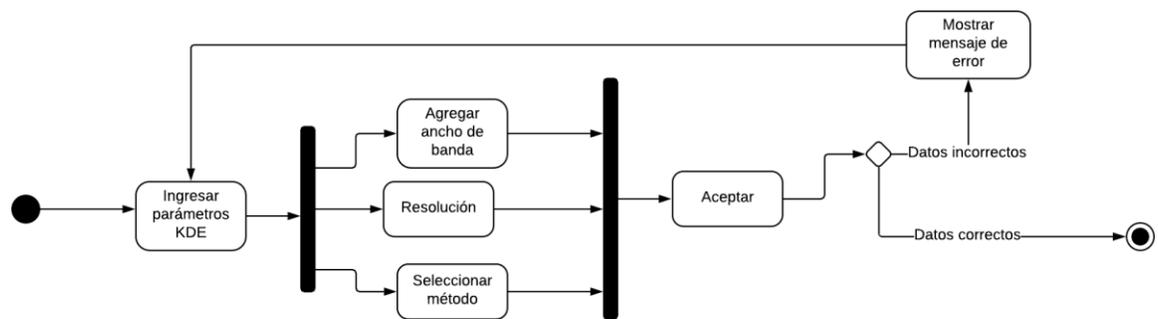
¹⁹ Fuente el autor

Flujo alternativo 1:

En el paso 2:

- 4) Se encuentra algún error en los datos ingresados se muestra un error
- 5) Se muestra nuevamente el formulario de ingreso de parámetros KDE

Ilustración 21: Diagrama de actividad, ingresar parámetros KDE²⁰



4. INGRESAR PARÁMETROS K-RIPLEY

Este caso de uso permitirá al usuario ingresar los parámetros necesarios, para que la función K de Ripley puede realizar el cálculo de clúster.

Precondiciones

El usuario debe haber cargado los archivos KML de puntos y del polígono previamente para que se habilite la opción de ingreso de parámetros.

Actores

Este caso de uso es realizado por el usuario

Flujo básico:

- 1) El usuario ingresa los 3 parámetros que pide el software para realizar el cálculo.
- 2) El usuario presiona el botón siguiente o aceptar.
- 3) Si los datos son correctos se finaliza se procede a hacer el cálculo.

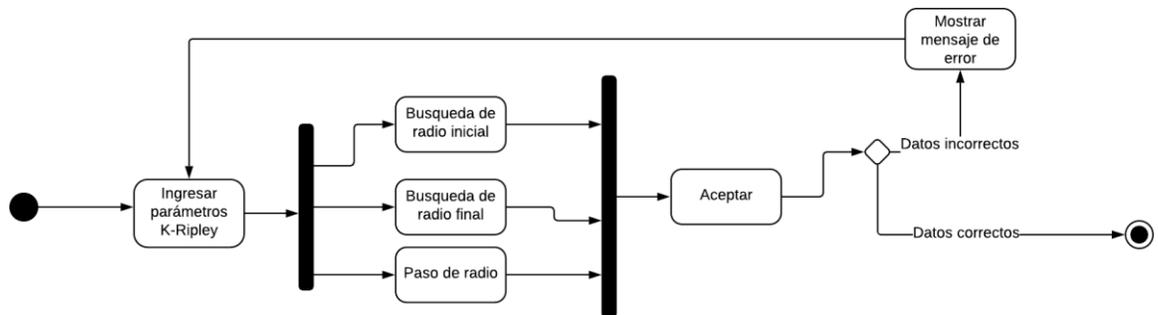
Flujo alternativo 1:

²⁰ Fuente el autor

En el paso 2:

- 4) Si los datos son incorrectos o se encuentra algún error.
- 5) Se muestra un mensaje de error y se le pide al usuario ingresar los datos nuevamente.

Ilustración 22: Diagrama de actividad: ingresar parámetros K de Ripley



5. VER RESULTADOS

Este caso de uso permitirá al usuario ver los resultados de los diferentes cálculos de los métodos KDE y función K de Ripley por medio del mapa de calor y análisis de clúster respectivamente.

Precondiciones

Se debe haber cargado los archivos KML y haber ingresado los parámetros KDE y K de Ripley.

Actores

Este caso de uso es realizado por el usuario

Flujo básico:

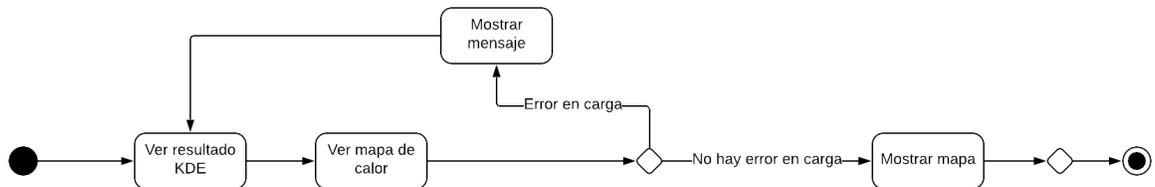
- 1) El usuario seleccionará la opción de ver resultados
- 2) El software se direcciona a cargar el mapa de calor
- 3) Si no hay error en carga de mapa, se muestra el mapa de calor

Flujo alternativo 1:

En el paso 2

- 4) Si se produce algún error en el momento de carga, se muestra un mensaje de error
- 5) Se muestra nuevamente la opción de ver resultado.

Ilustración 23: Diagrama de actividad: ver resultado KDE²¹



6. EXPORTAR RESULTADOS

Este caso de uso permitirá al usuario exportar los resultados obtenidos por el software a través de los 3 tipos de archivos permitidos por el software los cuales son KML, PNG y XLSX.

Precondiciones

Se debe haber cargado los archivos KML.

Actores

Este caso de uso es realizado por el usuario

Flujo básico:

Flujo alternativo 1:

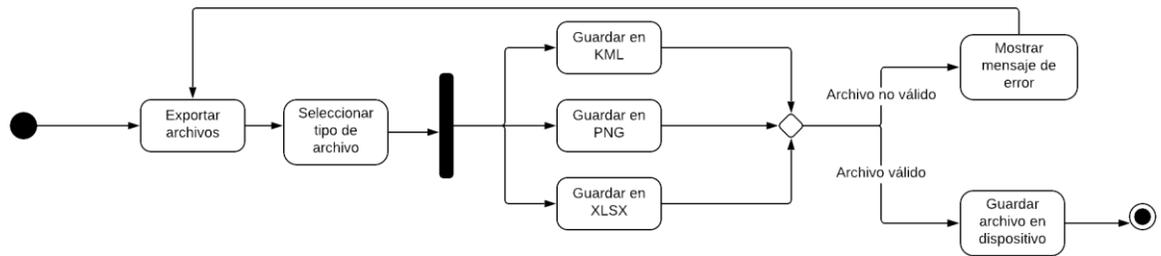
- 1) El usuario selecciona exportar archivos
- 2) El usuario selecciona el tipo de archivo
- 3) Si el archivo es válido y no se produce ningún error
- 4) Se guarda el archivo en el disco del dispositivo

En el paso 2

- 5) Si se produce algún error, se muestra un mensaje de error
- 6) Se muestra nuevamente la opción de exportar resultados

²¹ Fuente el autor

Ilustración 24: Diagrama de actividad: exportar resultados²²



4.3. DIAGRAMAS DE SECUENCIA

1. CARGAR ARCHIVOS

Una vez que el usuario abra la aplicación software, le aparecerá la interfaz de carga de archivos, en la cual el usuario deberá cargar el archivo KML del caso a estudiar, luego procederá a cargar el archivo KML del polígono para que se habilite el botón siguiente para que la aplicación software haga la interpretación de los archivos por medio de métodos y algoritmos hechos para este proceso. Luego la aplicación procederá a guardar los datos en memoria, luego de hacer esto, el software redireccionará al formulario de filtro de información.

²² Fuente los autores

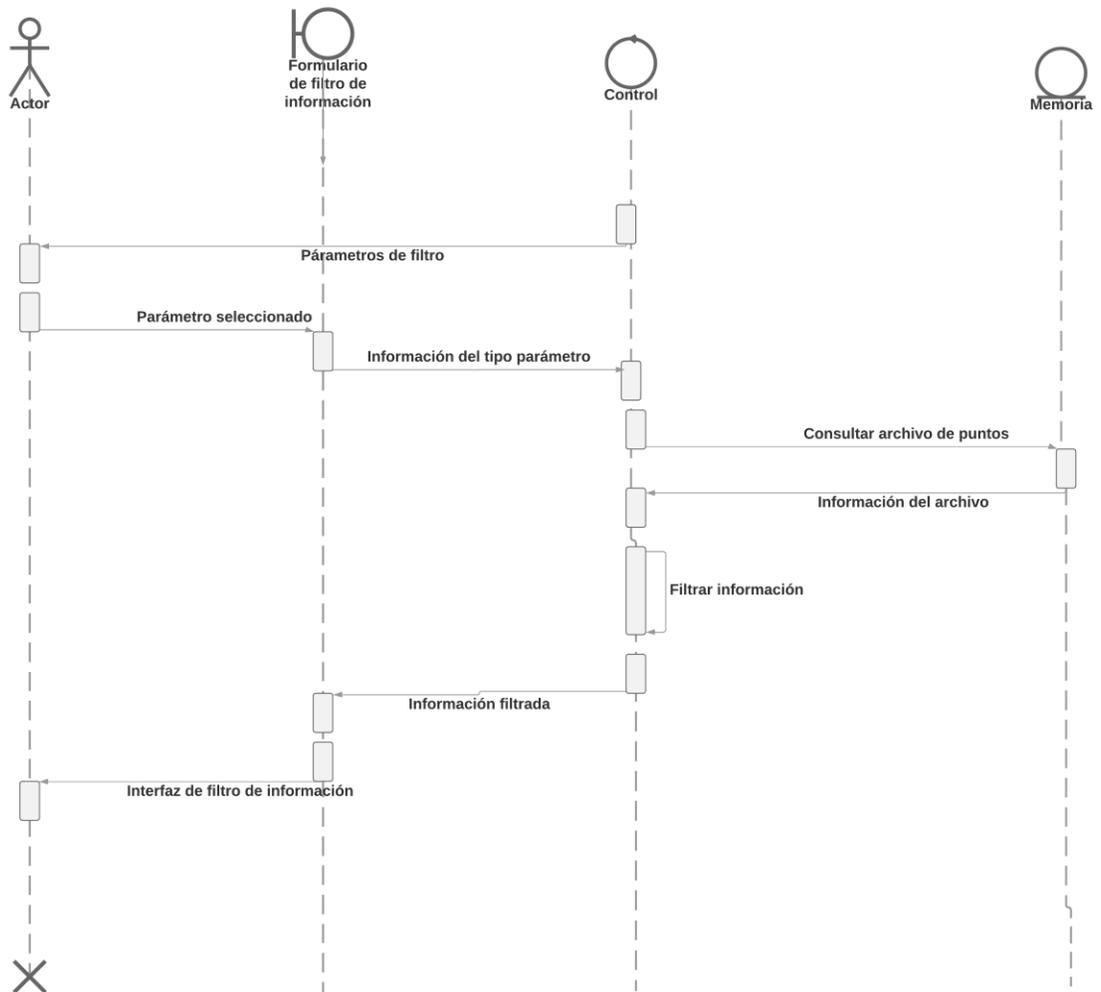
Ilustración 25: Diagramas de secuencia, cargar archivos²³



²³ Fuente los autores

2. FILTRO DE INFORMACIÓN

Ilustración 26: Diagrama de secuencia, filtrar información²⁴



El filtro de información será en la cual el usuario podrá filtrar los datos del archivo KML de puntos por medio de los parámetros como el sexo, edad, localización, fecha, para que el software haga el cálculo y muestre el resultado sobre los datos que se seleccionaron, una vez que el usuario llegue a la interfaz de filtro de información, el control hará una consulta de parámetros que contiene el archivo KML para luego mostrárselos al usuario y que él pueda seleccionar el que desee,

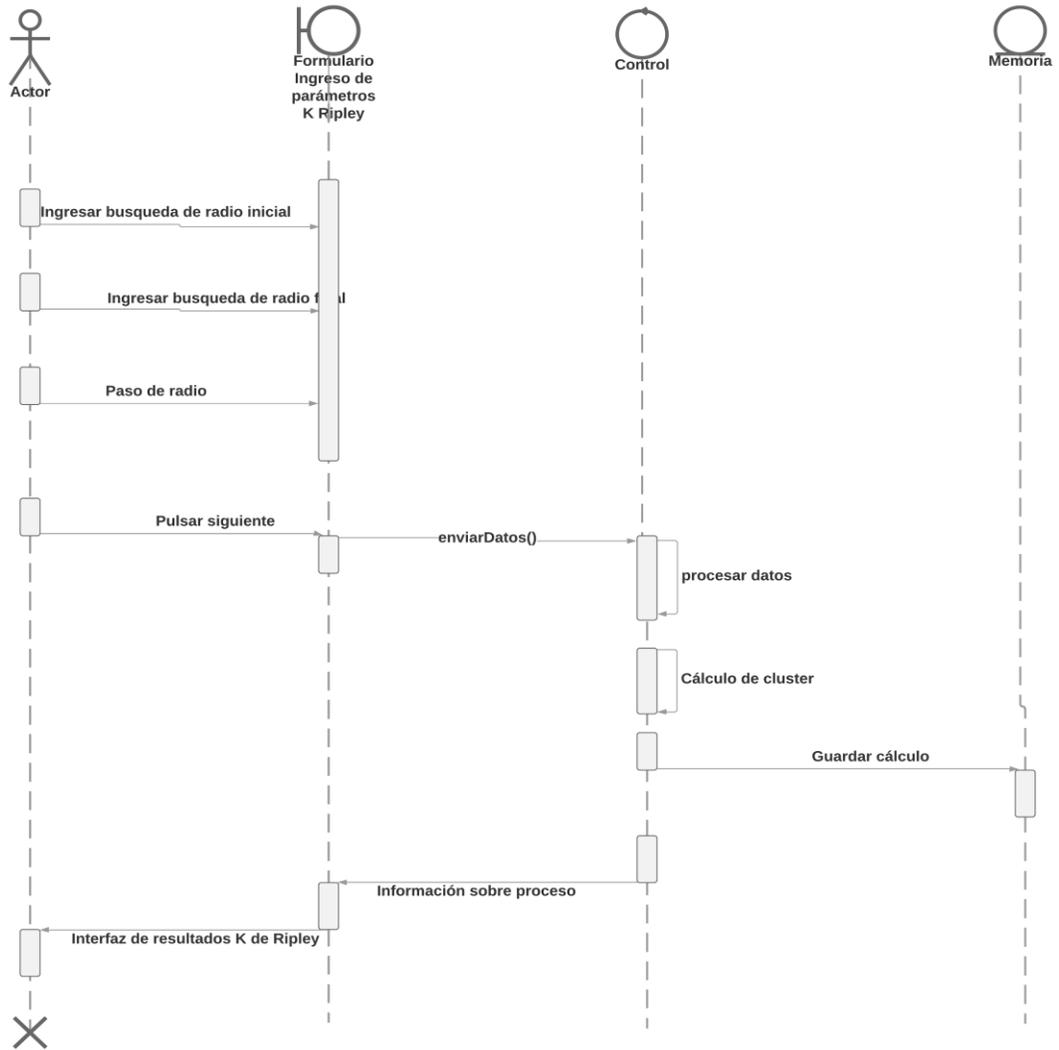
²⁴ Fuente los autores

una vez seleccionado el parámetro o parámetros, el software hará el filtro de información para proceder a hacer el cálculo sobre esos datos.

3. INGRESAR PARÁMETROS K DE RIPLEY

Este diagrama de secuencia muestra el funcionamiento del formulario ingreso de parámetros K de Ripley radio de búsqueda inicial, radio de búsqueda final y el paso de radio, al pulsar siguiente los datos serán enviados al control para su procesamiento y cálculo de clúster, luego se guardará el cálculo para mostrarlo en la interfaz de ver resultados de K de Ripley.

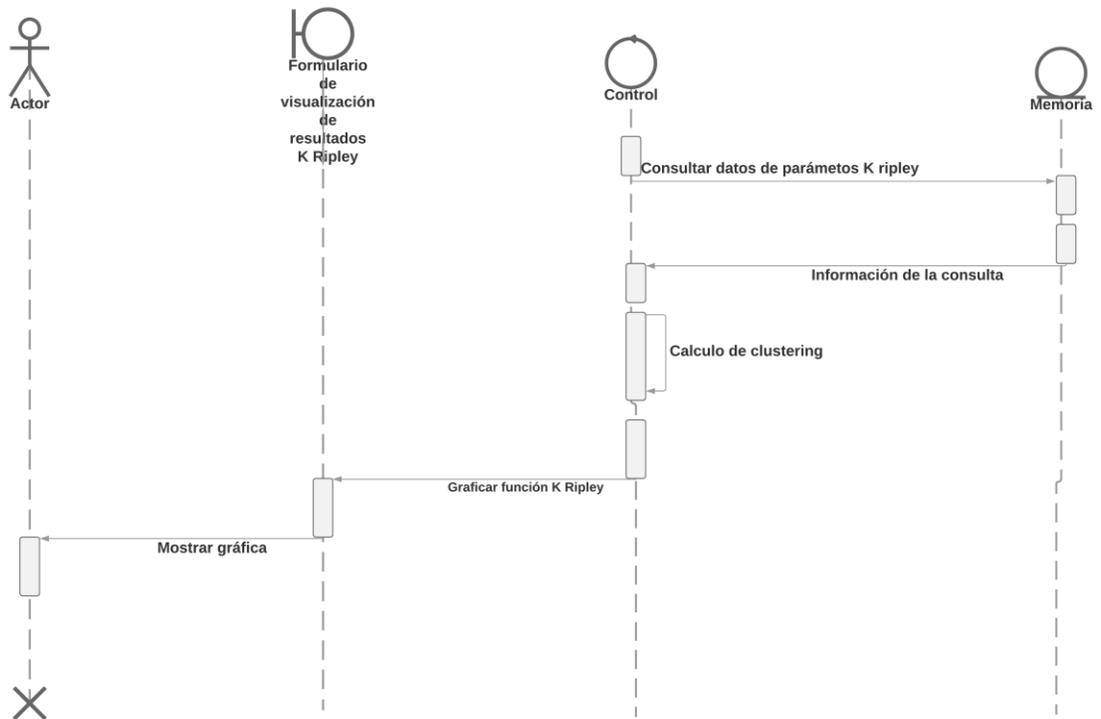
Ilustración 27: Diagrama de secuencia, ingresar parámetros K-Ripley²⁵



²⁵ Fuente los autores

4. VER RESULTADO K DE RIPLEY

Ilustración 28: Diagrama de secuencia, Ver resultados K de Ripley²⁶

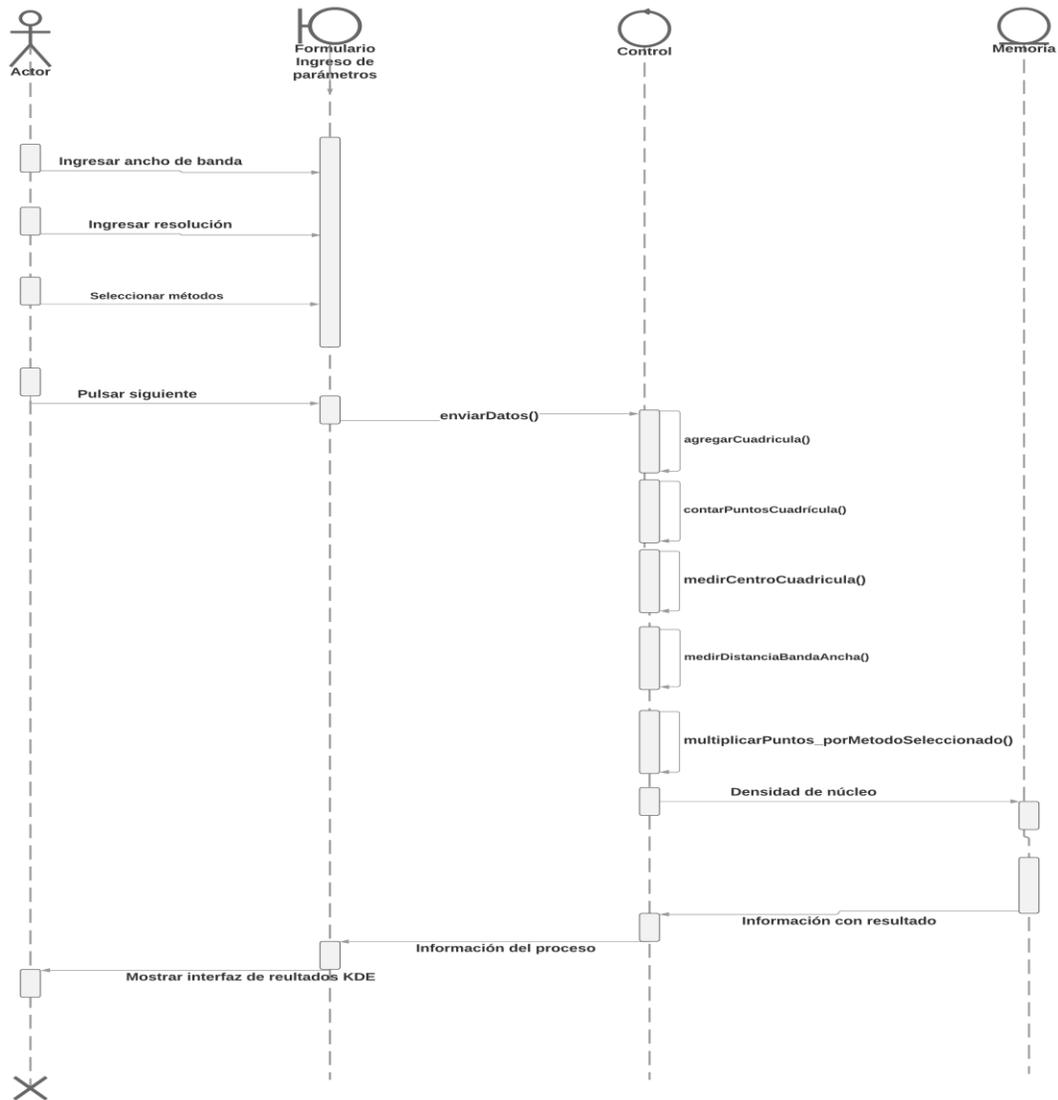


Este diagrama de secuencia explica cómo funciona la visualización de resultados de K de Ripley, de nuestro software, el control será el encargado de consultar los parámetros K de Ripley ingresados por el usuario anteriormente, con dichos datos se podrá hacer el cálculo de clúster y la respectiva gráfica de K de Ripley.

²⁶ Fuente los autores

5. INGRESAR PARÁMETROS KDE Y CÁLCULO DE DENSIDAD DE NÚCLEO

Ilustración 29: Diagrama de secuencia, ingreso de parámetros KDE²⁷



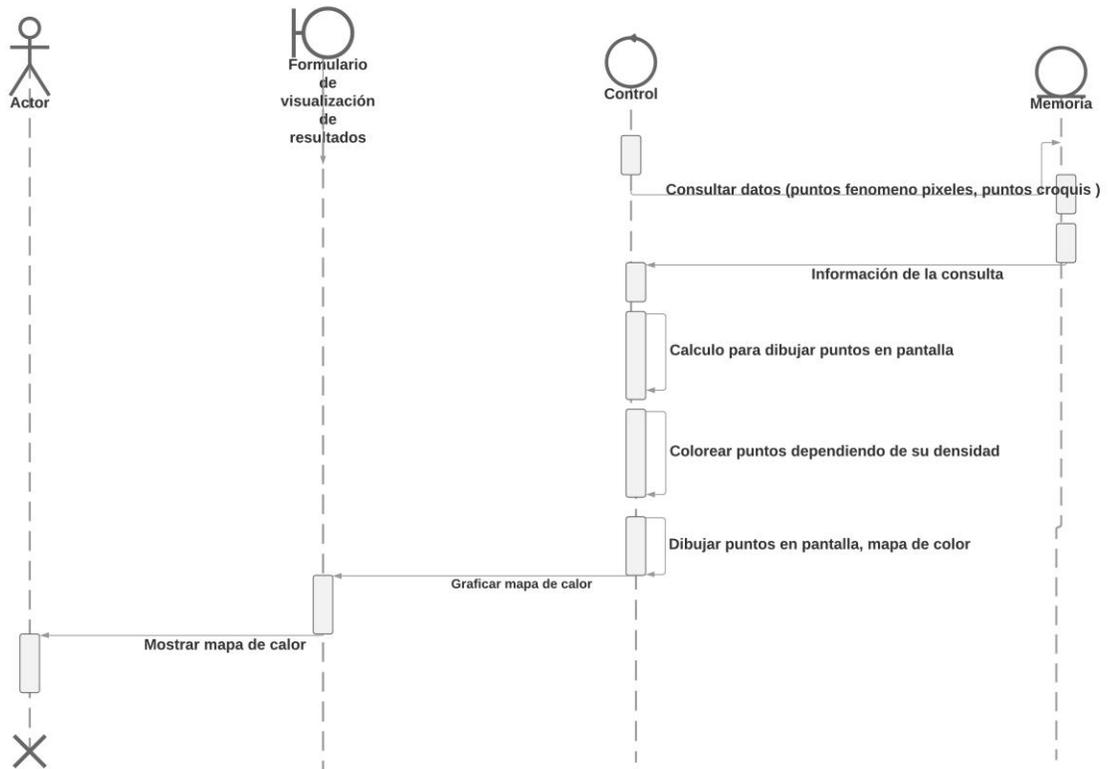
Este diagrama de secuencia muestra el funcionamiento del formulario de ingreso de parámetros de KDE, por medio de los cuales se podrá agregar una cuadrícula, luego medir se contará los puntos que se hallan en cada cuadro realizado, luego

²⁷ Fuente los autores

se cuentan los puntos que se hallan dentro del ancho de banda para poder hacer el cálculo de densidad de núcleo.

6. VER RESULTADOS KDE

Ilustración 30: Diagrama de secuencia, ver resultados KDE²⁸

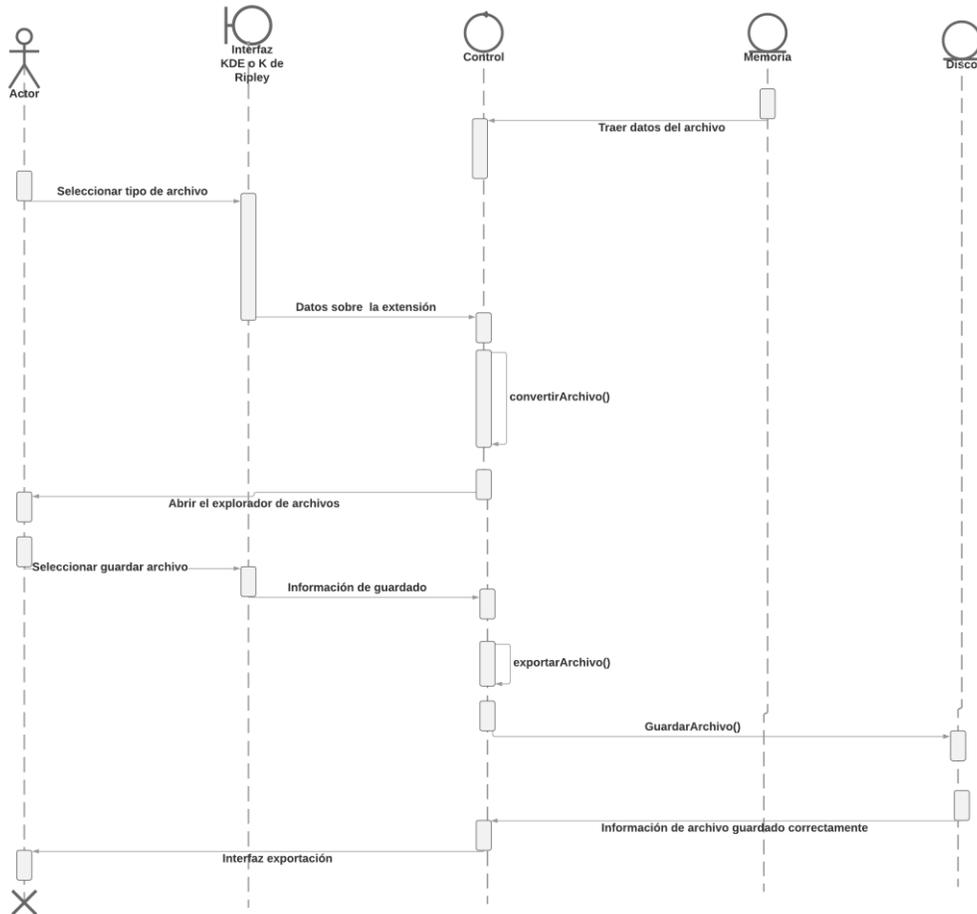


Este diagrama de secuencia explica cómo funciona la visualización de resultados de KDE de nuestro software, es decir, el mapa de calor, el control será el encargado de consultar los cálculos del método KDE hechos anteriormente para poder dibujar los puntos con sus respectivos colores asignados dependiendo de su aglomeración, también de dibujar el croquis del mapa a partir de los datos del archivo KML del polígono, una vez hecho esto, se mostrará el mapa de calor.

²⁸ Fuente los autores

7. EXPORTAR RESULTADOS

Ilustración 31: Diagrama de secuencia, exportar resultados²⁹



Este será el último paso que realizará el usuario en la aplicación software, la cual es la exportación de resultados por medio de los diferentes tipos de archivos (PNG, KML y XLSX), cuando el usuario esté en la interfaz de resultados KDE podrá guardar el mapa de calor por medio de archivos PNG y las densidades de cada celda por medio de KML.

Los resultados de la función K de Ripley, se podrán exportar por medio de archivos XLSX; para exportar un resultado, el usuario seleccionará la extensión o tipo de archivo en la que desea guardar sus resultados, luego se convertirá el archivo y se le permitirá al usuario salvar su archivo en el disco.

²⁹ Fuente los autores

4.4. MOCKUPS

1. CARGAR ARCHIVOS

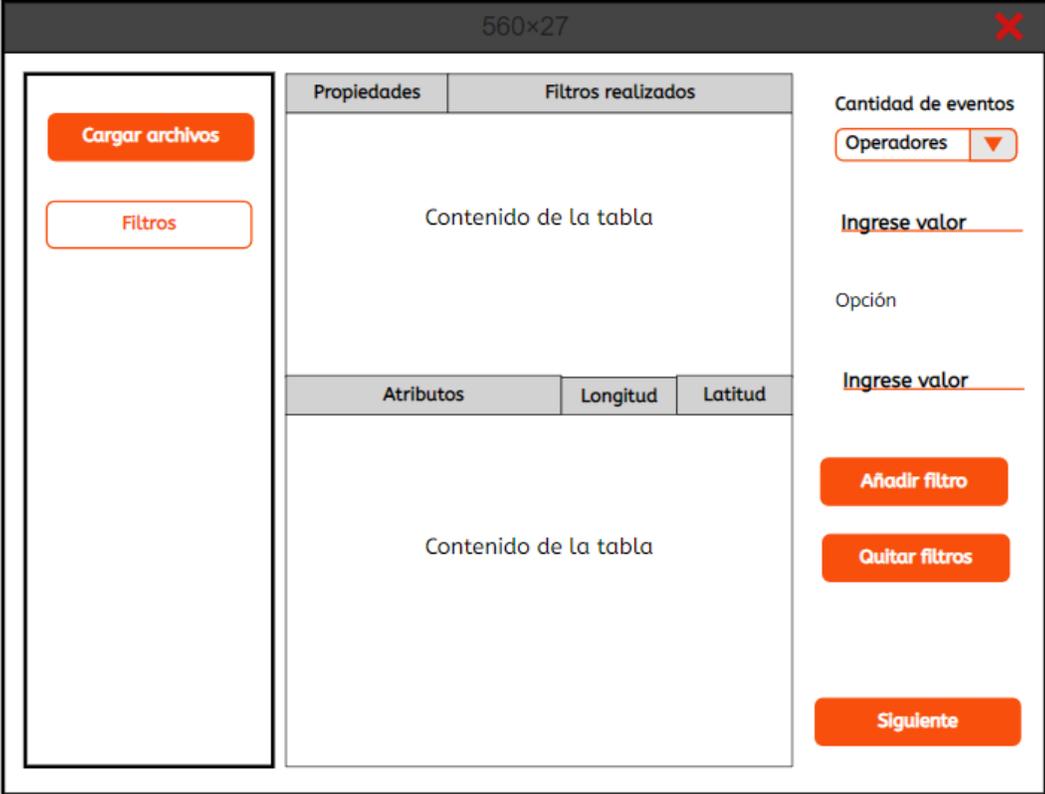
Ilustración 32: Mockups, cargar archivos³⁰

The image shows a software mockup for a file upload interface. It features a dark grey header bar with the text '560x27' and a red close button icon. The main content area is white and contains a vertical sidebar on the left with a red-bordered button labeled 'Cargar archivos'. The main area has two input fields: 'Archivo de puntos' and 'Archivo de poligono', each followed by an orange 'Abrir' button. At the bottom right, there is an orange 'Siguiete' button.

³⁰ Fuente los autores

2. FILTRAR INFORMACIÓN

Ilustración 33: Mockups, Filtro de información³¹



³¹ Fuente los autores

3. INGRESAR PARÁMETROS K DE RIPLEY

Ilustración 34: Mockup, ingresar parámetros K de Ripley³²

560x27

Cargar archivos

Filtros

Elegir Función

Radio de influencia mínimo

10

Radio de influencia máximo

100

Paso

10

Siguiente

³² Fuente los autores

4. VER RESULTADO K DE RIPLEY Y EXPORTAR RESULTADOS K RIPLEY

Ilustración 35: Mockup, ver resultado K de Ripley³³



³³ Fuente los autores

5. INGRESAR PARÁMETROS KDE

Ilustración 36: Mockup: Ingresar parámetros KDE³⁴

560x27

Cargar archivos

Filtros

Elegir Función

Ancho de banda

100

Resolución

100

Método

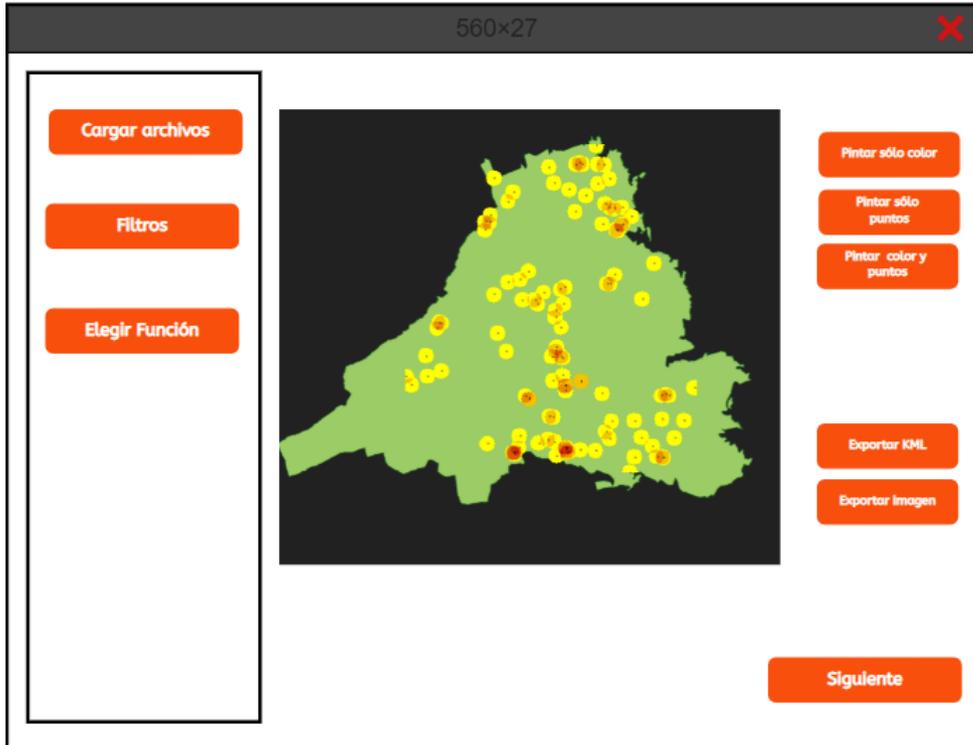
Función uniforme

Sigulente

³⁴ Fuente los autores

6. VER RESULTADOS KDE Y EXPORTAR RESULTADOS KDE

Ilustración 37: Mockup, ver resultado KDE³⁵



³⁵ Fuente los autores

5. RESULTADOS

5.1. ALISTAMIENTO DE LOS DATOS

Los datos de casos de dengue ocurridos en el municipio de Floridablanca en el año 2015, fueron recolectados en cada una de las IPS, unificados y depurados por la secretaría de salud de Santander, en total fueron recolectados un total de 990 de casos de dengue, luego la secretaria de salud de Santander notificó los casos a el SIVIGILA (Sistema Nacional de Vigilancia en Salud Pública). Dichos datos fueron georreferenciados y estructurados en un archivo de extensión KML, por parte de la FOSCAL, una versión simplificada de la estructura de este archivo se presenta a continuación:

```
<?xml version="1.0" encoding="utf-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
<Document id="root_doc">
<Placemark>
  <ExtendedData><SchemaData schemaUrl="#puntos_dengue_2015">
    <SimpleData name="ID">8</SimpleData>
    <SimpleData name="cod_eve">210</SimpleData>
    <SimpleData name="fec_not">08/01/2015</SimpleData>
    <SimpleData name="semana">1</SimpleData>
    <SimpleData name="edad_">27</SimpleData>
    <SimpleData name="sexo_">F</SimpleData>
    <SimpleData name="bar_ver_">SIN INFORMACION</SimpleData>
    <SimpleData name="dir_res_">CIRCUNVALAR 36A #140-
5</SimpleData>
    <SimpleData name="tip_ss_">C</SimpleData>
    <SimpleData name="con_fin_">1</SimpleData>
    <SimpleData name="ajuste_">7</SimpleData>
  </SchemaData></ExtendedData>
  <Point><coordinates>-73.08595276,7.084086895</coordinates></Point>
</Placemark>
```

Cada caso de dengue posee un total de 12 características, las cuales son: ID, Código del evento, fecha de notificación, semana, edad, sexo, barrio, dirección de residencia, tipo de seguridad social, condición final, y ajuste. Además, de las coordenadas de ocurrencia del caso, latitud y longitud, dichas características o parámetros permiten la inclusión de filtros, para el posterior análisis por medio de los métodos incluidos en la aplicación software.

Para el caso del polígono de la zona urbana del municipio de Floridablanca, fue necesario descargarlo, del marco Geoestadístico que se encuentra en la plataforma Geo portal del DANE, el cual generó un archivo KML que contiene los

diferentes puntos, que permite dibujar el polígono del mapa. A continuación, se presenta de forma simplificada la estructura del archivo.

```
<?xml version="1.0" encoding="utf-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2">
<Document id="root_doc">
<Schema name="mapa" id="mapa">
<SimpleField name="MPIO_CCDGO" type="string"></SimpleField>
</Schema>
<Placemark>
<Style><LineStyle><color>ff0000ff</color></LineStyle><PolyStyle><fill>0</fill></PolyStyle></Style>
<ExtendedData><SchemaData schemaUrl="#mapa">
<SimpleData name="DPTO_CCDGO">68</SimpleData>
<SimpleData name="MPIO_CCDGO">276</SimpleData>
<SimpleData name="CLAS_CCDGO">1</SimpleData>
<SimpleData name="SETR_CCDGO">000</SimpleData>
<SimpleData name="SECR_CCDGO">00</SimpleData>
<SimpleData name="SETR_CCNCT">682761000</SimpleData>
<SimpleData name="SECR_CCNCT">68276100000</SimpleData>
<SimpleData name="SECR_CAG">N/A</SimpleData>
<SimpleData name="SECR_NAREA">17.02510550380</SimpleData>
<SimpleData name="SECR_CSMBL">8</SimpleData>
<SimpleData name="SECR_NANO">2017</SimpleData>
<SimpleData name="Shape_Leng">0.29069340575</SimpleData>
<SimpleData name="Shape_Area">0.00139345626</SimpleData>
</SchemaData></ExtendedData>
<MultiGeometry><Polygon><outerBoundaryIs><LinearRing><coordinates>-
73.092255885,7.09887623600002 -73.092232194,7.09887038400007 -
73.0922175109999,7.09887356 -73.092201528,7.09887701700006 -
73.09219873,7.09885611900006 -73.092194664,7.09882575700004 -
73.092151203,7.09879500300002 -73.0921178899999,7.09875883500007
</coordinates></LinearRing></outerBoundaryIs></Polygon></MultiGeometry>
</Placemark>
```

5.2. ANÁLISIS DE RESULTADOS

Los resultados obtenidos a partir de los procesos de desarrollo e implementación de la aplicación software está dividida en dos partes, en la primera parte, se encuentran los hallazgos resultantes, debido a la implementación de la metodología de desarrollo, de igual manera, de los métodos que conforman la aplicación, el método KDE y función K de Ripley, que hacen posible el análisis de patrones de los puntos. En la segunda parte, el análisis de los resultados obtenidos en el proceso de aplicación de los métodos implementados a los datos de casos de dengue ocurridos en el municipio de Floridablanca, en el año 2015.

5.2.1. HALLAZGOS EN EL DESARROLLO E IMPLEMENTACIÓN DE LA APLICACIÓN SOFTWARE

La aplicación software fue programada aplicando la metodología de desarrollo ágil Crystal clear, por lo tanto, en cada una de las iteraciones iba produciendo resultados, es decir, módulos. Además, se iban realizando pruebas a cada uno de los módulos, junto con los datos reales de los casos de dengue, lo que permitió encontrar resultados previos arrojados por cada módulo.

A continuación, se detallan los diferentes hallazgos, encontrados en cada etapa de desarrollo de la aplicación software.

Tabla 3: Productos de la aplicación software por etapas

Fase	Productos desarrollados	Observaciones
1	Módulo carga de archivos	La carga de archivos incluyó tanto la carga de los puntos, como la del polígono y también la conversión de las coordenadas de los puntos y polígono a UTM.
2	Módulo cálculo de KDE	El cálculo de KDE es el método que permite conocer la densidad de núcleo de los datos, para luego pintar el mapa de calor, dicho cálculo se realiza por medio de los datos que ingresa el usuario, los cuales son: la resolución, el ancho de banda, y el tipo de función.
3	Módulo cálculo K de Ripley	El cálculo de K de Ripley es la que permite saber si los datos tienden agruparse o por el contrario se distribuyen de manera aleatoria, para los diferentes valores ingresados por el usuario, como lo son: radio de influencia mínimo y máximo, y el paso a paso, los cuales permiten el cálculo del valor K real y aleatorio.
4	Módulo visor Resultados	El visor de resultados consta de dos formularios: 1. El primero es la gráfica K de Ripley, la cual permite al usuario saber si los datos se agrupan o no como se menciona anteriormente, dicha gráfica consta de 3 curvas 1

		<p>de ellas representa los valores reales y las otras dos los valores generados aleatoriamente.</p> <p>2. El segundo es el mapa de calor que refleja de manera visual la distribución de los puntos en la pantalla, previo a mostrar el mapa, primero y teniendo la densidad, se calculan los colores, se hacen las conversiones necesarias y se ajustan y pintan los puntos en la pantalla.</p>
5	Módulo filtro de información	<p>Este módulo es el que permite al usuario filtrar la información para que luego, la aplicación software trabaje sobre esos datos filtrados, los filtros que se pueden aplicar dependen del atributo por el que se quiere hacer el filtro, ya que, si es un valor numérico en el caso de la edad podrá escoger valores, mayores, menos, iguales y entre un rango.</p>
6	Módulo exportación de resultados	<p>Este módulo fue el último que se implementó en la aplicación software, y es el que permite al usuario exportar los resultados de la gráfica K de Ripley por medio de archivos en formato XLSX y el mapa de calor por medio de archivos PNG y KML.</p>

5.2.1.1. MÓDULO CARGA DE ARCHIVOS

Los archivos de entrada que soporta la aplicación software son necesariamente de extensión KML, como se dijo en la sección de definición de la arquitectura del software, este formato de archivo es un lenguaje de marcado basado en XML y su carga y posterior lectura, se hizo por medio de la implementación de métodos que permiten la interpretación de las etiquetas que están definidas de manera estandarizada para los archivos KML.

Se tiene que el archivo KML de casos de dengue del municipio de Floridablanca, para el año 2015, posee un total de 990 casos, los cuales representan 990 puntos, a su vez cada caso posee un total de 12 características como se menciona en el

apartado de alistamiento de datos, el tiempo de carga de este archivo de 990 casos, es aproximadamente de 2590 milisegundos, para un archivo que contenía un total de 71460 puntos el tiempo de carga fue de 12800 milisegundos, lo que quiere decir que, los tiempos de carga e interpretación de los archivos KML son relativamente rápidos. Por otra parte, también fue necesario la carga de puntos del polígono del municipio de Floridablanca, el cual también es un archivo KML, por lo tanto, se implementaron métodos para la carga de dicho archivo, los métodos fueron similares a los que permiten la carga del archivo de puntos, aunque tienen algunas diferencias ya que, este archivo posee atributos y estructura diferentes.

5.2.1.2. MÓDULO CÁLCULO DE KDE

El método KDE como se ha mencionado múltiples veces es aquel que permite calcular o determinar la densidad de núcleo de los puntos, para dicho cálculo se toman los datos ingresados por el usuario como los son: ancho de banda, resolución y función de Kernel. La resolución indica el tamaño de cada una de las celdas en las que está dividida la cuadrícula y el ancho de banda indica la distancia máxima a la que se considera que dos casos están relacionados, esta relación se mide a través de los puntos centrales de cada celda en la cuadrícula.

El algoritmo funciona de la siguiente manera; primero se hace una cuadrícula, cada celda tendrá un tamaño igual a la resolución, para hacer el cálculo del cuadro se halla una coordenada mínima y máxima en el eje X y una coordenada mínima y máxima en el eje Y, luego de realizar la cuadrícula, se mide el centro de cada celda de la cuadrícula y se calcula la distancia desde dicho centro medio hasta todos los puntos. Los puntos que se encuentren dentro del ancho de banda son contados, para luego calcular la densidad con respecto a cada uno de los puntos acumulados y la función de Kernel que elija el usuario.

La aplicación software implementa 4 funciones de Kernel, las cuales son:

- **La función uniforme:**

$$K(u) = \frac{1}{2} \quad (8)$$

$$|u| \leq 1$$

Siendo u la distancia entre los puntos, dividido en el ancho de banda.

- **Función Triangular:**

$$K(u) = (1 - |u|) \quad (9)$$

$$|u| \leq 1$$

Siendo u la distancia entre los puntos, dividido en el ancho de banda.

- **Función de Epanechnikov:**

$$K(u) = \frac{3}{4}(1 - u^2) \quad (10)$$

$$|u| \leq 1$$

Siendo u la distancia entre los puntos, dividido en el ancho de banda.

- **Función Quartic:**

$$K(u) = \frac{15}{16}(1 - u^2)^2 \quad (11)$$

$$|u| \leq 1$$

Siendo u la distancia entre los puntos, dividido en el ancho de banda.

A los valores de densidad calculados, se les aplica una escala de color, que va desde el color amarillo al rojo, siendo el color amarillo asignado a los valores de densidad más bajos y el rojo a los valores de densidad más altos.

5.2.1.3. MÓDULO CÁLCULO DE K DE RIPLEY

La función K de Ripley, como se ha hecho mención anteriormente, es aquella que, permite saber si los datos del fenómeno estudiado tienden a formar aglomeraciones y por lo tanto no se comporta o distribuye de forma aleatoria en el espacio. Para realizar dicha estimación, se deben tomar los datos ingresados por el usuario en el formulario de K de Ripley, los cuales son: radio de influencia máximo y mínimo y el paso a paso.

El algoritmo comienza a medir la distancia que hay entre los puntos, si la distancia que existe entre un punto y otro, es menor a el radio de búsqueda, se hace un conteo de dichos puntos, si hay puntos fuera del radio de búsqueda estos punto se

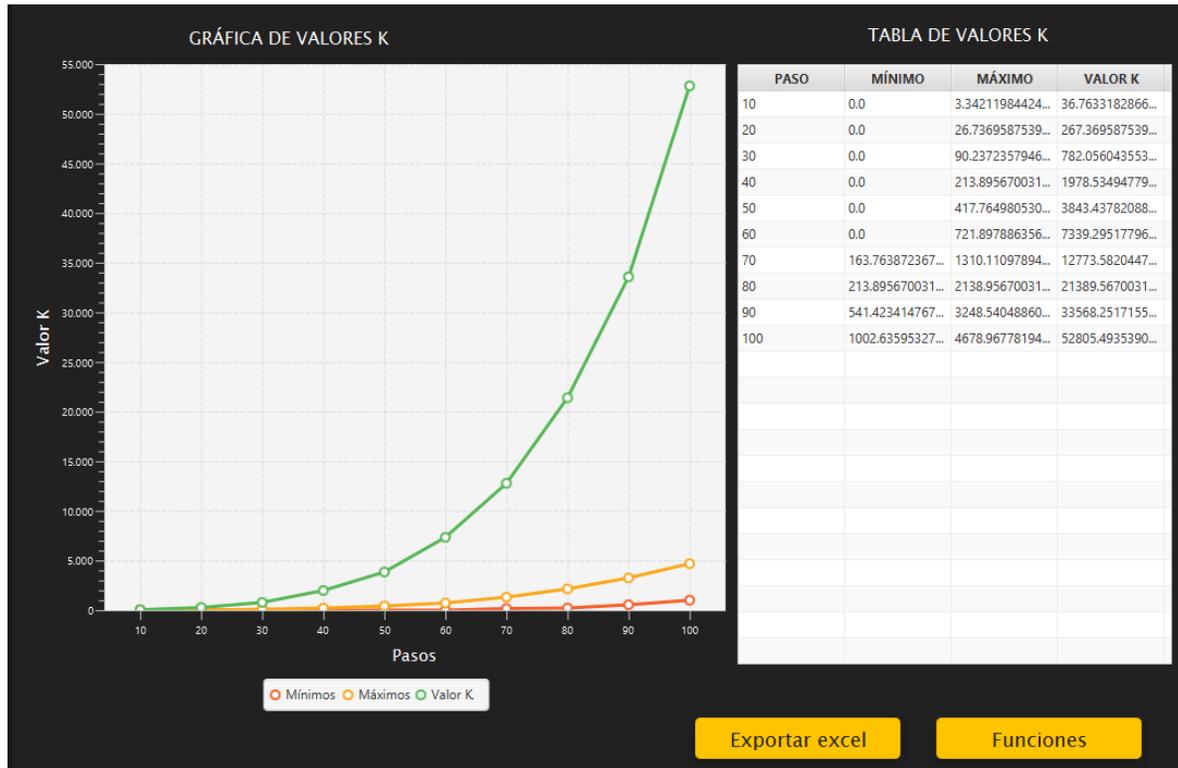
ignoran, al hacer el conteo me da un valor, dicho valor se multiplica por 2π y se divide por el número total de puntos, ese cálculo genera el valor K real, para cada una de las distancias es decir, el paso a paso que ingreso el usuario, luego, se genera la misma cantidad de puntos y se calcula el K de los respectivos puntos, esta operación se hace 999 veces. Luego se toma el K mínimo de los valores aleatorios el cual es el 2,5% y el K aleatorio máximo 97,5% y el K real y se realiza la gráfica de dichos valores.

5.2.1.4. MÓDULO VISOR RESULTADOS

El módulo visor de resultados consta de dos partes o interfaces, el primero es la gráfica de K de Ripley y el segundo es el mapa de calor. La interfaz de visualización de K de Ripley es la que permite conocer si los datos tienden a formar aglomeraciones o por el contrario se distribuyen de forma aleatoria, el resultado de esta función consta de la gráfica de los valores K reales, K mínimos y máximos generados; y una tabla que contiene cada uno de estos valores K calculados para los pasos ingresados por el usuario. Un ejemplo de resultado de la función K de Ripley se muestra en la ilustración 40.

La línea verde representa los valores K reales, la línea amarilla los valores de K máximo y la línea roja representa el valor K mínimo de los valores K generados.

Ilustración 38: Ejemplo de resultado de la Función K de Ripley.³⁶

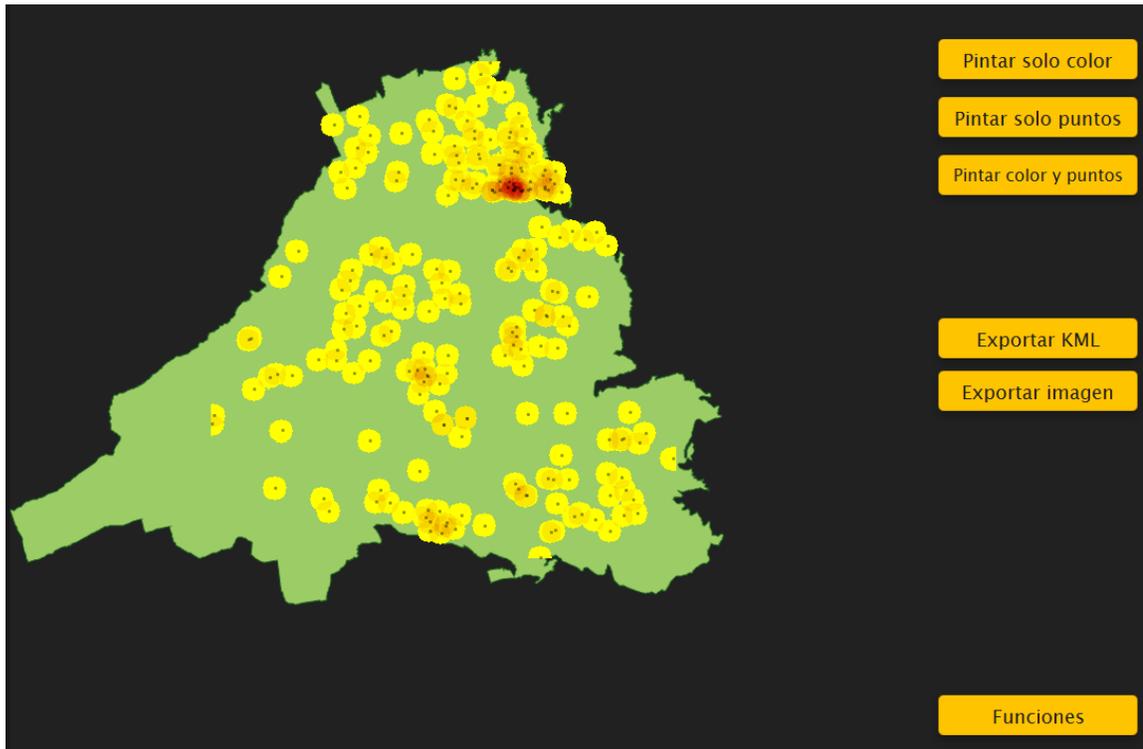


Por otra parte, la interfaz de visualización de resultados KDE consta de un mapa de calor que está conformado de dos archivos KML, el primero corresponde a los puntos de los casos del fenómeno y el segundo hace referencia al polígono del mapa, para la representación gráfica de cada archivo, se hizo una conversión del sistema de coordenadas que venía por defecto, al sistema de coordenadas UTM, luego, estas coordenadas son transformadas en coordenadas de pixeles en pantalla, los puntos y líneas dibujados en pantalla son representados a escala en metros.

Los colores de los puntos fueron determinados por medio de un componente desarrollado dentro de la aplicación software, el cual implementa el modelo de color HSV debido a que se pueden lograr degradados, este color se aplica dependiendo de la densidad calculada, además, se puede visualizar el mapa de calor mostrando solo los puntos, solo el color, o el color y puntos. Un ejemplo de resultado de KDE se presenta en la ilustración 41.

³⁶ Fuente los autores

Ilustración 39: Ejemplo de visualización de resultados KDE³⁷



La zona con el color más rojizo representa el área en la que se encuentra mayor concentración de casos de dengue, estas zonas son conocidas como puntos calientes.

5.2.1.5. MÓDULO DE FILTRO DE INFORMACIÓN

El módulo de filtro de información es el que permite al usuario filtrar los datos que pertenecen al KML de puntos, con el objetivo que la aplicación software trabaje con los datos seleccionados en el filtro o conjunto de filtros, ya que, la aplicación permite aplicar más de un filtro.

El módulo funciona de la siguiente manera; extrae los atributos que contiene el KML de puntos, para el caso del KML de puntos de dengue son 12 atributos que tiene cada punto (caso de dengue), los cuales se mencionan en el apartado de alistamiento de datos. Una vez que se hayan extraído los atributos, se le muestran

³⁷ Fuente los autores

a el usuario en el formulario de filtro, en el cual podrá seleccionar, el atributo o atributos por el cual desea filtrar la información, luego de hacer el filtro, se mostrará el número de puntos, además, en la tabla inferior, se muestran todos los casos de dengue filtrados, con sus respectivos atributos, con los cuales la aplicación software realizará los cálculos con los respectivos métodos. Un ejemplo de filtro de información se muestra en la ilustración 40.

Ilustración 40: Ejemplo de filtro de información³⁸

Propiedad	Filtros realizados		
ID			
cod_eve			
fec_not			
semana			
edad_	Entre 20 y 30		
sexo_	Igual M		
bar_ver_			
dir_res_			
tip_ss_			
con_fin_			
ajuste_			

Atributos	Longitud	Latitud
ID: 826 cod_eve: 210 fec_not: 24/07/2015 semana: 29 edad_: 29 sexo_: M bar_ver_: LOS ROBLES dir_res_: CALLE 56A #2-72 tip_ss_: C con_fin_: 1 ajuste_: 0	-73.090431...	7.09265995
ID: 843 cod_eve: 210 fec_not: 25/01/2015 semana: 3 edad_: 26 sexo_: M bar_ver_: SIN INFORMACIÓN dir_res_: CARRERA 2 #23-12	-73.095024...	7.072135448

Cantidad de eventos: 80

Operadores ▼

LA CUMBRE

Igual

Añadir filtros
Eliminar filtros

Filtrar eventos

Siguiente

5.2.1.6. MÓDULO DE EXPORTACIÓN DE RESULTADOS

El módulo de exportación de resultados es el que permite al usuario guardar los datos generados por la aplicación software en cada uno de los métodos, dichos datos o cálculos producidos se pueden guardar en 3 diferentes formatos de archivos. El primero de ellos, es para los datos de la función K de Ripley, este resultado es posible exportarlo en archivos XLSX, en el cual se guardarán los

³⁸ Fuente los autores

pasos, los valores K mínimos y máximos generados y los valores K reales, para cada uno de los pasos.

Por otra parte, se pueden exportar los resultados del método KDE en dos tipos de archivos, el primero es PNG, que permite guardar el mapa de calor de manera visual en este archivo de formato gráfico, además, se podrá guardar, las densidades y los respectivos polígonos de cada celda, en archivo de formato KML.

5.3. ANÁLISIS DE RESULTADOS DE LA APLICACIÓN SOFTWARE DE LOS MÉTODOS IMPLEMENTADOS

Los métodos implementados, aunque se aplican de forma separada dentro de la aplicación software, se complementan el uno con el otro, en el campo del análisis espacial de patrones de puntos y en este caso dentro de la misma problemática, el fenómeno del dengue, aunque hay que dejar en claro que, dichos métodos se pueden aplicar a otras problemáticas. Debido al funcionamiento de cada uno de los métodos se recomienda inicialmente aplicar la función K de Ripley, ya que, esta función es la que indica si los puntos tienden a formar aglomeración, es decir que el fenómeno estudiado no ocurre de forma aleatoria, entonces, valdría la pena aplicar el método KDE, para identificar en el mapa, las zonas en las que se encuentra mayor ocurrencia de eventos (puntos).

A continuación, se presenta una interpretación a los resultados generados por la aplicación software en cada uno de los métodos, se tuvieron en cuenta los datos, que generaban unas vistas interesantes o relevantes, incluso aplicando filtros a los datos, un análisis más detallado sería trabajo de las diferentes entidades expertas en temas de salud pública, en este caso el dengue.

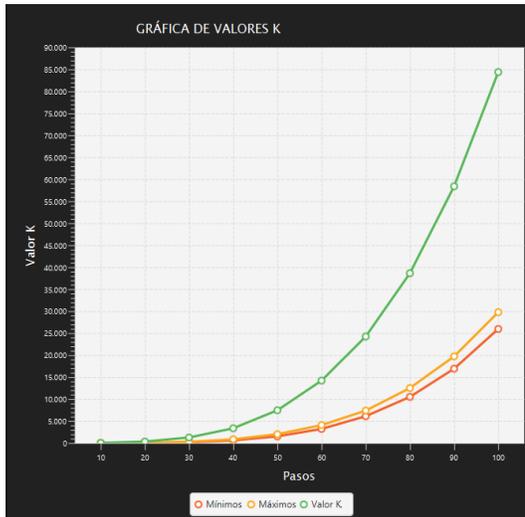
5.3.1. FUNCIÓN K DE RIPLEY

Los valores resultantes de la función K de Ripley arroja una estimación, la cual es que los datos de casos de dengue tienden a agruparse, pero en sí, los valores K calculados para cada uno de los pasos, por si solos no tiene ningún significado, por lo tanto, para poder interpretar este valor y sacar una conclusión o conjunto de conclusiones, se debe hacer una comparación de los valores K reales, con unos valores K que se generan aleatoriamente 999 veces, en el caso del dengue, se tiene un total de 990 casos de dengue, ocurridos en el municipio de Floridablanca, por lo tanto se generan 990 casos aleatoriamente dentro del mapa, 999 veces, dicho cálculo es un poco pesado para el programa, por lo que le toma

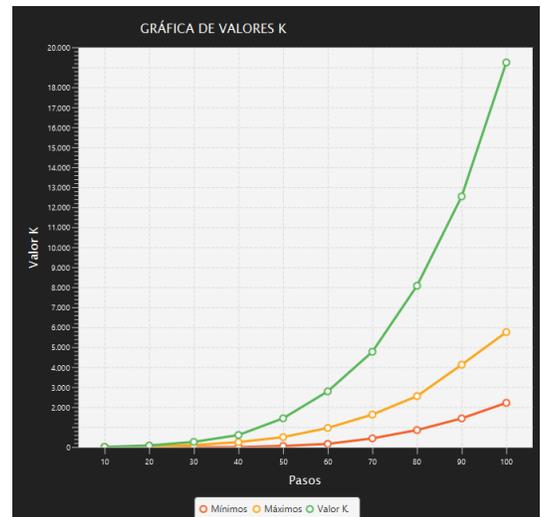
aproximadamente 323400 milisegundos, es decir, unos 5,39 minutos. Luego de hacer este proceso, se toma un intervalo de confianza que va desde 2,5 por ciento al 97,5 por ciento. Entonces, si el valor K de los valores reales, se encuentra por encima del intervalo de confianza, se concluye que los datos tienden a formar patrones de agrupación, si, por el contrario, el valor K real se encuentra por debajo del intervalo de confianza, entonces, significa que los datos tienden a dispersarse de forma uniforme, es decir, como si se repelieran el uno del otro.

En la ilustración 41, se muestran los resultados de la función K de Ripley de forma gráfica para los datos de casos de dengue que ocurrieron en el municipio de Floridablanca en el año 2015, los valores K fueron calculados para los diferentes pasos o distancias que van desde los 10 hasta los 100 metros, con intervalos de 10 metros, como se ve en la gráfica de K de Ripley, la línea de color verde, representa los valores K reales, la línea de color amarillo representa el valor K máximo del intervalo de confianza, el color rojo representa el valor K mínimo de intervalo de confianza.

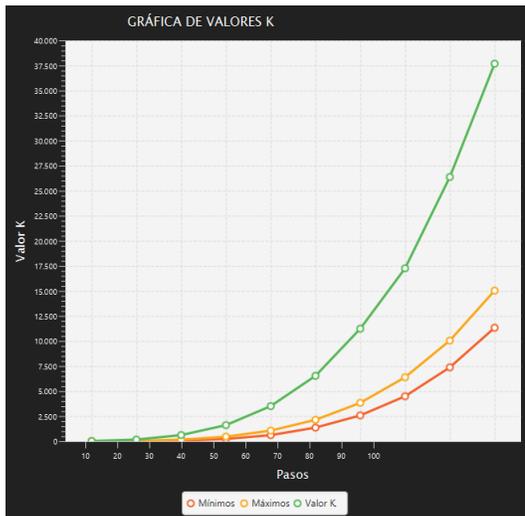
Ilustración 41: Grupo de resultados de la función K de Ripley – dengue Floridablanca 2015³⁹



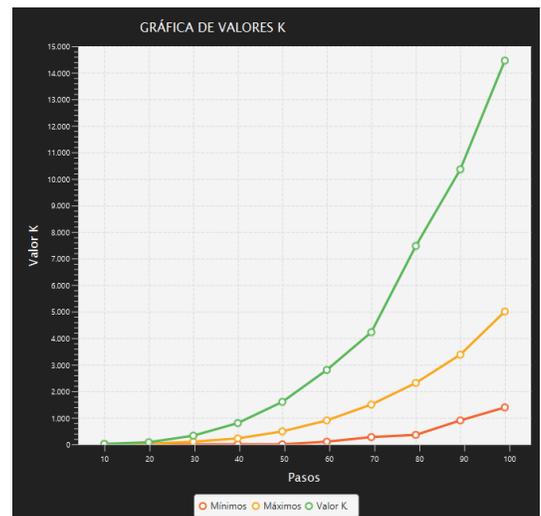
(a)



(b)

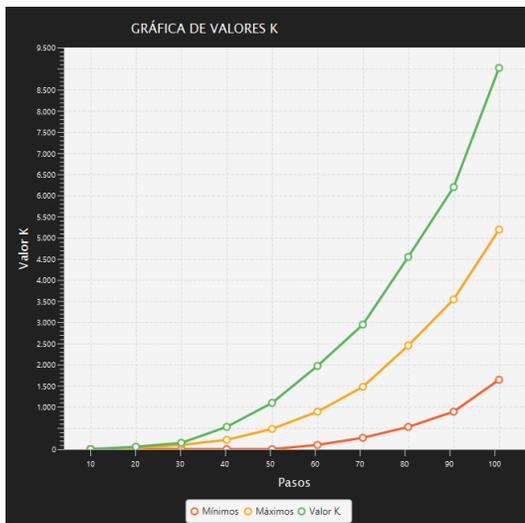


(c)

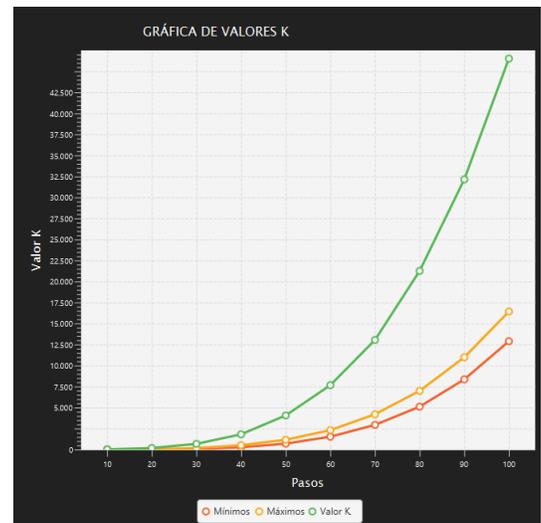


(d)

³⁹ Fuente los autores



(e)



(f)

La ilustración (a), representa los resultados al utilizar todos los datos, es decir, 990 casos de dengue, la ilustración (b), muestra los casos de dengue ocurridos a las personas de sexo masculino, la ilustración (c), corresponde a los casos de dengue ocurridos a las personas del sexo femenino, la ilustración (d), muestra los casos de dengue ocurridos a las personas con edades que van desde los 0 a los 10 años, la ilustración (e), corresponde a casos de dengue ocurridos a las personas mayores de 60 años, y por último la ilustración (f), muestra los casos de dengue ocurridos durante el mes de octubre, ya que, este fue el mes en el que más casos de dengue ocurrieron en el municipio de Floridablanca en el 2015, con un total de 113 casos reportados.

En los 6 casos presentados en las ilustraciones anteriores, se observa que los datos tienen una gran tendencia a agruparse, ya que se ve que los valores K reales, es decir, la línea verde, se encuentra por encima de los valores K generados, incluso para las distancias iguales a los 100 metros. Así mismo, se observa pequeñas variantes entre cada una de las ilustraciones. En la tabla número 4, se presentan los números de casos de dengue utilizados, según el filtro aplicado en cada resultado de la función K de Ripley.

Tabla 4: Número de casos de dengue por cada resultado⁴⁰

Filtro	Casos de dengue	Porcentaje (%)
Todos los datos	990	100

⁴⁰ Fuente los autores

Personas con sexo masculino	524	53
Personas con sexo femenino	466	47
Personas con edades entre 0 y 10 años	142	14
Personas mayores a 60 años	115	12,6
Mes de octubre	113	11,4

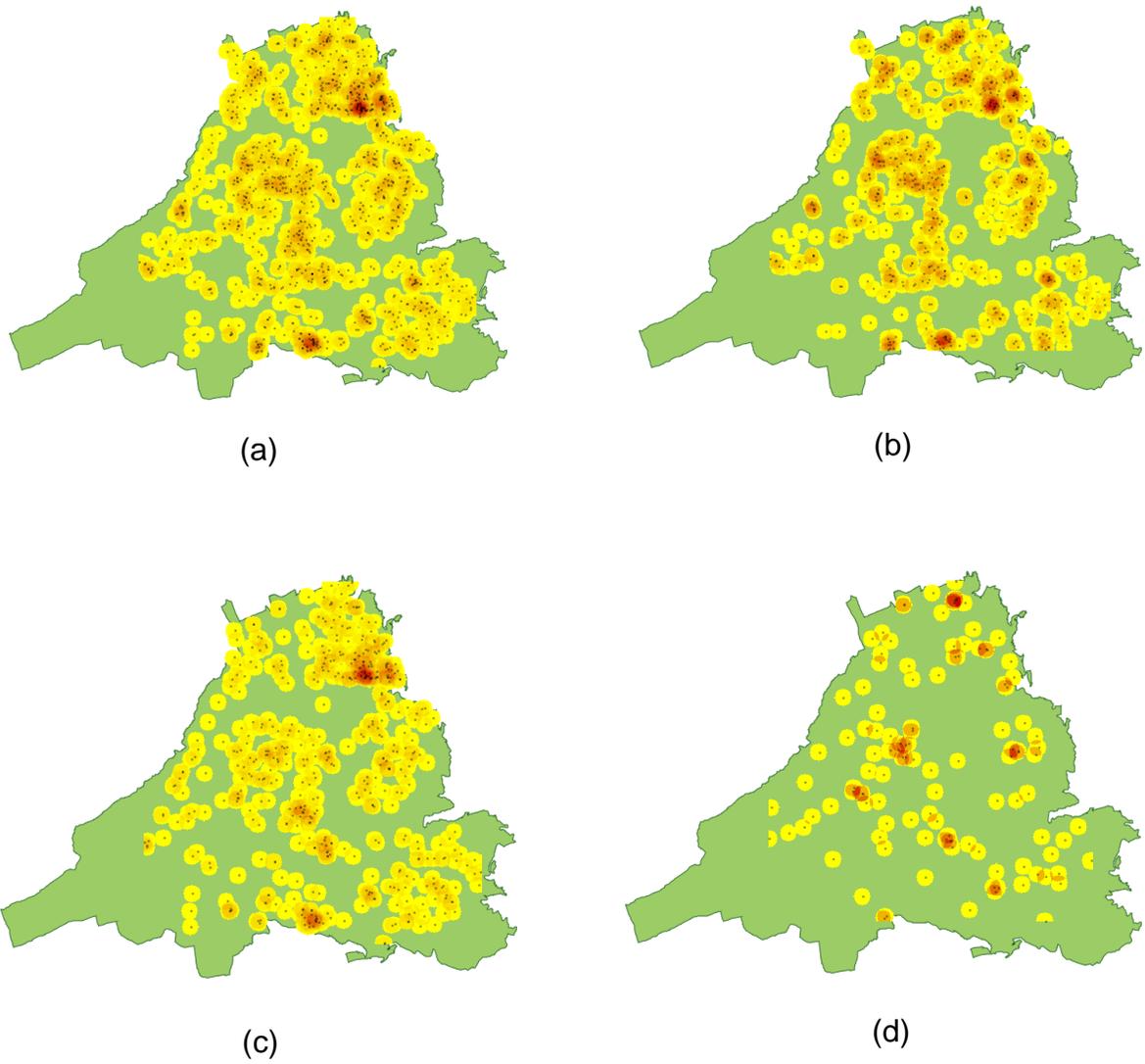
5.3.2. MÉTODO KDE

Los resultados que se obtuvieron gracias a la aplicación del método KDE a los datos de casos de dengue del año 2015, permiten detectar de una forma más clara y precisa, los lugares en los cuales se presenta mayor ocurrencia de casos de dengue dentro del municipio de Floridablanca. Hay que decir que, si al aplicar la función K de Ripley a los datos, esta nos arroja como resultado que los casos de dengue se distribuyen u ocurren de forma aleatoria, aunque, en el mapa se presenten concentraciones o aglomeraciones, esto no quiere decir que ocurren allí por alguna razón, ya sea, climática, social etc.

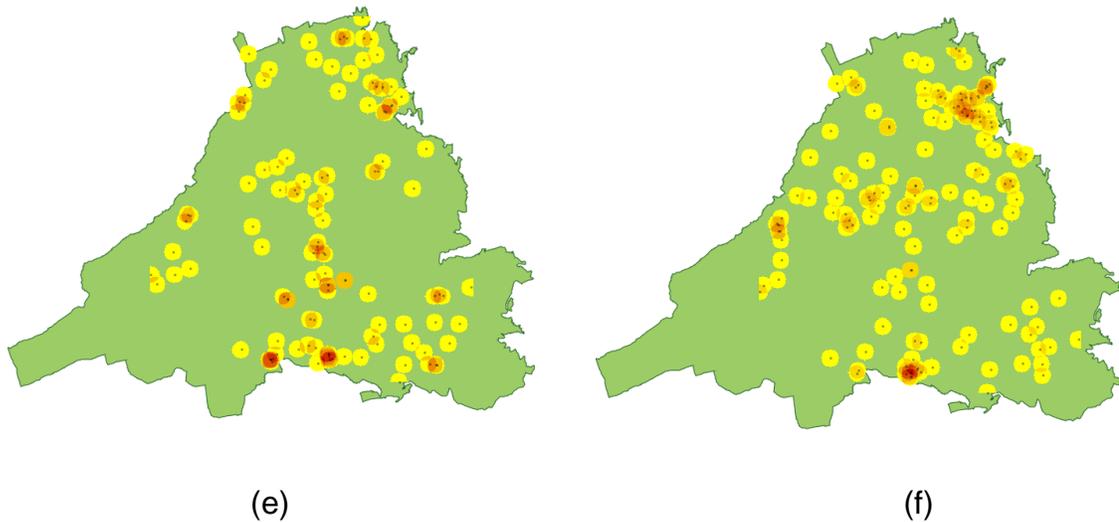
Lo que quiere decir que es recomendado, aplicar la función K de Ripley, antes de aplicar la función KDE, por lo tanto, una vez comprobado que los casos dengue no se distribuyen de forma aleatoria, si no que tienden a formar aglomeraciones, como reflejaban las gráficas de la función K de Ripley, se procede a aplicar la función KDE, con el objetivo de detectar las zonas en las que existe mayor ocurrencia de casos de dengue, es decir, los famosos puntos o zonas calientes, en el municipio de Floridablanca, para el año 2015.

En la ilustración 42, se muestran de forma gráfica los resultados de la función KDE, aplicando las mismas agrupaciones de datos o filtros que se usaron en la función K de Ripley.

Ilustración 42: Grupo de resultados del método KDE, dengue Floridablanca 2016⁴¹



⁴¹ Fuente los autores



Para todos los resultados mostrados anteriormente, se utilizó como función de núcleo la distribución uniforme, la cual se menciona en el apartado 5.2.1.2, se seleccionó un ancho de banda de 100 metros y una resolución de 10 metros.

La ilustración (a), representa los resultados al aplicar el método KDE a todos los datos, es decir 990 casos de dengue, lo cual arroja un total de 3 puntos o zonas calientes, bien marcados, la primera zona está aproximadamente ubicada en los alrededores del sector del Reposo en los barrios de Villaluz y José A. Morales, cerca de ese hay otro, en el barrio Laureles, el tercero está en el barrio La Paz, al sur del municipio. Para la ilustración (b), en la cual se le aplicó un filtro para que la aplicación software realizara el cálculo sobre los casos de dengue en los cuales los afectados fueron personas con sexo masculino, arrojó cerca de 10 zonas calientes, 5 de ellas se encuentran en la zona norte del municipio y 3 en la zona sur, las zonas de mayor concentración de puntos están ubicadas aproximadamente en el barrio Santa Ana, Villabel, barrio La Cumbre, sector de lagos I y lagos IV, y al sur del municipio en los barrios, La Paz, González Chaparro y Bucarica.

Para la ilustración (c), se seleccionó el filtro de casos de dengue para el sexo femenino, la cual arrojó 3 zonas calientes, en las que se puede notar un color rojizo más fuerte, dichas zonas están ubicadas aproximadamente en el sector del Reposo, lagos IV y otra se localiza al sur del municipio, en el barrio La Paz, nuevamente. La ilustración (d), muestra el resultado de filtrar los casos de dengue que afectaron a las personas de edades entre los 0 y los 10 años, es decir, niños; mostrando un total de 7 zonas calientes bastante marcadas, ubicadas

respectivamente en el barrio el Campanazo, el Reposo, La Cumbre, Villabel, lagos IV y el barrio Primavera. La ilustración (e), muestra los resultados del método KDE, aplicando un filtro para los casos de dengue que afectaron a personas con edades mayores a los 60 años, es decir, adultos mayores, arrojando como resultado cerca de 2 zonas relevantes o notoriamente marcadas, las cuales están localizadas en el barrio La Paz y el sector de Versalles, existen otras 4 zonas de menor relevancia pero igual se pueden observar aglomeraciones medianamente altas, ubicadas en el sector del Reposo, el san Bernardo, lagos IV y villa España. Para la ilustración (f), se hizo un filtro para los casos de dengue que ocurrieron entre las fechas 01 de octubre de 2015 al 31 de octubre de 2015, se escogió dicho filtro, porque se detectó que fue el mes en el cual ocurrieron más casos de dengue de ese año, con un total de 113 casos reportados, arrojando 2 zonas calientes, una de ellas localizadas en el sector del reposo, y la otra al sur del municipio en el barrio La Paz.

5.4. CONFRONTACIÓN DE RESULTADOS

El dengue en el municipio de Floridablanca no había sido abordado, por medio de métodos de análisis espacial, como los realizados en este proyecto, por lo tanto, no existen estudios que abarquen este tema en este municipio, o al menos no publicados a la fecha, lo que hace que no sea posible hacer una comparación de resultados, pero, si se puede efectuar una observación de resultados con respecto a otros estudios que realizaron un análisis del dengue; por ejemplo, Brasil, en el cual se hallaron 2 artículos, el primero de ellos se titula, análisis espacial de la aparición del dengue y las condiciones de vida en Nova Iguaçu, estado de Río de Janeiro, Brasil.

Dicho estudio observa la relación que existe entre las condiciones de vida y la probabilidad de que ocurran casos de dengue en un área investigada, ahora bien, la relación entre las clases socio económicas, el cambio de clima y como esto afecta la ocurrencia y distribución de la enfermedad en los distintos barrios del municipio de Nova entre los años comprendidos de 1996 al 2004. (Machado, Oliveira, & Souza-Santos, 2009). Lo que hallaron fue que la población fue afectada de forma similar cuando compararon los sexos y que los barrios más perjudicados fueron aquellos que se encontraban cercanos a las vías de acceso, en la ilustración 43 se destacan los barrios que tuvieron mayor incidencia de casos de dengue en el municipio de Nova Iguaçu, utilizando el método de Kernel.

Para el caso de Floridablanca, los resultados arrojaron un total 990 casos de dengue, de los cuales el 53% fueron personas con sexo masculino y 47% del sexo femenino, lo que quiere decir que hubo mayor incidencia de hombres afectados, a

pesar de que según del DANE en 53,2% de la población de Floridablanca son mujeres. Por otra parte, el mes con más casos reportados fue octubre con un total de 113 casos reportados.

Ilustración 43: Incidencia acumulada de dengue en Nova Iguaçu, Río de Janeiro, Brasil, 1996 a 2004.⁴²

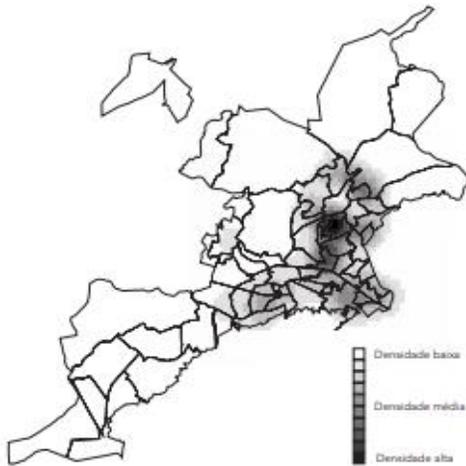
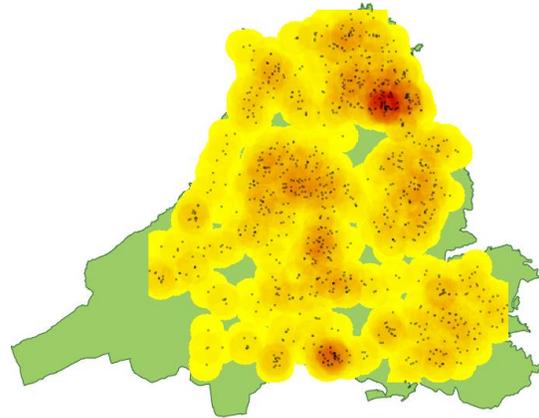


Ilustración 44: Incidencia de casos de dengue en Floridablanca, Santander 2015



El segundo artículo que seleccionó para realizar esta confrontación de resultados con respecto al análisis del dengue es; patrón de propagación de la primera epidemia de dengue en la ciudad de Salvador, Brasil, el objetivo para el cual se realizó dicho estudio fue el de detectar la distribución espacial a través del tiempo de la primera epidemia de dengue que se presentó en dicha ciudad, (Barreto, Teixeira, Costa, Carvalho, & Barreto, 2008), para lo cual aplicaron un método de estimación de densidad de núcleo, para las diferentes semanas epidemiológicas y los casos de dengue reportados en cada una de estas semanas, en la ilustración 45, muestra la distribución del dengue en el año 1995 en la ciudad de Salvador, siendo los barrios marginados y la península de Itapagipe, las zonas más afectadas, además, se percataron de que los distritos con mayor poder socio-económico se reportaron menos casos de dengue.

El análisis realizado en Floridablanca arrojó que las zonas más afectadas, por el dengue en el municipio en el año 2015, también conocidas como puntos calientes fueron tres; dos de ellos se encuentran en la zona norte del municipio, el primero de ellos, en el sector de Reposo, exactamente entre el barrio Villaluz y José A

⁴² Tomado de (Machado et al., 2009), página 6, gráfica 3c.

Morales, el segundo en el barrio los Laureles, estos 3 barrios se encuentran clasificados en el estrato socioeconómico 1. La tercera zona se halla en la región sur de Floridablanca, exactamente en el barrio La Paz.

Ilustración 45: Distribución espacial de casos notificados de dengue por área censal.

Salvador-Bahía, Brasil, 1995.⁴³

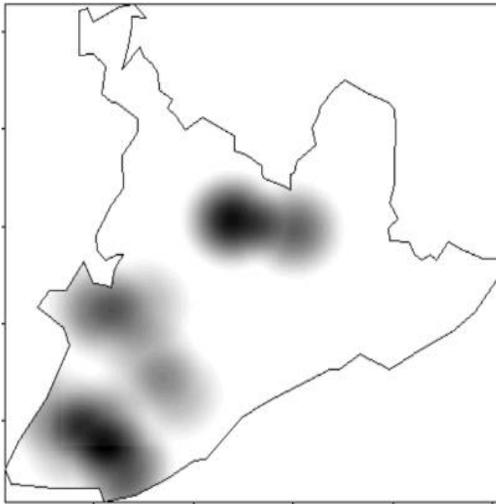
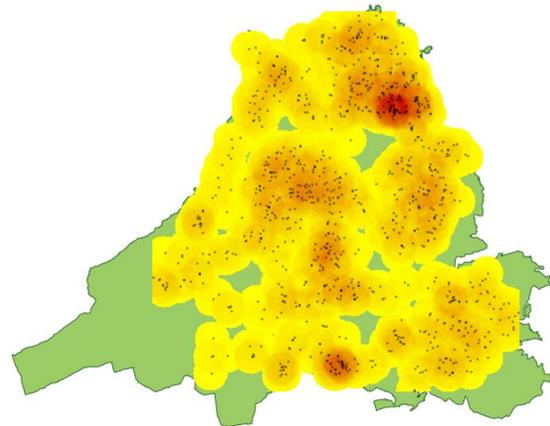


Ilustración 46: Distribución espacial de los casos de dengue en Floridablanca Santander 2015



Otro artículo relevante e interesante para realizar una confrontación de resultados es; distribución espacial del dengue basado en herramientas del Sistema de Información Geográfica, Valle de Aburrá, Colombia, debido a que es un artículo realizado en Colombia, más específicamente la región Andina, misma región en la cual se encuentra el departamento de Santander y por supuesto el municipio de Floridablanca.

Para determinar la distribución espacial del dengue mes a mes en los años comprendidos entre 2008 al 2011, omitiendo el año 2010, en el Valle de Aburrá, utilizaron ArcGIS 10. Por otra parte, utilizaron Slope, para corroborar la autocorrelación del fenómeno. Para realizar dichos análisis tuvieron un total de 2393 casos (Londoño C. et al., 2014), siendo el año 2009, el que presentó mayor número de casos, un total de 842, y a su vez el mes de diciembre de ese año el que presentó más casos de dengue reportados, además, la comuna 16 Belén, fue en la que se detectó mayores niveles de concentración de casos, en la ilustración 47, se presenta dicha aglomeración. Por otra parte, detectaron que existía un

⁴³ Tomado de (Barreto et al., 2008), página 7.

patrón de comportamiento, y era que la mayor cantidad de casos de dengue se presentaban en el segundo semestre del año.

En el caso de Floridablanca, se detectó que el mes más endémico del año 2015 en lo que corresponde a dengue fue octubre como se ha mencionado anteriormente, este mes pertenece al segundo semestre del año y se reportaron un total de 113 casos, siendo los barrios La Paz y el sector de Versalles, Reposo, el san Bernardo, lagos IV y villa España, los que presentaron mayor aglomeración de sucesos, todo puede relacionarse a que el mes de octubre tiende a ser un mes que presenta fuertes lluvias.

Ilustración 45: Distribución del dengue en el mes de diciembre en el año 2009, Valle de Aburrá.⁴⁴

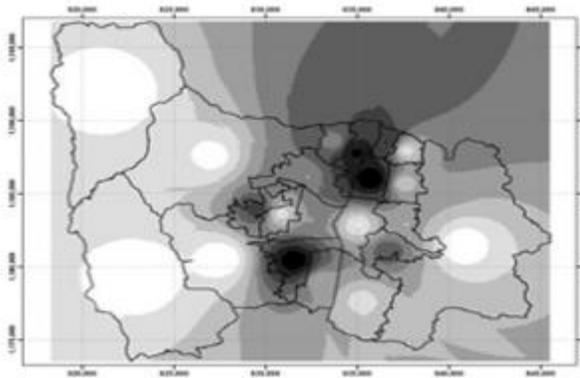
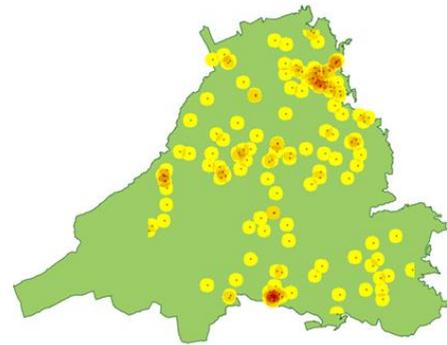


Ilustración 44: Distribución del dengue en el mes de octubre 2015, Floridablanca.



Por otra parte, se procedió a aplicar el método KDE a los datos de casos de dengue ocurridos en el municipio de Floridablanca en 2015, es decir 990 reportes, por medio del software ArcGIS versión 10.4, con el objetivo de comparar los resultados obtenidos en la aplicación software desarrollada en este proyecto y los producidos por el software mencionado anteriormente. Y con ello corroborar el debido funcionamiento de la aplicación software con respecto a ArcGIS, siendo este software un conjunto de herramientas de Sistemas de Información Geográfico bastante robusto y muy usado en el campo de análisis espacial.

A continuación, se presentan los resultados KDE obtenidos, tanto del software ArcGIS, como de la aplicación software desarrollada, aplicando diferentes anchos de banda, los cuales son 100 metros, 200 metros, 500 metros y 1000 metros, mostrando mapas de calor muy similares entre sí, para cada uno de los anchos de banda seleccionado, pero con una diferencia en la cantidad de colores mostrada,

⁴⁴ Tomado de (Londoño C. et al., 2014), página 11, figura 3.

ya que en este proyecto por cada densidad se asigna un color diferente y en ArcGIS, estos colores son asignados a un rango de densidad lo que disminuye la cantidad de colores a mostrar y hace ver franjas más uniformes de un mismo color.

Ilustración 46: Distribución del dengue en el municipio de Floridablanca, ancho de banda 100 metros⁴⁵

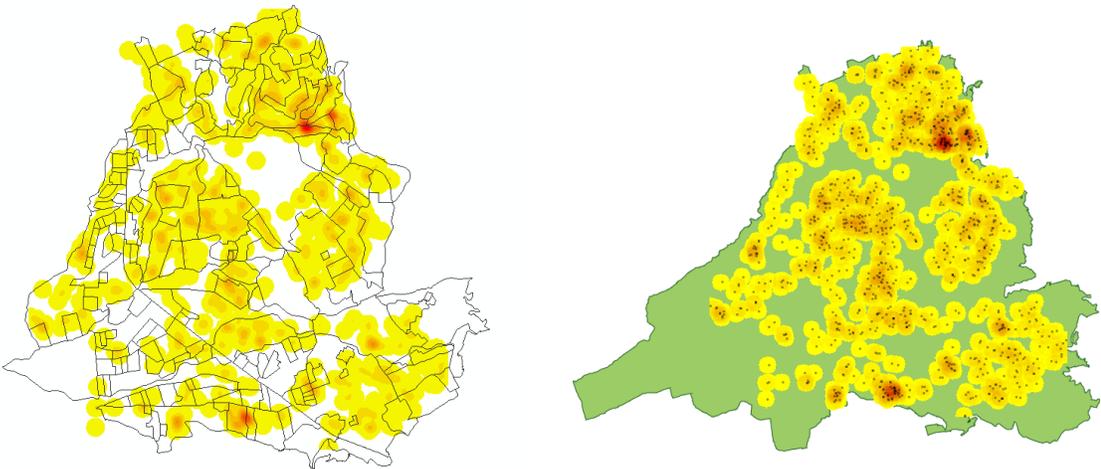
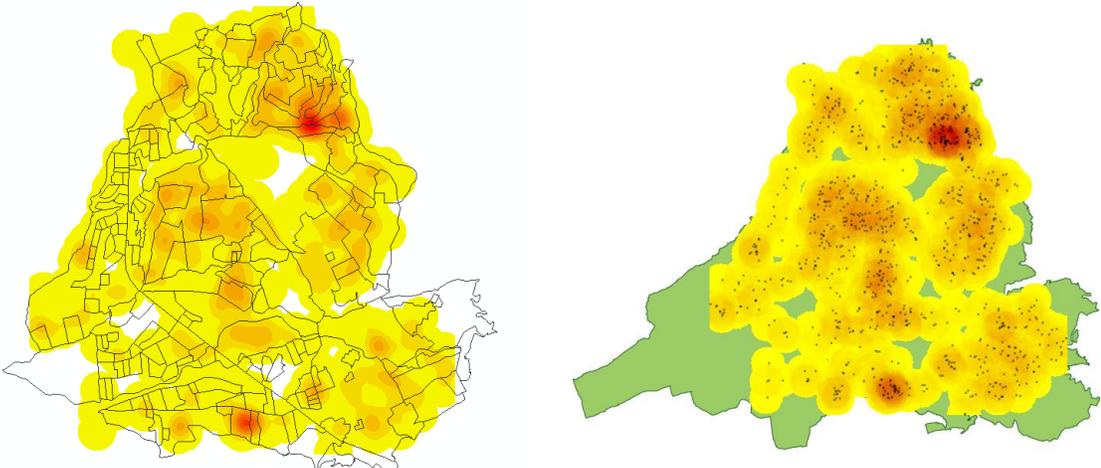


Ilustración 47: Distribución del dengue en el municipio de Floridablanca, ancho de banda 200 metros⁴⁶



⁴⁵ Fuente los autores

⁴⁶ Fuente los autores

Ilustración 48: Distribución del dengue en el municipio de Floridablanca, ancho de banda 500 metros⁴⁷

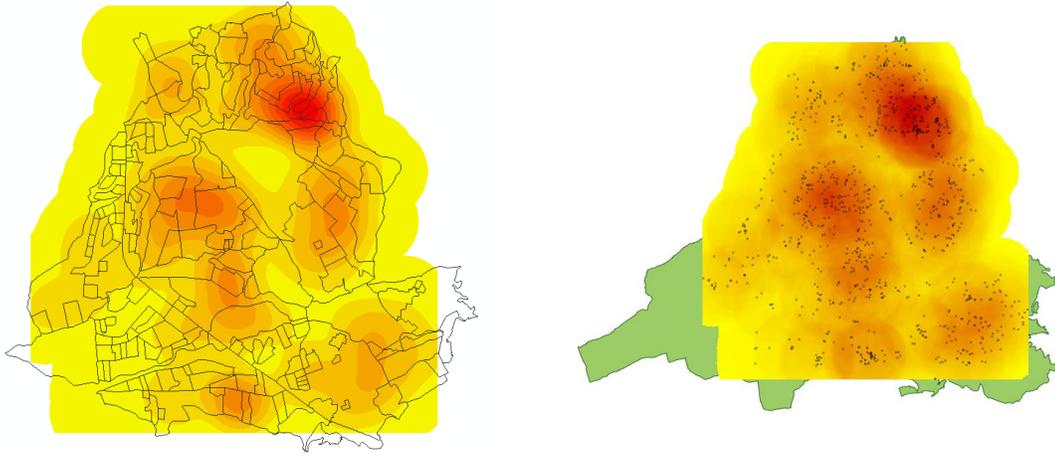
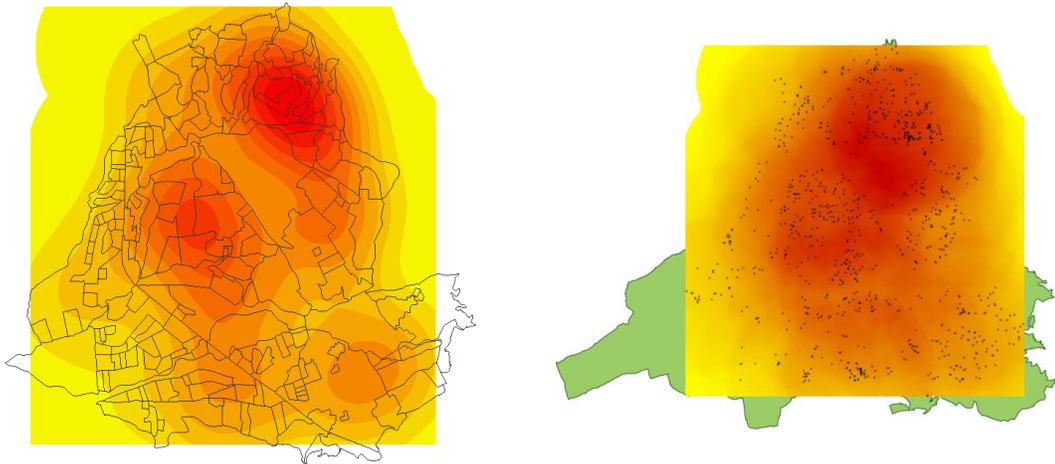


Ilustración 49: Distribución del dengue en el municipio de Floridablanca, ancho de banda 1000 metros.⁴⁸



⁴⁷ Fuente los autores

⁴⁸ Fuente los autores

6. CONCLUSIONES

Al emplear la Función K de Ripley, sobre los datos de casos de dengue ocurridos en el municipio de Floridablanca en el año 2015, se pudo determinar que dichos casos presentan fuertes tendencias de aglomeración o agrupación, y por lo tanto, no ocurren de forma aleatoria producto del azar, asimismo, como se pudo comprobar dicha aglomeración también se produce para los casos de dengue a los cuales se les aplicaron diferentes filtros, con el objetivo de corroborar que para dichos grupos de datos; la tendencia a aglomerarse también existía.

Después de que se identificó que los casos de dengue no se distribuyen de forma aleatoria, por medio de la Función K de Ripley, en el municipio de Floridablanca, identificar las zonas más afectadas por este fenómeno se hace muy importante, ya que, se pueden conocer los focos de aglomeración de los casos de dengue, por lo tanto, ahora entra en funcionamiento el método KDE, el cual permite detectar de forma gráfica y por medio de un mapa, los puntos o zonas de calor y como se encuentran distribuidas dichos puntos dentro del mapa, esta detección, posiblemente ayude a encontrar similitudes entre los barrios en los cuales existe mayor concentración de casos de dengue, ya sea por su localización geográfica, estratos social de sus habitantes, actores u objetos aledaños que puedan ayudar a promover la proliferación de *Aedes Aegypti*, entre otros factores.

Al utilizar la función K de Ripley junto con el método KDE, arrojó los resultados necesarios, para que en manos de la entidad responsable de la salud pública, puede ayudar notablemente a la reducción de ocurrencia del dengue en el municipio de Floridablanca, ya que, permite identificar las zonas o barrios más vulnerables y así promover campañas para obtener un impacto positivo que ayude a reducir el dengue; como se ha visto reflejado en otros lugares del mundo, en los cuales, por medio de herramientas similares a la aplicación software desarrollada en este proyecto y por medio del trabajo conjunto con las respectivas entidades responsables de la atención de dichos eventos, ayudaron notablemente en la reducción de la frecuencia de ocurrencia de estos eventos, como por ejemplo, casos de hurtos a residencias, casos de accidentalidad, casos de crímenes callejeros, casos de minas antipersonales, entre muchos más. Lo que quiere decir, que la implementación de la aplicación software en el campo de análisis espacial, con el objetivo de detectar patrones de aglomeración y ocurrencia de un fenómeno como el estudiado en este proyecto, puede ser de gran apoyo en la identificación de dichos patrones de aglomeración y, por lo tanto, ser de gran ayuda.

7. TRABAJO FUTURO

En el módulo de visualizador de resultados KDE, si bien es cierto que los puntos calientes se ubican de forma correcta y precisa sobre el polígono del mapa, se recomienda agregar una división por barrios, distritos o sectores, con el objetivo de que sea más rápida y exacta la identificación de las regiones más vulnerables y, por lo tanto, no sea necesario recurrir a otras herramientas externas para dicho proceso, ya que, puede generar confusión al momento de ubicar las regiones más afectadas. Con respecto a este módulo, también se recomienda implementar una opción en la cual se le permita al usuario hacer zoom sobre el mapa de calor, para que, por ejemplo, se pueda observar con más detalle la cantidad de puntos en alguna zona en la que se desea conocer dicho dato.

La herramienta software fue desarrollada para que soportara únicamente archivos KML, aunque dichos archivos tienen un gran soporte por la compañía Google, lo que los hace altamente difundidos, no son el tipo de archivo más común en lo que corresponde a la georreferenciación o cartografía digital, ya que, los archivos SHP son los más comunes para estas tareas, por lo que es recomendable permitir que la aplicación software soporte estos y otros tipos de archivos utilizados para estas tareas.

Otra temática que tratar podría ser la exportación de resultados del módulo KDE, por medio de archivos KML, ya que, el proceso para realizar la exportación a este formato tarda alrededor de 2 horas, por lo que se recomienda, de ser posible, buscar la manera de hacer una optimización a este algoritmo, con el objetivo de reducir el tiempo de exportación, ya que, para archivos con un mayor número de datos, puede llevar tiempos exagerados para realizar dicho proceso. Por otra parte, y con respecto al módulo de exportación, se podría implementar la opción que permita exportar archivos KMZ con imágenes ráster incluidas, Así como brindarle al usuario la opción de elegir la gama de color de su preferencia, también poder escoger aplicar un color diferente para cada densidad como se hace en este proyecto, o crear rangos de densidad asignando para cada rango un diferente color, permitiendo así disminuir la cantidad de colores mostrada, y ver el mapa con franjas más uniformes de un mismo color como se muestra normalmente en los mapas de calor.

Por otra parte, sería útil implementar una barra de progreso que muestre el estado del proceso que se está ejecutando y así el usuario pueda tener una idea más clara del tiempo aproximado que podría tomar dicho proceso y de esa forma mejorar la experiencia de usuario.

8. BIBLIOGRAFÍA

- Alegria, A. C., Sahli, H., & Zimanyi, E. (2011). Application of density analysis for landmine risk mapping. *Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services*, 223–228. <https://doi.org/10.1109/ICSDM.2011.5969036>
- ArcGis. (2014). Color—ArcGIS Pro | ArcGIS Desktop. Retrieved October 18, 2018, from <https://pro.arcgis.com/es/pro-app/help/mapping/layer-properties/color.htm>
- ArcGIS. (2017). Cómo funciona la densidad kernel—Ayuda | ArcGIS Desktop. Retrieved October 20, 2018, from <https://pro.arcgis.com/es/pro-app/tool-reference/spatial-analyst/how-kernel-density-works.htm>
- Barreto, F. R., Teixeira, M. G., Costa, M. D. C. N., Carvalho, M. S., & Barreto, M. L. (2008). Spread pattern of the first dengue epidemic in the city of Salvador, Brazil. *BMC Public Health*, 8, 1–20. <https://doi.org/10.1186/1471-2458-8-51>
- Castrillón, J. C., Castaño, J. C., & Urcuqui, S. (2015). Dengue en Colombia: diez años de evolución. *Revista Chilena de Infectología*, 32(2), 142–149. <https://doi.org/10.4067/S0716-10182015000300002>
- Cockburn, A. (2004). Crystal Clear. A Human-Powered Methodology For Small Teams, including The Seven Properties of Effective Software Projects. *Integration The Vlsi Journal*, (January 2004), 39. <https://doi.org/10.1111/j.1600-0560.2011.01732.x>
- Colorizer. (2014). Colorizador - Selector y convertidor de color (RGB HSL HSB / HSV CMYK HEX LAB). Retrieved October 18, 2018, from <http://colorizer.org/>
- CrimeStat. (2005). About CrimeStat. Retrieved October 20, 2018, from <https://www.icpsr.umich.edu/CrimeStat/about.html>
- Dai, P., Jing, C., Du, M., & Zhou, W. (2015). A method based on spatial analyst to detect hot spot of urban component management events. *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*, 55–59. <https://doi.org/10.1109/ICSDM.2015.7298025>
- Dixon, P. M., El-shaarawi, A. H., & Piegorsch, W. W. (2002). *Ripley ' s K function Ripley ' s K function*. 3, 1796–1803.
- Duncombe, J., Clements, A., Hu, W., Weinstein, P., Ritchie, S., & Espino, F. E. (2012). Geographical information systems for dengue surveillance. *The American Journal of Tropical Medicine and Hygiene*, 86(5), 753–755. <https://doi.org/10.4269/ajtmh.2012.11-0650>
- Esri. (2004). ArcGIS Spatial Analyst | Overview. Retrieved October 20, 2018, from <http://www.esri.com/software/arcgis/extensions/spatialanalyst>
- Esri. (2018). About ArcGIS | Mapping & Analytics Platform. Retrieved October

- 20, 2018, from <https://www.esri.com/en-us/arcgis/about-arcgis/overview>
- Fernández-Coppel, I. a. (2001). Las Coordenadas Geográficas y la Proyección UTM. *Geoasbuilt.Es*, 1–52. Retrieved from <http://www.cartesia.org/data/apuntes/cartografia/cartografia-geograficas-utm-datum.pdf%5Cnhttp://geoasbuilt.es/tutoriales/topografia/utm.pdf>
- Gallego Salguero, Á., & Sánchez Marco, M. (2013). *Manual de topografía en ingeniería*. Retrieved from <http://ebookcentral.proquest.com/lib/unabsp/detail.action?docID=3219389>
- GeoDa. (2016). GeoDa on Github. Retrieved October 20, 2018, from <http://geodacenter.github.io/>
- Google. (2014). Introducción a la documentación de KML | Keyhole Markup Language | Google Developers. Retrieved October 18, 2018, from <https://developers.google.com/kml/documentation/?hl=es-419>
- gvSIG. (2015). gvSIG Desktop - Portal gvSIG. Retrieved October 20, 2018, from <http://www.gvsig.com/es/productos/gvsig-desktop>
- Halstead, S. B. (2007). Dengue. *The Lancet*, 370(9599), 1644–1652. [https://doi.org/10.1016/S0140-6736\(07\)61687-0](https://doi.org/10.1016/S0140-6736(07)61687-0)
- Hu, Z., Baynard, C. W., Hu, H., & Fazio, M. (2015). GIS mapping and spatial analysis of cybersecurity attacks on a florida university. *2015 23rd International Conference on Geoinformatics*, 1–5. <https://doi.org/10.1109/GEOINFORMATICS.2015.7378714>
- JC, M. A. C. N. F. G. y S. M. L. (2005). Spatial analysis of dengue transmission in a medium-sized city in Brazil. TT - [Análisis espacial de la transmisión del dengue en una ciudad de tamaño mediano en Brasil]. *Revista de Saude Publica / Journal of Public Health*, 39(3), 444–451. Retrieved from http://www.scielosp.org/pdf/rsp/v39n3/en_24799.pdf
- Johansson, E., Gåhlin, C., & Borg, A. (2015). Crime Hotspots: An Evaluation of the KDE Spatial Mapping Technique. *2015 European Intelligence and Security Informatics Conference*, 69–74. <https://doi.org/10.1109/EISIC.2015.22>
- Kosmo. (2013). Kosmo-Plataforma SIG libre corporativa - Kosmo Desktop. Retrieved October 20, 2018, from http://www.opensig.es/index.php?option=com_content&task=view&id=12&Itemid=46
- Londoño C., L. A., Restrepo E., C., & Marulanda O., E. (2014). Distribución espacial del dengue basado en herramientas del Sistema de Información Geográfica, Valle de Aburrá, Colombia. (Spanish). *Spatial Distribution of Dengue Based on Geographic Information Systems Tools, Aburra Valley. (English)*, 32(1), 7–15. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=94336343&lang=es&site=ehost-live>

- Machado, J. P., Oliveira, R. M. de, & Souza-Santos, R. (2009). Análise espacial da ocorrência de dengue e condições de vida na cidade de Nova Iguaçu, Estado do Rio de Janeiro, Brasil. *Cadernos de Saúde Pública*, 25(5), 1025–1034. <https://doi.org/10.1590/s0102-311x2009000500009>
- Maling, D. H. (1992). *Coordinate systems and map projections*. Pergamon Press.
- Moreno, F. (2016). *Aplicación para el análisis espacial de datos en espacios de redes con NetKDE y función K para redes*.
- Moreno, L. (2004). Modelos de color. Retrieved October 18, 2018, from <https://desarrolloweb.com/articulos/1483.php>
- Nyerges, T. L. (1991). Analytical Map Use. *Cartography and GIS*, 18. Retrieved from http://courses.washington.edu/geog469/Lab_2010/criticalquestions_amu.pdf
- Ocazonez-Jiménez, R. (2013). Virus del dengue de serotipo 1 (DENV-1) de Colombia: su contribución a la presentación del dengue en el departamento de Santander. *Biomédica*, 33, 22–30. Retrieved from http://www.scielo.org.co/scielo.php?pid=S0120-41572013000500004&script=sci_abstract
- Organización Mundial de la Salud. (2018, February 2). Dengue y dengue grave. Nota descriptiva. Retrieved September 6, 2018, from Organización mundial de la salud. Centro de prensa website: <http://www.who.int/es/news-room/fact-sheets/detail/dengue-and-severe-dengue>
- Padilla, J. ., Rojas, D. ., & Sáenz, G. (2012). *Dengue en Colombia: epidemiología de la reemergencia a la hiperendemia*. Retrieved from [https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/INEC/INV/Dengue en Colombia.pdf](https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/INEC/INV/Dengue%20en%20Colombia.pdf)
- Palaniyandi, M. (2014). Web mapping GIS : GPS under the GIS umbrella for Aedes species dengue and chikungunya vector mosquito surveillance and control. *International Journal of Mosquito Research*, 1(3), 18–25.
- Peña Llopis, J. (2005). *Sistemas de información geográfica aplicados a la gestión del territorio: entrada, manejo, análisis y salida de datos espaciales: teoría general y práctica para ESRI ArcGIS 9 (4a. ed.)*. Retrieved from <http://ebookcentral.proquest.com/lib/unabsp/detail.action?docID=3191662>
- Pérez Navarro, A., Botella Plana, A., & Muñoz Bollas, A. (2011). *Introducción a los sistemas de información geográfica y geotelemática*. Retrieved from <http://ebookcentral.proquest.com/lib/unabsp/detail.action?docID=3201076>
- Polonczyk, A., & Lesniak, A. (2018). A Spatial Analysis of Selected Categories of Offences in Krakow Based on Data from the National Safety Risk Map. *2018 Baltic Geodetic Congress (BGC Geomatics)*, 17–22. <https://doi.org/10.1109/BGC-Geomatics.2018.00010>

- QGis. (2013). QGIS Desktop. Retrieved October 20, 2018, from <https://www.qgis.org/es/site/about/features.html#qgis-desktop>
- Rosenberg, M. S., & Anderson, C. D. (2011). PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods in Ecology and Evolution*, 2(3), 229–232. <https://doi.org/10.1111/j.2041-210X.2010.00081.x>
- SANET. (2009). Welcome to SANET. Retrieved October 20, 2018, from <http://sanet.csis.u-tokyo.ac.jp/>
- Serrato Álvarez, P. K. (2009). *Referencias y proyección empleadas en la cartografía colombiana. Perspectiva Geográfica, Vol. 14: 67-98, 2009.* Retrieved from <http://ebookcentral.proquest.com/lib/unabsp/detail.action?docID=3194531>
- Snyder, J. P. (1987). Map Projections: A Working Manual. *U.S. Geological Survey Professional Paper 1395*, 383. <https://doi.org/10.2307/1774978>
- Utm Zones. (2015). Retrieved October 10, 2018, from [http://www.jaworski.ca/utmzones.htm#utm zones](http://www.jaworski.ca/utmzones.htm#utm%20zones)
- Vanguardia. (2016). *72 mil casos de dengue se han registrado desde 2010 en Santander.* Retrieved from <http://www.vanguardia.com/area-metropolitana/bucaramanga/363496-71936-casos-de-dengue-se-han-reportado-desde-2010>
- Vanguardia. (2018a). *30 personas se contagian de dengue cada semana en Santander.* Retrieved from <http://www.vanguardia.com/area-metropolitana/bucaramanga/428982-30-personas-se-contagian-de-dengue-cada-semana-en-santander>
- Vanguardia. (2018b). *Santander, sexto departamento en reporte de casos de dengue.* Retrieved from <https://www.vanguardia.com/area-metropolitana/bucaramanga/445129-santander-sexto-departamento-en-reporte-de-casos-de-dengue>
- Zuo Zhang, Yanzhong Liu, Biao Chen, & Kunlun Chen. (2013). Using GIS and KDE analysis spatial distribution on public housing households: A case study. *2013 8th International Conference on Computer Science & Education*, 925–930. <https://doi.org/10.1109/ICCSE.2013.6554044>

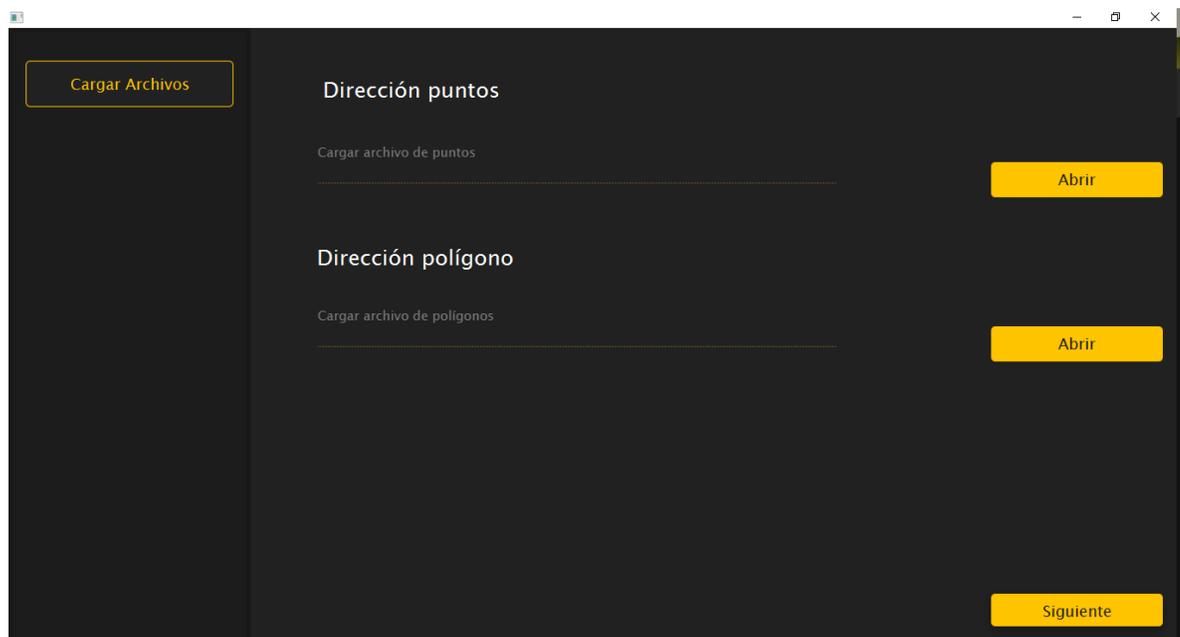
ANEXO A. MANUAL DE USUARIO

La aplicación fue desarrollada principalmente para el sistema operativo Windows, aunque puede ser ejecutado por cualquier sistema operativo que tenga Java. A continuación, se presentan los diferentes módulos incluidos dentro de la aplicación software, con los detalles de su respectivo uso.

A.1. INTERFAZ CARGAR ARCHIVOS

Esta interfaz es la que permite al usuario realizar la carga de archivos KML de los puntos y del polígono.

Ilustración 50: Interfaz de carga de archivos KML⁴⁹



El botón de la izquierda indica la interfaz activa o en la que se está posicionado, cada uno de los botones, permite cargar los archivos KML, el primero de ellos es para importar el archivo de puntos y el segundo es para la carga el archivo del polígono, luego de haber cargado los archivos correspondiente, podrá seleccionar el botón siguiente para que el software procese los archivos, y pasar a la interfaz

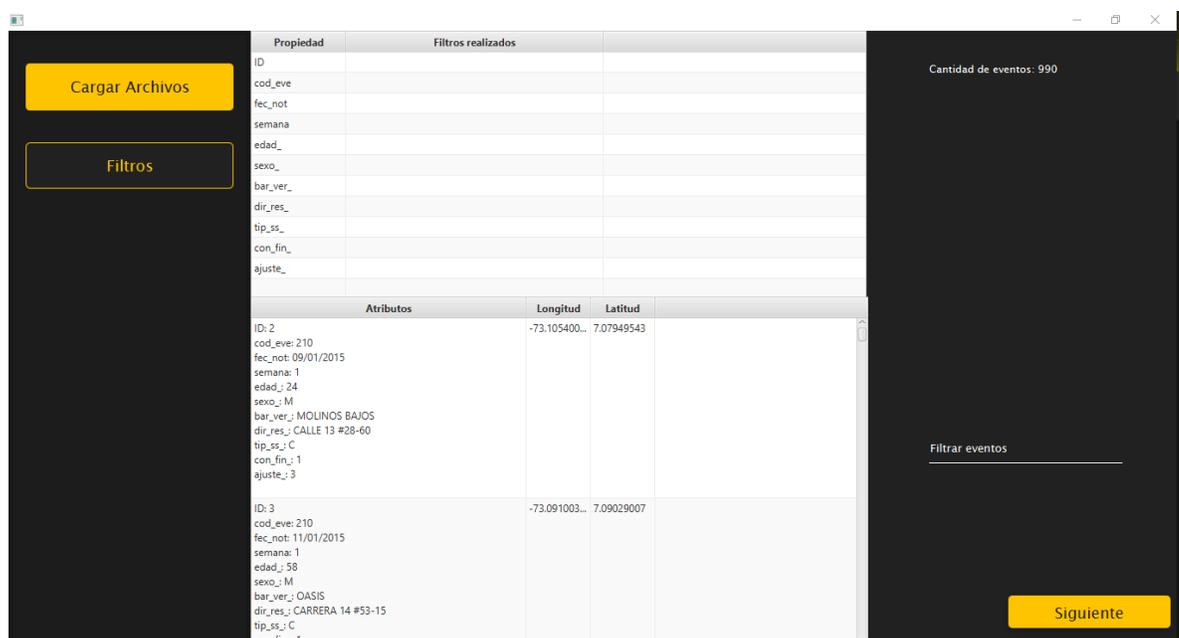
⁴⁹ Fuente los autores

de filtro, en el caso de que no se cargue uno de los archivos no se permitirá pasar a la interfaz de filtros.

A.2. INTERFAZ FILTRO DE INFORMACIÓN

Esta interfaz permite al usuario filtrar la información, medio de los diferentes atributos que tengan los casos, para los casos de dengue son un total de 11 atributos.

Ilustración 51: Interfaz de filtro de información⁵⁰

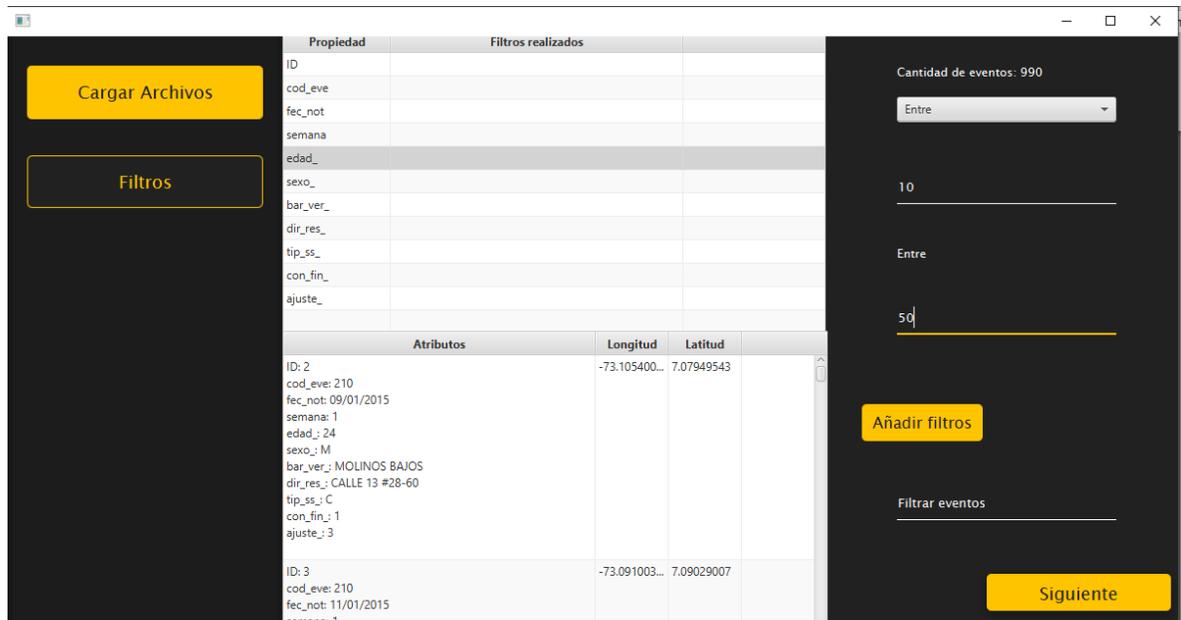


En la parte superior izquierda se ve la ventana o interfaz que está activa, en este caso filtros, estando en este paso, se puede regresar a la interfaz de carga de archivos.

En el medio, se pueden ver 2 tablas, en la superior se encuentran los parámetros por los cuales se hacen los filtros, en la tabla inferior están todos los puntos que se filtraron con sus respectivos atributos, en la parte superior derecha se muestra la cantidad de puntos con los que va a trabajar el software, y en la parte inferior derecha se encuentra un *input text*, para que el usuario ingrese el atributo.

⁵⁰ Fuente los autores

Ilustración 52: Interfaz de filtro de información, seleccionando parámetros⁵¹

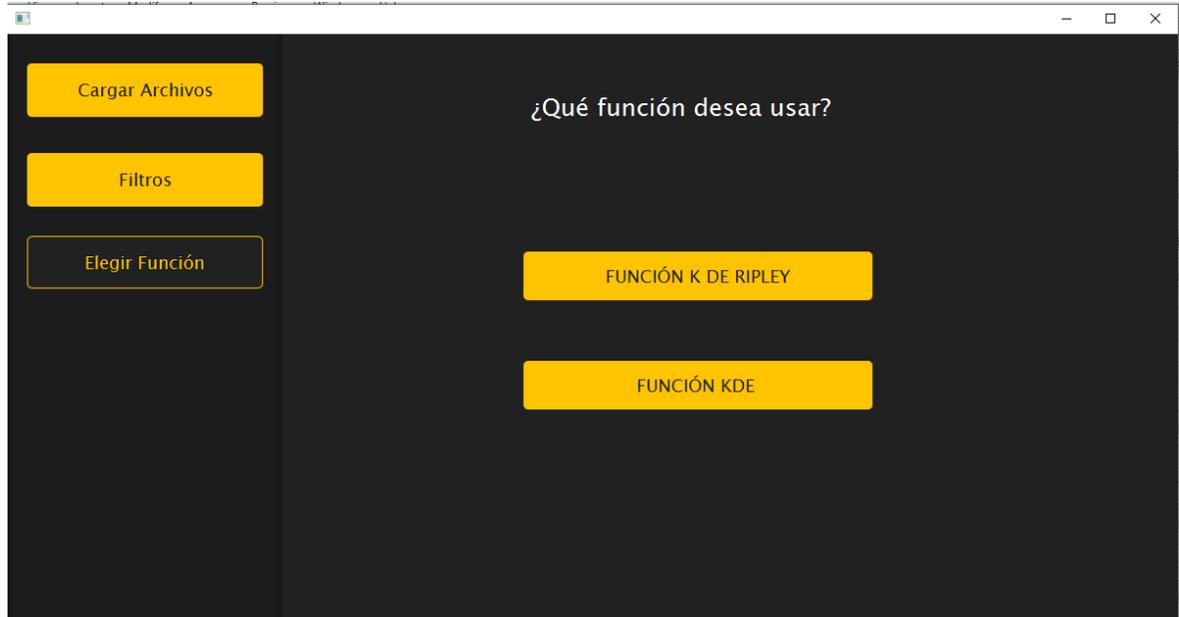


Para realizar un filtro, solo se debe hacer clic en el parámetro por el cual se quiere hacer dicho filtro, para este caso se seleccionó la edad, después, se habilitará un *menu button*, con 5 opciones, que son: mayor que, menor que, igual a, entre y diferente, debido a que, la edad es un valor numérico; en el caso de seleccionar por ejemplo, el sexo las opciones cambiarán a: igual, diferente y contiene. Una vez seleccionada la opción se escriben los valores que desea y oprime el botón siguiente, así, para todos filtros desee realizar, una vez agregado un filtro también puede borrarlo, en el botón borrar filtros; al dar al botón siguiente el software los redireccionará a la interfaz selección de función.

⁵¹ Fuente los autores

A.3. INTERFAZ SELECCIÓN DE MÉTODO

Ilustración 53: Interfaz selección de función⁵²



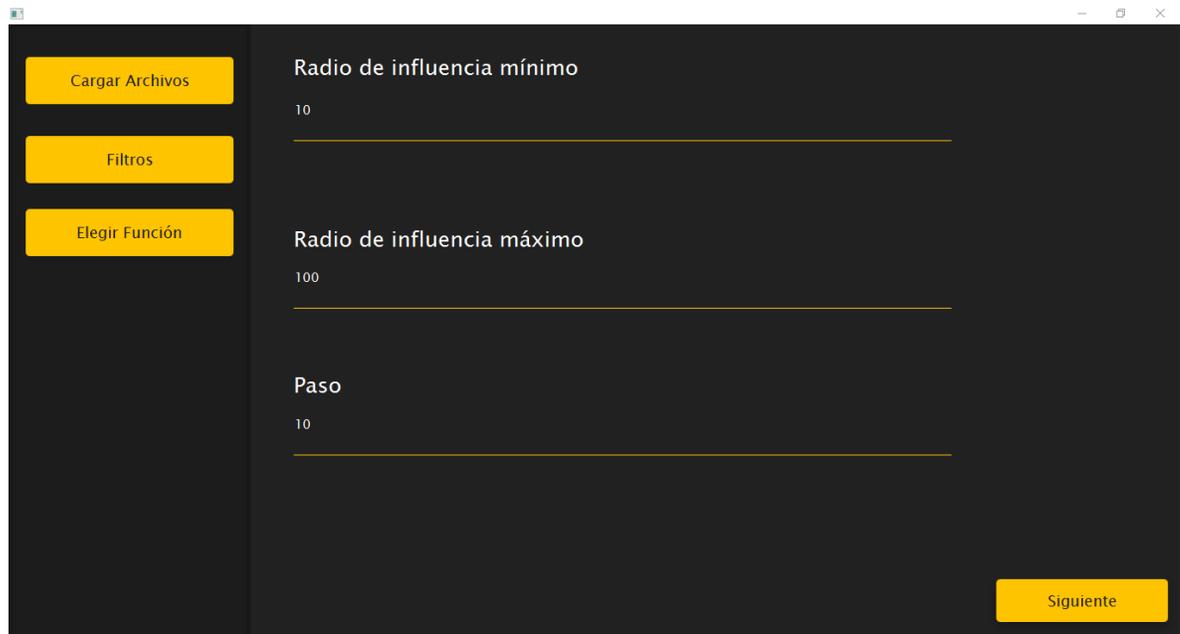
Esta interfaz está compuesta por los 3 botones ubicados en el apartado superior izquierda, que permite el cambio entre interfaces, además en la zona central de la pantalla se encuentran dos botones, por los cuales el usuario podrá seleccionar cual función desea ver primero, aunque se recomienda aplicar la función K de Ripley antes de la función KDE, por lo motivos ya mencionados, en la sección 5.2.1.3, pero las dos funcionan de forma separada, sin ningún inconveniente.

⁵² Fuente los autores

A.4. FORMULARIO INGRESO DE PARÁMETROS K DE RIPLEY

Este formulario permite al usuario ingresar los 3 parámetros que se requieren para que la función K de Ripley se ejecute de forma correcta, estos parámetros son: radio de influencia mínimo, radio de influencia máximo y el paso, cada uno de estos parámetros se explica en la sección 5.2.1.3.

Ilustración 54: Formulario de ingreso de parámetros KDE⁵³



The image shows a dark-themed web form with a sidebar on the left containing three yellow buttons: 'Cargar Archivos', 'Filtros', and 'Elegir Función'. The main area contains three input fields with yellow borders and labels: 'Radio de influencia mínimo' with the value '10', 'Radio de influencia máximo' with the value '100', and 'Paso' with the value '10'. A yellow 'Siguiente' button is located at the bottom right of the form.

En la zona central de este formulario, se ubican 3 *input text*, los cuales son: radio de influencia mínimo, es decir, cual es la distancia mínima a la que desea hacer la correlación de los puntos, el radio de influencia máximo es la distancia hasta donde va a tomar la relación que existe de un punto con respecto a otro y por último el paso, que significa la distancia en la que a va calculando esa relación desde el radio de influencia mínimo, es este caso 10, al máximo 100. Estos valores permiten calcular los valores K

Al oprimir el botón siguiente, hará los cálculos correspondientes y mostrará la gráfica K de Ripley con los valores K, el tiempo de este proceso varía

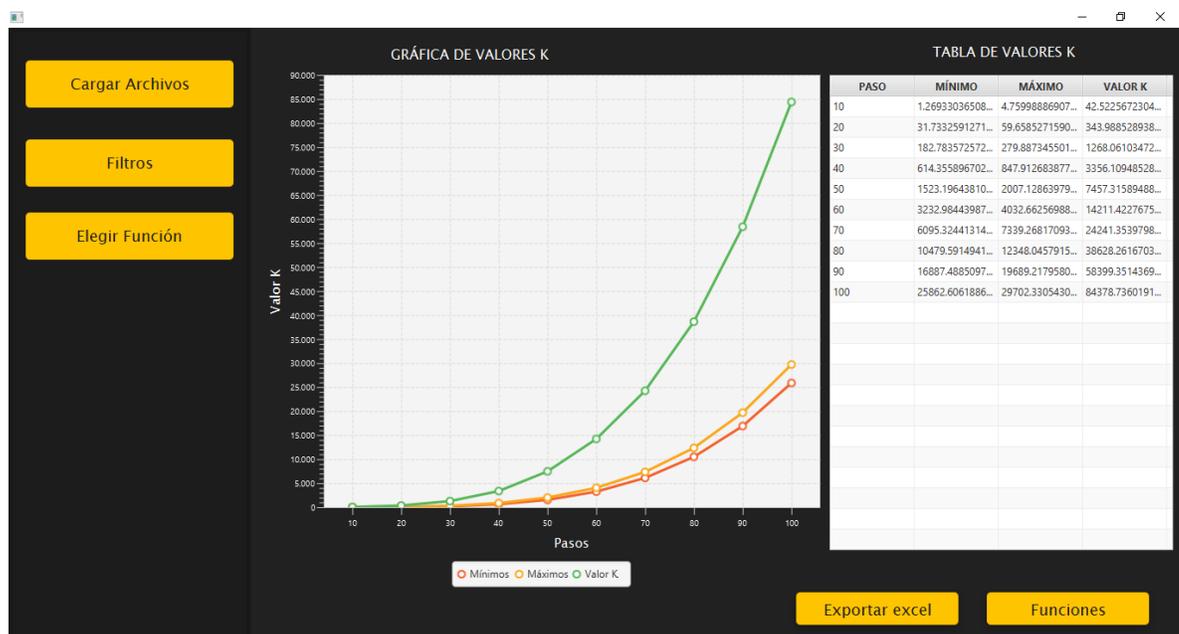
⁵³ Fuente los autores

dependiendo de la cantidad de datos, para 990 casos de dengue, tarda aproximadamente 303000 milisegundos.

A.5. INTERFAZ DE VISUALIZACIÓN RESULTADOS K DE RIPLEY

En esta interfaz muestra la gráfica K de Ripley con los valores K generados por el software, en base a los archivos KML cargados, los filtros realizados, y los parámetros K de Ripley ingresados anteriormente.

Ilustración 55: Interfaz de visualización de resultados K de Ripley⁵⁴



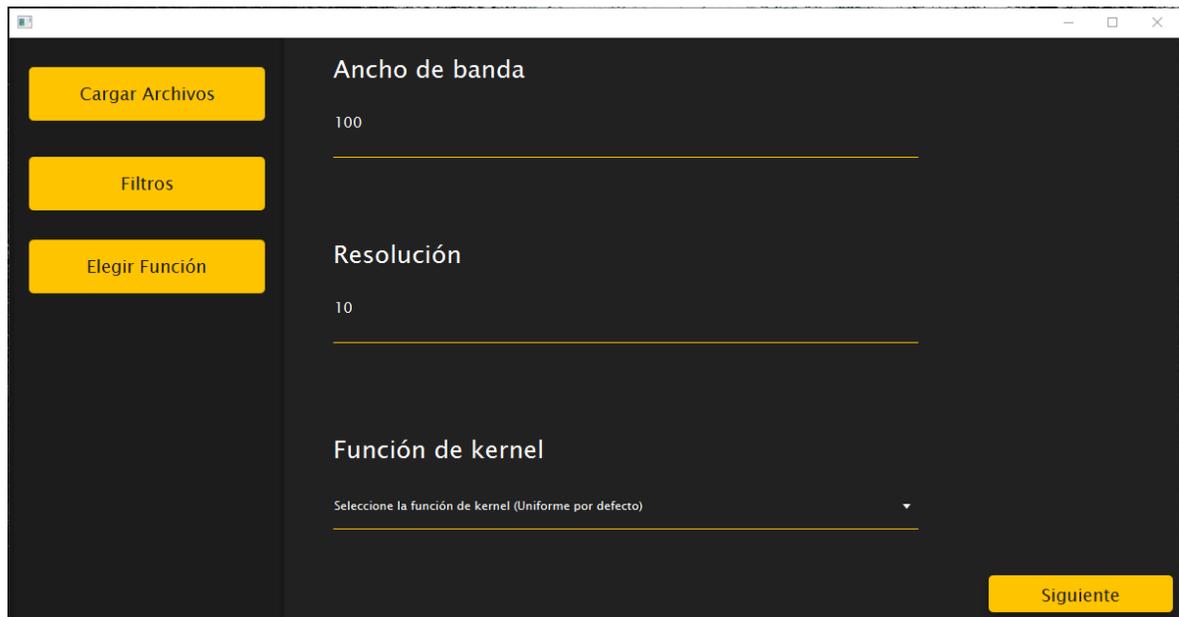
La interfaz de visualización resultados K de Ripley, está compuesta por la gráfica de los valores K, cada una de las tres líneas y su significado se explica en la sección 5.3.1, además, los pasos que ingresó el usuario en el formulario de ingreso de parámetros K de Ripley, en la parte derecha, se halla una tabla con los valores K mínimos y máximos generados para cada paso y los valores K reales de los casos de dengue para cada paso de igual forma. En la parte inferior se encuentra 2 botones, uno de ellos es para exportar la tabla de los valores K en formato XLSX y el otro es para regresar a la interfaz seleccionar función y realizar nuevamente el proceso.

⁵⁴ Fuente los autores

A.6. FORMULARIO INGRESO DE PARÁMETROS KDE

Este formulario permite al usuario ingresar los 3 parámetros que se requieren para el correcto funcionamiento del método KDE, los cuales son el ancho de banda, resolución y función de Kernel, cada uno de estos parámetros se explica detenidamente en la sección 5.2.1.2.

Ilustración 56: Formulario ingreso de parámetros KDE⁵⁵



The screenshot shows a web-based form for entering KDE parameters. It features a dark background with yellow accents. On the left side, there is a vertical sidebar with three yellow buttons: 'Cargar Archivos', 'Filtros', and 'Elegir Función'. The main content area is divided into three sections. The first section is titled 'Ancho de banda' and contains a text input field with the value '100'. The second section is titled 'Resolución' and contains a text input field with the value '10'. The third section is titled 'Función de kernel' and contains a dropdown menu with the text 'Seleccione la función de kernel (Uniforme por defecto)'. At the bottom right of the form, there is a yellow button labeled 'Siguiente'.

Este formulario está compuesto por los 3 botones ubicados en la parte superior izquierda que permite el cambio entre interfaces, en la parte central, se halla 2 *input text*, ancho de banda y resolución, en los cuales ingresará los valores que desee y 1 *menu button*, en el que usuario seleccionará la función con la que desea que se calcule la densidad de núcleo; cada función varía una de la otra y esto afectará un poco el mapa de calor.

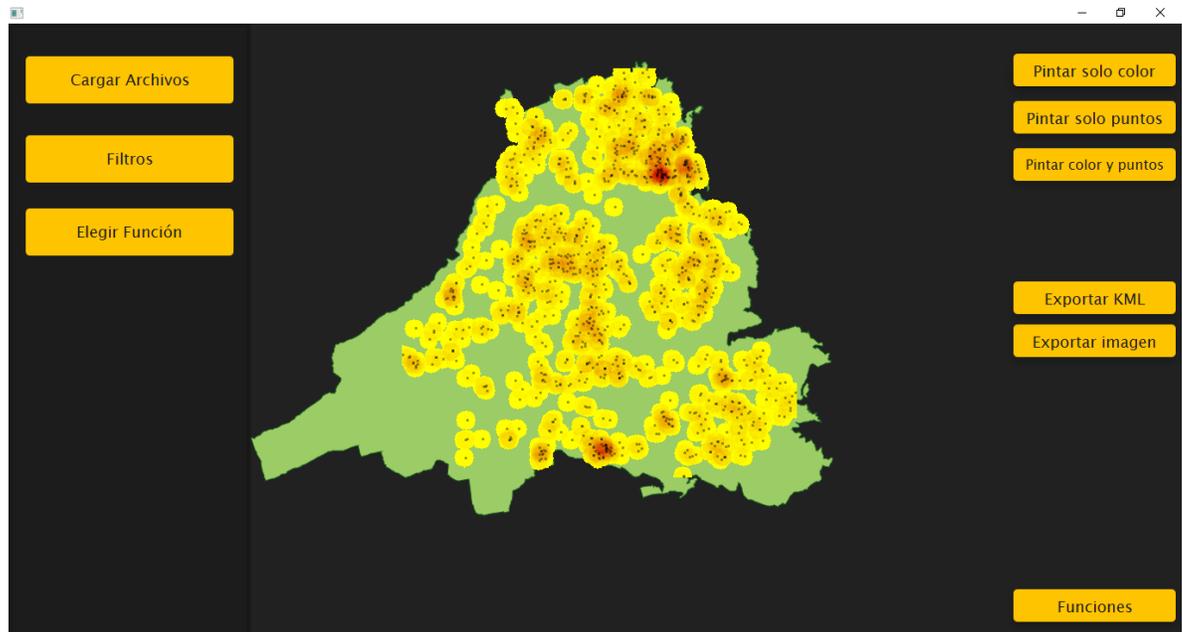
Al oprimir el botón siguiente, hará los cálculos correspondientes y mostrará el mapa de calor, el tiempo de este proceso varía dependiendo de la cantidad de datos, para 990 casos de dengue, tarda aproximadamente 13210 milisegundos.

⁵⁵ Fuente los autores

A.7. INTERFAZ VISUALIZACIÓN RESULTADOS KDE

En esta interfaz podrá ver el mapa de calor generado por el software en base a los archivos KML cargados, los filtros realizados, y los parámetros KDE ingresados anteriormente.

Ilustración 57: Interfaz de visualización de resultados KDE⁵⁶



La interfaz visualización de resultados KDE, está compuesta por los 3 botones ya mencionados en las anteriores interfaces, con su respectiva función, en la parte central está el mapa de calor generado y a la derecha se encuentran 3 botones que permiten cambiar la apariencia del mapa, ya sea pintando solo el color o solo los puntos o el color y los puntos, como se muestra en la ilustración 56, luego se hallan los botones de exportación de resultados de esta interfaz, exportar a archivos KML y exportar a imagen PNG, y en la parte inferior se encuentra el botón funciones para regresar a la interfaz de selección de métodos.

⁵⁶ Fuente los autores