

**APLICACIÓN PARA EL ANÁLISIS ESPACIAL DE DATOS EN ESPACIOS EN REDES
CON NETKDE Y FUNCIÓN K PARA REDES**

FEISAR ENRIQUE MORENO CORZO

**UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA – UNAB
FACULTAD DE INGENIERÍA DE SISTEMAS
MAESTRÍA EN GESTIÓN, APLICACIÓN Y DESARROLLO DE SOFTWARE
2017**

**APLICACIÓN PARA EL ANÁLISIS ESPACIAL DE DATOS EN ESPACIOS EN REDES
CON NETKDE Y FUNCIÓN K PARA REDES**

FEISAR ENRIQUE MORENO CORZO

Trabajo de Grado

**Director: Ingeniero Wilson Briceño Pineda
Codirector: Ingeniero Rafael Gustavo Ortiz Martínez**

**UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA – UNAB
FACULTAD DE INGENIERÍA DE SISTEMAS
MAESTRÍA EN GESTIÓN, APLICACIÓN Y DESARROLLO DE SOFTWARE
2017**

CONTENIDO

	pág.
RESUMEN	9
INTRODUCCIÓN	10
1. PLANTEAMIENTO DEL PROBLEMA	11
2. ESTADO DEL ARTE	13
2.1. ESPACIOS EUCLIDIANOS DE DOS DIMENSIONES	13
2.2. ESPACIOS DE REDES	14
2.3. HERRAMIENTAS COMPUTACIONALES	15
2.4. ANÁLISIS DE PATRONES DE PUNTOS – PPA	16
2.4.1. Kernel Density Estimator – KDE	17
2.4.2. La Función K	18
2.5. ANÁLISIS DE PATRONES DE PUNTOS EN REDES	21
2.5.1. NetKDE	21
2.5.2. Función K para Redes	23
3. OBJETIVOS	26
3.1. OBJETIVO GENERAL	26
3.2. OBJETIVOS ESPECÍFICOS	26
4. METODOLOGÍA DE LA INVESTIGACIÓN	27
4.1. DEFINICIÓN DE LA ARQUITECTURA DE DESARROLLO DE SOFTWARE	27
4.2. DEFINICIÓN DE CASOS DE USO	28

4.3. DESCRIPCIÓN DE CASOS DE USO	29
4.3.1. Visualizar Datos Geográficos	29
4.3.2. Cargar Archivo de Red	30
4.3.3. Cargar Archivo de Eventos	32
4.3.4. Realizar Cierre de Nodos	35
4.3.5. Realizar Proyección de Puntos	37
4.3.6. Generar Distribuciones Aleatorias de Puntos	39
4.3.7. Calcular Valores de Función K para Redes	41
4.3.8. Calcular Densidades de NetKDE	44
4.4. ALISTAMIENTO DE DATOS	47
5. ANÁLISIS DE RESULTADOS	52
5.1. HALLAZGOS EN LA IMPLEMENTACIÓN DE LA HERRAMIENTA SOFTWARE	52
5.1.1. Módulo de Carga de Archivos	52
5.1.2. Procesos de Normalización de Datos	54
5.1.3. Módulo de Función K para Redes	58
5.1.4. Módulo de NetKDE	60
5.1.5. Módulo de Visualización de Resultados Geográficos	61
5.1.6. Módulo de Creación y Verificación de Base de Datos Relacional Asociada	63
5.2. ANÁLISIS DE RESULTADOS DE LA APLICACIÓN DE LOS MÉTODOS IMPLEMENTADOS	64
5.2.1. Función K para Redes	64
5.2.2. NetKDE	68
6. CONCLUSIONES	75
7. RECOMENDACIONES	79

BIBLIOGRAFÍA	80
ANEXO A. ALGORITMO EN AVENUE PARA LA LOCALIZACIÓN DE DIRECCIONES	83
ANEXO B. MANUAL DE USUARIO	85
B.1. MENÚ CONFIGURACIÓN (CONFIGURATION)	85
B.1.1. Opción Base de datos (Database)	85
B.2. MENÚ CARGA (UPLOAD)	86
B.2.1. Opción cargar archivos de red (Upload network files)	86
B.2.2. Opción cargar archivos de eventos (puntos) [Upload event (point) files]	87
B.3. MENÚ PROCESOS (PROCESSES)	88
B.3.1. Opción Función K para redes (Network K function)	88
B.3.2. Opción NetKDE	91

LISTA DE TABLAS

	pág.
Tabla 1. Casos de Uso	29
Tabla 2. Direcciones ajustadas para Avenue	50
Tabla 3. Productos software por etapas	53
Tabla 4. Principales etiquetas KML	52
Tabla 5. Estimación de números de comparaciones de distancias (primera simplificación)	55
Tabla 6. Estimación de números de comparaciones de distancias (segunda simplificación)	56
Tabla 7. Número de accidentes de tránsito por resultado	67
Tabla 8. Defunciones por accidentes de tránsito en Colombia (2005-2014)	76

LISTA DE FIGURAS

	pág.
Figura 1. <i>Kernel Density Estimator</i> – KDE	19
Figura 2. Función K	20
Figura 3. Diferencia entre KDE y NetKDE	22
Figura 4. Diagrama de Casos de Uso	28
Figura 5. Diagrama de Actividades – Visualizar Datos Geográficos	29
Figura 6. Diagrama de Secuencia – Visualizar Datos Geográficos	31
Figura 7. Diagrama de Actividades – Cargar Archivo de Red	31
Figura 8. Diagrama de Secuencia – Cargar Archivo de Red	33
Figura 9. Diagrama de Actividades – Cargar Archivo de Eventos	33
Figura 10. Diagrama de Secuencia – Cargar Archivo de Eventos	35
Figura 11. Diagrama de Actividades – Realizar Cierre de Nodos	36
Figura 12. Diagrama de Secuencia – Realizar Cierre de Nodos	37
Figura 13. Diagrama de Actividades – Realizar Proyección de Puntos	38
Figura 14. Diagrama de Secuencia – Realizar Proyección de Puntos	39
Figura 15. Diagrama de Actividades – Generar Distribuciones Aleatorias de Puntos	40
Figura 16. Diagrama de Secuencia – Generar Distribuciones Aleatorias de Puntos	41
Figura 17. Diagrama de Actividades – Calcular Valores de Función K para Redes	42
Figura 18. Diagrama de Secuencia – Calcular Valores de Función K para Redes	44
Figura 19. Diagrama de Actividades – Calcular Densidades de NetKDE	45
Figura 20. Diagrama de Secuencia – Calcular Densidades de NetKDE	46
Figura 21. Actualización vial	47
Figura 22. Actualización vial – Extensión de vías	48

Figura 23. Proyección de puntos sobre líneas	57
Figura 24. Ejemplo de resultado de la función K para redes	60
Figura 25. Ejemplo de visualización de eventos (puntos)	61
Figura 26. Ejemplo de resultados de NetKDE	63
Figura 27. Resultados de la función K para redes – Bucaramanga 2014	65
Figura 28. Resultados de la función K para redes – Barcelona 2014	66
Figura 29. Resultados de la función K para redes – Accidentes de tránsito con víctimas mortales	67
Figura 30. Resultados de la función K para redes – Grupos de datos reducidos	68
Figura 31. Comparativo entre puntos de accidentes y resultados de NetKDE	69
Figura 32. Resultados de NetKDE – Bucaramanga 2014	71
Figura 33. Comparación entre resultados generales de NetKDE y resultados de accidentes con motocicletas involucradas	73
Figura 34. Comparación entre resultados generales de NetKDE y resultados de accidentes ocurridos durante el mes de diciembre	74
Figura 35. Opción base de datos (Database)	85
Figura 36. Opción cargar archivos de red (Upload network files)	86
Figura 37. Visualización de archivos de red	87
Figura 38. Opción cargar archivos de eventos (puntos) [Upload event (points) files]	88
Figura 39. Opción función K para redes (Network K function)	89
Figura 40. Cierre de nodos de red y proyección de puntos	90
Figura 41. Resultados de la función K para redes	91
Figura 42. Opción NetKDE	92
Figura 43. Resultados de NetKDE	93

RESUMEN

El presente documento se divide en seis capítulos que describen el desarrollo de la presente idea de trabajo de grado desde sus orígenes teóricos hasta el análisis de los resultados obtenidos con la aplicación del software desarrollado a un conjunto de datos reales que representan la situación de accidentalidad vial de las ciudades de Bucaramanga (Colombia) y Barcelona (España) durante el año 2014.

En el primer capítulo se hará una revisión de los dos métodos de análisis de patrones de puntos en redes seleccionados para su implementación software, buscando determinar de forma detallada sus contenidos y características, al tiempo que se señalan aquellos aspectos que las hacen idóneas para el proyecto en particular. De igual forma, se revisan los principales aportes y desarrollos teóricos que, sobre el tema de análisis espacial de patrones de puntos, han tenido lugar.

En el segundo capítulo se establecen, a partir de los contenidos revisados en el capítulo anterior, las bases teóricas y matemáticas sobre las cuales se desarrolló la aplicación software, al tiempo en que se determinan los aspectos cruciales que fueron tomados en cuenta a la hora de su planteamiento y diseño, para garantizar su idoneidad para la tarea mencionada, al igual que su eficacia y precisión.

En el tercer capítulo se detallan las metodologías aplicadas durante el proceso de desarrollo de la aplicación; se incluye en detalle la definición de la arquitectura, el desarrollo de los casos de uso y los subsiguientes objetos de diseño y de implementación; también se incluye una descripción detallada del alistamiento de los datos viales y de accidentes de tránsito utilizados, que posteriormente fueron utilizados para probar la aplicación desarrollada.

En el cuarto capítulo se exponen y analizan los resultados obtenidos al aplicar los métodos implementados sobre el conjunto de datos reales preparados y ajustados para tal fin; se presentan en detalle los hallazgos referentes a la accidentalidad en las ciudades mencionadas durante el año 2014.

En el capítulo quinto se presentan las conclusiones referentes al proceso de desarrollo de los diferentes componentes de la aplicación, y por último en el capítulo sexto se da una visión de lo que podría ser una continuación del presente trabajo.

INTRODUCCIÓN

El análisis de patrones de puntos no es un área nueva de estudio, sus orígenes se remontan a los años 50 y 60 del siglo XX. Una de las referencias más antiguas se encuentra en un artículo sobre relaciones espaciales llevado a cabo por Philip J. Clark y Francis C. Evans (Clark & Evans, 1954); otro ejemplo de estos primeros estudios es el llevado a cabo por David William Harvey centrado en el análisis de cuadrados (Harvey, 1966). Estos métodos, como la mayor parte de los métodos desarrollados, se centran en el análisis de espacios de dos dimensiones.

En contraste, el análisis de patrones de puntos en redes es un concepto bastante más reciente, que parte de una premisa que plantea que algunos eventos, como los accidentes de tránsito y los crímenes, ocurren casi exclusivamente en la red vial; la cual a su vez constituye un subespacio del espacio euclidiano de dos dimensiones y que por su topología hace que los eventos contenidos en una red se relacionen de forma diferente a los eventos contenidos en espacios planos. Una primera aproximación al análisis en espacio de redes se encuentra en el estudio llevado a cabo por William R. Black sobre autocorrelaciones en redes (Black, 1992).

En el entorno regional los antecedentes son bastante más recientes; en enero de 2015 en el Observatorio de Salud Pública de Santander – OSPS, grupo de investigación en salud pública adscrito a la Fundación Oftalmológica de Santander – FOSCAL, se conformó un grupo de interés en accidentalidad en el que confluían además profesionales de diferentes instituciones como la Secretaría de Salud de Santander y la Dirección de Tránsito de Bucaramanga. En este grupo se buscaba plantear estrategias de análisis de la accidentalidad vial tanto municipal como departamental.

En marzo de 2015 docentes de la Facultad de Ingeniería de Sistemas de la Universidad Autónoma de Bucaramanga – UNAB y miembros del área de sistemas de información del OSPS, organismo al que me encuentro vinculado, dieron inicio a una actividad de revisión bibliográfica en temas tecnológicos relacionados con el área de la salud por medio de un club de revistas (Deenadayalan, Grimmer-Somers, Prior, & Kumar, 2008) con el que buscaban componer un estado del arte de los temas mencionados y un banco de posibles proyectos; durante dos de las sesiones realizadas se presentó igual número de artículos de autoría de Zhixiao Xie y Jun Yan (Xie & Yan, 2008)(Xie & Yan, 2013) centrados en el análisis de patrones de puntos en redes vehiculares. De las discusiones subsiguientes a la presentación de estos artículos surgió la idea de crear una línea de investigación centrada en el análisis espacial de eventos en espacios de redes.

Lo anterior unido a la disponibilidad de acceso a información cartográfica de las ciudades de Bucaramanga y Barcelona e información de accidentes de tránsito de las mismas por parte del OSPS, llevaron al planteamiento del presente proyecto.

1. PLANTEAMIENTO DEL PROBLEMA

La accidentalidad es una de las problemáticas principales del manejo contemporáneo del flujo vehicular dentro de una ciudad, no sólo porque la ocurrencia de accidentes presupone una detención o ralentización de dicho flujo, con los correspondientes problemas de afectación a la movilidad, sino porque además cada incidente acarrea un riesgo directo contra la integridad física y económica de las personas que constantemente están utilizando las vías vehiculares ya sea como conductores, pasajeros o peatones y sin importar si se trata de los directamente implicados en un suceso o de aquellos que por circunstancias de tiempo y lugar puedan verse afectados o involucrados.

Por lo anterior, la prevención de la accidentalidad se convierte en un punto crítico a tener en cuenta dentro de las medidas que deben ser tomadas por las autoridades locales y regionales para garantizar la seguridad en las vías y la adecuada movilidad vehicular, necesarias para el bienestar de los habitantes. Dentro de las medidas encaminadas a evitar que los accidentes de tránsito tengan lugar, resulta de especial importancia el poder ubicar los sectores críticos dentro de la red vial, en los cuales tiene lugar el mayor número de accidentes, para así poder tomar medidas correctivas específicas a la problemática presentada en dichos lugares. Con la identificación de estos sectores críticos se busca tener el mayor impacto posible con respecto a la disminución de los accidentes de tránsito por medio de la aplicación de acciones correctivas.

Sin embargo, la detección de estos sectores críticos no resulta sencilla. En cada ciudad hay puntos que gran cantidad de personas perciben como de alta accidentalidad, estos lugares suelen estar relacionados con cruces viales que presentan un alto nivel de flujo vehicular, esto se debe en gran medida a que son puntos en los que la notoriedad de los accidentes allí ocurridos es alta ya sea porque son vistos por gran cantidad de testigos, causan un impacto negativo sobre la movilidad de un sector, o porque presentan víctimas mortales y son difundidos por medios locales; estos puntos pueden considerarse críticos, pero no son necesariamente los únicos existentes.

Esta falencia en la detección de puntos críticos ha sido abordada por medio del uso de software de simulación y medición, que mediante la revisión de los datos acumulados sobre accidentalidad determina los puntos de mayor concentración de la accidentalidad; sin embargo, muchas de estas aplicaciones no tienen en cuenta que los accidentes se circunscriben casi exclusivamente a una red vial y no a un espacio abierto, por lo que los resultados de dichos programas no necesariamente son certeros a la hora de crear una vista de la densidad de accidentes en un sector, zona o ciudad.

Es por esto que en el presente proyecto de investigación se plantea la creación de una herramienta software encaminada a la determinación, de forma mucho más precisa, de los puntos de mayor accidentalidad dentro de una red vial, recurriendo para ello a dos métodos de análisis de patrones de puntos que, por sus características particulares, resultan idóneos para el problema: la

función K para redes y NetKDE, ya que los mismos conciben dichas redes como esquemas unidimensionales interconectados, dando así una mayor exactitud a los resultados obtenidos.

Con el primer método, la Función K para redes, se busca la obtención de una prueba aleatoria que ayude a determinar si los accidentes de tránsito, y en general cualquier fenómeno que se vea delimitado por una red, se presentan de forma aleatoria o por el contrario tienden a agruparse, es decir a formar puntos críticos. El segundo método, NetKDE, aborda directamente el problema de la ubicación y delimitación de los puntos críticos dentro de la red.

2. ESTADO DEL ARTE

La creación de herramientas software de análisis de patrones de puntos ha tenido un desarrollo lento en los últimos años debido en parte a que no ha existido una gran demanda de este tipo de aplicaciones; sus avances en su mayoría se restringen al ámbito académico y de investigación por lo que es más común encontrar artículos sobre el desarrollo de modelos matemáticos y estadísticos y desarrollo de herramientas software no comerciales para probar estos modelos.

Aunque el análisis espacial de patrones de puntos no se limita solamente al análisis de accidentes de tránsito, sobre estos últimos se centra gran parte de los estudios desarrollados durante los últimos años.

2.1. ESPACIOS EUCLIDIANOS DE DOS DIMENSIONES

La aproximación más común a los fenómenos geográficos se basa en el uso de técnicas y métodos que tratan los espacios en los que ocurren los fenómenos como espacios euclidianos de dos dimensiones, a continuación, se presenta un breve resumen de las técnicas utilizadas en algunos de estos estudios.

- Uso de SIG¹ para identificar puntos calientes de choques entre vehículos y peatones, y paradas de bus inseguras (Truong & Somenahalli, 2014): En este artículo se describe el uso de la estadística G_i^* de Getis y Ord (Getis & Ord., 1992) para encontrar puntos calientes en datos recolectados en el Área metropolitana Adelaida, Australia, entre 1996 y 2008.
- Estimación de densidad de núcleo y agrupamiento por K medias para perfilar puntos calientes de accidentes en carreteras (Anderson, 2009): Este artículo se centra en el uso del método de estimación de densidad de núcleo (KDE por sus siglas en inglés) para la detección de puntos calientes de accidentes vehiculares; en este caso se utilizan datos recolectados en la ciudad de Londres, Reino Unido, entre los años 1999 y 2003.
- Sistemas de información geográfico en apoyo al sistema de análisis de accidentes de tráfico, caso de estudio: ciudad de Afyonkarahisar (Erdogan, Yilmaz, Baybura, & Gullu, 2008): Este artículo expone el uso de dos métodos, KDE y el análisis de repetibilidad, para identificar puntos calientes de accidentes en la ciudad de Afyonkarahisar, Turquía.
- Segmentación de líneas basada en densidades de puntos – Una optimización de la ubicación de señales de alerta de fauna silvestre en el sur de Finlandia (Krispa & Durot, 2007): En este artículo se propone la aplicación del método KDE aplicado a los registros existentes de accidente de vehículos con la fauna silvestre para optimizar la ubicación de

¹ Sistemas de Información Geográfica.

las señales de tráfico que alertan sobre la presencia de fauna silvestre en la región de Uusimaa, Finlandia. Nuevamente se observa la aplicación del método KDE.

- Ciclistas urbanos – un análisis espacial de la intensidad del peligro para el tráfico adulto y joven (Delmelle & Thill, 2008): En este artículo se plantea el uso de un modelo de regresión espacial ponderada para buscar dependencias entre algunas de las variables que describen los accidentes de tránsito que incluían ciclistas ocurridos en la ciudad de Búfalo, Estados Unidos, entre los años 2003 y 2004. En este estudio se utiliza una aproximación diferente a las abordadas en los estudios anteriores, ya que se centra más en modelos puramente estadísticos como los de regresiones.

2.2. ESPACIOS DE REDES

Una vertiente con una aproximación diferente, aunque basada en los métodos para espacios euclidianos en dos dimensiones, es la creación y adaptación de métodos restringidos a espacios de subconjuntos de redes unidimensionales contenidas en los espacios en dos dimensiones. En el estudio de los accidentes de tránsito estos espacios de redes sirven como una abstracción que representa las redes vehiculares que, a pesar de no ser espacios unidimensionales, a escalas urbanas pueden verse de esta forma. Los artículos que presentan estos métodos de redes usualmente resaltan el hecho de que con estudios basados en espacios de dos dimensiones normalmente se sobreestiman los puntos calientes dado que en muchos casos se pueden relacionar accidentes ocurridos en calles adyacentes que tal vez no tengan relación y que los accidentes de tránsito en su gran mayoría ocurren en las vías y no en todo el espacio bidimensional.

- Una estimación de densidad de núcleo para redes, su método computacional y una herramienta basada en SIG (Okabe, Satoh, & Sugihara, 2009): En este artículo se plantea que el uso de KDE para estimar densidades de patrones de puntos en redes tiende a sobreestimar las densidades alrededor de los nodos; por lo que se propone una adaptación para redes del método KDE que corrige estas sobreestimaciones. También se presenta una aplicación software que implementa el método presentado.
- Estimación de densidad de núcleo de accidentes de tráfico en un espacio de redes (Xie & Yan, 2008): Este artículo también menciona el problema de sobreestimación de densidades que pueden presentar los métodos de estimación de densidades en dos dimensiones cuando se aplican a fenómenos que ocurren solamente en el subconjunto de espacio de red. Junto con el desarrollo de esta idea se propone una adaptación de KDE para redes denominada NetKDE y se comparan los resultados de los dos métodos a un conjunto de datos de accidentes de tránsito ocurridos durante el año 2005 en Bowling Green, Estados Unidos.

- Detección de agrupaciones de accidentes de tráfico con la estimación de densidad de núcleo para redes y estadísticas locales espaciales: un acercamiento integrado (Xie & Yan, 2013): Este artículo es la continuación directa del artículo anterior, en él se plantea la aplicación de la I de Morán, una medida de autocorrelación, para darle significancia estadística a los resultados obtenidos por la aplicación de NetKDE. Los datos utilizados son los mismos del artículo anterior, accidentes de tránsito ocurridos en Bowling Green, Estados Unidos durante 2005.

2.3. HERRAMIENTAS COMPUTACIONALES

A continuación, se presenta una serie de herramientas computacionales comerciales y académicas que implementan métodos de análisis de patrones de puntos, es importante destacar que la mayoría de ellas se basan en métodos para espacios euclidianos en dos dimensiones y no espacios de redes.

- Spatial Analyst Extension of ESRI's ArcGIS (Esri, 2016): Esta extensión de ArcGIS, software comercial para sistemas de información geográfica de amplia difusión, incluye una implementación de KDE para análisis de patrones de puntos. En la extensión también se incluyen otras herramientas para realizar análisis de distancias y costos de viaje, operadores matemáticos para análisis geográficos y análisis estadísticos con datos basados en celdas.
- CrimeStat (National Institute of Justice, 2015): CrimeStat es una herramienta software para el mapeo y análisis geográfico de crímenes que incluye una implementación de KDE para el análisis espacial. CrimeStat fue desarrollado por Ned Levine & Asociados en 1996 con fondos del Instituto Nacional de Justicia (NIJ por sus siglas en inglés), perteneciente al Departamento de Justicia de los Estados Unidos, y ha venido siendo actualizado por medio de nuevas versiones desde entonces.
- SANET (Okabe et al., 2015): La herramienta de análisis espacial sobre una red denominada SANET fue desarrollada por investigadores de la Universidad de Tokio, Japón, en 2006. Esta aplicación incluye implementaciones de la función K para redes y un método propio de detección de puntos de concentración de eventos basado en KDE. SANET requería de ArcGIS para poder ser utilizada hasta que en octubre de 2015 se lanzó una versión beta autónoma.
- PASSaGE (Rosenberg & Anderson, 2011): PASSaGE es una aplicación desarrollada por Michael S. Rosenberg & Corey D. Anderson en 2001, investigadores del Centro de Medicina Evolutiva e informática / Instituto de Biodiseño de la Universidad del Estado de Arizona, Estados Unidos. Esta aplicación provee una serie de métodos estadísticos para análisis de datos espaciales tales como análisis de segundo orden bivariados y covariados,

análisis de vecindades de segundo orden, análisis wavelet angular, entre otros. También incluye herramientas para análisis de datos en redes. La última actualización de la herramienta fue publicada en junio de 2011.

- Programita (Centre for Environmental Research - UFZ, 2016): Esta herramienta fue publicada en 2004 en la Escuela para Graduados de la Facultad de Agronomía de la Universidad de Buenos Aires, Argentina. Su enfoque es el de análisis de patrones de puntos en ecología, incluye diferentes métodos estadísticos como patrones univariados, bivariados y multivariados, y patrones cualitativa y cuantitativamente marcados. La última actualización disponible fue publicada en diciembre de 2013.

2.4. ANÁLISIS DE PATRONES DE PUNTOS – PPA

Para poder reducir el número de accidentes de tránsito en una región determinada es necesario entender las características de dichos accidentes, el insumo principal para llegar a este conocimiento es el registro de los accidentes ocurridos en el pasado. La ocurrencia de los accidentes de tránsito no suele ser aleatoria en espacio y tiempo, y desde el punto de vista geográfico estos accidentes tienden a formar grupos, denominados “puntos calientes”. La ocurrencia de los accidentes está determinada en gran medida por el volumen de tráfico (Black, 1991). Además del volumen de tráfico hay otros factores que influyen en la ocurrencia de los accidentes de tránsito tales como las condiciones y características del terreno, el clima y las condiciones de visibilidad entre otros.

Para tratar de crear un modelo que describa los accidentes de tránsito se puede recurrir al análisis de patrones de puntos (PPA por sus siglas en inglés), el cual viene desarrollándose desde finales de los años 50 del siglo XX (Gatrell, Bailey, Diggle, & Rowlingson, 1996) y cuenta con diferentes técnicas estadísticas para el estudio de las características geográficas de fenómenos que se puedan representar como puntos en un espacio comúnmente euclidiano en dos dimensiones.

Los métodos asociados al PPA se pueden clasificar en dos grandes grupos (Bailey & Gatrell, 1995):

Métodos que examinan los efectos de primer orden de un proceso espacial: Se centran en hallar la probabilidad de ocurrencia de un fenómeno espacial en cada lugar dentro de un área de estudio. Se presupone que se presentan variaciones en la intensidad de ocurrencia de un proceso a través del espacio, estas variaciones se deben a las propiedades del entorno.

Métodos que examinan los efectos de segundo orden de un proceso espacial: Parten de la premisa de que la ocurrencia de un evento en un lugar incrementa la probabilidad de ocurrencia del mismo tipo de evento en sus inmediaciones, lo que resulta en la formación de grupos (clusters) dentro del área de estudio. En estos métodos se presupone que hay interdependencia espacial entre los lugares de los eventos.

Dos métodos de PPA ampliamente difundidos son KDE y la Función K, métodos de primer y segundo orden respectivamente. A continuación, se describen estos dos métodos.

2.4.1. Kernel Density Estimator – KDE

KDE es un método de análisis de patrones de puntos muy difundido, existen aplicaciones software que lo implementan como la extensión de análisis espacial de ArcGIS y CrimeStat (Levine, 2013). Este método se basa en el cálculo de densidades por sectores en un área que por simplicidad en los cálculos se toma como un espacio euclidiano en dos dimensiones sin barreras o accidentes geográficos. Para la aplicación del método el espacio se divide en una cuadrícula uniforme y cada uno de los puntos que representan el fenómeno estudiado se asigna a alguna de las divisiones (cuadrados). También se define un radio de influencia, denominado ancho de banda, que suele ser mayor en distancia a la longitud del lado de cada uno de los cuadrados. Desde el centro de cada uno de estos cuadrados se buscan todos los puntos que estén dentro del ancho de banda y se aplica una función de peso multiplicada por un coeficiente que incluye en radio de búsqueda.

Los valores calculados para cada punto dentro del ancho de banda se suman para hallar el valor de densidad correspondiente para cada uno de los cuadrados en los que se dividió el espacio. Al usar un ancho de banda mayor a las dimensiones de las divisiones (cuadrados) se llega a observar que divisiones que no tienen puntos en su interior pueden tener asociados valores de densidad significativos si hay concentración de puntos dentro del ancho de banda, lo que hace más homogénea la distribución de los valores de densidad.

La función de densidad para cada una de las divisiones realizadas es:

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

Donde $\lambda(s)$ es la función de densidad en el punto s , r representa el ancho de banda seleccionado, k es una función de peso relacionada con el ancho de banda y la distancia d_{is} entre el centro de la división s y cada uno de los puntos i .

La función de peso k , denominada función de núcleo, se usa para representar un efecto de decaimiento por distancia y debe estar construida de tal forma que tenga su valor máximo para las distancias d_{is} cercanas a cero (0), su valor decaiga a medida que la distancia se acerque al valor del ancho de banda r y una vez supere este ancho de banda su valor sea cero (0). Las funciones de peso más difundidas son la función Gaussiana, la función cuadrática y la función de varianza mínima (Xie & Yan, 2008) (Schabenberger & Gotway, 2005) (Gibin, Longley, & Atkinson, 2007). A continuación se describen estas tres funciones.

Función Gaussiana:

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

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

Función cuadrática:

$$k\left(\frac{d_{is}}{r}\right) = K\left(1 - \frac{d_{is}^2}{r^2}\right), \text{ para } d_{is} \leq r \quad (3)$$

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

Donde K puede tomar valores como $3/\pi$ o $3/4$ (función de Epanechnikov).

Función de varianza mínima:

$$k\left(\frac{d_{is}}{r}\right) = \frac{3}{8}\left(3 - 5\frac{d_{is}^2}{r^2}\right), \text{ para } d_{is} \leq r \quad (4)$$

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

En la figura 1 se muestran las divisiones (cuadrados) realizada sobre los puntos que representan un fenómeno y la demarcación del área de influencia, definida por el ancho de banda r , de una de las divisiones; para cada uno de los puntos i dentro del área de influencia se debe hallar la distancia d_{is} desde el centro de la división s .

2.4.2. La Función K

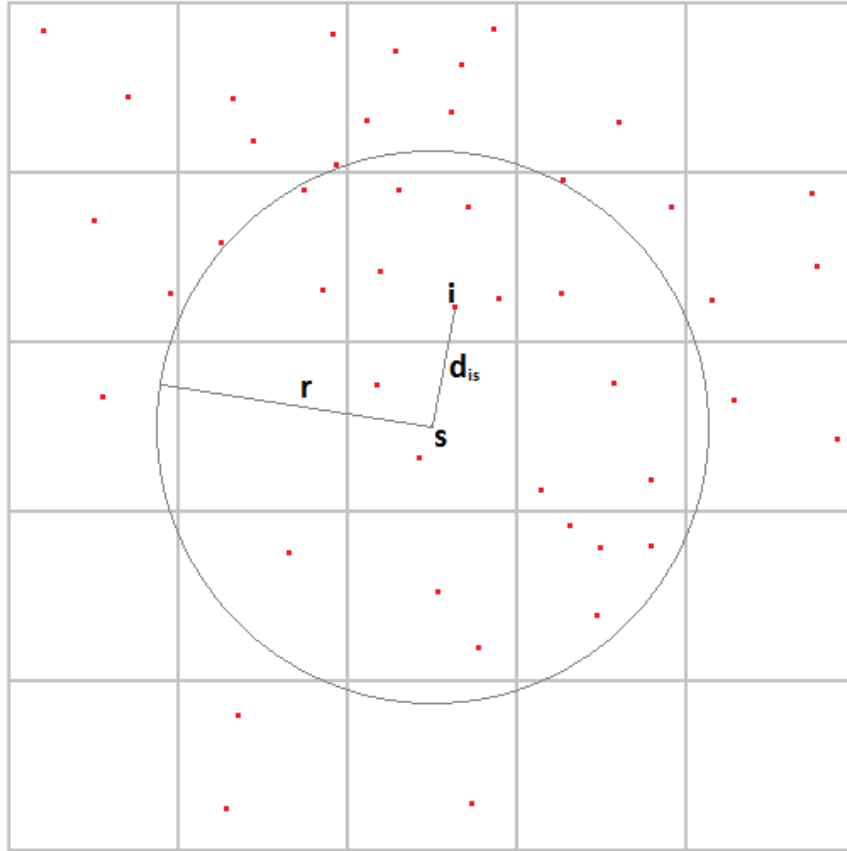
La función K, propuesta por Brian D. Ripley en 1976 (Dixon, El-Shaarawi, & Piegorsch, 2002), representa la media observada del número de puntos dentro de una distancia definida dividida por la densidad de puntos en el área de estudio.

La forma general de la función K está dada por la ecuación:

$$K(h) = \lambda^{-1} E(\text{Número de puntos con distancia Euclidiana 'h' o menor a un punto arbitrario P}) \quad (5)$$

Donde λ es la densidad de puntos en el área. El valor de la función K para un fenómeno aleatoriamente distribuido es igual a πh^2 .

Figura 1. *Kernel Density Estimator – KDE*



Fuente El autor.

Para estimar el valor de la función K se parte del valor estimado de E, denominado operador de expectación, que para espacios planos se estima mediante la ecuación:

$$\hat{E}(h) = \frac{1}{n} \sum_{i \neq j} I_h(d_{ij}) \quad (6)$$

Donde h es la distancia máxima entre los puntos a ser tomada en consideración, n es el número total de puntos, d_{ij} es la distancia euclidiana entre dos puntos i y j , $I_h()$ es una función de peso con valor uno (1) cuando la distancia entre los puntos es menor o igual a h y valor cero (0) cuando esta distancia es superior a h y R es el área del espacio estudiado. La densidad de puntos λ se puede estimar como:

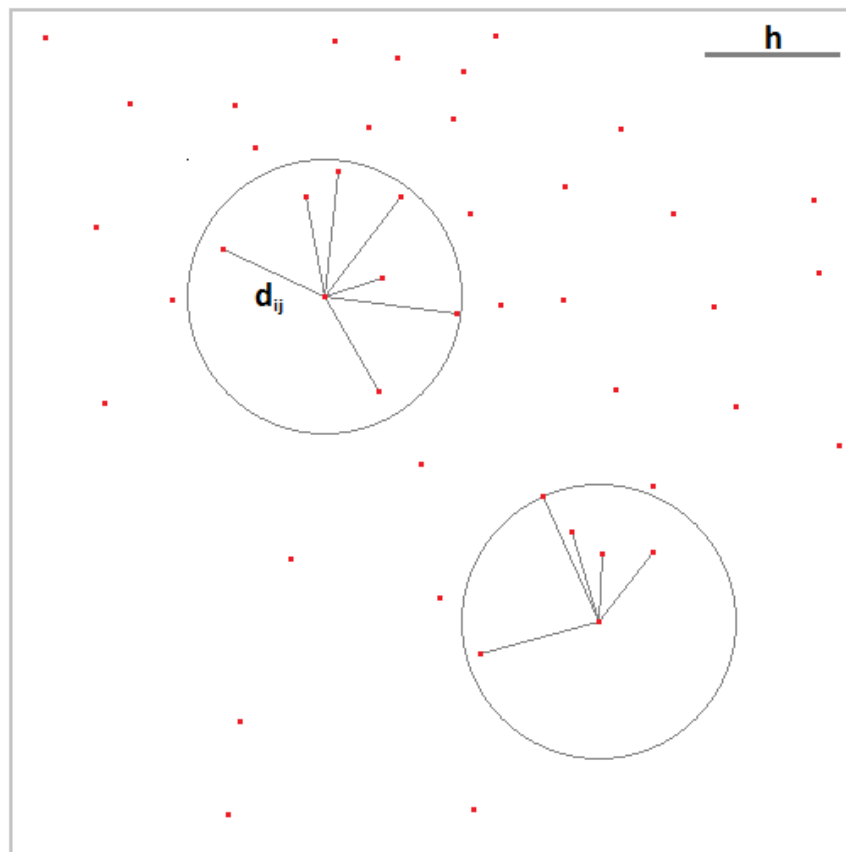
$$\hat{\lambda} = \frac{n}{R} \quad (7)$$

Al remplazar las funciones (6) y (7) sobre la función (5) se obtiene el estimador de la función K para espacios planos dado por la ecuación:

$$\hat{K}^{pl}(h) = \frac{1}{\lambda n} \sum_{i \neq j} \sum I_h(d_{ij}) = \frac{R}{n^2} \sum_{i \neq j} \sum I_h(d_{ij}) \quad (8)$$

La figura 2 muestra las distancias d_{ij} menores o iguales a h para dos puntos arbitrarios dentro del área de estudio.

Figura 2. Función K



Fuente El autor.

La función K puede ser utilizada para determinar si una serie de puntos ubicados sobre un espacio tienden a formar grupos o por el contrario se distribuye de forma aleatoria; una forma de poner a prueba estas afirmaciones es por medio de simulaciones de Monte Carlo. Para esto se genera una

gran cantidad de conjuntos aleatorios de puntos sobre el mismo espacio (por ejemplo 1.000 conjuntos), la función K se aplica con diferentes valores de h (distancia máxima entre puntos a tenerse en cuenta) sobre cada uno de los conjuntos de datos aleatorios generados y sobre el conjunto original de puntos. Con los resultados de la función K sobre los conjuntos aleatorios para cada valor de h se definen intervalos de confianza, por ejemplo del 95% (entre el valor que ocupa el percentil 2,5 y el valor que ocupa el percentil 97,5), si el valor de la función K para el conjunto de datos original es superior al intervalo de confianza definido se puede concluir que existe tendencia estadísticamente significativa a formar agrupaciones y por lo tanto la distribución de los datos sobre el espacio no es aleatoria.

Existen diversas aplicaciones software que implementan la función K; una de estas aplicaciones, basada en ArcGIS, es la denominada Análisis Espacial en Redes (SANET por sus siglas en inglés), desarrollada por investigadores de la Universidad de Tokio, Japón (Okabe et al., 2015).

2.5. ANÁLISIS DE PATRONES DE PUNTOS EN REDES

La ocurrencia de algunos de los eventos que pueden ser abordados desde el análisis de patrones de puntos se encuentra restringida a subconjuntos del espacio en dos dimensiones, uno de estos subconjuntos está representado por las carreteras y caminos dentro del área. Eventos como los accidentes de tránsito y los crímenes se ubican en su mayoría dentro de este subconjunto del espacio; este tipo de espacios se denominan subespacios de red. Aunque el espacio representado por las vías, carreteras y caminos es también un espacio en dos dimensiones, para efectos de simplicidad en los cálculos se puede representar como un espacio unidimensional representado por líneas y sus intersecciones.

La aplicación de métodos para espacios euclidianos en dos dimensiones como los detallados anteriormente puede llevar a resultados que no reflejan la realidad de los eventos que se pretenden modelar; como lo expresa Miller (Miller, 1999) “La uniformidad de espacios en dos dimensiones es una presunción muy fuerte para el análisis de eventos de puntos que ocurren en el espacio infinito de una dimensión”. Es por lo anterior que se han planteado para el manejo de eventos en espacios de red modelos alternativos, algunos derivados de modelos para espacios euclidianos en dos dimensiones.

A continuación, se describirán dos de estos modelos para espacios de redes, NetKDE, derivado a partir de KDE y una versión de la función K para redes.

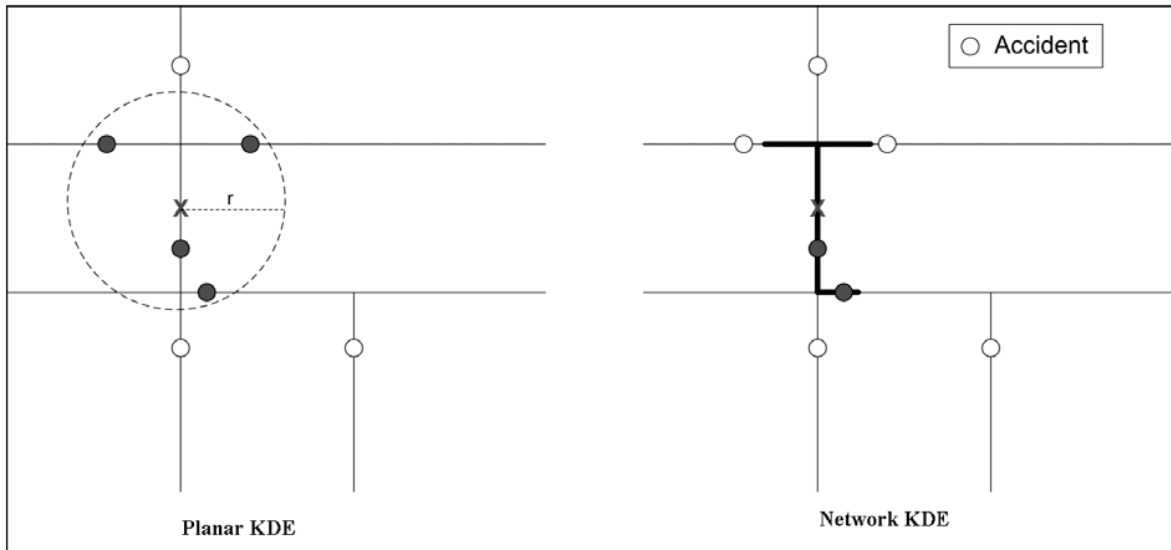
2.5.1. NetKDE

NetKDE (abreviación de Network KDE) es un método propuesto por Zhixiao Xie y Jun Yan en 2008 (Xie & Yan, 2008) como una extensión del método KDE para espacios de red. Conceptos generales aplicados en KDE como el de densidad se manejan en este modelo sobre unidades lineales en

lugar de unidades de área. Las diferentes opciones de funciones de decaimiento por distancia (k) también son usadas para redes. Una diferencia fundamental entre NetKDE y KDE es que la medición del ancho de banda y de las distancias entre puntos se lleva a cabo sobre el espacio de red en lugar de hacerlo sobre el espacio en dos dimensiones.

Dado que las unidades de medida con las que trabajan estos métodos son diferentes, unidades lineales y unidades de área respectivamente, los valores de densidad obtenidos mediante estos métodos pueden ser muy diferentes, aunque se trabaje con el mismo conjunto de datos. Lo anterior se debe en gran medida a que con KDE se suelen detectar agrupaciones con mayores concentraciones de puntos a las detectadas mediante NetKDE. La figura 3 muestra un conjunto de datos al que se le aplican los dos métodos, para el mismo ancho de banda en el caso de KDE se incluyen cuatro (4) puntos mientras que para NetKDE solo se incluyen dos (2).

Figura 3. Diferencia entre KDE y NetKDE



Fuente Kernel Density Estimation of traffic accidents in a network space (Xie & Yan, 2008).

La función de densidad en espacios de red se define como:

$$\lambda(s) = \sum_{i=1}^n \frac{1}{r} k\left(\frac{d_{is}}{r}\right) \quad (9)$$

Si se compara con la función equivalente de KDE se encontrará que la única diferencia radica en que aquí el “área” se representa únicamente por r , el ancho de banda, y no por πr^2 , el área del círculo con radio igual al ancho de banda.

Al utilizar un espacio más complejo que el utilizado por KDE, NetKDE requiere una fase de preparación de la información igualmente compleja; se debe crear un sistema de referencia

conformado por segmentos de línea, cada uno de estos segmentos representa la porción de vía que une dos intersecciones o esquinas adyacentes. Las vías que no llevan a una intersección (camino bloqueados, fin de la vía, límite del área de estudio, etc.) también se consideran segmentos. El listado de segmentos junto con las uniones entre dos o más de ellos, las intersecciones, conforman el conjunto de datos que representa la red.

Cada uno de los segmentos construidos debe a su vez dividirse en segmentos de longitud fija denominados *lixels* (unión entre las palabras línea y píxel), estos *lixels* son una traslación directa de los segmentos cuadrados utilizados en el método KDE. Durante el proceso de construcción de los *lixels* es muy probable que, sin importar que longitud de *lixel* se escoja, para cada segmento se obtenga un *lixel* de longitud inferior. Por ejemplo, si se tiene un segmento de vía de 63 metros y una longitud de *lixel* predefinida de 10 metros, se obtendrán 6 *lixels* de con el tamaño definido de 10 metros y un *lixel* de 3 metros de longitud denominado *lixel* residual. Si la longitud de *lixel* escogida es pequeña en comparación con la longitud de los segmentos, la presencia de los *lixels* residuales no afectará el patrón de densidad obtenido debido a que estos solamente representan una porción mínima del total de *lixels* construidos.

Una vez definido el conjunto de *lixels* se deben llevar los puntos que representan los eventos a analizar a la red, es probable que estos puntos no se ubiquen exactamente dentro de los *lixels*, si este es el caso, es necesario proyectar de forma perpendicular cada uno de los puntos al *lixel* más cercano.

Dentro de un espacio euclidiano en dos dimensiones la distancia más corta entre dos puntos es el segmento de línea recta que los une; para un espacio de red esta afirmación no siempre es válida dado que las distancias se deben medir a través de la red. Por lo general se pueden trazar varios “camino” entre dos puntos arbitrarios dentro de una red, por lo que se hace necesario implementar un algoritmo que busque entre estos posibles “camino” el de menor longitud.

Una vez se cuente con los datos y algoritmos anteriormente mencionados se puede proceder a calcular el valor de densidad para cada uno de los *lixels* utilizando la función (9).

2.5.2. Función K para Redes

En entornos que abarcan áreas pequeñas que contienen a su vez redes pequeñas en extensión el uso de la función K plana puede llevar a diferencias en longitud sustanciales entre las distancias medidas a través de la red y las distancias euclidianas. La versión de la función K para espacios en redes presentada a continuación fue desarrollada por Atsuyuki Okabe e Ikuho Yamada (Okabe & Yamada, 2000).

Para la implementación de la función K se define una red finita $N = N(L, Q)$ conformada por un conjunto de enlaces $L = \{L_1, \dots, L_{nL}\}$ y un conjunto de nodos $Q = \{q_1, \dots, q_{nQ}\}$; el conjunto total de enlaces se denomina L_T y la longitud total de todos los segmentos que conforman la red se

denomina l_T . En una red de vías un nodo del conjunto Q representará la intersección entre dos o más calles (esquinas) y un enlace del conjunto L representará el segmento de vía que une dos intersecciones. La distancia entre dos puntos arbitrarios dentro de L_T en \mathcal{N} se representa por la distancia más corta a través de la red entre ellos y se denomina *distancia de red*, para efectos de simplicidad en esta sección del documento al hablar de distancias se hace referencia a distancias de red.

Después se define un conjunto de objetos $P=\{p_1, \dots, p_n\}$ y se asume que cada uno de los objetos se encuentran ubicados en los enlaces L_T de \mathcal{N} .

Como hipótesis nula se asume que los puntos que representan los objetos del conjunto P están distribuidos uniformemente sobre L_T en \mathcal{N} . Lo anterior se representa mediante la función

$$f(p) = \begin{cases} \frac{1}{l_T}, & p \in L_T \\ 0, & p \notin L_T \end{cases} \quad (10)$$

Donde p es un punto de P . Siguiendo esta premisa, la probabilidad de que un punto p se encuentre dentro de un subconjunto L_S de L_T es proporcional a la longitud l_S de L_S . De lo anterior se llega a que la probabilidad de que el número de puntos de P que se encuentran ubicados en L_S sea igual a n_S está dada por la ecuación

$$\Pr[N(L_S) = n_S] = {}_n C_{n_S} \left(\frac{L_S}{L_T}\right)^{n_S} \left(1 - \frac{L_S}{L_T}\right)^{n-n_S} \quad (11)$$

Donde

$${}_n C_{n_S} = \frac{n!}{n_S!(n-n_S)!} \quad (12)$$

Bajo la presunción de la hipótesis nula planteada, la función K se define como

$$K(t) = \frac{1}{\omega} E \left(\begin{array}{l} \text{el número de puntos de } P \text{ dentro de una} \\ \text{distancia de red } t \text{ de un punto } p \text{ en } L_T \end{array} \right) \quad (13)$$

Donde E representa el valor esperado con respecto a todas las posibles localizaciones de p sobre L_T , y ω es la densidad de puntos P dada por

$$\omega = \frac{n}{l_T} \quad (14)$$

Defínase $L_p(t)$ como un subconjunto de L_T donde la distancia de red entre un punto p y cualquier punto de $L_p(t)$ es menor o igual a t ; sea $l_p(t)$ la longitud de $L_p(t)$. Dado que $\omega l_p(t)$ es el número esperado de puntos P en $L_p(t)$, se tiene

$$K(t) = \frac{1}{\omega} \cdot \frac{1}{l_T} \int_{p \in L_T} \omega \cdot l_p(t) dp = \frac{1}{l_T} \int_{p \in L_T} l_p(t) dp \quad (15)$$

Donde dp representa la integración de p sobre L_T .

Ahora, supóngase que $L_{p_i}(t)$ es un subconjunto de L_T donde la distancia de red entre un punto arbitrario p_i y cualquier punto de $L_{p_i}(t)$ es menor o igual a t . Reemplazando la definición de $L_{p_i}(t)$ de la ecuación número (13) y el valor de ω dado por la ecuación (14) se puede obtener el estimador de la función K para redes dado por la ecuación

$$\hat{K}(t) = \frac{l_T}{n^2} \sum_{i=1}^n (\text{el número de puntos de } P \text{ en } L_{p_i}(t)) \quad (16)$$

Una vez se obtiene el valor de la función K de red observada y la función K de red para un conjunto de datos uniformemente distribuidos dada por la función (15), al comparar los valores obtenidos se tiene que si $\hat{K}(t) > K(t)$, los puntos tienden a formar agrupaciones o clusters, y si $\hat{K}(t) < K(t)$, los puntos tienden a distribuirse de forma dispersa no aleatoria.

3. OBJETIVOS

3.1. OBJETIVO GENERAL

Construir una aplicación software que permita realizar el análisis espacial de conjuntos de datos en espacios de redes por medio de los algoritmos NetKDE y función K para redes.

3.2. OBJETIVOS ESPECÍFICOS

- Hacer una revisión detallada de los contenidos de los dos métodos de análisis de patrones de puntos a ser implementados, NetKDE y Función K para redes, junto a los desarrollos teóricos que se hayan hecho sobre los mismos, y que puedan resultar de utilidad para la creación de un sistema de simulación de redes viales.
- Definir estructuras de entrada y salida de información espacial tanto para patrones de puntos como para definición de espacios de redes.
- Construir un visor de resultados de análisis espaciales que permita comparar diferentes vistas de los conjuntos de datos iniciales de acuerdo a los parámetros adicionales contenidos en estos datos.
- Aplicar los algoritmos de análisis de patrones de puntos desarrollados a un conjunto de datos real que represente la red vehicular de una municipalidad y los accidentes de tránsito ocurridos en ella.

4. METODOLOGÍA DE LA INVESTIGACIÓN

4.1. DEFINICIÓN DE LA ARQUITECTURA DE DESARROLLO DE SOFTWARE

Para la determinación de la metodología de trabajo es necesario señalar que al ser el presente proyecto un trabajo realizado por una única persona no resultó necesario establecer una dinámica de trabajo grupal, ni un sistema de coordinación de que equipos de trabajo. Sin embargo, fue igualmente necesario establecer un proceso de desarrollo de software por etapas, siguiendo las fases tradicionales de desarrollo que se plantean dentro de un proceso iterativo.

Por lo anterior, y de acuerdo a las necesidades particulares del proyecto, se utilizó la metodología de desarrollo de software *Crystal Clear*, propuesta por Alistair Cockburn (Cockburn, 2004); ésta es una de las denominadas metodologías de desarrollo de software ágiles que se adhieren al manifiesto ágil publicado en 2001 (Beck et al., 2001). Se escogió esta metodología por dos razones principales: primero, por su capacidad para minimizar el riesgo de error en procesos de desarrollo de software que disponen de tiempo limitado y, segundo, porque su capacidad adaptativa permite que las soluciones diseñadas vayan evolucionando a la par de las nuevas necesidades que puedan ir surgiendo a lo largo del proyecto.

De igual forma, la documentación resultante de la aplicación de las diferentes fases del proceso de construcción de la herramienta software se llevó a cabo con base en el lenguaje de modelado UML. Escogido tanto por su amplia difusión y aceptación dentro de la comunidad de desarrollo de software, como por su carácter visual y expositivo, que lo hace idóneo para construir, describir y documentar un sistema de software, y por consiguiente presentarlo dentro de una investigación de tipo académico.

En cuanto a las necesidades específicas del presente proyecto, la aplicación de software construida maneja volúmenes de datos de tamaño medio (representaciones de redes de vehículos y coordenadas geográficas de puntos dentro de estas redes), al tiempo que una capacidad de procesamiento de datos alta. Es necesario tener en cuenta, además, que la información obtenida mediante la aplicación de NetKDE a los conjuntos de datos de entrada no requiere ser compartida masivamente en Internet, por lo cual no fue necesario implementar herramientas para tal propósito.

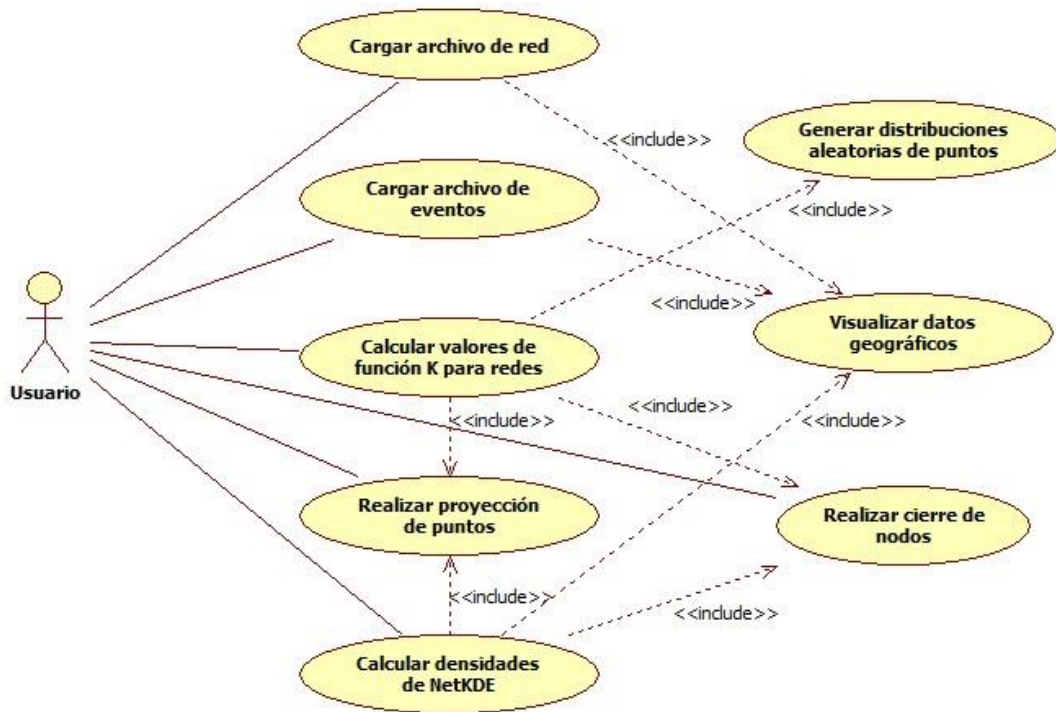
Teniendo en cuenta el conjunto de características mencionadas, especialmente por las altas necesidades de procesamiento de datos, los tiempos de respuesta del software planteado pueden llegar a ser extensos -del orden de segundos o minutos- dependiendo del tamaño del conjunto de datos de entrada. Es por esto que se planteó la construcción de la herramienta software en un entorno de escritorio, lo cual permite un mejor desempeño. La plataforma de codificación utilizada fue Java SE, lo que permitió crear una aplicación portable a equipos de cómputo con diferentes sistemas operativos, facilitando la compatibilidad.

Respecto al manejo de los diversos datos de entrada utilizados por la aplicación, para los datos que representan las redes de vías se utilizaron como punto de partida archivos con formato KML (*Keyhole Markup Language*), formato desarrollado inicialmente para Google Maps pero que a partir de 2008 es un estándar internacional². Los datos de los accidentes de tránsito, representados por puntos en un sistema de coordenadas geográficas, también se manejaron mediante archivos en formato KML.

Finalmente, y con la idea de disponer con facilidad de la información obtenida en la ejecución del proyecto, la información de las redes y puntos es almacenada en una base de datos relacional MySQL, los datos de salida, los resultados de la aplicación de los métodos NetKDE y la Función K para Redes, también se almacenan en dicha base de datos relacional. Se seleccionó este tipo de base de datos por sus características inherentes, especialmente por su facilidad de uso y su capacidad de trabajar con múltiples plataformas, al igual que por su amplia difusión.

4.2. DEFINICIÓN DE CASOS DE USO

Figura 4. Diagrama de Casos de Uso



Fuente El autor.

² <http://www.opengeospatial.org/pressroom/pressreleases/857>

Tabla 1. Casos de Uso

No	Caso de Uso	Incluye	Actor(es)
1	Visualizar datos geográficos		
2	Cargar archivo de red	1	Usuario
3	Cargar archivo de eventos	1	Usuario
4	Realizar cierre de nodos		Usuario
5	Realizar proyección de puntos		Usuario
6	Generar distribuciones aleatorias de puntos		
7	Calcular valores de función K para redes	4, 5, 6	Usuario
8	Calcular densidades de NetKDE	1, 4, 5	Usuario
Fuente El autor.			

4.3. DESCRIPCIÓN DE CASOS DE USO

A continuación, se describen todos los casos de uso identificados por medio de flujos básicos y alternativos apoyados en diagramas de actividades UML, se incluye información de los actores asociados, precondiciones y postcondiciones. Además de las descripciones anteriores se incluye la identificación de objetos de programación y mensajes entre objetos por medio de diagramas de secuencia UML.

4.3.1. Visualizar Datos Geográficos

Este caso de uso permite ver en pantalla la representación gráfica de los datos geográficos de las redes y/o eventos cargados a la aplicación y los resultados de la aplicación del método NetKDE sobre estos conjuntos de datos.

Actores

Este caso de uso no es llamado directamente por ningún actor, su llamado se hace a través de otros casos de uso que lo incluyen.

Precondiciones

El usuario debe haber ingresado a un caso de uso que haga llamado a este caso de uso.

Flujo Básico

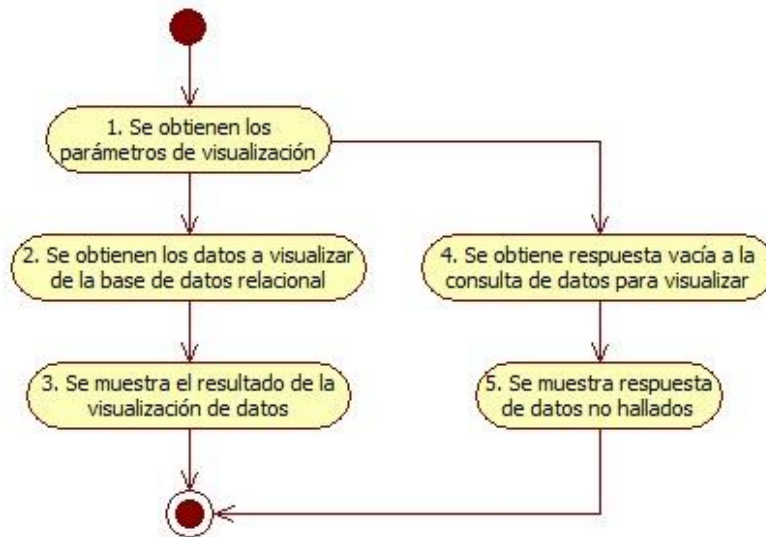
- 1) Se obtienen los parámetros de visualización.
- 2) Se obtienen los datos a visualizar de la base de datos relacional.
- 3) Se muestra el resultado de la visualización de datos.

Flujo Alternativo 1

En el paso 1:

- 4) Se obtiene respuesta vacía a la consulta de datos para visualizar.
- 5) Se muestra respuesta de datos no hallados.

Figura 5. Diagrama de Actividades – Visualizar Datos Geográficos



Fuente El autor.

Descripción de Objetos e Interacciones

Este caso de uso no tiene asociado un objeto tipo formulario, en su lugar especifica un objeto de la clase *VisorRed* que obtiene los datos de líneas de redes y puntos de eventos a través de instancias de las clases *DbRedes* y *DbEventos* respectivamente, y por último dibuja en pantalla los resultados mediante un objeto *Graphics2D* de Java. Este objeto puede ser agregado a un formulario (ver figura 6).

Las clases *DbRedes* y *DbEventos* antes mencionados son especializaciones de la clase *DbConexion*, la cual cuenta con atributos y métodos para establecer comunicación con una base de datos MySQL.

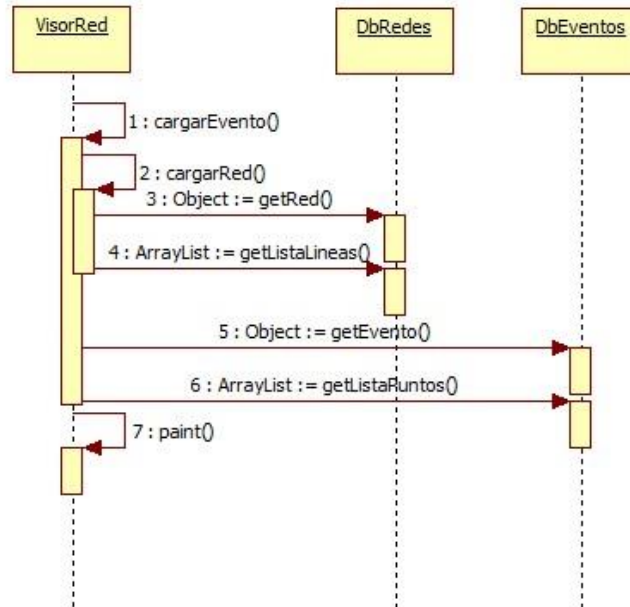
4.3.2. Cargar Archivo de Red

Este caso de uso permite la carga de un archivo que representa una red vial en formato KML a una base de datos relacional.

Actores

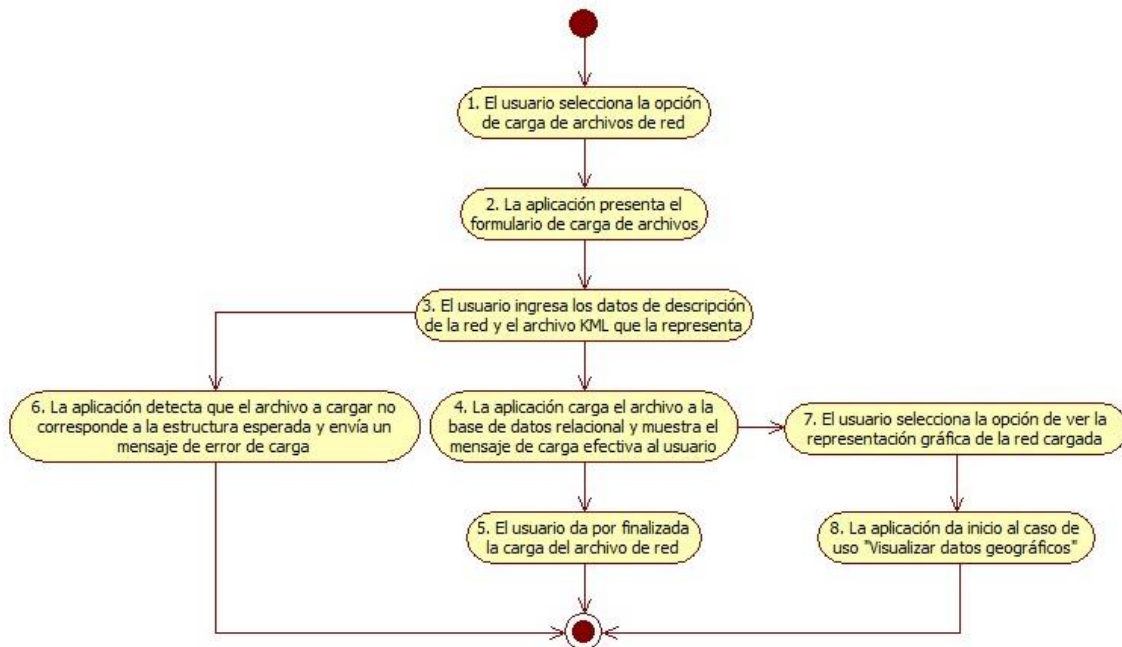
Usuario.

Figura 6. Diagrama de Secuencia – Visualizar Datos Geográficos



Fuente El autor.

Figura 7. Diagrama de Actividades – Cargar Archivo de Red



Fuente El autor.

Flujo Básico

- 1) El usuario selecciona la opción de carga de archivos de red.
- 2) La aplicación presenta el formulario de carga de archivos.
- 3) El usuario ingresa los datos de descripción de la red y el archivo KML que la representa.
- 4) La aplicación carga el archivo a la base de datos relacional y muestra el mensaje de carga efectiva al usuario.
- 5) El usuario da por finalizada la carga del archivo de red.

Flujo Alternativo 1

Después del paso 3:

- 6) La aplicación detecta que el archivo a cargar no corresponde a la estructura esperada y envía un mensaje de error de carga.

Flujo Alternativo 2

Después del paso 4:

- 7) El usuario selecciona la opción de ver la representación gráfica de la red cargada.
- 8) La aplicación da inicio al caso de uso “Visualizar datos geográficos”.

Descripción de Objetos e Interacciones

Este caso de uso implementa el formulario *FrmCargarRedes*, el cual permite solicitar al usuario los datos y archivos necesarios para crear una red. Los datos y archivos recibidos son entregados a una instancia de la clase *PrCargaRedes*, la cual tiene capacidad para procesar archivos KML de líneas (redes) y por medio de una instancia de la clase *DbRedes* puede agregar la información procesada a una base de datos MySQL (ver figura 8).

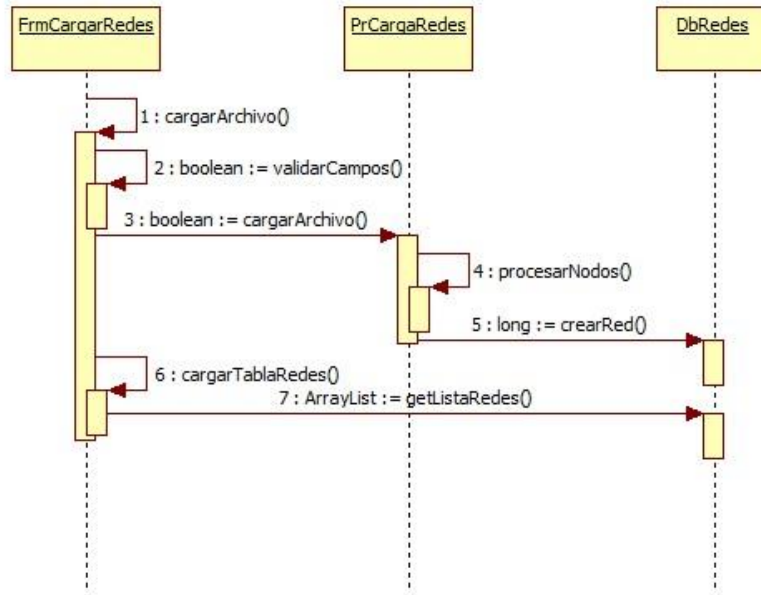
4.3.3. Cargar Archivo de Eventos

Este caso de uso permite la carga de un archivo que contiene los puntos que representan eventos ocurridos en un subespacio de red en formato KML a una base de datos relacional.

Actores

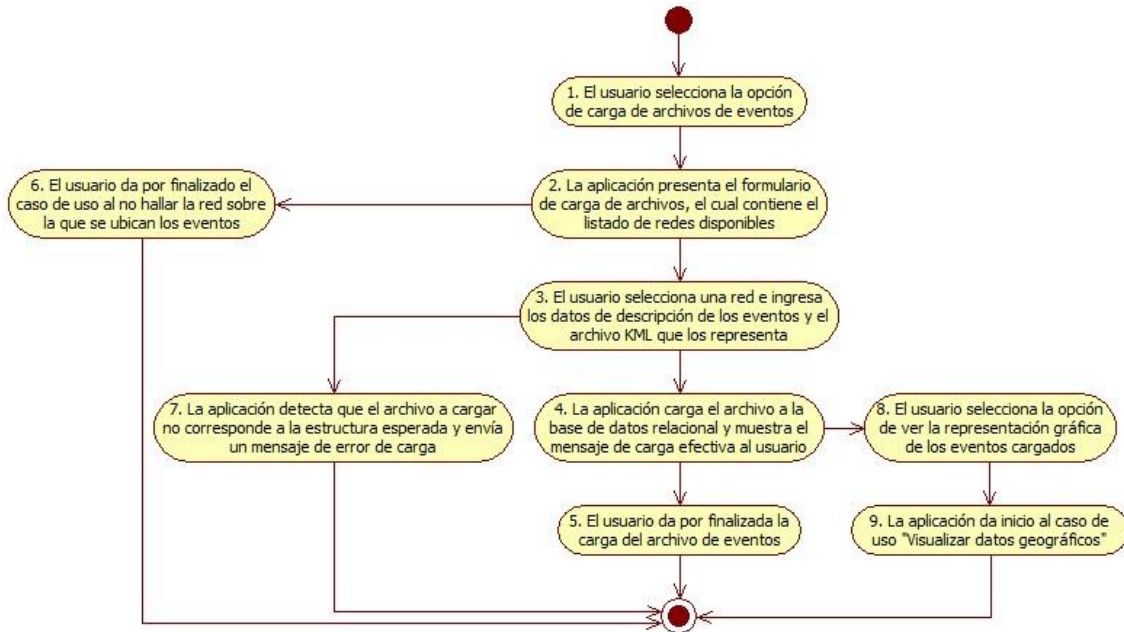
Usuario.

Figura 8. Diagrama de Secuencia – Cargar Archivo de Red



Fuente El autor.

Figura 9. Diagrama de Actividades – Cargar Archivo de Eventos



Fuente El autor.

Precondiciones

Se debe haber cargado anteriormente el archivo de la red sobre la que se ubican los eventos a cargar.

Flujo Básico

- 1) El usuario selecciona la opción de carga de archivos de eventos.
- 2) La aplicación presenta el formulario de carga de archivos, el cual contiene el listado de redes disponibles.
- 3) El usuario selecciona una red e ingresa los datos de descripción de los eventos y el archivo KML que los representa.
- 4) La aplicación carga el archivo a la base de datos relacional y muestra el mensaje de carga efectiva al usuario.
- 5) El usuario da por finalizada la carga del archivo de eventos.

Flujo Alternativo 1

Después del paso 2:

- 6) El usuario da por finalizado el caso de uso al no hallar la red sobre la que se ubican los eventos.

Flujo Alternativo 2

Después del paso 3:

- 7) La aplicación detecta que el archivo a cargar no corresponde a la estructura esperada y envía un mensaje de error de carga.

Flujo Alternativo 3

Después del paso 4:

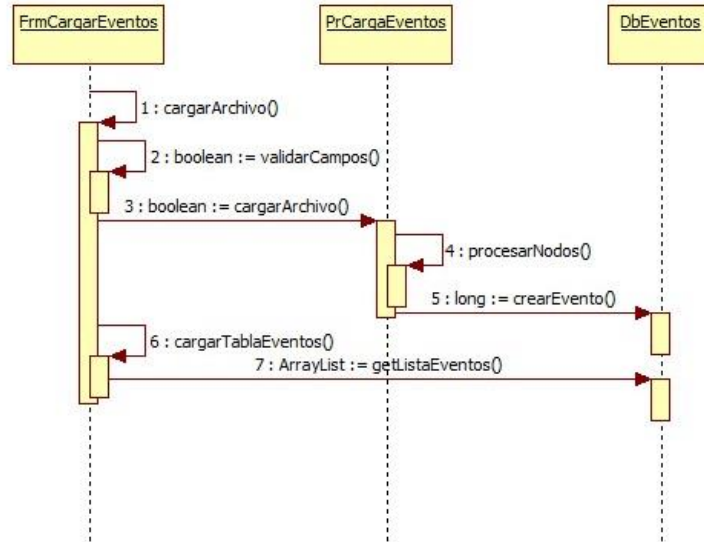
- 8) El usuario selecciona la opción de ver la representación gráfica de los eventos cargados.
- 9) La aplicación da inicio al caso de uso "Visualizar datos geográficos".

Descripción de Objetos e Interacciones

Este caso de uso es muy similar en estructura al caso de uso *Cargar Archivo de Red*, implementa el formulario *FrmCargarEventos*, el cual permite solicitar al usuario los datos y archivos necesarios para crear un evento. Los datos y archivos recibidos son entregados a una instancia de la clase *PrCargaEventos*, la cual tiene capacidad para procesar archivos KML de puntos y por medio de una

instancia de la clase *DbEventos* puede agregar la información procesada a una base de datos MySQL (ver figura 10).

Figura 10. Diagrama de Secuencia – Cargar Archivo de Eventos



Fuente El autor.

4.3.4. Realizar Cierre de Nodos

Este caso de uso permite el cierre (unión) de los nodos de una red que se encuentren a una distancia máxima predefinida.

Actores

Usuario.

Precondiciones

Se debe haber cargado anteriormente el archivo de la red sobre la que se realizarán las proyecciones.

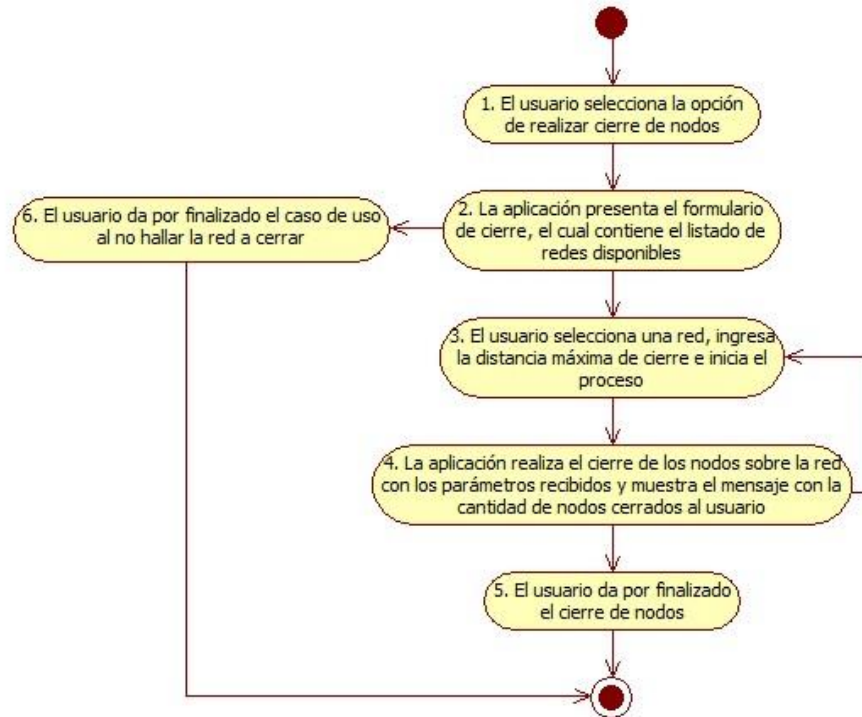
Flujo Básico

- 1) El usuario selecciona la opción de realizar cierre de nodos.
- 2) La aplicación presenta el formulario de cierre, el cual contiene el listado de redes disponibles.
- 3) El usuario selecciona una red, ingresa la distancia máxima de cierre e inicia el proceso.

4) La aplicación realiza el cierre de nodos sobre la red con los parámetros recibidos y muestra el mensaje con la cantidad de nodos cerrados al usuario.

5) El usuario da por finalizado el cierre de nodos.

Figura 11. Diagrama de Actividades – Realizar Cierre de Nodos



Fuente El autor.

Flujo Alternativo 1

Después del paso 2:

6) El usuario da por finalizado el caso de uso al no hallar la red a cerrar.

Flujo Alternativo 2

Después del paso 4 el usuario decide realizar nuevamente el cierre de nodos con diferentes parámetros por lo que el caso de uso vuelve al paso 3.

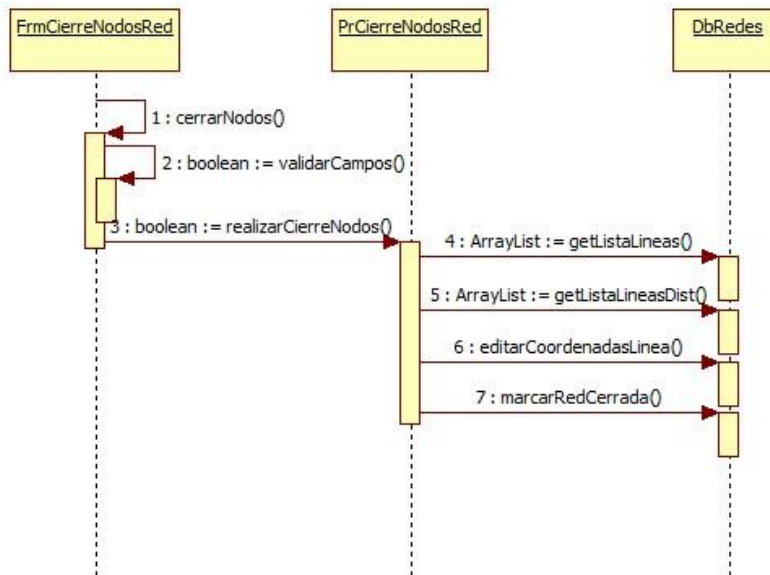
Requerimientos Especiales

Este caso de uso puede ser llamado dentro de la ejecución de los casos de uso "Calcular valores de función K para redes" y "Calcular densidades de NetKDE".

Descripción de Objetos e Interacciones

Este caso de uso implementa el formulario *FrmCierreNodos*, el cual solicita la distancia máxima de cierre de nodos sobre una red previamente cargada. Los datos recibidos son entregados a una instancia de la clase *PrCierreNodos*, este objeto realiza los cálculos necesarios para determinar cuáles nodos deben unirse y por medio de un objeto de tipo *DbRedes* obtiene los datos necesarios para sus operaciones y actualiza la información de los nodos de la red seleccionada en la base de datos MySQL (ver figura 12).

Figura 12. Diagrama de Secuencia – Realizar Cierre de Nodos



Fuente El autor.

4.3.5. Realizar Proyección de Puntos

Este caso de uso permite la proyección de cada uno de los puntos que representan eventos sobre la línea de la red más cercana a ellos dentro de un radio predefinido.

Actores

Usuario.

Precondiciones

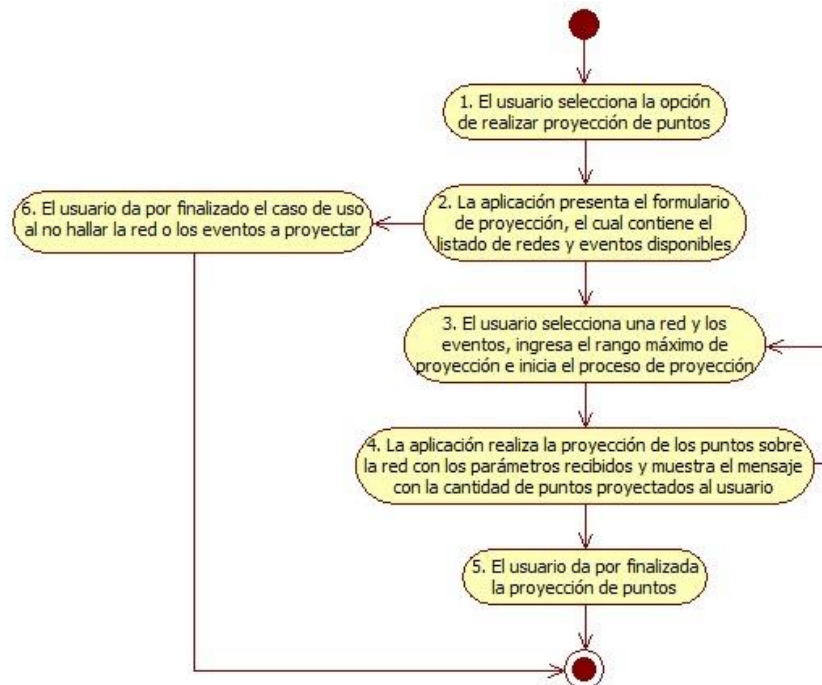
Se debe haber cargado anteriormente el archivo de la red sobre la que se realizarán las proyecciones.

Se debe haber cargado anteriormente el archivo de eventos (puntos) a proyectar.

Flujo Básico

- 1) El usuario selecciona la opción de realizar proyección de puntos.
- 2) La aplicación presenta el formulario de proyección, el cual contiene el listado de redes y eventos disponibles.
- 3) El usuario selecciona una red y los eventos, ingresa el rango máximo de proyección e inicia el proceso de proyección.
- 4) La aplicación realiza la proyección de los puntos sobre la red con los parámetros recibidos y muestra el mensaje con la cantidad de puntos proyectados al usuario.
- 5) El usuario da por finalizada la proyección de puntos.

Figura 13. Diagrama de Actividades – Realizar Proyección de Puntos



Fuente El autor.

Flujo Alternativo 1

Después del paso 2:

- 6) El usuario da por finalizado el caso de uso al no hallar la red o los eventos a proyectar.

Flujo Alternativo 2

Después del paso 4 el usuario decide realizar nuevamente la proyección de los puntos con diferentes parámetros por lo que el caso de uso vuelve al paso 3.

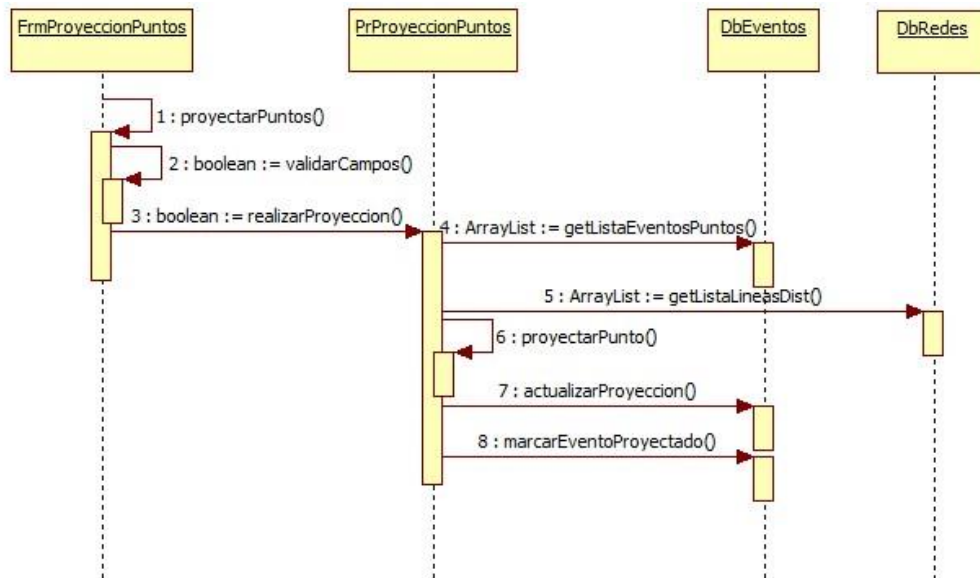
Requerimientos Especiales

Este caso de uso puede ser llamado dentro de la ejecución de los casos de uso “Calcular valores de función K para redes” y “Calcular densidades de NetKDE”.

Descripción de Objetos e Interacciones

Este caso de uso implementa el formulario *FrmProyeccionPuntos*; por medio de este formulario se solicita la distancia máxima de proyección de puntos sobre una la red asociada al evento. Los datos suministrados por el usuario son entregados a una instancia de la clase *PrProyeccionPuntos*, la cual realiza las proyecciones perpendiculares o de distancia más cercana de los puntos sobre la red. Por medio de un objeto de tipo *DbRedes* obtiene los datos de la red correspondiente y por medio de un objeto de tipo *DbEventos* obtiene los datos de los puntos y actualiza las coordenadas proyectadas de los puntos en la base de datos MySQL (ver figura 14).

Figura 14. Diagrama de Secuencia – Realizar Proyección de Puntos



Fuente El autor.

4.3.6. Generar Distribuciones Aleatorias de Puntos

Este caso de uso permite la generación aleatoria de grupos de puntos sobre una red dada, estos grupos serán utilizados posteriormente en los cálculos de la función K para redes.

Actores

Este caso de uso no es llamado directamente por ningún actor, su llamado se hace a través del caso de uso *Calcular Valores de Función K para Redes*.

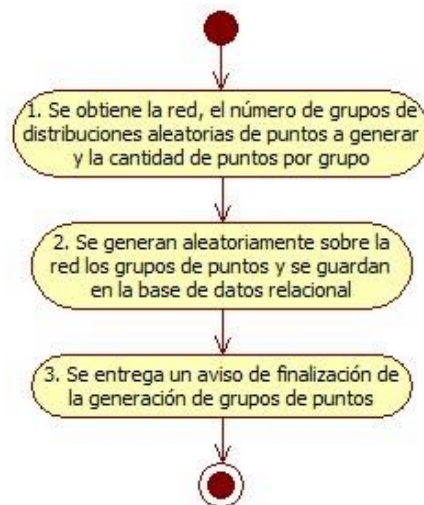
Precondiciones

El usuario debe haber ingresado al caso de uso *Calcular Valores de Función K para Redes*.

Flujo Básico

- 1) Se obtiene la red, el número de grupos de distribuciones aleatorias de puntos a generar y la cantidad de puntos por grupo.
- 2) Se generan aleatoriamente sobre la red los grupos de puntos y se guardan en la base de datos relacional.
- 3) Se entrega un aviso de finalización de la generación de grupos de puntos.

Figura 15. Diagrama de Actividades – Generar Distribuciones Aleatorias de Puntos



Fuente El autor.

Requerimientos Especiales

Este caso de uso puede ser llamado dentro de la ejecución del caso de uso "Calcular valores de función K para redes".

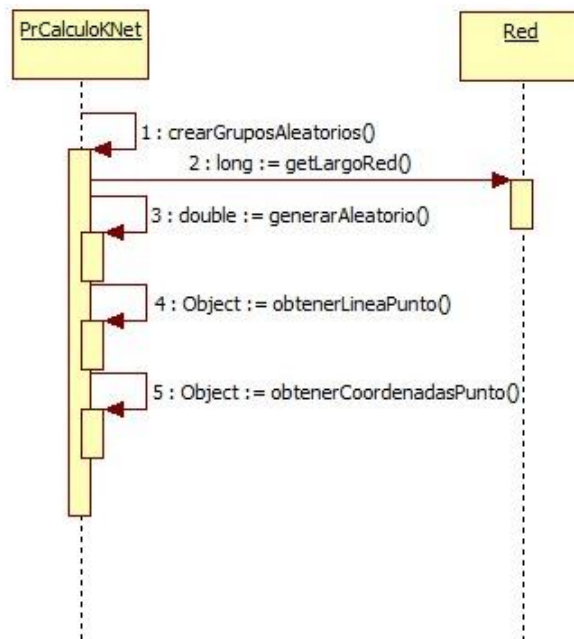
Descripción de Objetos e Interacciones

Este caso de uso no tiene asociado un formulario propio, se invoca desde la ejecución del caso de uso *Calcular Valores de Función K para Redes*, dicho caso de uso provee el número de grupos aleatorios a generar y el número de puntos para cada grupo a una instancia de la clase *PrCalculoKNet*, la cual obtiene la longitud de la red por medio de una instancia de la clase *DbRedes* y genera aleatoriamente, por medio de una distribución uniforme, los grupos de datos solicitados (ver figura 16).

4.3.7. Calcular Valores de Función K para Redes

Este caso de uso realiza el cálculo de valores de la función K para redes sobre un grupo de eventos proyectados sobre un subespacio de red.

Figura 16. Diagrama de Secuencia – Generar Distribuciones Aleatorias de Puntos



Fuente El autor.

Actores

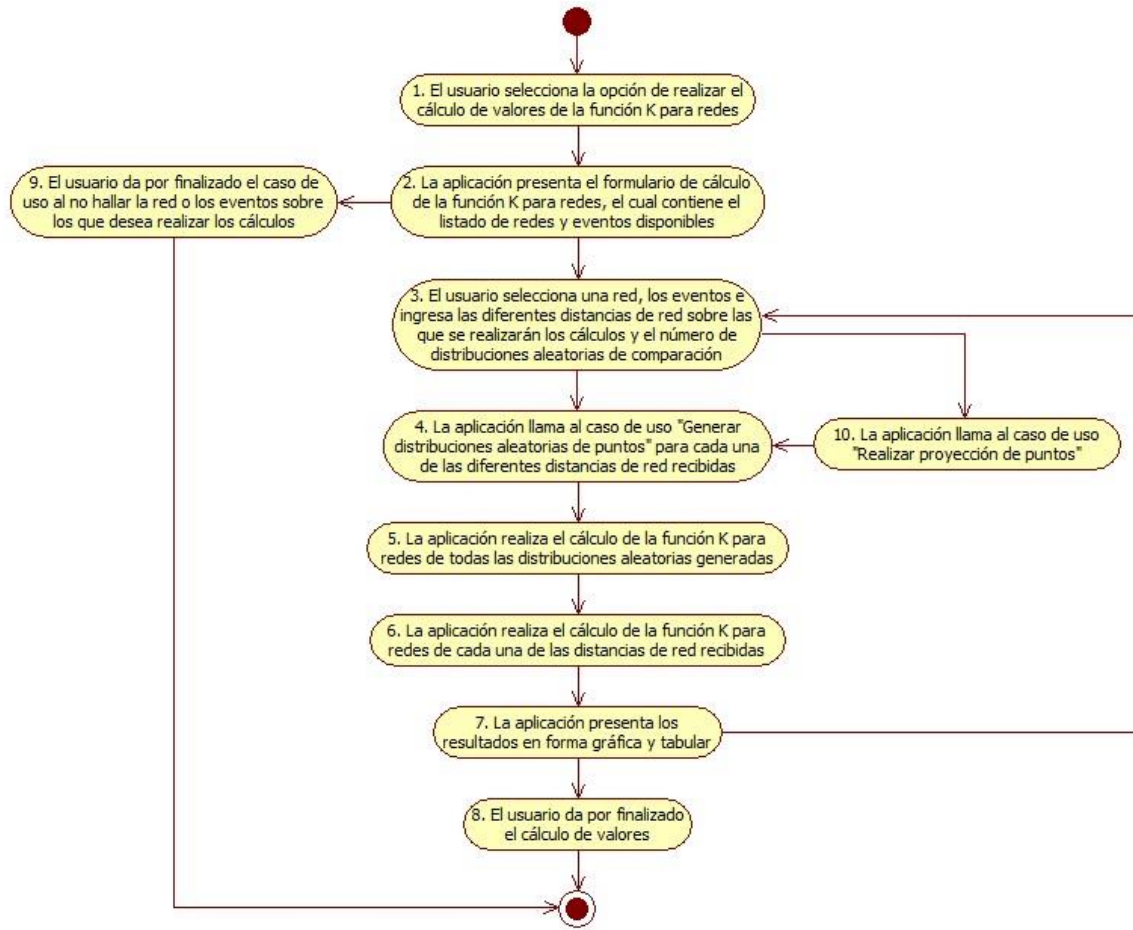
Usuario.

Precondiciones

Se debe haber cargado anteriormente el archivo de la red sobre la que se realizarán las proyecciones.

Se debe haber cargado anteriormente el archivo de eventos (puntos) a proyectar.

Figura 17. Diagrama de Actividades – Calcular Valores de Función K para Redes



Fuente El autor.

Flujo Básico

- 1) El usuario selecciona la opción de realizar el cálculo de valores de la función K para redes.
- 2) La aplicación presenta el formulario de cálculo de la función K para redes, el cual contiene el listado de redes y eventos disponibles.
- 3) El usuario selecciona una red, los eventos e ingresa las diferentes distancias de red sobre las que se realizarán los cálculos y el número de distribuciones aleatorias de comparación.
- 4) La aplicación llama al caso de uso "Generar distribuciones aleatorias de puntos" para cada una de las diferentes distancias de red recibidas.

5) La aplicación realiza el cálculo de la función K para redes de todas las distribuciones aleatorias generadas.

6) La aplicación realiza el cálculo de la función K para redes de cada una de las distancias de red recibidas.

7) La aplicación presenta los resultados en forma gráfica y tabular.

8) El usuario da por finalizado el cálculo de valores.

Flujo Alternativo 1

Después del paso 2:

9) El usuario da por finalizado el caso de uso al no hallar la red o los eventos sobre los que desea realizar los cálculos.

Flujo Alternativo 2

Después del paso 3, la aplicación detecta que no se ha realizado el proceso de proyección de puntos sobre la red.

10) La aplicación llama al caso de uso “Realizar proyección de puntos”.

La aplicación continúa en el paso 4.

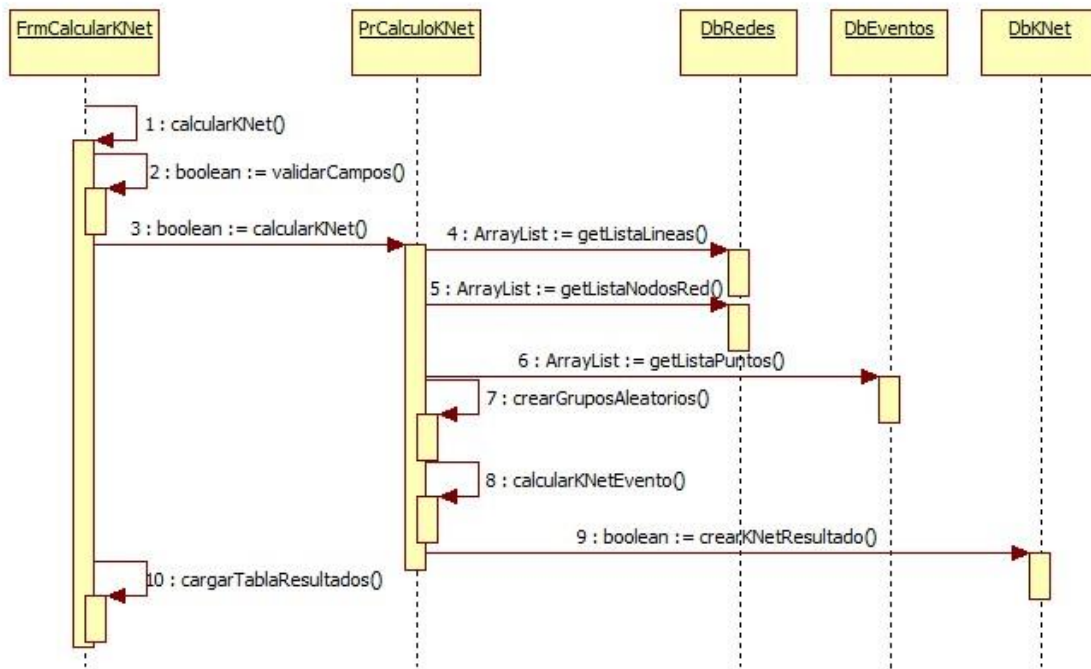
Flujo Alternativo 3

Después del paso 7 el usuario decide realizar nuevamente el proceso de cálculo de valores con diferentes parámetros por lo que el caso de uso vuelve al paso 3.

Descripción de Objetos e Interacciones

Este caso de uso implementa el formulario *FrmCalcularKNet*; por medio de este formulario se solicitan al usuario los datos necesarios para el cálculo de la función K para redes, estos datos son entregados a una instancia de la clase *PrCalculoKNet*, la cual realiza los cálculos correspondientes la función. Por medio de una instancia de la clase *DbRedes* se obtienen los datos de la red correspondiente, los datos del evento asociado se obtienen por medio de una instancia de la clase *DbEventos*. Los datos resultantes del cálculo de la función se almacenan en la base de datos relacional MySQL por medio de una instancia de *DbKNet*. Por último, el formulario muestra el listado de resultados de la función K almacenados en la base de datos (ver figura 18).

Figura 18. Diagrama de Secuencia – Calcular Valores de Función K para Redes



Fuente El autor.

4.3.8. Calcular Densidades de NetKDE

Este caso de uso realiza el cálculo de las densidades del método NetKDE sobre un grupo de eventos proyectados sobre un subespacio de red.

Actores

Usuario.

Precondiciones

Se debe haber cargado anteriormente el archivo de la red sobre la que se realizarán las proyecciones.

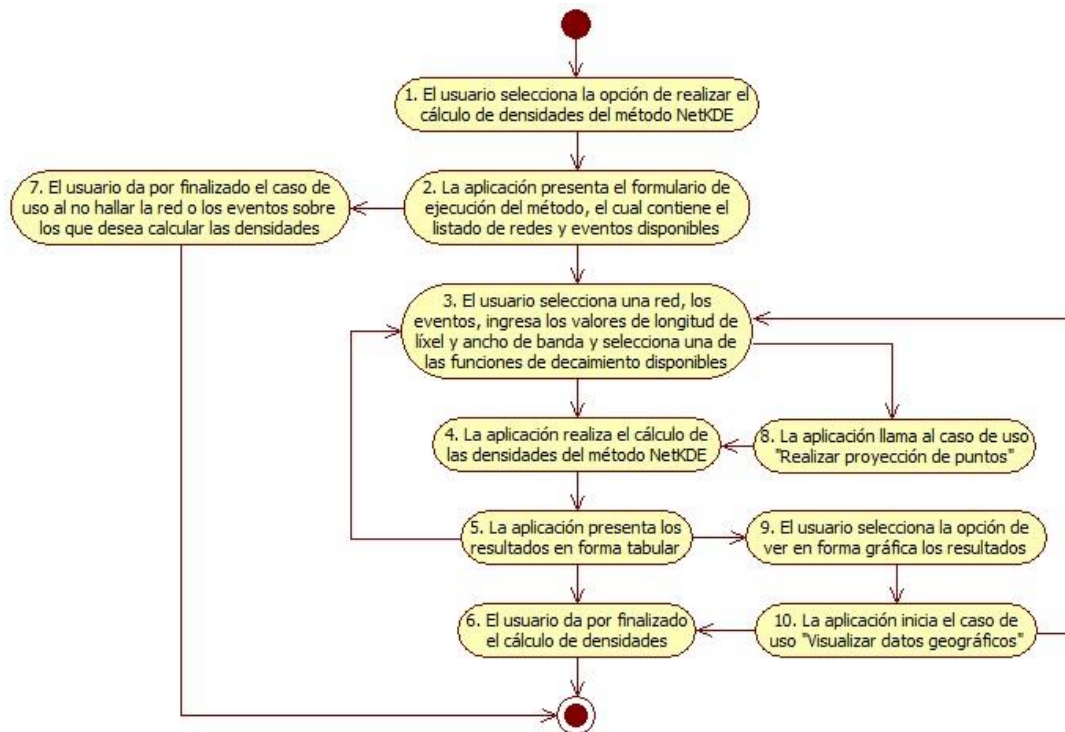
Se debe haber cargado anteriormente el archivo de eventos (puntos) a proyectar.

Flujo Básico

- 1) El usuario selecciona la opción de realizar el cálculo de densidades del método NetKDE.
- 2) La aplicación presenta el formulario de ejecución del método, el cual contiene el listado de redes y eventos disponibles.

- 3) El usuario selecciona una red, los eventos, ingresa los valores de longitud de *líxel* y ancho de banda y selecciona una de las funciones de decaimiento disponibles.
- 4) La aplicación realiza el cálculo de las densidades del método NetKDE.
- 5) La aplicación presenta los resultados en forma tabular.
- 6) El usuario da por finalizado el cálculo de densidades.

Figura 19. Diagrama de Actividades – Calcular Densidades de NetKDE



Fuente El autor.

Flujo Alternativo 1

Después del paso 2:

- 7) El usuario da por finalizado el caso de uso al no hallar la red o los eventos sobre los que desea calcular las densidades.

Flujo Alternativo 2

Después del paso 3, la aplicación detecta que no se ha realizado el proceso de proyección de puntos sobre la red.

8) La aplicación llama al caso de uso “Realizar proyección de puntos”.

La aplicación continúa en el paso 4.

Flujo Alternativo 3

Después del paso 5:

9) El usuario selecciona la opción de ver en forma gráfica los resultados.

10) La aplicación inicia el caso de uso “Visualizar datos geográficos”.

El caso de uso continúa en el paso 6.

Flujo Alternativo 4

Después del paso 5 el usuario decide realizar nuevamente el proceso de cálculo de densidades con diferentes parámetros por lo que el caso de uso vuelve al paso 3.

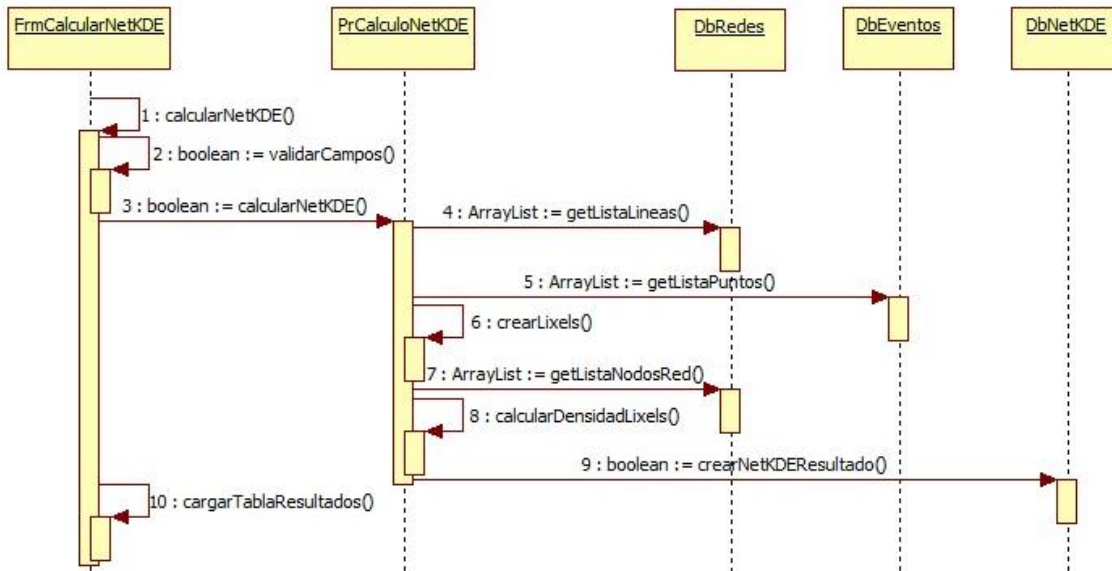
Flujo Alternativo 5

Después del paso 10 el usuario decide realizar nuevamente el proceso de cálculo de densidades con diferentes parámetros por lo que el caso de uso vuelve al paso 3.

Descripción de Objetos e Interacciones

Este caso de uso implementa el formulario *FrmCalcularNetKDE*; en este formulario se solicitan los datos necesarios para el cálculo del método NetKDE, estos datos son entregados a una instancia de la clase *PrCalculoNetKDE* que crea los *lixels* y realiza los cálculos de densidades. Por medio de un objeto de tipo *DbRedes* se obtienen los datos de la red correspondiente, los datos del evento asociado se obtienen por medio de un objeto de tipo *DbEventos*. Las densidades de *lixels* resultantes se almacenan en la base de datos relacional MySQL por medio de un objeto de tipo *DbNetKDE*. Por último, el formulario muestra el listado de resultados de NetKDE almacenados en la base de datos (ver figura 20).

Figura 20. Diagrama de Secuencia – Calcular Densidades de NetKDE



Fuente El autor.

4.4. ALISTAMIENTO DE DATOS

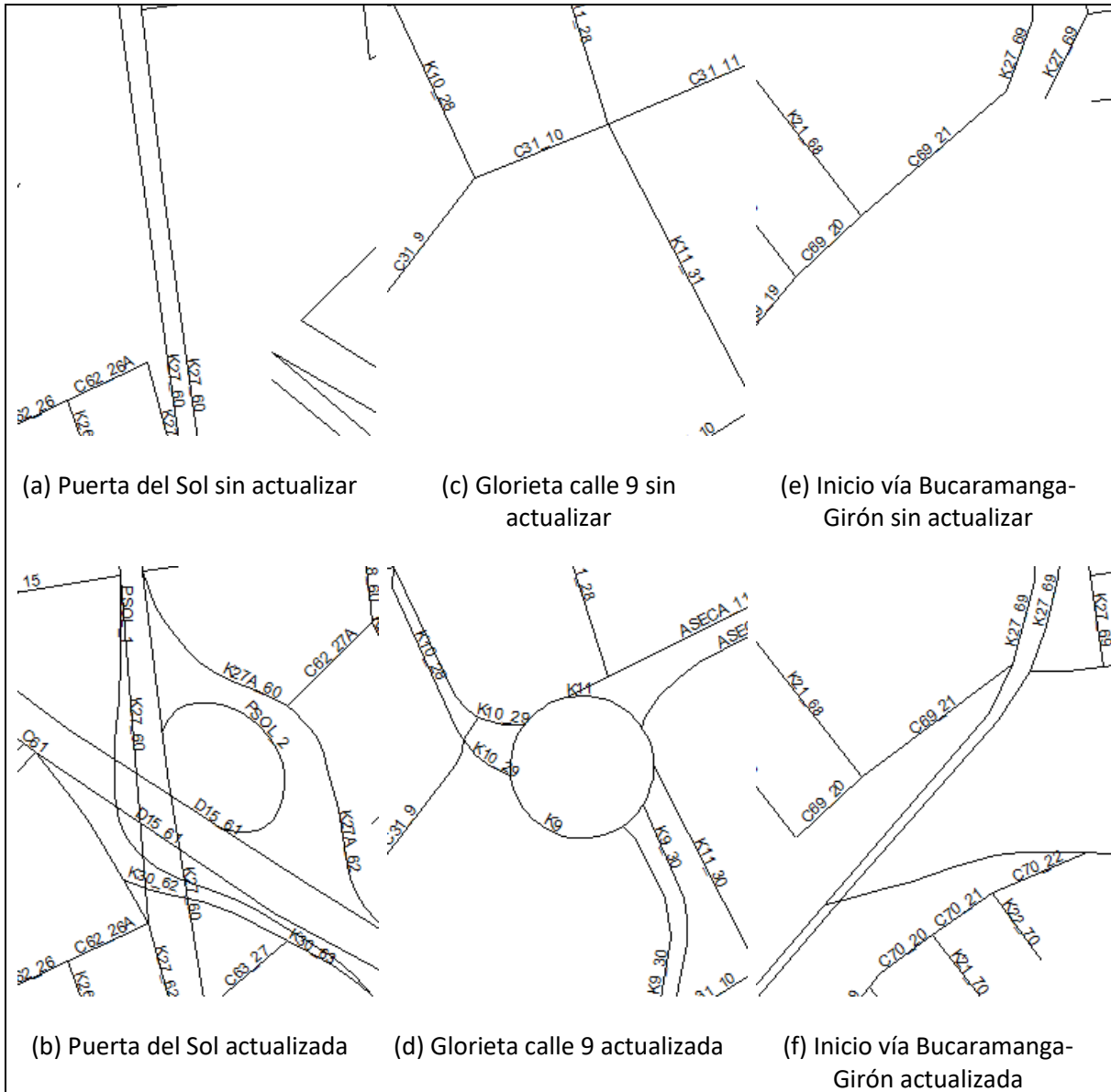
Para la obtención de los datos que representan la red vehicular del municipio de Bucaramanga se tomó como punto de partida los archivos de cartografía alojados en el Observatorio de Salud Pública de Santander – FOSCAL. Estos archivos están representados principalmente por un archivo de extensión SHP (ESRI Shapefile), para el uso en aplicaciones software de sistemas de información geográfica como ArcGIS y Arch/Info; y un archivo DBF que contiene los datos estructurados que describen cada uno de los registros que representan el sistema de vías (direcciones, identificadores de líneas asociadas, etc.).

Esta información cartográfica fue creada a finales de los años 90 por lo que no incluía algunas de las estructuras relativamente recientes de la red vial de Bucaramanga como el intercambiador de la Puerta del Sol y la glorieta de la calle 9 con avenida Quebradaseca. Además de estas estructuras se detectó que la cartografía tenía algunas vías faltantes; entre estas vías se encontraban la diagonal 15, la vía Bucaramanga-Girón, la vía a Chimitá y la vía al mar entre otras. Para el proceso de actualización vial se utilizó la aplicación ArcMap 10.2.2³ de ESRI y se tomó como referencia la información geográfica disponible en Google Maps⁴, algunos ejemplos de las actividades incluidas en la actualización vial se pueden ver en la figura 21.

³ <http://desktop.arcgis.com/es/desktop/>

⁴ <http://maps.google.com>

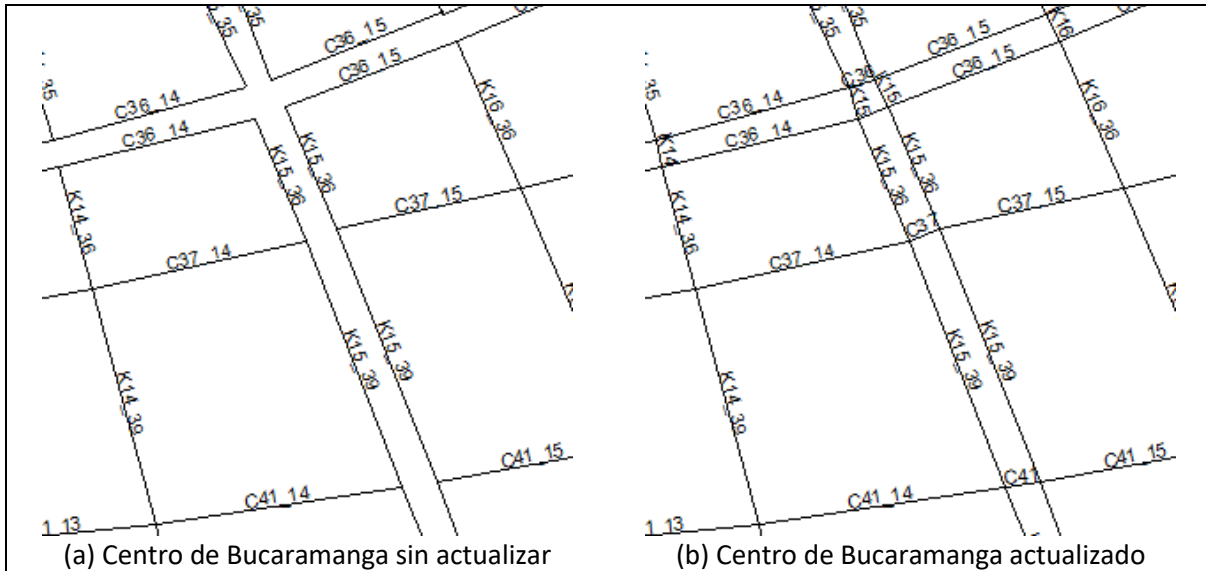
Figura 21. Actualización vial



Fuente El autor.

Además de la actualización vial fue necesario incluir extensiones de las vías en los lugares en los que se cruzan con vías dobles y en los lugares en los que las vías dobles no tienen separadores, esto debido a que para los dos métodos es necesario determinar las distancias más cortas a través de la red vial (ver figura 22).

Figura 22. Actualización vial – Extensión de vías



Fuente El autor.

La información vial actualizada fue exportada posteriormente a formato KML por medio del servicio web de MyGeoData⁵. 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><Folder><name>vial_buc</name>
  <Schema name="vial_buc" id="vial_buc">
    <SimpleField name="LINE_" type="float"></SimpleField>
    <SimpleField name="DIRECCION"
type="string"></SimpleField>
  </Schema>
  <Placemark>
    <Style>
      <LineStyle><color>ff0000ff</color></LineStyle>
      <PolyStyle><fill>0</fill></PolyStyle>
    </Style>
    <ExtendedData><SchemaData schemaUrl="#vial_buc">
      <SimpleData name="LINE_">1</SimpleData>
      <SimpleData name="DIRECCION">A39 48</SimpleData>
    </SchemaData></ExtendedData>
    <LineString>
      <coordinates>
        -89.6640625,78872.390625 -14.34375,78714.140625
```

⁵ <http://converter.mygeodata.eu/>

```

        </coordinates>
    </LineString>
</Placemark>
</Folder></Document></kml>

```

El método descrito por (Xie & Yan, 2008) para NetKDE indica que se debe realizar una proyección de los puntos que representan los accidentes de tránsito sobre la red vehicular representada por líneas en la cartografía, pero dado que la información sobre accidentes de tránsito para el año 2014 suministrada por la Dirección de Tránsito de Bucaramanga no contaba con el componente de ubicación geográfica (latitud y longitud del punto que representa el accidente), fue necesario ubicar cada uno de los accidentes de tránsito tomando como base su dirección. Para lo anterior fue necesario crear un algoritmo en el lenguaje de programación Avenue (ver anexo 1), incluido en el software ArcView 3.2 de ESRI, que tomara cada una de las direcciones y las ubicara dentro de la red vehicular. En la tabla 2 se dan algunos ejemplos de los ajustes que se realizaron sobre las direcciones para que pudieran ser ubicadas por medio del algoritmo en Avenue.

De los 3.731 datos de accidentes de tránsito correspondientes al año 2014 en el municipio de Bucaramanga se ubicaron 3.105 por medio del algoritmo mencionado anteriormente. De los restantes 626 registros se logró una ubicación manual de 556 por medio de ArcMap 10.2.2. Los 70 registros restantes no pudieron ser ubicados debido a que no se contaba con información completa del lugar del accidente, estos registros se descartaron de la base de datos de trabajo. El proceso de ubicación automática con el algoritmo de localización produjo un conjunto de archivos cartográficos (SHP, DBF, etc.); estos archivos fueron a su vez utilizados como insumo de entrada en el proceso de ubicación manual, con lo que se obtuvo al finalizar este proceso un conjunto de archivos con los puntos (automáticos y manuales) georreferenciados.

Tabla 2. Direcciones ajustadas para Avenue

Dirección original	Dirección ajustada
Carrera 33 con calle 48	cr_33_48-0
Calle 56 #17A-12	cli_56_17a-12
Avenida Quebrada Seca #28-25	av_qseca_25-25
Fuente	El autor.

Una vez se obtuvieron las coordenadas geográficas de cada uno de los accidentes de tránsito, se procedió a la elaboración del correspondiente archivo KML por medio del servicio web de MyGeoData mencionado anteriormente. Una versión simplificada de la información contenida en este archivo se presenta a continuación:

```

<?xml version="1.0" encoding="utf-8" ?>
<kml xmlns="http://www.opengis.net/kml/2.2">
<Document><Folder><name>puntos_unif</name>

```

```

<Schema name="puntos_unif" id="puntos_unif">
  <SimpleField name="ID" type="float"></SimpleField>
  <SimpleField name="FECHA" type="string"></SimpleField>
  <SimpleField name="HORA" type="string"></SimpleField>
  <SimpleField name="DIRECC" type="string"></SimpleField>
</Schema>
<Placemark>
  <ExtendedData><SchemaData schemaUrl="#puntos_unif">
    <SimpleData name="ID">1.00000</SimpleData>
    <SimpleData name="FECHA">2014/08/22</SimpleData>
    <SimpleData name="HORA">11:00:00 a.m.</SimpleData>
    <SimpleData name="DIRECC">CRA 15 CALLE 19A</SimpleData>
  </SchemaData></ExtendedData>
  <Point>
    <coordinates>
      -40.635767152,80110.5758702823
    </coordinates>
  </Point>
</Placemark>
</Folder></Document></kml>

```

Para los datos de la red vial de la ciudad de Barcelona no fue necesario realizar un proceso de adecuación cartográfica dado que la fecha de elaboración de la cartografía recibida era bastante reciente, databa del año 2014, y su proceso de elaboración fue muy exhaustivo en lo referente a la precisión en comparación con la red vial que representa. Todos los registros de accidentes de tránsito recibidos incluían las coordenadas geográficas respectivas por lo que no fue necesario realizar un proceso de georreferenciación.

5. ANÁLISIS DE RESULTADOS

Los resultados obtenidos como parte de los procesos de implementación y utilización de la herramienta software desarrollada se dividen en dos grandes grupos; el primer grupo hace referencia a los hallazgos realizados con respecto a la implementación de los diferentes componentes metodológicos y de software que conforman los dos métodos de análisis de patrones de puntos, la función K para redes y NetKDE. El segundo grupo se centra en el análisis de los resultados obtenidos en el proceso de aplicación de los métodos implementados sobre los grupos de datos reales de accidentes de tránsito de las ciudades de Bucaramanga (Colombia) y Barcelona (España) ocurridos durante el año 2014.

5.1. HALLAZGOS EN LA IMPLEMENTACIÓN DE LA HERRAMIENTA SOFTWARE

Dado que la construcción de la herramienta software se llevó a cabo aplicando una metodología de desarrollo ágil que requería de entregas frecuentes [24], éste se dividió en iteraciones que producían resultados aplicables a los grupos de datos de accidentes de tránsito previamente obtenidos y revisados. Las fases de desarrollo centradas en los productos obtenidos se detallan en la tabla 3.

A continuación, se detallan los hallazgos pertenecientes a cada una de las etapas del desarrollo de software.

5.1.1. Módulo de Carga de Archivos

Los archivos de entrada seleccionados corresponden al formato KML, como se detalló en la sección 3.1; este formato está basado en la notación XML y su importación se realizó por medio de métodos recursivos que interpretan las etiquetas definidas en el estándar KML para datos de atributos, líneas, puntos y coordenadas, en la tabla 4 se resumen las principales etiquetas KML.

La prueba sobre la versión inicial de carga de archivos de red tomó aproximadamente 105 minutos con el archivo de red vial de la ciudad de Bucaramanga. En esta primera versión se agregaba cada registro de definición de atributo, línea, atributo de línea y detalle de línea a la base de datos relaciona de la aplicación de forma separada, es decir cada registro contaba con su propia sentencia *INSERT* independiente.

Si se tiene en cuenta que el archivo de red de la ciudad de Bucaramanga contiene 95.381 registros y que el archivo correspondiente de la ciudad de Barcelona cuenta con 385.293, se puede estimar por medio de una regla de 3 simple su tiempo de carga a aproximadamente 424,1 minutos. Estos

tiempos de carga se redujeron drásticamente con la aplicación de sentencias *bulk INSERT*⁶ para grupos de 100 registros, el tiempo de carga del archivo de red de la ciudad de Bucaramanga pasó a ser de aproximadamente 2 minutos y 25 segundos, y el de la ciudad de Barcelona a 9 minutos y 50 segundos; una reducción de tiempo del 97,7% en ambos casos.

Tabla 3. Productos software por etapas

Fase	Producto Incluidos	Observaciones
1	Módulo de carga de archivos	La carga de archivos incluyó los procesos de conversión de sistemas de coordenadas y medición de segmentos de vías.
2	Procesos de normalización de datos	Los procesos de normalización de datos implementados fueron el cierre de nodos de red y la proyección de puntos de eventos sobre segmentos de red.
3	Módulo de función K para redes	La implementación de este método requirió la implementación previa de un componente software de generación aleatoria de grupos de datos localizados en redes.
4	Módulo de NetKDE	La implementación de este método requirió el desarrollo previo de un componente de búsqueda de distancias más cortas en redes.
5	Módulo de visualización de resultados geográficos	Este módulo incluye componentes de visualización de datos cartográficos comunes como las escalas de tamaño, leyendas y etiquetas de datos. También se incluyen componentes para la visualización de puntos calientes y la comparación de diferentes grupos de resultados.
6	Módulo de creación y verificación de base de datos relacional asociada	Este módulo no se tuvo en cuenta en la elaboración del cronograma del proyecto pero se incluyó por considerar que daba una mayor facilidad de distribución al software desarrollado.
Fuente El autor.		

Además de la estructura general de los archivos KML, en el caso específico de los atributos de los puntos se incluyó la definición del formato de fecha y hora, esto con el fin de permitir posteriormente la inclusión de filtros en los conjuntos de datos a analizar por medio de los métodos incluidos en la aplicación; estos valores de fecha y hora pueden estar unidos en un único atributo o pueden estar separados, se incluyó manejo para los principales formatos tanto de fecha (dd/mm/yyyy, mm/dd/yyyy, etc.) como de hora (am/pm o 00-23 horas).

Por último, en el proceso de creación de las líneas de una red se incluyó el cálculo de la longitud de estas, que a su vez incluyó el cálculo de la longitud de cada uno de los segmentos de recta que

⁶ <https://dev.mysql.com/doc/refman/5.5/en/optimizing-innodb-bulk-data-loading.html>

conforman la línea; este cálculo soporta el manejo tanto de coordenadas geográficas como de coordenadas planas (proyecciones de Mercator) (Osborne, 2008).

Tabla 4. Principales etiquetas KML

Etiqueta	Descripción
Schema	Contenedor de atributos
SimpleField	Definición de un atributo específico, permite definir el nombre y tipo de dato del atributo
Folder	Contenedor de todos los elementos representados por el archivo, ya sean líneas o puntos
Placemark	Contenedor de los componentes que definen una única línea o un punto
SchemaData	Contenedor de los valores de todos los atributos de una línea o punto
SimpleData	Contenedor del valor de un atributo específico de una línea o punto
LineString	Contenedor de todas las coordenadas de una línea
Point	Contenedor de las coordenadas de un punto
Coordinates	Entrada de una dupla de valores de latitud y longitud que definen una coordenada, esta dupla se debe separar con una coma (,). En esta entrada se puede incluir un tercer atributo que representa la altitud sobre el nivel del mar
Fuente KML Reference (Google, 2016).	

5.1.2. Procesos de Normalización de Datos

Los procesos de normalización de los datos geográficos de redes (líneas) y eventos (puntos) cargados en la aplicación se describirán a continuación. Aunque estos procesos no se encuentran especificados en los métodos de análisis de patrones de puntos implementados, hacen parte integral del tratamiento de datos geográficos y son necesarios para llevar a cabo la aplicación de los métodos implementados.

El primer proceso de normalización es el de cierre de nodos de red, este proceso consiste en comprobar si dos puntos (nodos) de dos líneas que se encuentran dentro de una distancia dada en realidad se encuentran conectados, de estarlo es necesario modificar las coordenadas de uno de los nodos asignándole los valores de coordenadas del segundo nodo. Este en principio es un proceso sencillo, pero se deben tener en cuenta algunos factores que pueden hacer de este un proceso que demande mucho tiempo de procesamiento.

El primer factor tiene que ver con la cantidad teórica posible de pares de nodos que requerirían ser comparados para verificar si se encuentran dentro de la distancia máxima definida para aplicar el proceso de corrección de coordenadas. Teniendo en cuenta las cantidades de nodos presentes en las redes utilizadas, 14.811 para Bucaramanga y 45.742 para Barcelona, y considerando que dos nodos pertenecientes a una misma línea no se deben comparar entre sí, se puede estimar que el número de comparaciones necesarias ascendería a 219.289.480 para la red vial de Bucaramanga y

2.092.047.366 para la de Barcelona⁷. Estas magnitudes en sí mismas son lo suficientemente grandes como para que sea necesario considerar la creación de procesos alternativos de comparación de nodos sin tener en cuenta factores adicionales.

El segundo factor está relacionado con el proceso de cálculo necesario para hallar la distancia entre dos nodos arbitrarios. Este proceso incluye el manejo de diferentes sistemas de coordenadas y los cálculos necesarios para hallar distancias entre puntos pertenecientes a dichos sistemas. El proceso de cálculo de distancia entre dos puntos no es complejo ni extenso, pero incluye el manejo de variables numéricas decimales de doble precisión y flujos de decisión. Si estas características se unen a las del primer factor expuesto se tiene como resultado un proceso que requeriría una cantidad excesiva de tiempo de proceso. En contraste, la cantidad de memoria RAM necesaria para llevar a cabo un proceso de estas proporciones puede no ser un factor determinante dado que se pueden crear funciones modulares, lo que ayudará a disminuir la cantidad de memoria necesaria para el proceso.

Debido a estos factores se hizo necesaria la elaboración de un proceso alternativo en el que se inicia con la aplicación de algunas simplificaciones para disminuir drásticamente el número de comparaciones necesarias. La primera simplificación consistió en asumir que todas las intersecciones involucran por lo menos un nodo terminal de línea, esto puede ocasionar que algunos nodos de unión reales no sean revisados; pero teniendo en cuenta la estructura típica de una red vial, en la que dos tramos de vía se conectan directamente por sus extremos o que un tramo desemboca en otro, el impacto de esta decisión se puede considerar como mínimo.

Al aplicar este criterio el número de comparaciones necesarias se puede estimar así, primero se parte del número total de líneas, 5.754 para Bucaramanga y 14.776 para Barcelona, lo que implica un total de nodos terminales igual al doble del número de líneas; si estos números de nodos se multiplican por la diferencia entre el número total de nodos y el promedio de nodos por línea se tienen 170.415.366 y 1.351.676.099 comparaciones para las redes viales de Bucaramanga y Barcelona respectivamente, lo que representa en ambos casos una reducción cercana al 30%. Los datos detallados se incluyen en la tabla 5.

Tabla 5. Estimación de números de comparaciones de distancias (primera simplificación)

Red Vial	Nodos	Líneas	Nodos/Línea	Comparaciones requeridas	Comparaciones reducidas	Reducción
Bucaramanga	14.811	5.754	2,574	219.289.480	170.415.366	22,29%
Barcelona	45.742	14.776	3,0957	2.092.047.366	1.351.676.099	35,39%
Fuente El autor.						

⁷ Estos valores estimados se obtienen elevando al cuadrado la diferencia entre el número total de nodos y el promedio de nodos por línea, que es de 2,574 para Bucaramanga y de 3,0957 para Barcelona.

La segunda simplificación al método consistió en la aplicación de un proceso reducido de cálculo de distancia previo. La aplicación de este proceso parte de la estimación de unos deltas de latitud y longitud en una coordenada arbitraria dentro de la red vial y en el sistema de coordenadas de la red que se esté procesando; estos deltas equivalen al 110% de la distancia máxima de ajuste y solamente se calculan una única vez para la red. Una vez calculados se utilizan como filtro previo al cálculo real de distancia, solamente se compararán los nodos cuyas latitudes y longitudes se encuentren dentro de las coordenadas menos delta y coordenadas más delta de un nodo terminal.

Con esta segunda simplificación solamente se compara cada nodo terminal en promedio con otros 3,67 nodos para distancias máximas de ajuste de un (1) metro, lo que lleva a reducciones en el número total de comparaciones superiores al 99% en ambos casos. Los totales de comparaciones se presentan en la tabla 6.

Tabla 6. Estimación de números de comparaciones de distancias (segunda simplificación)

Red Vial	Nodos terminales	Comparaciones requeridas	Comparaciones reducidas (1)	Comparaciones reducidas (2)	Reducción total
Bucaramanga	11.508	219.289.480	170.415.366	42.234	99,98%
Barcelona	29.552	2.092.047.366	1.351.676.099	108.455	99,99%
Fuente El autor.					

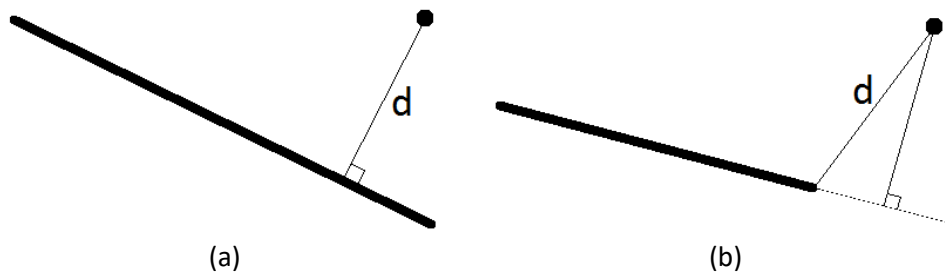
Es importante tener en cuenta que, aunque las reducciones en números de comparaciones son muy significativas, los procesos aún son extensos en tiempo ya que se agrega una gran cantidad de comparaciones simplificadas previas como filtros a los cálculos de distancias reales.

El segundo proceso de normalización de datos se centra en la proyección de los eventos (puntos) sobre las líneas de red más cercanas; este proceso es mencionado como requisito en el numeral 6 de la implementación de NetKDE (Xie & Yan, 2008) pero no se detalla una forma de llevarlo a cabo. Para la implementación de este proceso se partió del hecho que la proyección de un punto sobre una línea debe ser tal que no exista otra proyección posible con una distancia euclidiana menor, esto teóricamente se puede lograr proyectando cada punto sobre todas las líneas de la red y seleccionando la proyección de menor distancia, pero esta aproximación a la proyección presentaría los mismos problemas del proceso teórico de cierre de nodos presentado anteriormente en esta sección, es decir, sería necesario realizar un número muy alto de operaciones que incluirían cálculos para hallar puntos de proyección y cálculos de distancias entre puntos. Por lo anterior se hizo necesario plantear estrategias alternativas que redujeran la cantidad total de cálculos necesaria para realizar la proyección de todos los puntos de eventos.

El proceso implementado parte del siguiente postulado: la distancia más cercana entre un punto y una línea recta de longitud infinita está dada por la proyección perpendicular del punto sobre

dicha línea; en el caso de líneas rectas de longitud finita se debe hallar la distancia del punto a la continuación de la línea y comprobar si la proyección perpendicular se encuentra dentro de los límites reales de la línea, si la proyección efectivamente se encuentra dentro de estos límites se puede concluir que se ha hallado la proyección de menor distancia entre el punto y la línea específicos, de no ser así, se deben calcular las distancias desde el punto a los dos vértices de la línea y seleccionar el vértice que corresponda a la menor longitud como el punto de proyección. Una representación gráfica del proceso se presenta en la figura 23; en (a) la proyección se encuentra dentro de los límites de la línea por lo que esta proyección representa la distancia más corta d , en (b) la proyección se encuentra fuera de la línea por lo que la distancia más corta d se mide con respecto al vértice (nodo) más cercano.

Figura 23. Proyección de puntos sobre líneas



Fuente El autor.

Este postulado no disminuye el número de comparaciones necesarias para hallar la proyección más cercana por lo que fue necesario incluir un proceso previo de selección de líneas a comparar con cada punto, este proceso inicia con la estimación de la extensión real en metros de una unidad de latitud y una de longitud del sistema de coordenadas de la red en un punto arbitrario de esta; estos valores se utilizan para estimar de forma más rápida la distancia entre un punto y cada uno de los segmentos de línea de la red. Posteriormente sobre estas distancias estimadas se aplica un filtro de distancia de proyección máxima permitida y sobre los segmentos que superen este filtro se aplica el criterio definido en el postulado del párrafo anterior.

De la aplicación de este proceso pueden resultar algunos puntos que no se encuentren a una distancia menor a la distancia máxima definida de proyección, estas discrepancias pueden atribuirse al hecho de que no todos los accidentes que involucren vehículos tienen lugar sobre la red vial o a errores en los datos arrojados por los dispositivos GPS utilizados en la recolección de información de los accidentes de tránsito. Los puntos sin proyección no son tomados en cuenta para los procesos de cálculo de los métodos implementados.

Para los datos de accidentes de tránsito de la ciudad de Bucaramanga se utilizó una distancia máxima de proyección de 10 metros, con lo que se logró la proyección de la totalidad de los puntos. Este resultado se debe a que, como se describió en la sección 3.4, la gran mayoría de los

eventos fueron ubicados directamente sobre la red vial por medio de un algoritmo para la localización de direcciones y los eventos restantes fueron ubicados de forma manual sobre dicha red. Para el caso de la proyección de eventos de accidentes ocurridos en Barcelona se utilizó una distancia máxima de proyección de 50 metros con la que se proyectaron 8.615 puntos de los 8.776 existentes; los restantes 151 puntos no ubicados fueron descartados. Esta diferencia en la efectividad de la proyección con respecto a los datos de Bucaramanga se explica por el hecho de que la ubicación de los eventos de Barcelona fue tomada por medio de dispositivos GPS y no se tuvieron que realizar procesos de ubicación de puntos sobre la red.

5.1.3. Módulo de Función K para Redes

La función K para redes se define en términos generales como una sumatoria de distancias entre puntos medidas a través de una red aplicando un límite de distancia máxima a tener en cuenta y multiplicada por un factor de densidad lineal; un valor K calculado no tiene significado en sí mismo, su significado se obtiene al comparar series de valores K reales con series de valores K calculado sobre grupos de datos aleatorios. El método implementado contempla cuatro procesos a saber, el primero consiste en la generación de grupos de datos aleatorios, el segundo abarca los cálculos necesarios para determinar la distancia más corta entre dos puntos a través de la red, el tercero se centre en el cálculo de los valores de la función propiamente dicha y el cuarto cubre la presentación de los resultados en pantalla. Cada uno de estos procesos presentó retos de implementación los cuales se exponen a continuación.

Para el primer proceso, el de generación de datos (puntos) aleatorios sobre una red no se contó con una descripción detallada sobre su implementación dentro del método de la función K para redes. La aproximación utilizada consiste en tomar la red como una sucesión de segmentos de recta continua en una sola dimensión, sin bifurcaciones, para esto se calcula la extensión total de la red en metros basada en las longitudes de líneas calculadas durante el proceso de carga de los archivos KML de red (ver sección 4.1.1 *Módulo de Carga de Archivos*), posteriormente se generan grupos de números aleatorios de distribución uniforme entre cero (0) y el valor total de longitud calculado; cada grupo contiene una cantidad de números igual a la cantidad de puntos sobre los que se calculará la función K para redes y el número total de grupos es determinado por el usuario tomándose como valor por defecto 100 grupos.

Cada uno de los valores generados es utilizado para determinar el segmento de línea en el que se encontrará un punto aleatorio, para esto se toman todos los segmentos en orden uno tras otro y se utiliza el valor aleatorio como medida de longitud que determinará el segmento y la distancia dentro de este a partir de uno de sus vértices a la que se encontrará el punto. Con las coordenadas de los nodos del segmento seleccionado y la distancia a uno de estos nodos se determinan las coordenadas del punto aleatorio generado.

El segundo proceso, cuyo objetivo es el cálculo de la distancia mínima a través de la red entre todos los pares de puntos que se encuentren a una distancia menor a una distancia máxima de búsqueda definida por el usuario, se implementó de forma recursiva. El algoritmo creado se describe a continuación:

- a) Como parámetro inicial se recibe una distancia máxima de búsqueda.
- b) Se carga el listado total de puntos a comparar.
- c) Para cada uno de los puntos se busca el segmento de línea que lo contiene junto con los demás puntos que comparten el mismo segmento, se determina la distancia del punto inicial a cada punto y si esta distancia es menor a la distancia máxima de búsqueda, se carga el punto junto con la distancia en una tabla hash.
- d) Se determina la distancia del punto a cada uno de los vértices (nodos) del segmento de línea y se buscan los segmentos conectados a cada vértice.
- e) Para cada segmento encontrado se recibe como parámetro de entrada la distancia recorrida desde el punto base y el vértice inicial.
- f) Se buscan los puntos contenidos en el segmento y se determina la distancia de cada uno de estos al vértice inicial, estas distancias sumadas a la distancia recorrida se comparan con la distancia máxima de búsqueda y de ser menores se agregan a la tabla hash; si en la tabla ya se encuentra registrado el punto se determina cuál de las distancias medidas es menor y se guarda este valor.
- g) Se agrega la longitud del segmento a la distancia acumulada y si este resultado es menor que la distancia máxima de búsqueda se buscan los segmentos conectados al vértice final del segmento; para cada segmento hallado se repite el paso (e) entregando como distancia recorrida la nueva distancia acumulada. Si la nueva distancia acumulada es mayor a la distancia máxima de búsqueda se finaliza el subproceso de búsqueda correspondiente al segmento actual.

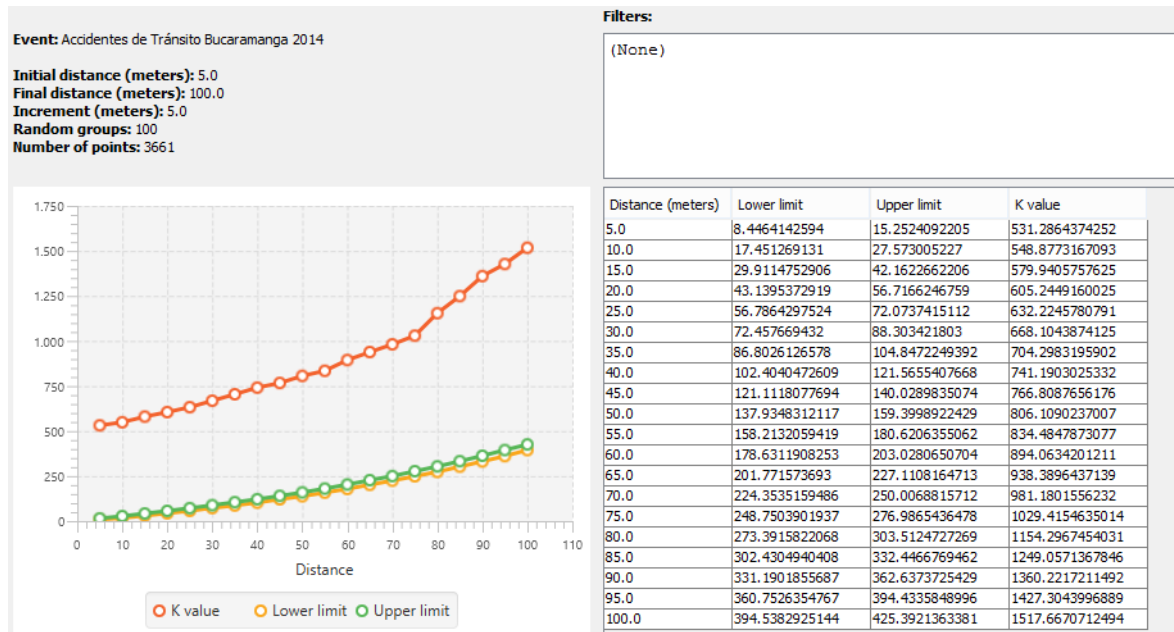
El resultado final del algoritmo es una lista de tablas hash con todos los puntos que se encuentra dentro de la distancia máxima de búsqueda de cada uno de los puntos.

El tercer proceso realiza el cálculo del valor de la función K para redes para cada uno de los grupos de puntos aleatorios y para el grupo de puntos real, dentro del cálculo se utiliza el segundo proceso y con los datos obtenidos se realizan los conteos de puntos y las operaciones aritméticas propias de la función. Este proceso es el más sencillo de los tres dado que una vez generados los puntos aleatorios y calculadas las distancias mínimas solamente queda pendiente la realización de operaciones aritméticas.

El cuarto y último proceso se ocupa de la presentación de datos en pantalla, para esto se implementó un formulario en el que se presenta la tabla respectiva de resultados obtenidos, en esta se incluyen los valores K correspondientes a los puntos reales y los valores que corresponden a los percentiles 2,5 y 97,5 de los valores K correspondientes a los grupos de puntos aleatorios. El

formulario también incluye un componente de gráficas de líneas de los valores K obtenidos, para esto se utilizó un componente embebido JFXPanel de JavaFx, el cual facilitó en gran medida la construcción de las gráficas. Un ejemplo de resultado de función K para redes se presenta en la figura 24.

Figura 24. Ejemplo de resultado de la función K para redes



Fuente El autor.

5.1.4. Módulo de NetKDE

El algoritmo para la implementación del método NetKDE parte de la creación de *lixels*, este proceso se realiza determinando coordenadas con base en cálculos de proporcionalidad con respecto a las longitudes y coordenadas de los segmentos de línea, estos cálculos incluyen la determinación de las coordenadas centrales de cada *lixel* denominadas *lxcenter*. Una vez se tienen identificados todos los *lixels* de la red con sus respectivos *lxcenter* se procede con la asignación de puntos (eventos) a cada uno de los *lixels*. En este punto del algoritmo se debe determinar cuáles *lixels* se encuentran dentro del ancho de banda de cada *lixel* que contenga puntos asignados, para esto se utiliza el algoritmo de determinación de distancia mínima a través de la red entre dos puntos teniendo en cuenta que en lugar de hallar distancias entre puntos de eventos se hará entre los *lxcenter* de los *lixels*. Los datos obtenidos de los cálculos realizados junto con las coordenadas de los *lixels* se almacenan en la base de datos relacional MySQL para su posterior representación gráfica.

Aunque la descripción anteriormente citada es más reducida que la dedicada al método de la función K para redes, esto no se debe a que este método sea más sencillo, en realidad es más

extenso. La diferencia en la extensión de la presentación de resultados se debe en buena medida a que los dos procesos comparten algoritmos complejos como el del cálculo de distancias a través de la red y las normalizaciones de datos detalladas en la sección 4.1.2. Otra razón de la aparente falta de complejidad es que la representación gráfica de los resultados se cubre en la siguiente sección.

5.1.5. Módulo de Visualización de Resultados Geográficos

Los datos geográficos a visualizar se dividieron en tres tipos: el primer tipo corresponde a la representación gráfica de las redes de líneas; el segundo se centra en la representación gráfica de los puntos (eventos) incluyendo también la representación gráfica de la red correspondiente; y el tercer tipo hace referencia a la representación gráfica de los resultados de NetKDE. Para los tres tipos de gráficas se implementó un único componente basado en la clase Graphics2D de Java⁸, que se incluye en formularios específicos con datos y componentes adicionales propios de cada tipo de dato geográfico.

Los tres tipos de datos incluyen la representación de líneas definidas por diferentes sistemas de coordenadas por lo que comparten elementos comunes. Para la representación gráfica de estos grupos de líneas es necesario realizar la conversión de su sistema de coordenadas a un sistema que tenga en cuenta la resolución de la pantalla del dispositivo utilizado, en todos los casos se aplica una conversión intermedia que convierte las coordenadas a un sistema de coordenadas planas en metros, y estas coordenadas intermedias son transformadas en coordenadas de píxeles en pantalla. Como guía y complemento se incluye un componente gráfico de escala en metros o kilómetros dependiendo de las dimensiones reales de los datos representados. Un ejemplo de visualización de puntos (eventos) se presenta en la figura 25.

También se incluye en el componente el manejo de eventos de ratón para permitir el desplazamiento (eventos de arrastrar y soltar), acercamiento y alejamiento (eventos de rueda del ratón) a través de la representación gráfica de los diferentes tipos de datos y la posibilidad de incluir datos propios de la red (etiquetas) en la gráfica.

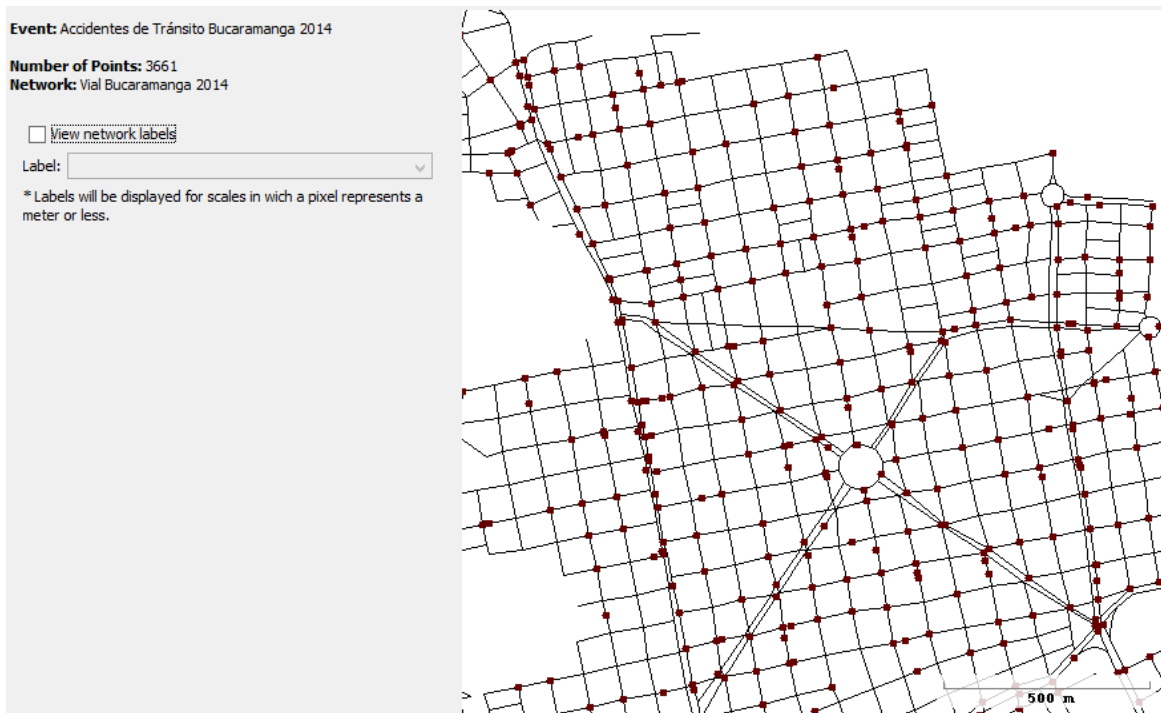
De los tres tipos de datos el que presentó más retos para su representación gráfica fue el de resultados de NetKDE, en este no solo era necesario presentar la red de datos, también se requería la representación de los diferentes valores de densidad a través de la red y de los puntos calientes de concentración de puntos. Para la representación de los valores de densidad se utilizó una escala de 10 colores que van desde el amarillo hasta el rojo oscuro⁹ y que representan los

⁸ <https://docs.oracle.com/javase/8/docs/api/java/awt/Graphics2D.html>

⁹ Los códigos hexadecimales de los colores son #FFFF00 y #AA0000 respectivamente.

deciles de las densidades calculadas desde las más bajas hasta las más altas. Los colores fueron determinados por medio de un componente desarrollado dentro de este proyecto que implementa el modelo de color HSV (Agoston, 2005). Para los *lixels* con densidades iguales a cero se utiliza el color negro. Para complementar la información representada por estos colores se incluyó una leyenda de datos que incluye los colores junto a los respectivos rangos de valores de densidad que representan.

Figura 25. Ejemplo de visualización de eventos (puntos)



Fuente El autor.

Para la representación gráfica de los puntos calientes se optó por la utilización de una sombra de color rojo pálido que ayuda a resaltar los límites de cada punto caliente. Un ejemplo de resultado de NetKDE que incluye una leyenda de datos y puntos calientes resaltados se presenta en la figura 26.

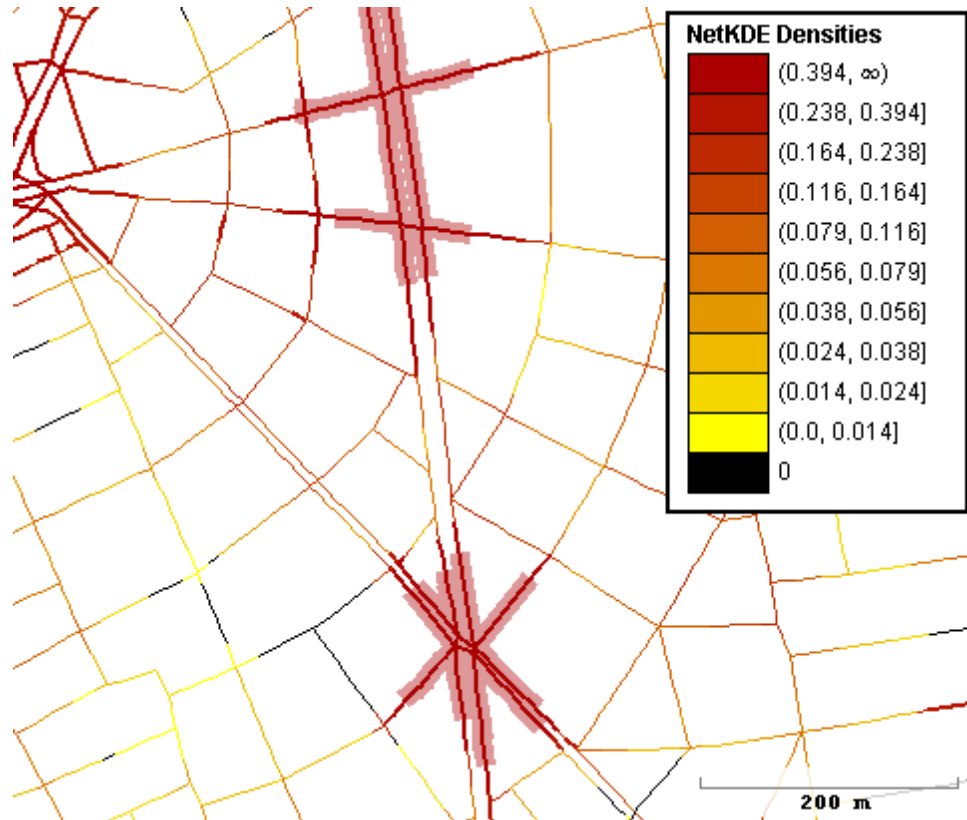
Es importante resaltar que la definición del método NetKDE no incluye una medida de identificación de puntos caliente, solamente los define como puntos en los que la densidad calculada es mayor a las densidades de las zonas circundantes; esto es debido a que la definición de los puntos calientes guarda una relación directa con el tipo de fenómeno representado. Debido a esto se incluyó un componente de deslizador (slider) graduado entre 0 y 100 que representa y modifica el porcentaje de valores de densidad resaltados que representan los puntos calientes, el valor de resalte por defecto es del 5%. Por último, se incluyó en el formulario la opción de ver dos

resultados simultáneamente en pantalla, con esta funcionalidad se pueden realizar comparaciones gráficas de resultados obtenidos de un mismo fenómeno con filtros de valores diferentes.

5.1.6. Módulo de Creación y Verificación de Base de Datos Relacional Asociada

Durante las primeras iteraciones del ciclo de desarrollo del presente proyecto no se contempló la necesidad de incluir un componente de creación de la base de datos relacional asociada, se asumió que el proceso de instalación debía incluir la creación de la base de datos directamente en el motor por medio de un script SQL. Pero una vez finalizada la implementación de las funcionalidades principales se tomó la decisión de construir un módulo que se encargara de esta tarea, además de la tarea de verificación de conexión, tareas necesarias para la detección y corrección de posibles errores de portabilidad y conexión. Este módulo se encarga de comprobar y actualizar en un archivo de propiedades las credenciales de conexión a un motor de bases de datos MySQL, además de ejecutar el script de creación de base de datos cuando sea requerido.

Figura 26. Ejemplo de resultados de NetKDE



Fuente El autor.

5.2. ANÁLISIS DE RESULTADOS DE LA APLICACIÓN DE LOS MÉTODOS IMPLEMENTADOS

Aunque los dos métodos implementados se enfocan en temas diferentes dentro de un mismo campo del conocimiento, el análisis espacial de patrones de puntos, y en este caso también dentro de una misma problemática, los accidentes de tránsito, su aplicación es complementaria. En general se recomienda realizar inicialmente la aplicación de la función K para redes, y si los resultados obtenidos indican que los eventos (puntos) tienden a formar grupos entonces se puede aplicar el método NetKDE.

A continuación, se presenta una interpretación a los resultados generales y de algunas vistas de datos consideradas relevantes, un análisis más detallado sería labor de un experto o grupo de expertos en temas viales o de movilidad.

5.2.1. Función K para Redes

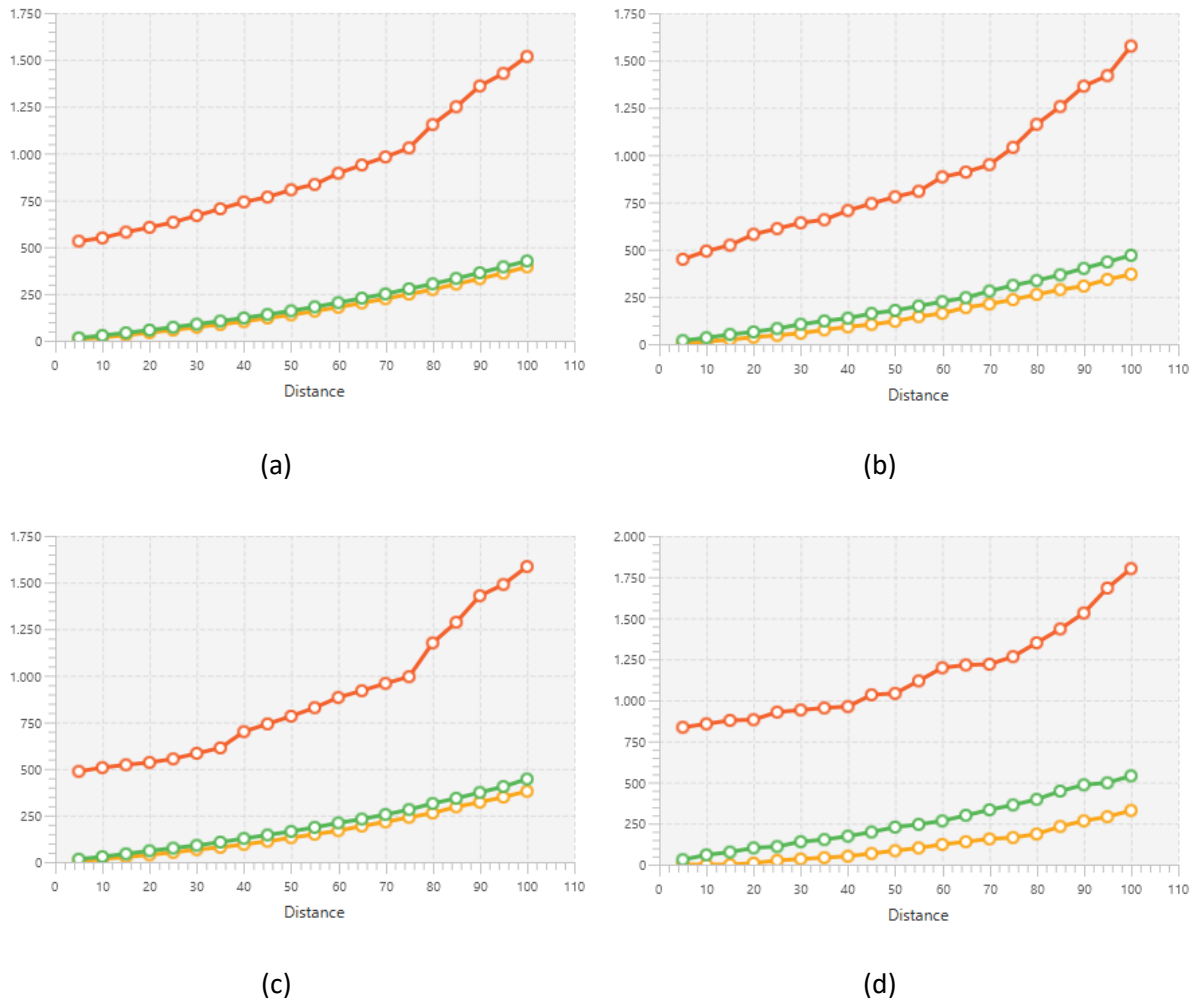
Los valores de la función K para redes dan una medida de la tendencia de los datos distribuidos a lo largo de una red a agruparse. Su valor en sí mismo no tiene un significado, para poder sacar conclusiones al respecto este valor se debe comparar con valores de la función calculados sobre datos aleatorios. En general el procedimiento de cálculo consiste en generar una gran cantidad de grupos de datos aleatorios sobre la misma red, cada uno de estos grupos debe tener la misma cantidad de puntos que el grupo de datos original; posteriormente se calcula el valor K para cada grupo de datos y se selecciona un intervalo de confianza de los resultados aleatorios que va desde el percentil 2,5 hasta el percentil 97,5; si el valor K correspondiente al grupo de datos reales se encuentra dentro del intervalo de confianza se concluye que los datos tienden a distribuirse de forma aleatoria, si por el contrario el valor K se encuentra por encima del intervalo de confianza entonces se concluye que los datos tienden a agruparse, por último, si el valor K se encuentra por debajo del intervalo de confianza se concluye que los datos tienden a dispersarse a través de la red de forma más o menos uniforme.

En la figura 27 se muestran los resultados gráficos de la función K para redes de los datos de accidentes de tránsito de la ciudad de Bucaramanga durante el año 2014 así como algunas vistas relevantes de los mismos datos. Los valores se calculan para distancias entre 5 y 100 metros con intervalos de 5 metros. En todas las gráficas el valor K se representa en color rojo, el límite inferior del intervalo de confianza en amarillo y el límite superior en verde. La gráfica (a) muestra los resultados al utilizar en conjunto completo de datos de accidentes de tránsito, la gráfica (b) corresponde a los accidentes ocurridos entre las 6:00 pm y las 6:00 am, la gráfica (c) corresponde a los accidentes en los que se vieron involucradas motocicletas y la gráfica (d) corresponde a los accidentes ocurridos durante el mes de diciembre.

En los cuatro casos representados se observa una fuerte tendencia de los datos a agruparse, incluso para distancias máximas de 5 metros. Dado que los datos de accidentes de tránsito de la

ciudad de Bucaramanga se ubicaron directamente sobre la cartografía de la red vial, ya fuera por medio de un algoritmo o de forma manual, a partir de las direcciones de ocurrencia suministradas por la dirección de tránsito municipal, las cuales en muchos casos correspondían a intersecciones entre vías, se presentó gran cantidad de casos en los que dos o más accidentes compartían exactamente las mismas coordenadas; lo anterior puede tener un efecto en los valores K reales sobre todos para distancias pequeñas.

Figura 27. Resultados de la función K para redes – Bucaramanga 2014



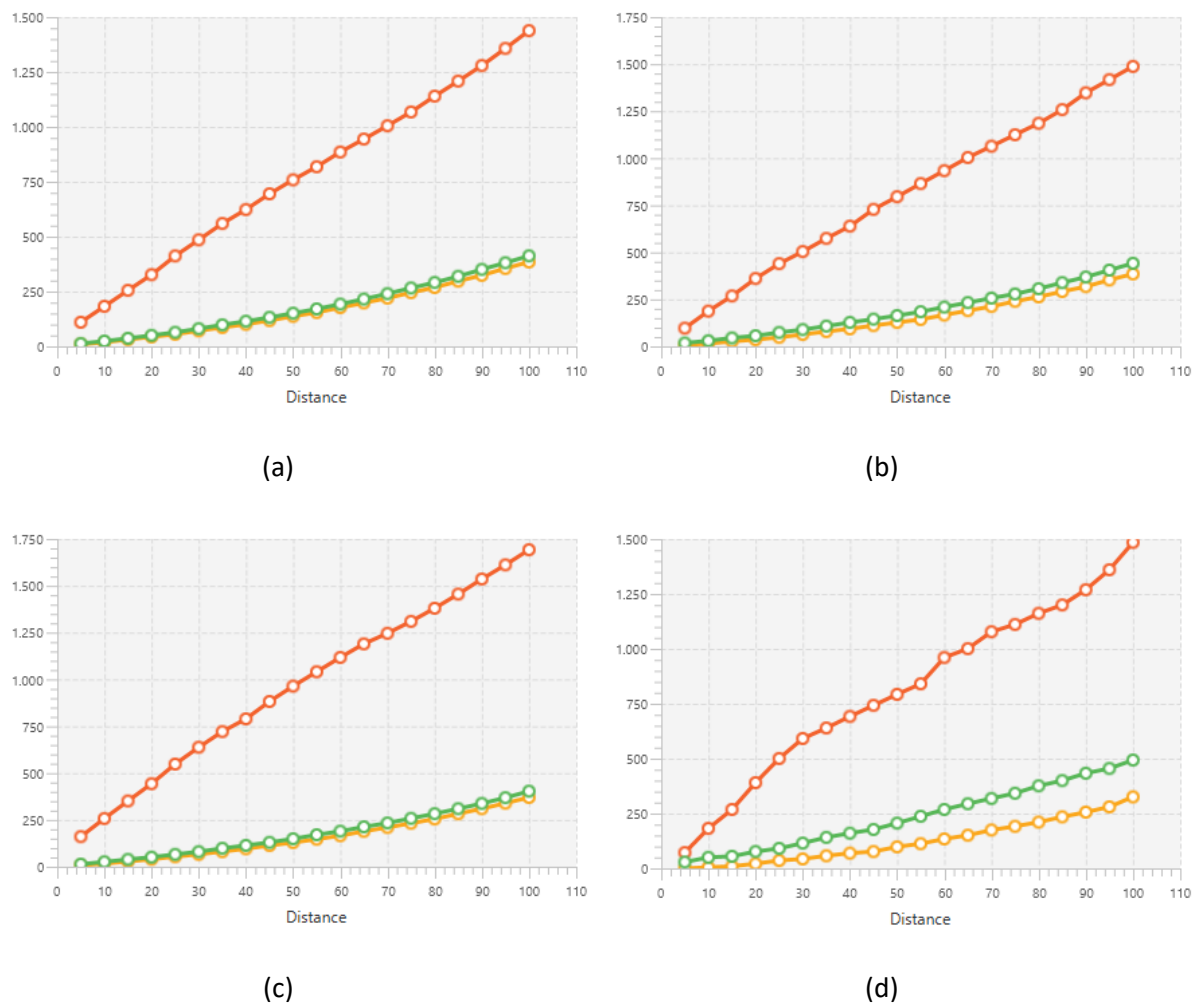
Fuente El autor.

Para descartar esta posible influencia sobre los valores K se aplicó la función general y las diferentes vistas a los datos de accidentes de tránsito de la ciudad de Barcelona ocurridos durante 2014 (ver figura 28). A diferencia de los datos de la ciudad de Bucaramanga, las ubicaciones de

estos accidentes de tránsito fueron recogidas directamente con dispositivos GPS en lugar de ser posicionadas directamente sobre la cartografía de la red vial correspondiente.

Al comparar los grupos de datos se observa que los valores K de los datos de la ciudad de Barcelona están más cerca del intervalo de confianza para distancias pequeñas que los datos correspondientes de la ciudad de Bucaramanga, por lo que se puede concluir que la forma de ubicación de los puntos tiene un impacto importante. A pesar del efecto mencionado se observa que el segundo grupo de resultados presenta la misma tendencia fuerte al agrupamiento incluso desde las distancias menores, aunque de una forma menos marcada en estas.

Figura 28. Resultados de la función K para redes – Barcelona 2014



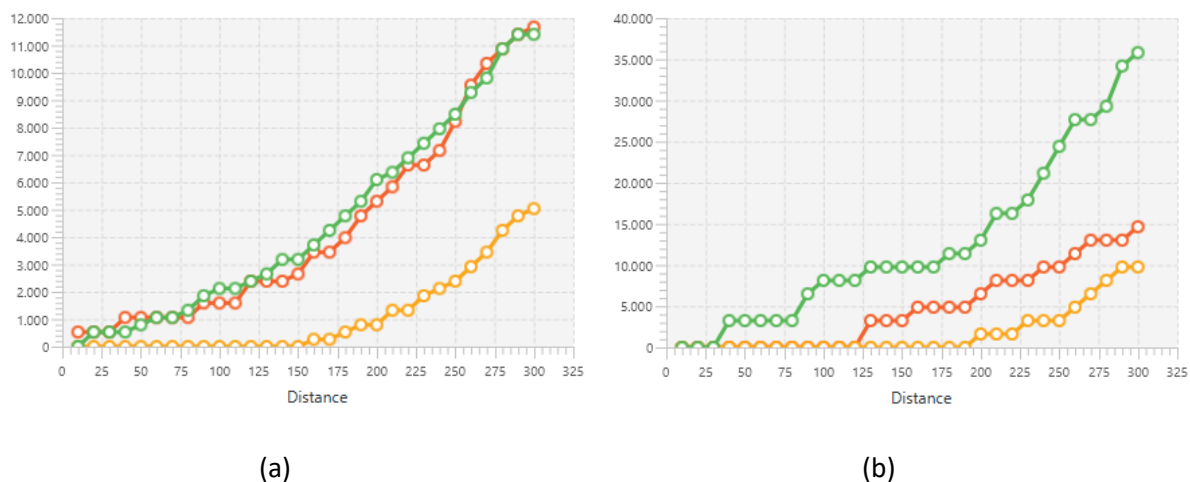
Fuente El autor.

Aunque en la gran mayoría de los grupos de datos a los que se aplicó la función K para redes hubo una tendencia clara hacia el agrupamiento, se encontró una excepción notable en lo referente a

accidentes de tránsito con víctimas mortales. Tanto para los datos de la ciudad de Bucaramanga como para los de la ciudad de Barcelona los resultados no mostraron una tendencia fuerte a agruparse, de hecho, los valores obtenidos para los datos de Bucaramanga fueron muy similares a los correspondientes valores del límite superior del intervalo de confianza. Buscando dar mayor claridad a los datos calculados se extendió el cálculo de los mismos hasta una distancia de 300 metros y para una mayor claridad en las gráficas de resultados utilizaron valores de distancias cada 10 metros. Los resultados se presentan en la figura 29, la gráfica (a) representa los valores K para la ciudad de Bucaramanga y la gráfica (b) los valores de la ciudad de Barcelona.

Una posible explicación a los resultados obtenidos con respecto a los accidentes de tránsito con víctimas mortales es que el número total de datos manejados es muy inferior al de los otros escenarios analizados, en la ciudad de Bucaramanga se presentaron 42 casos que corresponden al 1,1% del total y en la ciudad de Barcelona se presentaron 29 casos que representan el 0,3% del total. En la tabla 7 se presentan los números de accidentes de tránsito utilizados en cada grupo de resultados.

Figura 29. Resultados de la función K para redes – Accidentes de tránsito con víctimas mortales



Fuente El autor.

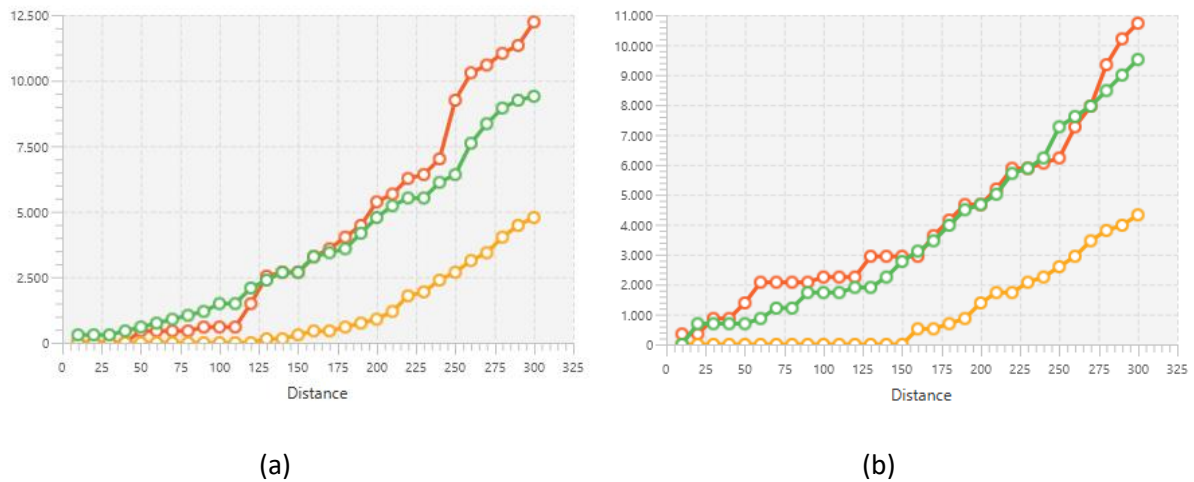
Tabla 7. Número de accidentes de tránsito por resultado

Grupo de datos	Bucaramanga		Barcelona	
	Accidentes	%	Accidentes	%
Completo	3.661	100,0	8.615	100,0
De 6:00 pm a 6:00 am	1.048	28,6	2.775	32,2
Con motocicletas involucradas	1.808	49,4	4.801	55,7
Mes de diciembre	333	9,1	756	8,8
Con víctimas mortales	42	1,1	29	0,3

Fuente El autor.

Para tratar de determinar la influencia que puede tener el tamaño del grupo de dato analizado sobre los resultados se realizaron dos cálculos con tamaños de grupo similares a los de accidentes con víctimas mortales sobre los datos de la ciudad de Bucaramanga. El primer grupo corresponde a los accidentes de tránsito ocurridos durante la semana del 9 al 15 de junio de 2014 entre las 6:00 pm y las 6:00 am, con un total de 56 accidentes; el segundo grupo está conformado por los accidentes de tránsito ocurridos durante el mes de diciembre entre las 6:00 pm y las 6:00 am que involucraban motocicletas, con un total de 52 accidentes. Para los dos grupos se utilizaron distancias entre los 10 y los 300 metros con incrementos de 10 metros. Los valores K de estos dos grupos de datos se presentan en la figura 30. Los grupos antes descritos se representan en las gráficas (a) y (b) respectivamente.

Figura 30. Resultados de la función K para redes – Grupos de datos reducidos



Fuente El autor.

Los resultados de estos dos últimos grupos también se presentan cercanos al límite superior del intervalo de confianza, aunque con características diferentes. El primer grupo se encuentra claramente dentro del intervalo de confianza para valores de distancias menores o iguales a los 120 metros y es claramente superior a partir de los 250 metros. El segundo grupo presenta valores claramente superiores entre los 50 y 80 metros y para valores a partir de los 280 metros. Aunque los resultados no pueden considerarse concluyentes se puede inferir que el tamaño del grupo de datos analizado tiene una influencia importante sobre los resultados de la función K para redes.

5.2.2. NetKDE

Los resultados obtenidos por medio de la aplicación del método NetKDE ayudan a identificar de forma clara y precisa dentro de una red los lugares en los que se presentan mayores concentraciones de los eventos analizados. Es importante tener en cuenta que, si la distribución

de los datos utilizados es aleatoria, aunque puedan identificarse altas concentraciones de eventos denominadas puntos calientes, estas no podrán ser atribuidas a condiciones geográficas específicas o características de las redes, en su lugar se deberán al azar.

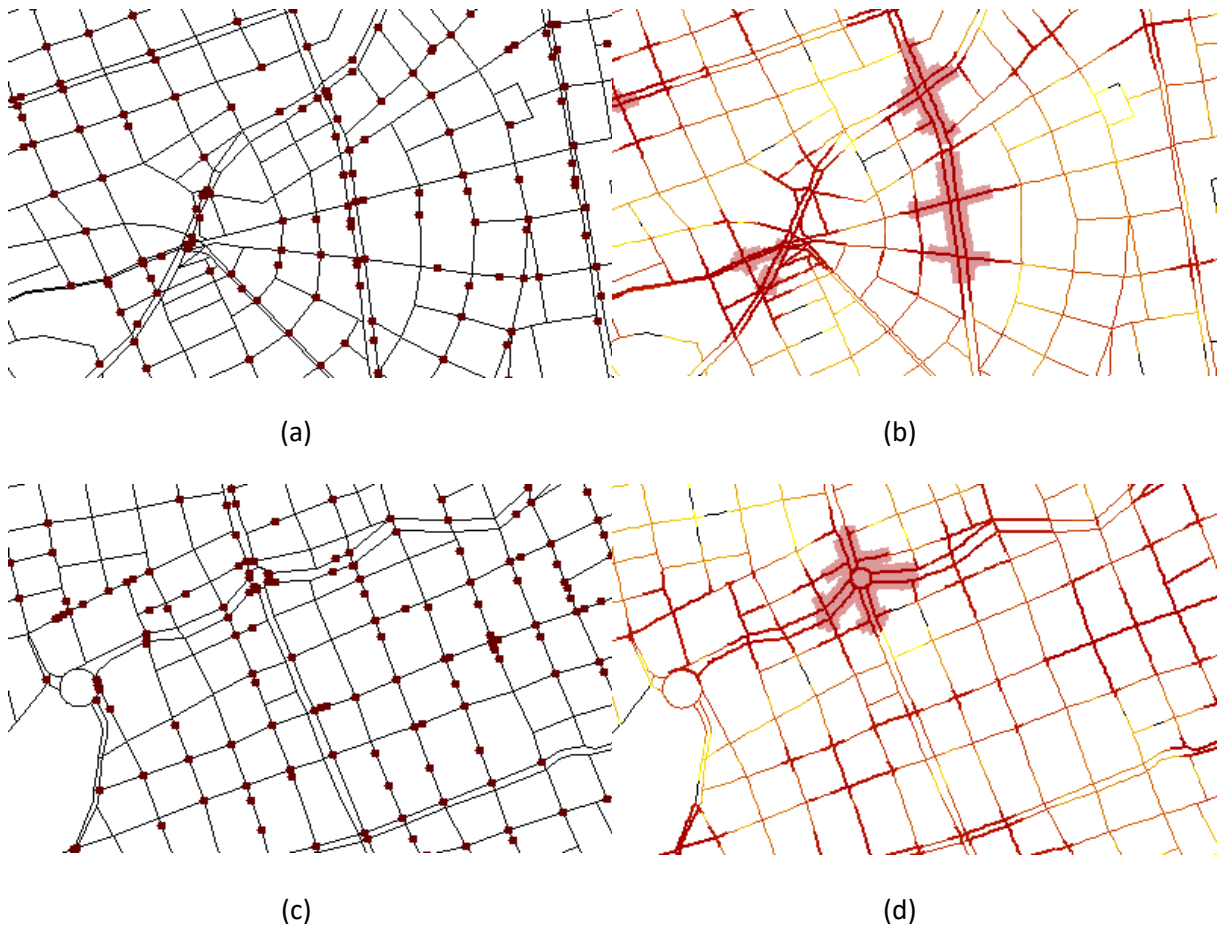
Una vez comprobado por medio de la aplicación de la función K para redes que los datos de accidentes de tránsito ocurridos durante el año 2014 en la ciudad de Bucaramanga no se distribuyen a través de la red vial de forma aleatoria y por el contrario tienden a agruparse, se procedió a aplicar el método NetKDE al conjunto total de datos y a las vistas específicas detalladas en la sección 4.2.1. No se incluyen las vistas de accidentes con víctimas mortales debido a que la aplicación de la función K para redes no arrojó resultados concluyentes con respecto a la tendencia de estos eventos a formar grupos. Tampoco se presenta un análisis de los resultados correspondientes a los datos de la ciudad de Barcelona porque las interpretaciones de los mismos son muy similares a las de los datos de la ciudad de Bucaramanga.

En la figura 31 se muestran dos ejemplos de comparación de gráficas de puntos con resultados de NetKDE, las gráficas de puntos (a) y (c) se corresponden con las gráficas de resultados (b) y (d) respectivamente. En los dos ejemplos se puede observar que en (a) y (c) no es posible determinar a simple vista en donde se encuentran las mayores concentraciones de puntos, esto se debe principalmente a que en algunos de los casos dos o más puntos ocupan la misma posición o posiciones muy cercanas, lo que impide contar con exactitud el número de puntos que hay en cada lugar o incluso en algunos casos malinterpretar la información presentada creyendo identificar las concentraciones mayores en lugares incorrectos. En las gráficas (b) y (d) por el contrario los lugares de mayor concentración, los puntos calientes, están claramente delimitados y con la ayuda de la nomenclatura de direcciones propia de cada lugar es posible describirlos con precisión.

La descripción de los puntos calientes identificados en la ciudad de Bucaramanga se detalla a continuación, en la figura 32 se muestran estos puntos calientes divididos en 4 grupos que recorren la ciudad de norte a sur.

A continuación, se presenta una comparación general entre la vista de los accidentes de tránsito ubicados como puntos sobre la red vial y la vista general de resultados de NetKDE. Después se presentan los resultados generales obtenidos para los datos de la ciudad de Bucaramanga, seguidos de las comparaciones de estos resultados generales con los resultados correspondientes a las vistas de datos de accidentes entre las 6:00 pm y las 6:00 am, accidentes que involucraron motocicletas y accidentes ocurridos durante el mes de diciembre. Para todos los resultados analizados se aplicó como función de núcleo la distribución de Epanechnikov (ver sección 1.4.1), el ancho de banda seleccionado fue de 100 metros, la longitud de *lixel* de 10 metros y se tomó como medida para determinar los puntos calientes los *lixels* con densidades iguales o superiores al percentil 98 (2% más altas).

Figura 31. Comparativo entre puntos de accidentes y resultados de NetKDE



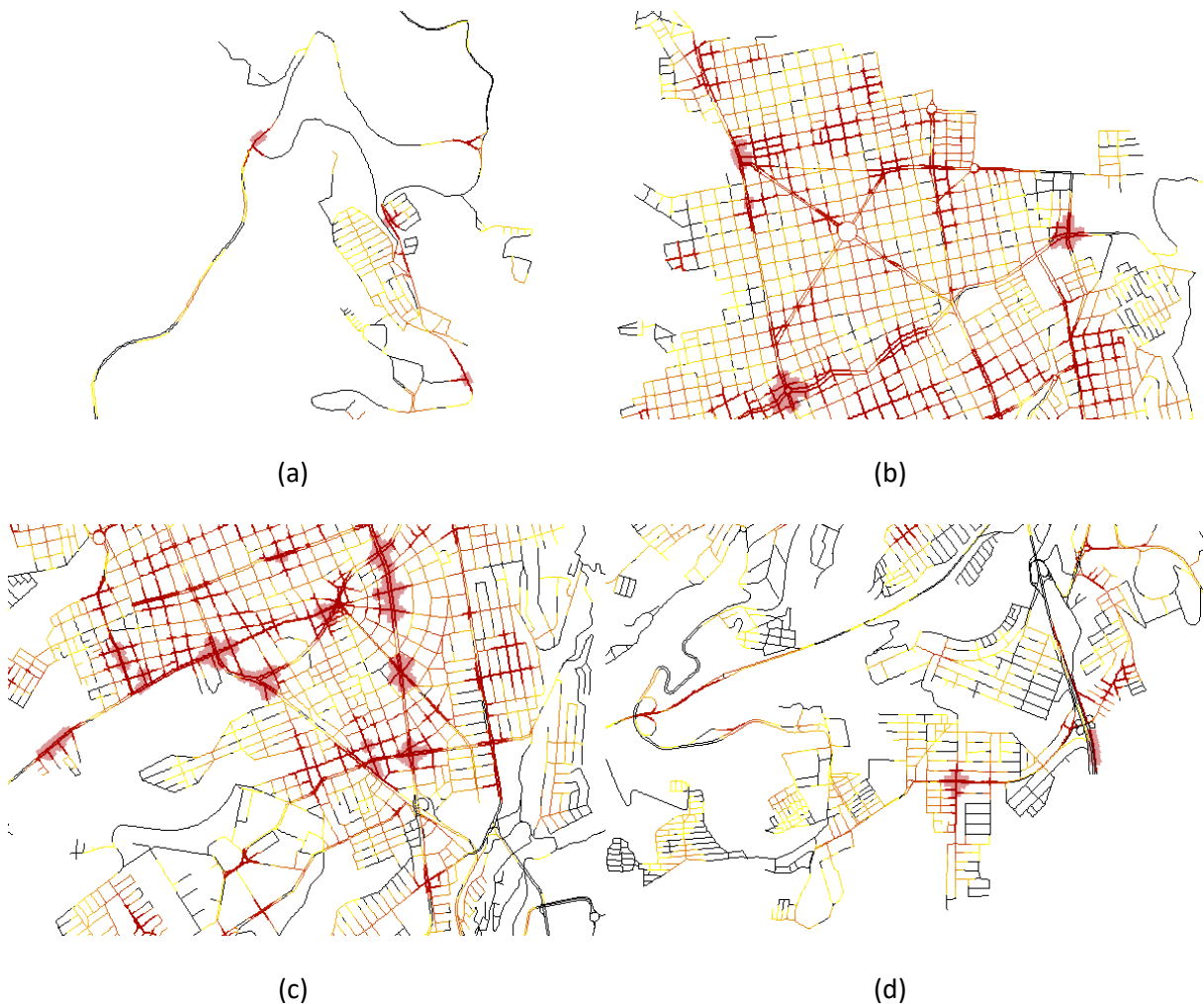
Fuente El autor.

El grupo (a) muestra 3 puntos calientes, el primero ubicado sobre la vía a Chimitá desde los 40 hasta los 80 metros al norte del cruce con la vía al Café Madrid. El segundo se encuentra en la salida del barrio Olas Bajas en el cruce con la vía a Rionegro. El tercer punto se encuentra en el cruce de la avenida Libertador con la calle 5N. En este primer grupo todos los puntos calientes se ubican sobre o muy cerca de cruces entre vías de tráfico de transporte de carga pesada con vías secundarias que conducen a zonas residenciales.

El grupo (b) presenta 4 puntos calientes, el primero, bastante extenso, ubicado en entre las calles 10 y 12 con carrera 15, para esta última en ambos sentidos del tráfico. El segundo se encuentra ubicado en la intersección de la calle 14 con carrera 15 en sentido sur-norte. El tercer punto, también bastante extenso, se ubica en la intersección de la carrera 33A y la avenida Quebradaseca, legando en su extremo norte más allá de la calle 19, y adentrándose tanto en la avenida Quebradaseca hacia el oriente como hacia el occidente y en la carrera 33A hacia el sur cerca de 40 metros en todos los casos. El cuarto punto, aún más extenso, se ubica en la glorieta

de la avenida Quebradaseca con carrera 15 y abarca las calles adyacentes tanto hacia el norte como hacia el sur, oriente y occidente. En este caso solamente los puntos 1 y 3 están relacionados con vehículos de transporte de carga pesada; todos los puntos se ubican sobre vías principales en lugares comerciales.

Figura 32. Resultados de NetKDE – Bucaramanga 2014



Fuente El autor.

El grupo (c) es el que presenta mayor cantidad de puntos calientes, un total de 15. El primer punto se encuentra ubicado en la intersección de la carrera 21 con calle 36. El segundo punto, también de gran tamaño, se encuentra en la intersección de la carrera 27 con avenida la Rosita y abarca las calles adyacentes. El tercer punto, el de mayor extensión en la ciudad, se encuentra muy cercano al anterior en la carrera 27 con calles 42, 45 y 48, llegando por la calle 45 hasta la carrera 26. El cuarto punto se encuentra en el triángulo formado por la calle 45 con avenida la Rosita y la carrera 22. El quinto punto, también de gran extensión, abarca la calle 45 entre carreras 14A y 16 y llega

por el norte hasta la calle 43. El sexto punto, bastante cercano al anterior y también de gran extensión corresponde a la intersección entre la diagonal 15 y la avenida la Rosita, llegando por el oriente hasta la carrera 17. El séptimo punto se ubica en la calle 45 entre carreras 9 y 10. El octavo punto se encuentra en la intersección entre la carrera 10 y la calle 44. El noveno punto se ubica en la intersección de la carrera 9 con calle 42. El décimo punto se encuentra en la intersección de la carrera 9 con calle 43. El punto número 11, también extenso, se ubica a lo largo de la calle 45 entre carreras 2 y 3, llegando por el oriente casi hasta la carrera 4. El punto 12, también extenso, se ubica en la intersección entre la carrera 27 y la avenida González Valencia y se adentra en las dos vías en todos los sentidos. El punto número 13, bastante extenso, se ubica en la intersección entre la carrera 27 y la calle 56. El punto 14 se ubica en la intersección entre la carrera 21 y la calle 55. El punto 15, el último del grupo, se ubica en la carrera 21 entre la calle 56 y la diagonal 15. Todos los puntos de este grupo, con excepción del octavo punto, se ubican sobre vías de alto tráfico vehicular; todas las zonas son comerciales.

El grupo (d), que corresponde a la zona sur de la ciudad, solamente presenta 2 puntos calientes a saber, el primero se ubica sobre la autopista a Floridablanca sobre el costado oriental en sentido sur-norte 50 metros al sur del puente de Provenza; es importante resaltar que este punto se encuentra en el límite sur del municipio de Bucaramanga por lo que no se cuenta con datos de accidentes ocurridos en el límite norte del municipio de Floridablanca, lo que podría cambiar la configuración del punto caliente. El segundo y último punto se ubica en la intersección de carrera 22 con calle 105. En los dos casos se trata de vías de alto flujo vehicular y en el segundo caso se trata de la única zona residencial y comercial en la que se ubica un punto caliente.

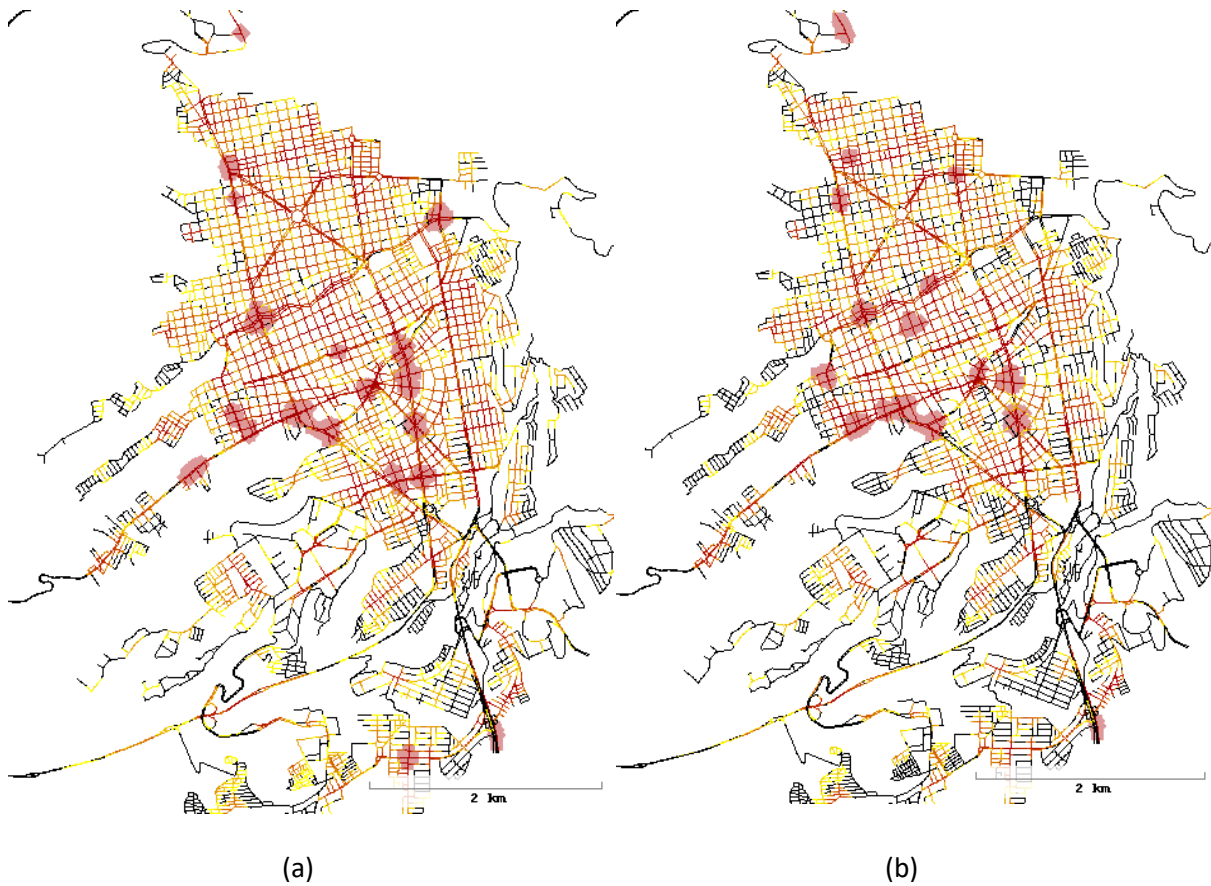
Como complemento a la descripción de los puntos caliente se puede agregar que las vías de ingreso a la ciudad como la vía a Chimitá, la vía a Cúcuta y la vía a Girón no presentan las mayores concentraciones de accidentes de tránsito; específicamente se puede apreciar sobre la vista de puntos que la vía a Girón, cerca del terminal de transportes, contiene una alta concentración de accidentes de tránsito, pero en comparación con otros lugares no se clasifica como punto caliente.

Al comparar la vista general presentada con la vista de accidentes ocurridos entre las 6:00 pm y las 6:00 am, se observa que en general los puntos calientes se corresponden, con las notables ausencias de los puntos calientes ubicados sobre la calle 45 hacia el occidente y la carrera 9, al igual que los puntos ubicados hacia el norte de la ciudad. Como novedad aparecen dos nuevos puntos calientes, el primero ubicado en el cruce entre la carrera 28 y la calle 15, zona estudiantil y residencial. El segundo punto caliente se ubica en la intersección entre la carrera 36 y la calle 35, zona residencial y comercial.

La comparación entre la vista general de resultados y la vista de accidentes que involucran motocicletas se presenta en la figura 33, la gráfica (a) corresponde a los resultados generales y la gráfica (b) a los accidentes con motocicletas involucradas. Los dos resultados presentan bastantes diferencias individuales con la tendencia general de los accidentes del grupo (b) concentrados

principalmente en lo que se denomina zona de meseta y no presentando puntos calientes en las zonas periféricas.

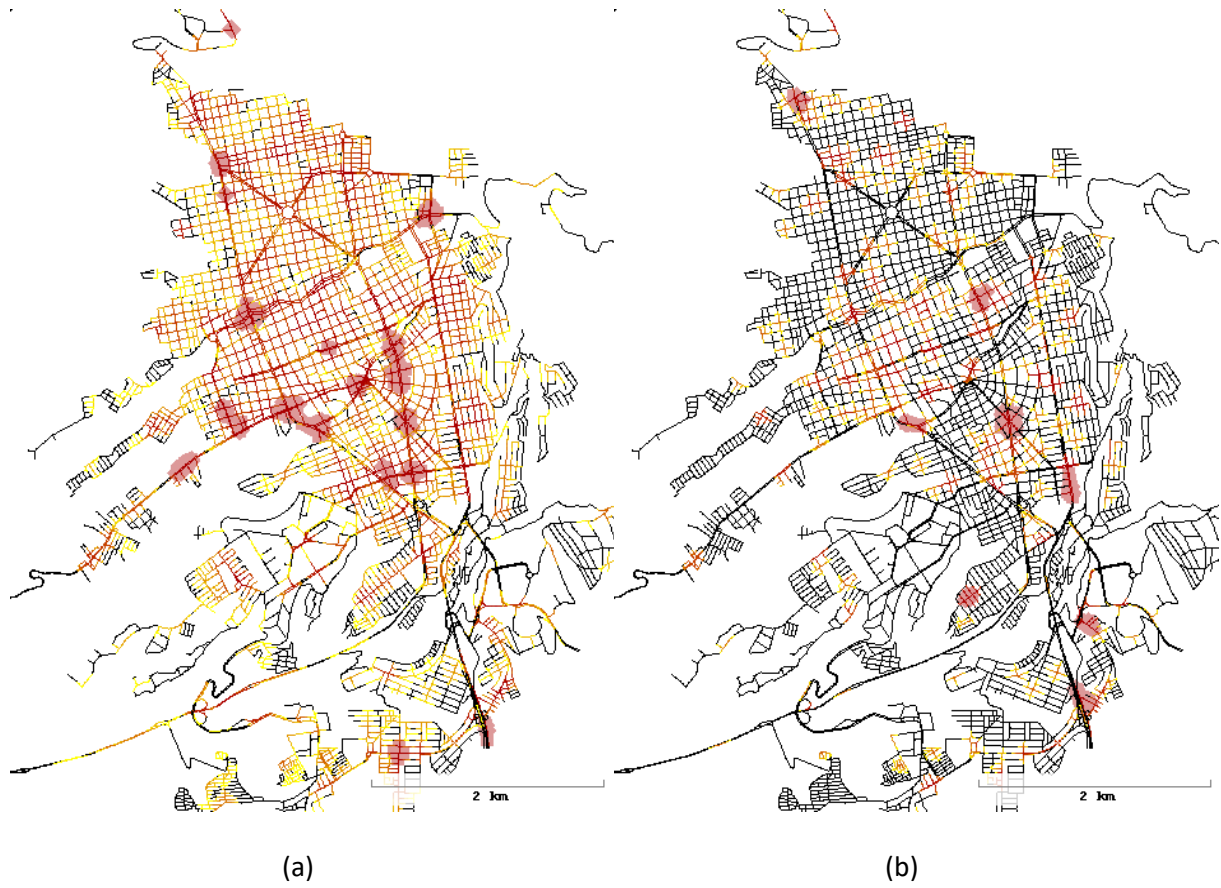
Figura 33. Comparación entre resultados generales de NetKDE y resultados de accidentes con motocicletas involucradas



Fuente El autor.

Por último, la comparación de la vista general con la vista de accidentes ocurridos durante el mes de diciembre también presenta diferencias sustanciales. En la figura 34 se presenta gráficamente esta comparación, la gráfica (a) corresponde a los resultados de la vista general y la gráfica (b) a los accidentes de tránsito ocurridos en el mes de diciembre. El grupo (b) solamente presenta 8 puntos calientes, 6 de los cuales no están presentes en la vista general. Uno de los puntos calientes, el ubicado en la intersección entre la carrera 13 y la calle 67 del barrio la Victoria, se encuentra en una zona residencial sobre una vía de tráfico medio; descripción que no se corresponde con la de los demás puntos calientes. A priori se puede inferir que durante el mes de diciembre los hábitos de transporte de muchos de los habitantes de la ciudad cambian debido a que es época de vacaciones escolares y en muchos casos también laborales, pero sería útil un análisis mes a mes para determinar si puede existir una explicación diferente a este comportamiento particular.

Figura 34. Comparación entre resultados generales de NetKDE y resultados de accidentes ocurridos durante el mes de diciembre



Fuente El autor.

6. CONCLUSIONES

Los esfuerzos y estrategias concentrados en la disminución en la ocurrencia de los accidentes de tránsito han tomado gran protagonismo en los últimos años. En septiembre de 2015 la Asamblea General de las Naciones Unidas adoptó la Agenda 2030 para el Desarrollo Sostenible (Organización de las Naciones Unidas, 2015), que en su objetivo 3.6 plantea la búsqueda de la disminución a la mitad en el número de muertes y lesiones causadas por accidentes de tránsito en el mundo para el año 2020. Este objetivo obedece en buena medida al gran impacto que tienen estos eventos sobre la calidad de vida de las personas; la Organización Mundial de la Salud (OMS) en su informe sobre la situación mundial de la seguridad vial publicado en 2015 (World Health Organization, 2015a) pone de manifiesto que los accidentes de tránsito son la principal causa de mortalidad en el mundo en personas entre 15 y 29 años por encima incluso de los suicidios, los homicidios y el VIH/SIDA. En este informe también se afirma que, aunque el número de muertes causadas por accidentes de tránsito se ha estabilizado en los años recientes (alrededor de 1.250.000 muertes por año en el mundo), en 68 países éstas han aumentado y de estos el 84% corresponde a países de ingresos bajos y medios; además las tasas de mortalidad en países de bajos ingresos son muy superiores a las tasas en países con ingresos altos.

La OMS en una nota descriptiva publicada en noviembre de 2016 sobre lesiones causadas por el tránsito (World Health Organization, 2015b) da algunas cifras adicionales que ponen de manifiesto el gran impacto que puede llegar a tener una mejor comprensión de la forma en que se presentan los accidentes de tránsito, estimando que de no tomarse medidas para su disminución, en 2030 estos serán la séptima causa general de muerte. Además del impacto en la salud y la calidad de vida de las personas víctimas de estos accidentes y de sus familias, en este informe se muestra que por medio de una investigación llevada a cabo en el año 2010 se estimó que los accidentes de tránsito cuestan a los países cerca del 3% de su producto interno bruto, llegando al 5% en algunos países de bajos ingresos.

La situación en Colombia no difiere de la situación mundial en general, en nuestro país la tasa de mortalidad por accidentes de tránsito se ha estabilizado en los últimos años, en la tabla 8 se presentan los totales de defunciones y tasas de mortalidad para los años 2005 a 2014. Al comparar estos datos con la información presentada en el informe de la OMS mencionado anteriormente se observa que la tasa de mortalidad por accidentes de tránsito en Colombia (14,0) es inferior a la tasa mundial (17,4) y a la tasa de las américas (15,9) para el año 2013 pero es muy superior a la tasa europea (9,3) para el mismo año.

De aquí se puede concluir que estos datos de mortalidad, aunque no son especialmente altos al ser comparados con los datos del resto del mundo, sí presentan la oportunidad de ser mejorados de forma sustancial. La disminución sistemática del número de accidentes de tránsito requiere de la participación de diferentes grupos de interés como las autoridades municipales, regionales y nacionales, las agremiaciones de transportadores y conductores en general, las instituciones

educativas y de investigación y la ciudadanía en general. El presente trabajo presenta una alternativa, apoyada en tecnologías de la información, para la comprensión de la forma en que los accidentes de tránsito se distribuyen a lo largo de la red vial de un municipio o región. De la aplicación de la herramienta desarrollada se obtuvo información que puede ayudar a comprender esta problemática a nivel local, aunque se recomienda su aplicación a grupos de datos de otros años y en general diferentes estrategias de recolección y organización de la información que se presentan a continuación en este capítulo y en el capítulo posterior.

Tabla 8. Defunciones por accidentes de tránsito en Colombia (2005-2014)

Año	Número de defunciones	Tasa por 100.000 habitantes
2005	6.149	14,3
2006	6.426	14,8
2007	6.722	15,3
2008	6.554	14,7
2009	6.626	14,7
2010	6.115	13,4
2011	5.918	12,9
2012	6.359	13,7
2013	6.610	14,0
2014	6.813	14,3
Fuente Bodega de datos del SISPRO - Ministerio de Salud y Protección Social		

Las redes viales continuamente están siendo renovadas, modificadas y ampliadas por lo que la identificación y control de los lugares más propensos a la ocurrencia de accidentes de tránsito se convierte en una labor compleja. Este trabajo de identificación generalmente se realiza basado en la intuición y experiencia, o apoyándose en datos analizados por medio de procesos no formales. La estructuración de estrategias para la toma de decisiones en temas de prevención de accidentes de tránsito está por fuera del marco de este proyecto, pero se pueden realizar labores de apoyo que ayuden a este proceso. Una de estas labores es el proceso de actualización de la cartografía vial de la ciudad de Bucaramanga, que en los últimos años ha venido cambiando drásticamente en varios lugares; ejemplos de estos son el puente de la calle 9, la ampliación de la autopista a Floridablanca y el intercambiador del Mesón de los Búcaros. Como parte de este proyecto se realizó un proceso de actualización cartográfica de red vial de la ciudad de Bucaramanga al año 2014, pero esta labor debe ser llevada a cabo de forma continua.

Los procesos de normalización de datos implementados, el cierre de nodos de red y la proyección de puntos de eventos sobre la red, son un paso previo fundamental a la aplicación de métodos en los que sea necesario calcular distancias a través de redes. Aunque la calidad inherente a los datos cartográficos utilizados sea alta, no se debe prescindir de estos procesos, ya que en muchos de los

casos la construcción de estas estructuras de información no es llevada a cabo teniendo en cuenta que pueden ser utilizadas en procesos de análisis de patrones de puntos sobre redes.

La aplicación de la función K para redes a los datos de accidentes de tránsito de las ciudades de Bucaramanga y Barcelona ayudó a corroborar que la ocurrencia de estos accidentes presenta una fuerte tendencia a formar grupos desde el punto de vista geográfico y no obedece simplemente a procesos aleatorios. Se comprobó además que diferentes vistas o subconjuntos de estos datos también presentan el mismo tipo de tendencia a agruparse, con la notable excepción de los accidentes con víctimas mortales, de los que no se obtuvieron resultados concluyentes; la explicación a esto último puede estar relacionada con el bajo número de datos utilizados. Para dar claridad a esta temática se recomienda llevar a cabo nuevos procesos de cálculo de valores K incluyendo datos de diferentes años para ampliar el número de eventos analizados. Es importante resaltar que un proceso de este tipo puede presentar otros problemas no contemplados con anterioridad en lo referentes a los cambios que se presentan en la red vial a través del tiempo.

Una vez comprobado que los accidentes de tránsito no ocurren al azar a través de la red vial, la búsqueda de los puntos en los que estos eventos tienden a concentrarse toma una importancia central. La estimación de estos puntos de concentración o puntos calientes, realizada directamente sobre la representación gráfica de los puntos de los eventos, no arroja resultados con límites precisos y en algunos casos puede producir identificaciones erróneas. Es por esto que la inclusión de métodos formales de detección de puntos calientes, como NetKDE, debe primar sobre la detección directa sobre los datos de ubicación de los eventos.

Dado que dentro de la metodología de NetKDE no se especifica una medida exacta de lo que se puede considerar un punto caliente, la inclusión de un componente de deslizador o *slider* que controle las cantidades de *lixels* identificados como parte de puntos calientes ayuda a visualizar de forma dinámica la forma en la que se distribuyen y tienden a concentrarse los accidentes dentro de la red vial. Así mismo la inclusión de una vista de comparación de resultados de NetKDE ayuda a comprender la forma en que dos vistas de un mismo evento se distribuyen a la vez que muestra las similitudes y diferencias que puedan existir entre estas.

Los resultados obtenidos al aplicar los dos métodos desarrollados sobre los datos de accidentes de tránsito ocurridos en el año 2014 en las ciudades de Bucaramanga y Barcelona dan una visión detallada de la forma en la que estos eventos se distribuyen a lo largo de la red vial de estas dos ciudades, previa comprobación de su tendencia a formar grupos. Este entendimiento de la distribución geográfica de los accidentes de tránsito, en manos de la entidad responsable de su atención, aplicado a la prevención de los mismos podría tener un impacto positivo al disminuir las frecuencias de ocurrencia de estos eventos especialmente en las zonas consideradas críticas. En general la aplicación de los métodos de análisis espacial apropiados sobre un fenómeno o evento específico pueden dar una perspectiva y un entendimiento más claros del mismo desde el punto

de vista geográfico; este nuevo conocimiento aunado a estrategias de intervención podría ayudar a cambiar la frecuencia y la distribución del evento o fenómeno que estudia.

7. RECOMENDACIONES

Los procesos de normalización de datos implementados fueron probados y optimizados, pero aún son demandantes en tiempo y capacidad de cómputo, por lo que se hace necesario un nuevo proceso de revisión y posible optimización de los mismos. Si bien las ciudades analizadas no son pequeñas, existen ciudades sensiblemente más grandes; en el caso específico de la ciudad de Bucaramanga, si se toma su población estimada al año 2016¹⁰ como medida indirecta del tamaño de su malla vial y se compara con la población de la ciudad de Bogotá estimada para el mismo año, proporcionalmente solo representaría el 6,6% del tamaño de esta última.

Otra temática relacionada directamente con la información cartográfica que sirve como base para la aplicación de los métodos de análisis de patrones de puntos implementados es la del manejo de diferentes sistemas de coordenadas. En el marco de desarrollo del presente proyecto se incluyó el manejo de coordenadas geográficas y un manejo general de sistemas de coordenadas planas basados en diferentes proyecciones cartográficas; una segunda fase de implementación de la herramienta software aquí presentada podría incluir soporte para proyecciones cartográficas específicas como la proyección de Mercator (Osborne, 2008) en sus diferentes representaciones.

La herramienta software implementada soporta el manejo de archivos KML, los cuales, aunque ampliamente difundidos, no son el tipo de archivo más común en cartografía digital, esta distinción le corresponde a los archivos SHP (ESRI Shapefile); por lo que se recomienda incluir en una siguiente fase de implementación soporte para este tipo de archivos.

Como se detalló en la sección 4.2.1, la ubicación de los datos de eventos (puntos) directamente sobre la cartografía de la red vial puede presentar un sesgo que posiblemente desembocará resultados no confiables de la función K para redes para distancias pequeñas, al ocasionar que en muchos casos dos o más puntos ocupen las mismas coordenadas; para prevenir la presencia de este tipo de errores se recomienda que la recolección de datos geográficos se realice por medio de dispositivos GPS.

Junto a NetKDE se han propuesto otros métodos basados en KDE para la detección de puntos calientes en redes, un ejemplo sobresaliente de este tipo de métodos es el incluido en la aplicación SANET (Okabe et al., 2015). Con el fin de ampliar y corroborar los resultados obtenidos tras la aplicación de NetKDE, se recomienda realizar un comparativo de los resultados arrojados por las implementaciones de estos dos métodos, lo que ayudará a establecer comparaciones de rendimiento y efectividad.

¹⁰ http://dane.gov.co/files/investigaciones/poblacion/proyepobla06_20/Municipal_area_1985-2020.xls

BIBLIOGRAFÍA

- Agoston, M. K. (2005). *Computer Graphics and Geometric Modelling* (1st ed.). Londres: Springer-Verlag.
- Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hot spots. *Accident Analysis & Prevention*, 41(3), 359–364.
- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Essex: LONGMAN.
- Beck, K., Beedle, M., Bennekum, A. van, Cockburn, A., Cunningham, W., Fowler, M., ... Thomas, D. (2001). Principles behind the Agile Manifesto. Recuperado de <http://agilemanifesto.org/principles.html>
- Black, W. R. (1991). Highway accidents: a spatial and temporal analysis. *Transport Research Record: Journal of the Transportation Research Board*, 75–82.
- Black, W. R. (1992). Network Autocorrelation in Transport Network and Flow Systems. *Geographical Analysis*, 24(3), 207–222.
- Centre for Environmental Research - UFZ. (2016). Programita. Recuperado de <http://www.ufz.de/index.php?en=41413>
- Clark, P. J., & Evans, F. C. (1954). Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology*, 35(4), 445–453.
- Cockburn, A. (2004). *Crystal Clear: A Human-Powered Methodology for Small Teams*. Upper Saddle River, New Jersey: Addison-Wesley Professional.
- Deenadayalan, Y., Grimmer-Somers, K., Prior, M., & Kumar, S. (2008). How to run an effective journal club: a systematic review. *Journal of Evaluation in Clinical Practice*, 14(5), 898–911.
- Delmelle, E., & Thill, J.-C. (2008). Urban bicyclists – a spatial analysis of adult and youth traffic hazard intensity. *Transportation Research Record: Journal of the Transportation Research Board*, 2074, 31–39.
- Dixon, P. M., El-Shaarawi, A. H., & Piegorsch, W. W. (2002). Ripley's K function. *Encyclopedia of Environmetrics*, 3, 1796–1803.
- Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Accident Analysis & Prevention*, 40(1), 174–181.

- Esri. (2016). ArcGIS Spatial Analyst. Recuperado de <http://www.esri.com/software/arcgis/extensions/spatialanalyst>
- Gatrell, A. C., Bailey, T. C., Diggle, P. J., & Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21(1), 256–274.
- Getis, A., & Ord., J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), 189–207.
- Gibin, M., Longley, P., & Atkinson, P. (2007). Kernel Density Estimation and Percent Volume Contours in General Practice Catchment Area Analysis in Urban Areas. In *Proceedings of the GIScience research UK conference (GISRUK)* (pp. 270–276). Recuperado de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.228.9937&rep=rep1&type=pdf#page=282>
- Google. (2016). KML Reference. Recuperado de <https://developers.google.com/kml/documentation/kmlreference>
- Harvey, D. W. (1966). Geographical Processes and the Analysis of Point Patterns: Testing Models of Diffusion by Quadrat Sampling. *Transactions of the Institute of British Geographers*, (40), 81–95.
- Krispa, J. M., & Durot, S. (2007). Segmentation of lines based on point densities – An optimisation of wildlife warning sign placement in southern Finland. *Accident Analysis & Prevention*, 39(1), 38–46.
- Levine, N. (2013). *CrimeStat IV: a spatial statistics program for the analysis of crime incident locations, version 4.0*. Washington D.C.: National Institute of Justice.
- Miller, H. J. (1999). Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geographical Analysis*, 31(4), 373–399.
- National Institute of Justice. (2015). CrimeStat: Spatial Statistics Program for the Analysis of Crime Incident Locations. Recuperado de <http://nij.gov/topics/technology/maps/pages/crimestat.aspx>
- Okabe, A., Okunuki, K., Satoh, T., Shiode, S., Okabe, K., Sugihara, K., ... Morita, M. (2015). SANET - Spatial analysis along networks. Recuperado de <http://sanet.csis.u-tokyo.ac.jp/>
- Okabe, A., Satoh, T., & Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science*, 23(1), 7–32.

- Okabe, A., & Yamada, I. (2000). The K-function method on a network and its computational implementation. *Geographical Analysis*, (33), 271–290.
- Organización de las Naciones Unidas. (2015). Proyecto de documento final de la cumbre de las Naciones Unidas para la aprobación de la agenda para el desarrollo después de 2015. *Africa*, 17852(10), 1–17.
- Osborne, P. (2008). *The Mercator Projection* (1st ed.). Edinburgh: Peter Osborne.
- Rosenberg, M. S., & Anderson, C. D. (2011). PASSaGE - Pattern Analysis, Spatial Statistics and Geographic Exegesis. Recuperado de <http://www.passagesoftware.net/index.php>
- Schabenberger, O., & Gotway, C. A. (2005). *Statistical methods for spatial data analysis*. Boca Raton, Florida: Chapman & Hall/CRC.
- Truong, L., & Somenahalli, S. (2014). Using GIS to Identify Pedestrian- Vehicle Crash Hot Spots and Unsafe Bus Stops. *Journal of Public Transportation*, 14(1), 99–114.
- World Health Organization. (2015a). *Global status report on road safety 2015*. Geneva. Recuperado de http://apps.who.int/iris/bitstream/10665/189242/1/9789241565066_eng.pdf?ua=1
- World Health Organization. (2015b). Road traffic injuries. Recuperado de <http://www.who.int/mediacentre/factsheets/fs358/en/>
- Xie, Z., & Yan, J. (2008). Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, 32(5), 396–406.
- Xie, Z., & Yan, J. (2013). Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport Geography*, 31, 64–71.

ANEXO A. ALGORITMO EN AVENUE PARA LA LOCALIZACIÓN DE DIRECCIONES

```
'Se obtiene la vista sobre la que se realizará la ubicación de
puntos
theView = av.GetProject.FindDoc("View1")
mTheme = theView.FindTheme("Vial_buc.shp")

def = av.GetProject.MakeFileName("theme", "shp")
def = FileDialog.Put(def, "*.shp", "New Theme")
if (def <> nil) then
    tbl = FTab.MakeNew(def, Point)

    if (tbl.HasError) then
        if (tbl.HasLockError) then
            MsgBox.Error("No se pudo obtener el bloqueo de
escritura del archivo " + def.GetBaseName, "")
        else
            MsgBox.Error("No se pdo crear el archivo " +
def.GetBaseName, "")
        end
        return nil
    end
end

fld = Field.Make("ID", #FIELD_DECIMAL, 8, 0)
fld.SetVisible(TRUE)
tbl.AddFields({fld})
tbl.SetEditable(False)
theTheme = FTheme.Make(tbl)
theView.AddTheme(theTheme)
theTheme.SetActive(TRUE)
theTheme.SetVisible(TRUE)
theView.SetEditableTheme(theTheme)
av.GetProject.SetModified(true)
end

'Se inicialzan los objetos de comparación
aMatchSource = mTheme.GetMatchSource
aMatchKey = MatchKey.Make("us_addr.stn")
aMatchKey.AllowIntersections(aMatchSource.GetXStanRules,
aMatchSource.GetXDelimiter)
aMatchCase = MatchCase.Make(aMatchSource, aMatchKey)
aMatchCase.AllowIntersections(aMatchSource, aMatchKey)
aMatchPref = MatchPref.Make 'Se utilizan las preferencias por
defecto

'Se abre el índice de comparación
aMatchSource.OpenIndex
```

```

'Se abre el archivo de direcciones
addrFile = LineFile.Make("address.txt".AsFileName,
#FILE_PERM_READ)

count_id = 0
address = addrFile.ReadElt
while (address <> nil)
    count_id = count_id + 1
    aMatchKey.SetKey(address)
    numCand = aMatchSource.Search(aMatchKey, 80, aMatchCase)
    if (numCand = 0) then
        'MsgBox.Info("No se halló coincidencia para la dirección
"++address, "No hallado")
        address = addrFile.ReadElt
        continue
    end

    'Se halla el mejor candidato
    aMatchCase.ScoreCandidates
    bestCand = aMatchCase.GetBestCand

    'Se extrae la ubicación y se agrega al archivo de resultado
    pntLocation = aMatchSource.GetPoint(bestCand)

    theView.SetEditableTheme(theTheme)

    thePrj = theView.GetProjection
    if (thePrj.IsNull.Not) then
        pntLocation = pntLocation.ReturnUnprojected(thePrj)
    end
    theIdField = theTheme.GetFTab.FindField("ID")
    theShpField = theTheme.GetFTab.FindField("Shape")
    theTheme.GetFTab.BeginTransaction
    rec = theTheme.GetFTab.AddRecord
    theTheme.GetFTab.SetValue(theIdField, rec, count_id)
    theTheme.GetFTab.SetValue(theShpField, rec, pntLocation)
    theTheme.GetFTab.EndTransaction
    theTheme.GetFTab.GetSelection.ClearAll
    theTheme.GetFTab.GetSelection.Set(rec)
    theTheme.GetFTab.UpdateSelection

    theView.SetEditableTheme(NIL)

    address = addrFile.ReadElt
end

aMatchSource.EndMatch
addrFile.Close

```

ANEXO B. MANUAL DE USUARIO

La aplicación *Point Patter Analysis on Network Spaces* requiere para su funcionamiento de acceso a un servidor de bases de datos relacionales MySQL 5.4 o superior. La presente compilación está realizada para ejecutarse en sistemas operativos Windows, pero al ser desarrollada sobre la plataforma Java puede ser compilada para ejecutarse en otros sistemas operativos.

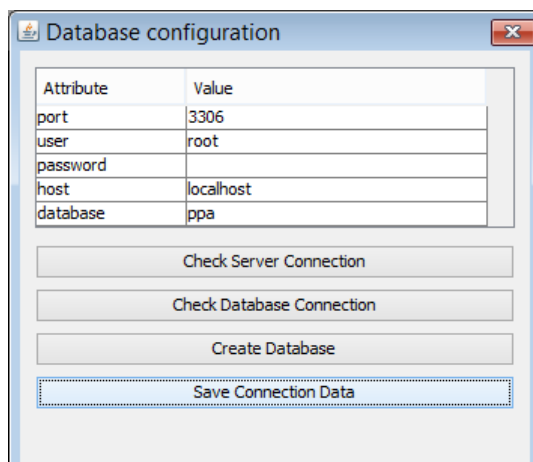
A continuación, se describen las diferentes opciones incluidas en la aplicación.

B.1. MENÚ CONFIGURACIÓN (CONFIGURATION)

B.1.1. Opción Base de datos (Database)

Esta opción permite la configuración de las credenciales de acceso a la base de datos relacional MySQL asociada, y de no existir esta, permite su creación. En la tabla de configuración se deben agregar las credenciales de acceso a la base de datos (puerto, usuario, contraseña, servidor y nombre de la base de datos) (ver figura 35).

Figura 35. Opción base de datos (Database)



Attribute	Value
port	3306
user	root
password	
host	localhost
database	ppa

Check Server Connection

Check Database Connection

Create Database

Save Connection Data

Fuente El autor.

El botón *Check Server Connection* permite verificar si se tiene acceso al servidor y si las credenciales de acceso son válidas, en esta validación no se tiene en cuenta el valor dado para el campo database (nombre de la base de datos). El botón *Check Database Connection* permite verificar si se tiene acceso a la base de datos configurada con las credenciales dadas. El botón *Create Database* ejecuta un proceso de creación de la base de datos con el nombre dado en el

servidor configurado. En los tres casos, de ocurrir un error se mostrará el respectivo mensaje arrojado por el servidor de bases de datos.

Por último, el botón *Save Connection Data* permite almacenar las credenciales de acceso a la base de datos en la aplicación para ser utilizada en los procesos llevados a cabo por las demás opciones de la aplicación.

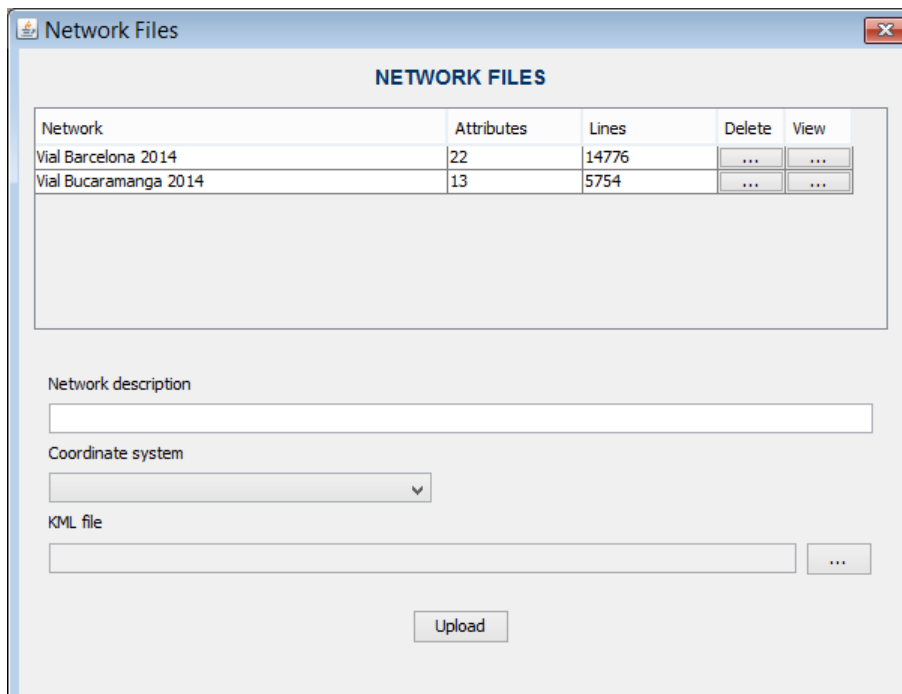
Importante: Esta opción no requiere ser ejecutada cada vez que se ingrese a la aplicación, dado que una vez almacenadas las credenciales de conexión estas se utilizarán de forma automática.

B.2. MENÚ CARGA (UPLOAD)

B.2.1. Opción cargar archivos de red (Upload network files)

Esta opción permite la carga de archivos KML de redes a la base de datos relacional MySQL asociada a la aplicación (ver figura 36).

Figura 36. Opción cargar archivos de red (Upload network files)

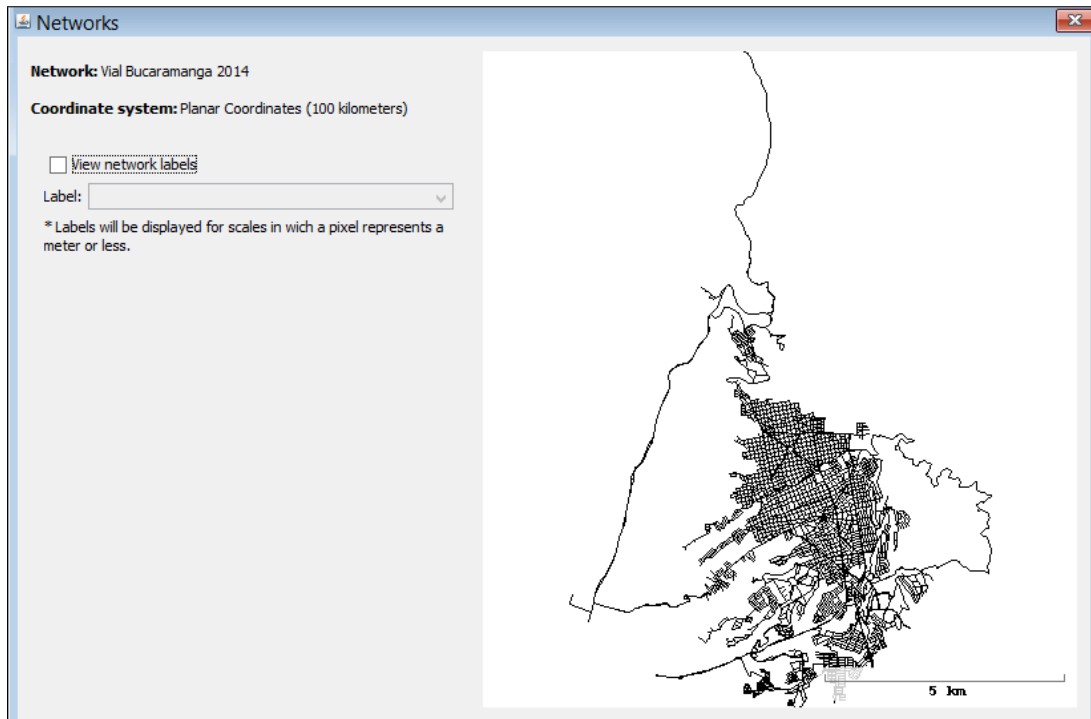


Fuente El autor.

En la tabla ubicada en la parte superior de la ventana se listan los archivos de red cargados anteriormente en la aplicación, para estos archivos existen las opciones de borrar y visualizar; un

ejemplo de visualización de redes se muestra en la figura 37 Sobre esta opción de visualización se pueden realizar ampliaciones y reducciones por medio de la rueda del ratón, y se pueden realizar desplazamientos utilizando la acción de arrastrar y soltar con el ratón. Si el archivo KML de red cargado tiene datos adicionales que permitan identificar las vías, estos datos se pueden agregar como etiquetas marcando el cuadro de chequeo *View network labels* y seleccionando la etiqueta deseada en la lista desplegable.

Figura 37. Visualización de archivos de red



Fuente El autor.

Para cargar un nuevo archivo de redes se debe suministrar un nombre, indicar el sistema de coordenadas asociado en la lista desplegable y seleccionar el archivo KML respectivo; una vez suministrados estos datos se debe oprimir el botón *Upload* con lo que se realizará el proceso de carga.

B.2.2. Opción cargar archivos de eventos (puntos) [Upload event (point) files]

Esta opción permite la carga de archivos KML de puntos que representan eventos. En general esta opción es muy similar a la opción de carga de archivos de redes; en la parte superior de la ventana se presenta una tabla que contiene los archivos de eventos cargados anteriormente, para cada uno de estos archivos se pueden realizar las acciones de borrar y visualizar (ver figura 38).

Para agregar un nuevo archivo de eventos se debe dar una descripción del mismo, seleccionar en la lista desplegable la red a la que estarán asociados los eventos (es importante resaltar que para la carga del archivo de puntos se utilizará como sistema de coordenadas el asignado a la red seleccionada), seleccionar el archivo KML a ser cargado e indicar los atributos de fecha y hora asociados a los eventos del archivo junto con sus respectivos formatos de datos. Una vez suministrados los datos requeridos se debe oprimir el botón *Upload* con lo que se realizará el proceso de carga.

Figura 38. Opción cargar archivos de eventos (puntos) [Upload event (points) files]

The screenshot shows a window titled "Event (Points) Files" with a table of existing events and a form for uploading new ones. The table has columns for Event, Network, Attributes, Points, Delete, and View. The form includes fields for Event description, Network (dropdown), KML file (with a browse button), Date attribute (dropdown), Date format (dropdown), Time attribute (dropdown), and Time format (dropdown). An Upload button is at the bottom.

Event	Network	Attributes	Points	Delete	View
Accidentes de Tránsito Barcelona 2014	Vial Barcelona 2014	31	8766
Accidentes de Tránsito Bucaramanga 2014	Vial Bucaramanga 2014	28	3661

Event description

Network

KML file
 ...

Date attribute
 (None) ▾

Date format
 (None) ▾

Time attribute
 (None) ▾

Time format
 (None) ▾

Upload

Fuente El autor.

B.3. MENÚ PROCESOS (PROCESSES)

B.3.1. Opción Función K para redes (Network K function)

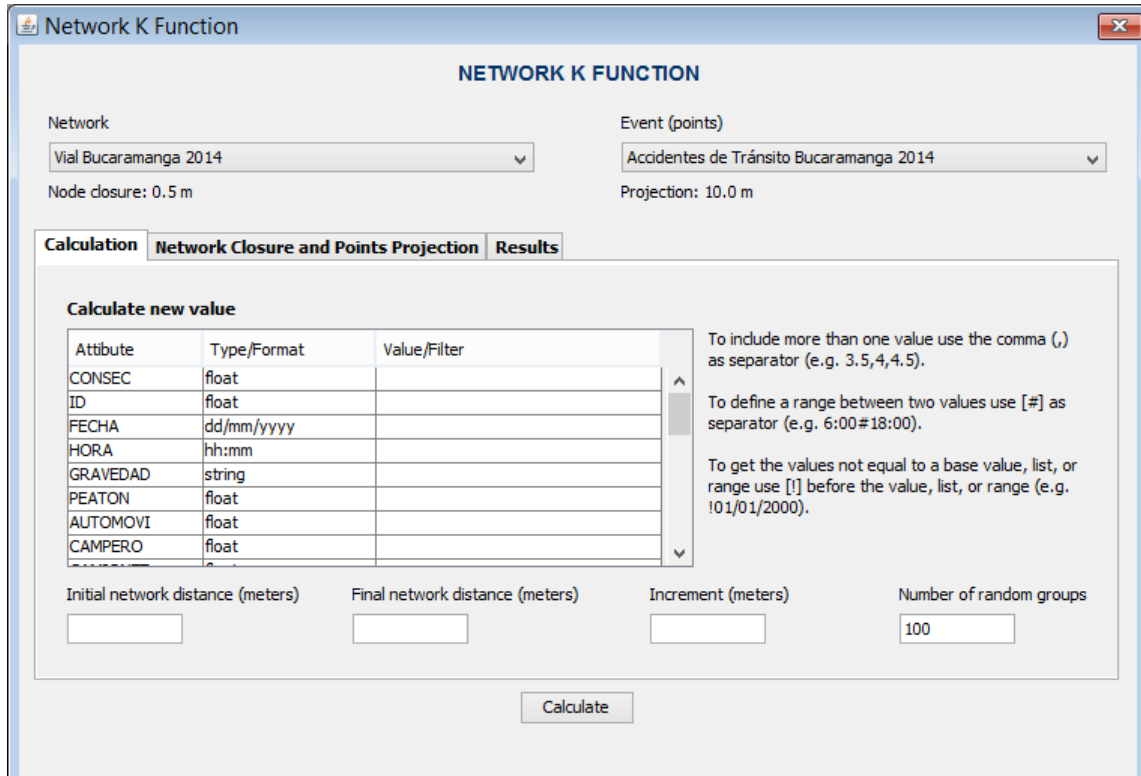
Esta opción permite el cálculo de series de valores de la función K para redes sobre el conjunto de puntos asociados a un evento en una red, o un subconjunto de estos datos definido por un grupo de filtros.

Para poder utilizar las herramientas presentes en esta opción es necesario haber realizado la carga de por lo menos un archivo KML de redes y un archivo de eventos asociado, estos pueden ser

seleccionados en las listas desplegables que se encuentran ubicados en la parte superior de la ventana (ver figura 39).

Las herramientas incluidas en esta opción se encuentran divididas en tres secciones: Cálculo (calculation), Cierre de nodos de red y proyección de puntos (network closure and points projection), y Resultados (results).

Figura 39. Opción función K para redes (Network K function)



Fuente El autor.

En la sección de cálculo es donde se calcula la función K para redes propiamente dicha. Para su aplicación se pueden seleccionar diferentes filtros en la tabla ubicada en la esquina superior izquierda de la sección, esta tabla muestra los diferentes atributos contenidos en el archivo KML de eventos seleccionado, junto a cada atributo se presenta el tipo de dato correspondiente, y en el caso de los atributos de fecha y hora los formatos de datos respectivos. La aplicación de filtros es opcional y se pueden seleccionar tantos filtros como sea necesario. Los tipos de filtros se dividen en tres grupos, valores individuales, listados de valores separados por comas (,) y rangos de valores demarcados por medio del símbolo numeral (#); a estos tres grupos se puede agregar el símbolo de admiración (!) al inicio para indicar que se desea aplicar un filtro con los valores diferentes al valor, lista o rango indicado.

A continuación, se solicitan los valores necesarios para realizar el cálculo de la función K. Se deben indicar las distancias de red inicial y final, y el incremento entre distancias; por ejemplo, si se indica una distancia inicial de 10 metros, una distancia final de 30 metros y un incremento de 5 metros, se realizarán los cálculos para las distancias de 10, 15, 20, 25 y 30 metros. También se debe indicar la cantidad de grupos de datos aleatorios que se utilizarán para determinar los intervalos de confianza de la función, por defecto se toma un número de 100 grupos. Una vez se hayan indicado los filtros deseados y suministrado los datos requeridos se debe oprimir el botón *Calculate* con lo que se iniciará el proceso de cálculo de la función.

La sección de cierre de nodos de red y proyección de puntos (ver figura 40) permite la realización de dos procesos de depuración de datos necesarios para la aplicación de la función K para redes. El primero de los procesos es el de cierre de nodos, el cual busca hacer que los nodos de red que se encuentren unos de otros a una distancia definida o menos pasen a estar unidos, esto con el fin de que los procesos de cálculos de distancias a través de la red tengan en cuenta estos nodos. El segundo proceso, el de proyección de puntos, reubica los puntos de los eventos sobre los segmentos de red más cercanos tomando como base una distancia de proyección máxima; los nodos que no se encuentran ubicados sobre la red no serán tomados en cuenta en los cálculos.

Figura 40. Cierre de nodos de red y proyección de puntos

The screenshot shows a software window titled "Network K Function". At the top, it says "NETWORK K FUNCTION". Below this, there are two dropdown menus: "Network" set to "Vial Bucaramanga 2014" and "Event (points)" set to "Accidentes de Tránsito Bucaramanga 2014". Below these are two text fields: "Node closure: 0.5 m" and "Projection: 10.0 m".

The main area of the window has three tabs: "Calculation", "Network Closure and Points Projection" (which is selected), and "Results". Under the selected tab, there are two sections:

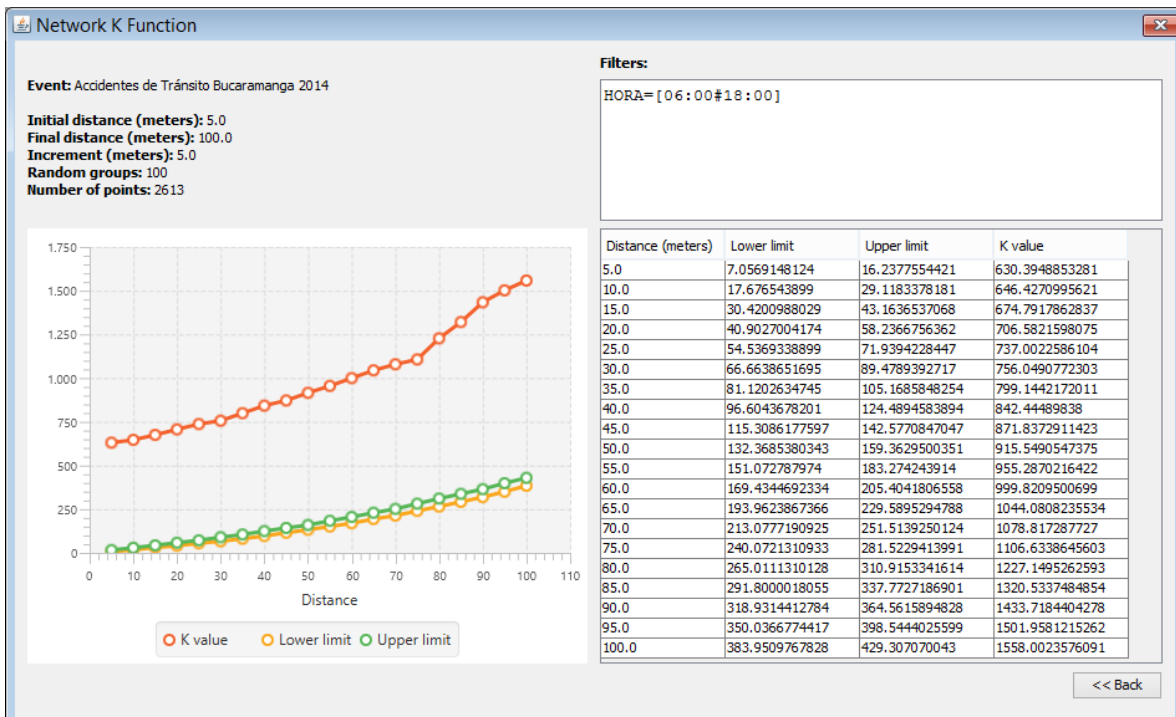
- Network:** "Vial Bucaramanga 2014". It has a text field for "Maximum node closure distance (meters)" with the value "0.5" and a "Close nodes" button. To the right of the button, it says "Node closure: 0.5 m".
- Event (points):** "Accidentes de Tránsito Bucaramanga 2014". It has a text field for "Maximum projection distance (meters)" with the value "10.0" and a "Project points" button. To the right of the button, it says "Projection: 10.0 m".

At the bottom center of the window is a "Calculate" button.

Fuente El autor.

La sección de resultados presenta una tabla con los cálculos de la función K redes realizados anteriormente sobre el evento seleccionado, estos cálculos pueden ser borrados o visualizados en detalle. La figura 41 presenta un ejemplo de visualización de resultados, esta se divide en cuatro partes: en la parte superior izquierda de la ventana se ubican los datos del evento y los parámetros de entrada utilizados en los cálculos, en la parte superior derecha se muestran los filtros aplicados, en la parte inferior izquierda se presenta la gráfica de resultados, y en la parte inferior derecha se muestra la tabla de resultados de la función.

Figura 41. Resultados de la función K para redes



Fuente El autor.

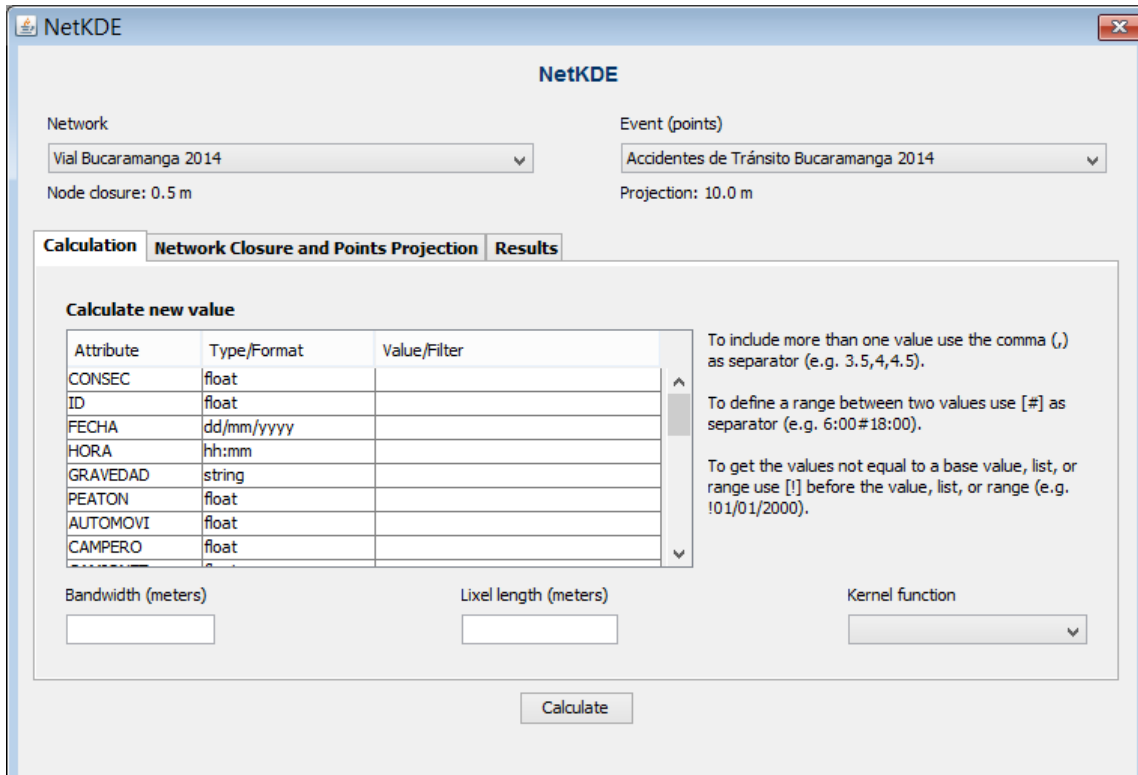
B.3.2. Opción NetKDE

Esta opción permite la aplicación del método NetKDE sobre el conjunto de puntos que conforman un evento, o un subconjunto de estos puntos definido por una serie de filtros.

Al igual que con la función K para redes, para poder utilizar las herramientas de NetKDE es necesario haber realizado la carga de un archivo KML de redes y un archivo de eventos asociado, estos datos pueden ser seleccionados en las listas desplegables que se encuentran en la parte superior de la ventana (ver figura 42).

Esta opción se encuentra dividida en las mismas secciones que la opción de cálculo de la función K para redes: Cálculo (calculation), Cierre de nodos de red y proyección de puntos (network closure and points projection), y Resultados (results).

Figura 42. Opción NetKDE



Fuente El autor.

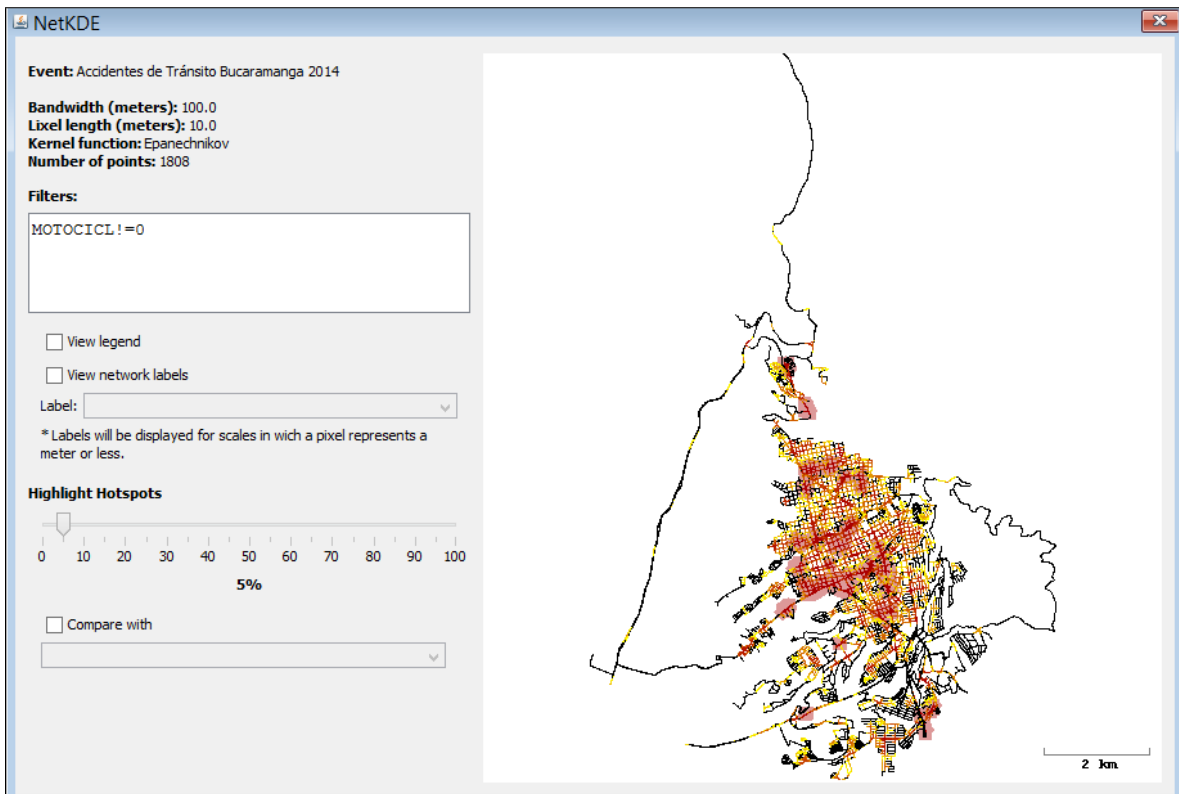
En la sección de cálculo se aplica el método NetKDE al conjunto de puntos seleccionado. Los filtros disponibles son los mismos descritos en la sección anterior (Opción función K para redes). Los valores necesarios para la aplicación del método son el ancho de banda (bandwidth) de búsqueda de puntos en metros, la longitud en metros de los segmentos de línea en los que se dividirá la red denominados *lixels* y la función de decaimiento por distancia o función de núcleo a utilizar. Una vez ingresados los filtros deseados y los datos requeridos se debe oprimir el botón *Calculate* para iniciar la aplicación del método.

La sección de cierre de nodos de red y proyección de puntos se encuentra descrita en la opción función K para redes.

Por último, la sección de resultados muestra una tabla con los cálculos del método NetKDE realizados anteriormente sobre el evento seleccionado, estos cálculos a su vez pueden ser borrados o visualizados. En la figura 43 se muestra un ejemplo de resultados de NetKDE, en la

parte izquierda de la ventana se presentan los datos específicos del evento y los parámetros utilizados en los cálculos junto con los filtros aplicados, seguidos de opciones de visualización de la leyenda de datos que delimita los códigos de colores y una opción para incluir etiquetas de atributos de la red, también se incluye un componente de deslizador o *slider* que permite controlar el porcentaje de *lixels* con densidades altas que se resaltan como puntos calientes. En la parte inferior se presenta un componente de selección o *checkbox* junto a un componente de lista desplegable que permiten comparar visualmente los resultados seleccionados con otros resultados de NetKDE realizados sobre la misma red.

Figura 43. Resultados de NetKDE



Fuente El autor.

En la parte derecha de la ventana se presenta el resultado gráfico de NetKDE, el color negro indica ausencia de eventos cercanos y la escala de colores del amarillo al rojo indica niveles de concentración de eventos correspondiendo los colores amarillos a niveles bajos y los colores rojos a niveles altos de densidad. Al igual que las visualizaciones de redes y puntos de eventos, sobre esta visualización se pueden realizar ampliaciones y reducciones utilizando la rueda del ratón, y se pueden realizar desplazamientos por medio la acción de arrastrar y soltar con el ratón.