

Desarrollo de una aplicación para el análisis espacial de datos en espacios de redes

Feisar Enrique Moreno Corzo

Facultad de Ingeniería, Universidad Autónoma de Bucaramanga – UNAB
Bucaramanga, Colombia

fmoreno4@unab.edu.co

Resumen— Aunque el análisis de patrones de puntos no es un área de estudio nueva, mantiene su vigencia al permitir analizar diferentes tipos de fenómenos desde una perspectiva espacial por medio de diversos métodos. Entre los métodos de análisis de patrones de puntos existentes se encuentran algunos enfocados en el análisis sobre espacios de redes, definidas como espacios lineales contenidos dentro de los espacios euclidianos en dos dimensiones. Un ejemplo de espacio de redes es el formado por la red vial de una ciudad o región. Sobre estos espacios de redes se pueden realizar análisis de fenómenos como accidentes de tránsito y crímenes. El presente artículo se centra en el desarrollo de una herramienta software que implementa dos de estos métodos, la función K para redes y NetKDE, y su aplicación sobre los accidentes de tránsito ocurridos en las ciudades de Bucaramanga (Colombia) y Barcelona (España) durante el año 2014.

I. INTRODUCCIÓN

La creación de herramientas software de análisis de patrones de puntos (PPA por sus siglas en inglés) ha tenido un desarrollo lento en los últimos años, esto debido en gran medida a que no existe una gran demanda de este tipo de aplicaciones; sus avances se ubican principalmente en el ámbito académico y de investigación por lo que generalmente se encuentran artículos sobre el desarrollo de modelos matemáticos y estadísticos. El desarrollo de herramientas software está principalmente enfocado a la validación de estos modelos.

La aproximación más común al estudio de eventos con componente geográfico se basa en el uso de técnicas y métodos que tratan los espacios en los que ocurren estos eventos como espacios euclidianos de dos dimensiones. Una vertiente con una aproximación diferente, aunque basada en estos métodos, es la centrada en la creación y adaptación de métodos restringidos a espacios de subconjuntos de redes unidimensionales contenidas en los espacios de dos dimensiones.

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.

II. ANTECEDENTES

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 en 1954 [1]; otro ejemplo de estos primeros estudios es el llevado a cabo por David William Harvey en 1966 centrado en el análisis de cuadrados [2]. Estos artículos, y muchos de los desarrollados posteriormente, se centran en el análisis de patrones en espacios de dos dimensiones.

La ocurrencia de varios de los eventos que pueden ser abordados desde el análisis de patrones de puntos se no se encuentra suscrita a la totalidad del espacio en dos dimensiones, algunos de los eventos se presentan de forma casi exclusiva en el subespacio representado por las carreteras y caminos dentro de un área; eventos como los accidentes de tránsito y los crímenes sirven como ejemplo; este tipo de subespacio se denomina espacio 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.

El análisis de patrones de puntos en redes es un concepto bastante más reciente que su equivalente en dos dimensiones. El subespacio de redes por su topología hace que los eventos contenidos en él se relacionen de forma diferente a los eventos contenidos en espacios planos en dos dimensiones. Una primera aproximación al análisis en espacios de redes se encuentra en el estudio llevado a cabo por William R. Black en 1992 sobre auto correlaciones en redes [3].

En general, los métodos asociados al PPA se pueden clasificar en dos grandes grupos [4]:

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, y que 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 en 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. Se presupone que hay interdependencia espacial entre los lugares de los eventos.

Dos métodos de PPA ampliamente difundidos son la Estimación de Densidades por Núcleos (KDE por sus siglas en inglés) y la Función K, métodos de primer y segundo orden respectivamente.

KDE es un método de análisis de patrones de puntos muy difundido, existen aplicaciones software comerciales que lo implementan como la extensión de análisis espacial de ArcGIS [5] y CrimeStat [6]. 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 euclideo 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 evento estudiado se asigna a alguna de las divisiones (cuadrados). También se define un radio de influencia entre eventos, denominado ancho de banda.

La función K, propuesta por Brian D. Ripley en 1976 [7], representa la media observada del número de puntos dentro de una distancia definida dividida por la densidad de puntos en el área de estudio. Esta función puede ser utilizada para determinar si una serie

de puntos sobre un espacio tienden a formar grupos o por el contrario se distribuye de forma 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 [8].

La aplicación de métodos para espacios euclidianos en dos dimensiones, como los detallados anteriormente, sobre eventos que ocurren dentro de subespacios de redes puede llevar a resultados que no reflejan la realidad que se pretenden modelar; como lo expresa Miller [9] “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 esto que se han planteado algunos modelos alternativos para el manejo este tipo de eventos, algunos de estos derivados de modelos para espacios euclidianos en dos dimensiones. Dos de estos modelos para espacios de redes son NetKDE, derivado a partir de KDE y una versión de la función K para redes.

III. MÉTODOS IMPLEMENTADOS

El primero de los métodos de análisis de patrones de puntos para redes implementados es la función K para redes, la cual puede llevar a diferencias sustanciales entre las distancias medidas a través de la red y las distancias euclidianas, sobre todo en entornos que abarcan áreas pequeñas que contienen a su vez redes pequeñas en extensión. Esta versión de la función K para espacios en redes fue desarrollada por Atsuyuki Okabe e Ikuho Yamada en 2000 [10].

El segundo método implementado es NetKDE (abreviación de Network KDE), un método propuesto por Zhixiao Xie y Jun Yan en 2008 [11] 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. 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 trabaja este método son diferentes a las utilizadas por KDE, 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. Además de lo anterior, con KDE se suelen detectar agrupaciones con mayores concentraciones de puntos a las detectadas mediante NetKDE.

IV. ARQUITECTURA DE SOFTWARE

Partiendo de las necesidades particulares del proyecto se utilizó la metodología de desarrollo de software *Crystal Clear*, propuesta por Alistair Cockburn [12]; ésta es una de las llamadas metodologías de desarrollo de software ágiles que se adhieren al manifiesto ágil publicado en 2001 [13]. La selección de esta metodología obedeció a su capacidad para minimizar el riesgo de error en procesos de desarrollo de software que disponen de tiempo limitado y a su capacidad adaptativa que permite que las soluciones diseñadas vayan evolucionando a la par de las nuevas necesidades que puedan ir surgiendo a lo largo del proyecto.

La documentación resultante de las diferentes fases del proceso de construcción de la herramienta software se llevó a cabo apoyándose 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 de la herramienta software, se detectó una necesidad de manejo de volúmenes de datos de tamaño medio (representaciones de redes de vehículos y coordenadas geográficas de puntos dentro de estas redes) y una capacidad de procesamiento de datos alta. Además de lo anterior, la información obtenida mediante la aplicación de NetKDE a los conjuntos de datos de entrada no requiere ser compartida masivamente en la web, 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 implementado pueden llegar a ser extensos -del orden de minutos en algunos casos- dependiendo del tamaño del conjunto de datos de entrada, se planteó la construcción de la herramienta software en un entorno de escritorio, lo cual permite un mejor desempeño. La plataforma de codificación seleccionada fue Java SE, lo que permite crear una aplicación portable a equipos de cómputo con diferentes sistemas operativos, facilitando la compatibilidad.

En lo referente al manejo de los diversos datos de entrada utilizados por la aplicación, para los datos que representan las redes de vías se utilizan archivos con formato KML (*Keyhole Markup Language*), formato desarrollado inicialmente para Google Maps pero que a partir de 2008 es un estándar internacional [14]. Los datos de los accidentes de tránsito, representados por puntos en un sistema de coordenadas, también se manejan mediante archivos 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.

V. CASOS DE PRUEBA

Al utilizar un espacio más complejo que el espacio euclidiano de dos dimensiones, la función K para redes y NetKDE requieren una fase de preparación de la información igualmente compleja. Como casos de prueba se seleccionaron dos conjuntos de datos de red vehicular y accidentes de tránsito, ambos correspondientes al año 2014.

El primer grupo de datos pertenece a la ciudad de Bucaramanga, Colombia. Los datos que representan la red vehicular fueron tomados de archivos de cartografía suministrados por 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 más 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 se encontraban la diagonal 15, la vía Bucaramanga a Girón, la vía a Chimitá y la vía al mar entre otras. Para el proceso de actualización vial se utilizó la aplicación ArcMap 10.2.2 [15] de ESRI y se tomó como referencia la información geográfica disponible en Google Maps [16], algunos ejemplos de las

actividades incluidas en la actualización vial se pueden ver en la figura 1. La información vial revisada y actualizada fue exportada posteriormente a formato KML por medio del servicio web de MyGeoData [17].

TABLA I
DIRECCIONES AJUSTADAS PARA AVENUE

Dirección original	Dirección ajustada
Carrera 33 con calle 48	cr_33_48-0
Calle 56 #17A-12	c11_56_17a-12
Avenida Quebrada Seca #28-25	av_qseca_25-25

En lo referente a la información sobre accidentes de tránsito para el año 2014 suministrada por la Dirección de Tránsito de Bucaramanga, esta no contaba con el componente de ubicación geográfica (latitud y longitud del punto que representa el accidente), por lo que fue necesario ubicar sobre la cartografía de la red vial 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 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 I se dan algunos ejemplos de los ajustes que se debieron realizar sobre las direcciones para que pudieran ser ubicadas por medio del algoritmo en Avenue.

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 el proceso un conjunto de archivos con los puntos (automáticos y manuales) georreferenciados.

Una vez se obtuvieron las coordenadas geográficas de cada uno de los accidentes de tránsito, se procedió a la elaboración de un archivo KML por medio del servicio web de MyGeoData mencionado anteriormente. El sistema de coordenadas de este primer grupo de datos es una proyección de Mercator [18].

El segundo grupo de datos corresponde a accidentes de tránsito ocurridos en la ciudad de Barcelona, España. La información cartográfica de la red vial utilizada, recibida en formato SHP, no fue ajustada por ser reciente y completa. De igual forma no fue necesario un proceso de ubicación de los accidentes de tránsito ya que estos fueron recibidos en el mismo formato cartográfico. La elaboración de los correspondientes archivos KML fue llevada a cabo por medio del servicio web antes mencionado. Este segundo grupo de datos estaba representado en el sistema de coordenadas geográficas.

VI. ANÁLISIS DE RESULTADOS

Aunque los dos métodos implementados se enfocan en temas diferentes dentro de un mismo campo del conocimiento, 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.

A. Función K para Redes

El valor de la función K en sí mismo no tiene un significado, para poder sacar conclusiones al respecto de este valor, debe ser comparado 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 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 2 se muestran los resultados gráficos de la función K para redes aplicada a los datos de accidentes de tránsito de las ciudades de Bucaramanga en la gráfica (a) y Barcelona en la gráfica (b). En las dos gráficas el valor K se representa en color rojo, el límite inferior del intervalo de confianza en amarillo y el límite

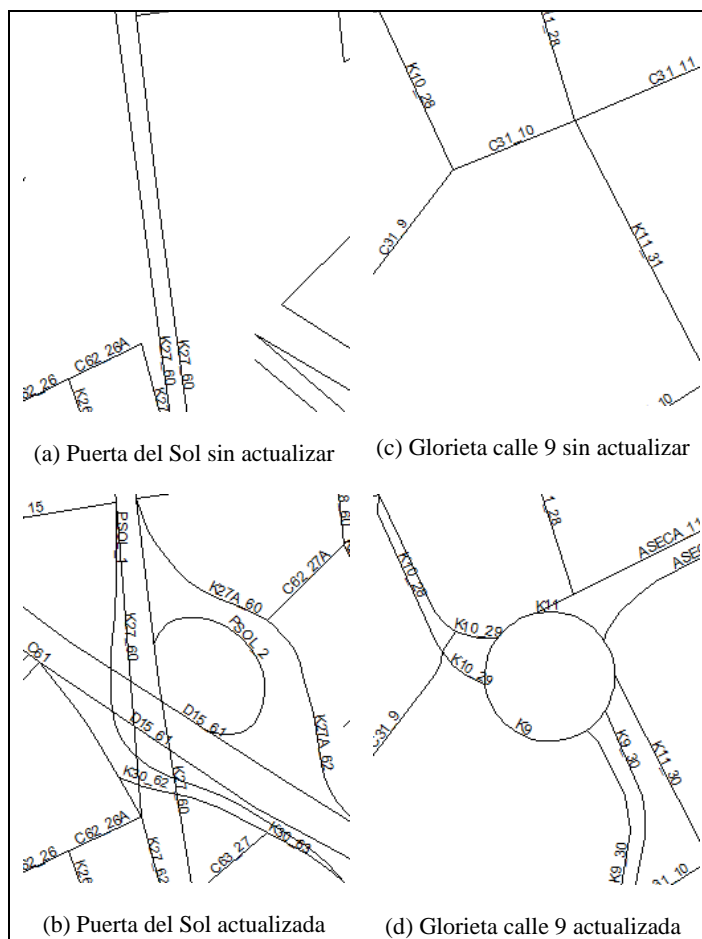


Fig. 1 Actualización vial

De los 3.731 datos de accidentes de tránsito recibidos, se ubicaron 3.105 por medio del algoritmo mencionado anteriormente. De los

superior en verde. En los dos casos se observa una fuerte tendencia de los datos a agruparse, incluso para distancias máximas de 5 metros.

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.

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 una gran cantidad de casos en los que dos o más accidentes compartían exactamente las mismas coordenadas. De lo anterior se concluye que la forma de ubicación de los puntos tiene un impacto importante sobre los valores K calculados, sobre todos para distancias pequeñas.

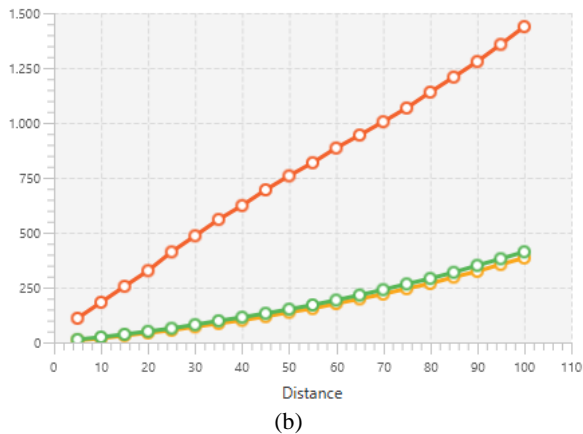
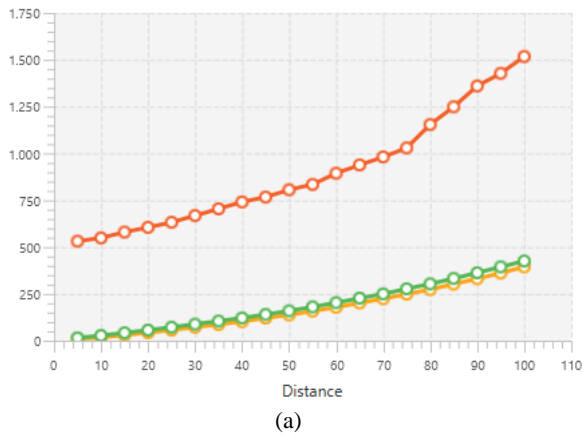


Fig. 2 Resultados generales de la función K para redes

Un análisis similar se aplicó sobre diferentes vistas de los datos para ambas ciudades como los accidentes ocurridos entre las 6:00 pm y las 6:00 am, los accidentes en los que se vieron involucradas motocicletas, los accidentes ocurridos durante el mes de diciembre y los accidentes con víctimas mortales. En todas las vistas los resultados fueron muy similares, una fuerte tendencia al agrupamiento, con una notable excepción en el grupo de accidentes con víctimas mortales.

Tanto para los datos de la ciudad de Bucaramanga como para los de la ciudad de Barcelona los resultados no mostraron tendencia 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 se utilizaron valores de distancias cada 10 metros. Los resultados se presentan en la figura 3, la gráfica (a) representa los valores K para la ciudad de Bucaramanga y la gráfica (b) los valores de la ciudad de Barcelona.

TABLA II
NÚMERO DE ACCIDENTES DE TRÁNSITO POR RESULTADO

Grupo de datos	Bucaramanga		Barcelona	
	Eventos	%	Eventos	%
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

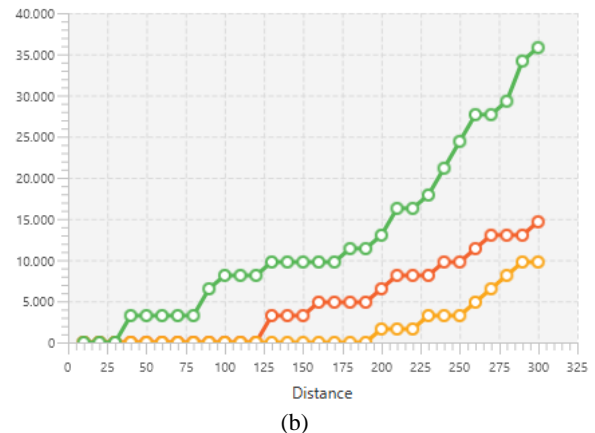
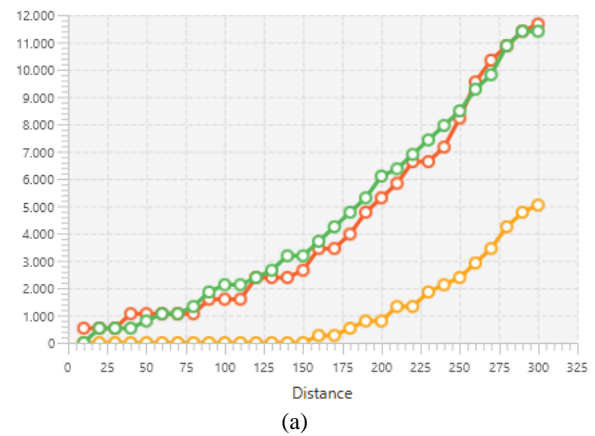


Fig. 3 Resultados para accidentes de tránsito con víctimas mortales

Una posible explicación a los resultados obtenidos con respecto a los accidentes de tránsito con víctimas mortales se centra en 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 2 se presentan los números de accidentes de tránsito utilizados en cada grupo de resultados.

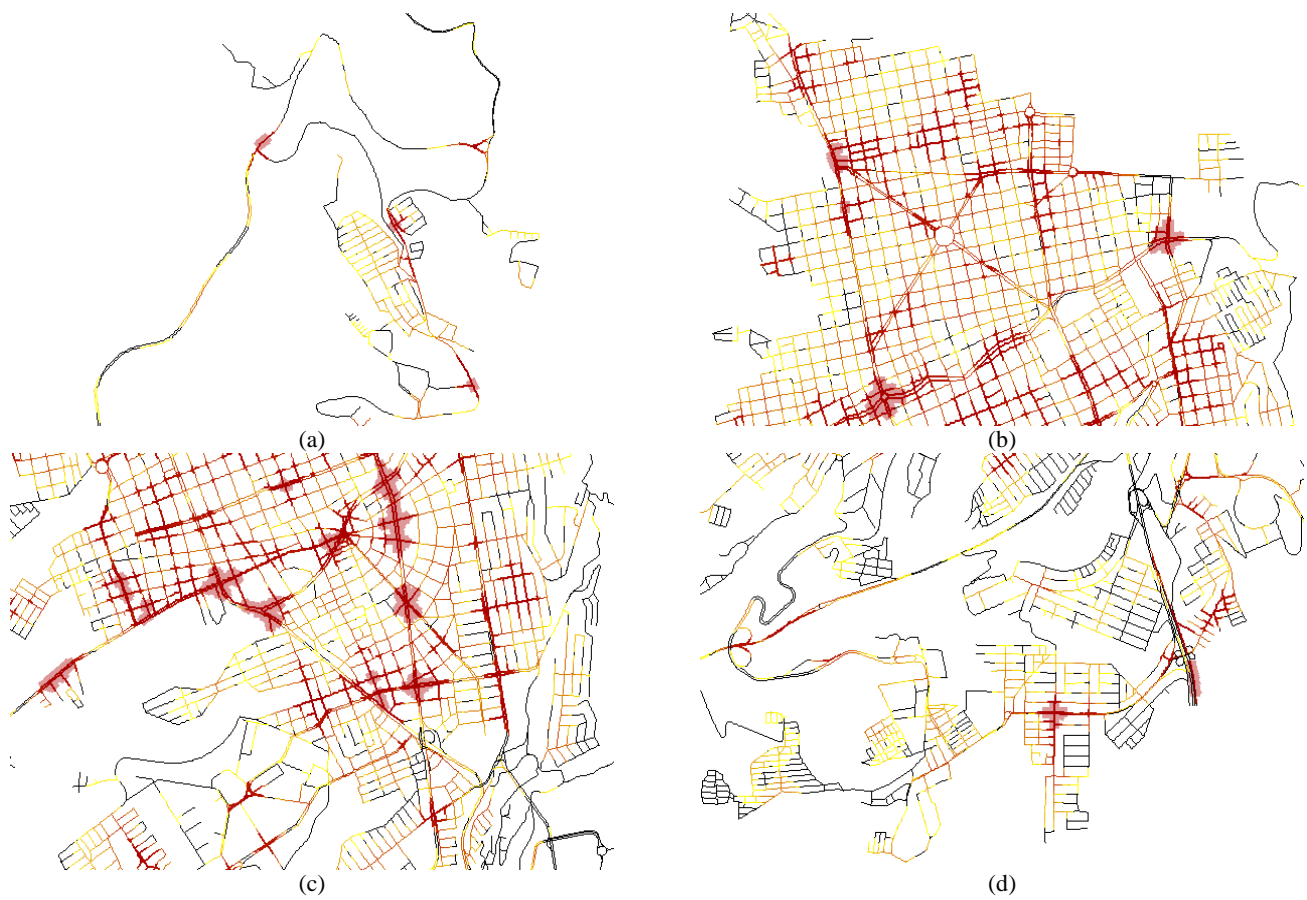


Fig. 4 Resultados de NetKDE – Bucaramanga 2014

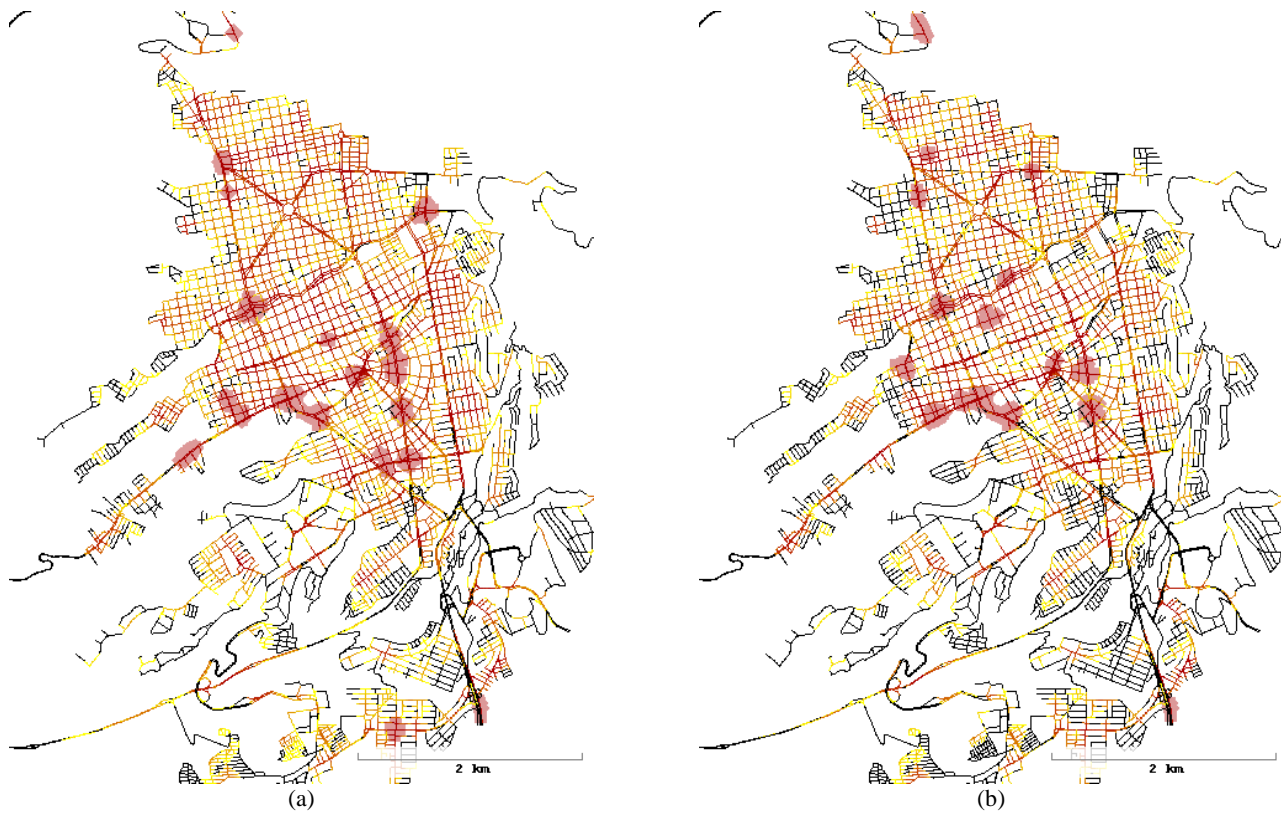


Fig. 5 Comparación entre resultados generales de NetKDE y resultados de accidentes con motocicletas involucradas

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 correspondió 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 estuvo 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.

Los resultados de estos dos últimos grupos también se presentaron cercanos al límite superior del intervalo de confianza, aunque con características diferentes. El primer grupo se encontraba claramente dentro del intervalo de confianza para valores de distancias menores o iguales a los 120 metros y fue claramente superior a partir de los 250 metros. El segundo grupo presentaba 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.

B. 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 las mayores concentraciones de los eventos analizados.

TABLA III
PUNTOS CALIENTES NETKDE – BUCARAMANGA 2014

#	Ubicación
1	Vía a Chimitá al norte del cruce con la vía al Café Madrid.
2	Salida del barrio Olas Bajas en el cruce con la vía a Rionegro.
3	Intersección de la avenida Libertador con calle 5N.
4	Entre calles 10 y 12 con carrera 15.
5	Intersección de la calle 14 con carrera 15 en sentido sur-norte.
6	Intersección de carrera 33A con avenida Quebradaseca.
7	Glorieta de la avenida Quebradaseca con carrera 15 y calles adyacentes.
8	Intersección de la carrera 21 con calle 36.
9	Intersección de la carrera 27 con avenida la Rosita y calles adyacentes.
10	Carrera 27 con calles 42, 45 y 48.
11	Triángulo formado por la calle 45 con avenida la Rosita y la carrera 22.
12	Calle 45 entre carreras 14A y 16.
13	Intersección entre la diagonal 15 y la avenida la Rosita.
14	Calle 45 entre carreras 9 y 10.
15	Intersección de la carrera 10 con calle 44.
16	Intersección de la carrera 9 con calle 42.
17	Intersección de la carrera 9 con calle 43.
18	Calle 45 entre carreras 2 y 3.
19	Intersección entre la carrera 27 y la avenida González Valencia.
20	Intersección entre la carrera 27 y la calle 56.
21	Intersección entre la carrera 21 y la calle 55.
22	Carrera 21 entre calle 56 y diagonal 15.
23	Autopista a Floridablanca costado oriental en sentido sur-norte 50 metros al sur del puente de Provenza.
24	Intersección de carrera 22 con calle 105.

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. Por esta razón no se incluyen las vistas de accidentes con víctimas mortales. 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.

Para todos los resultados analizados se aplicó como función de núcleo la distribución de Epanechnikov [19], 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 *lixiels* con densidades iguales o superiores al percentil 98 (2% más altas). En la figura 4 se muestran los puntos calientes detectados en la ciudad de Bucaramanga divididos en 4 grupos que recorren la ciudad de norte a sur. En total se detectaron 24 puntos calientes detallados en la tabla III.

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 puntos calientes diferentes, el primero ubicado en el cruce entre la carrera 28 y la calle 15. El segundo punto caliente se ubica en la intersección entre la carrera 36 y la calle 35.

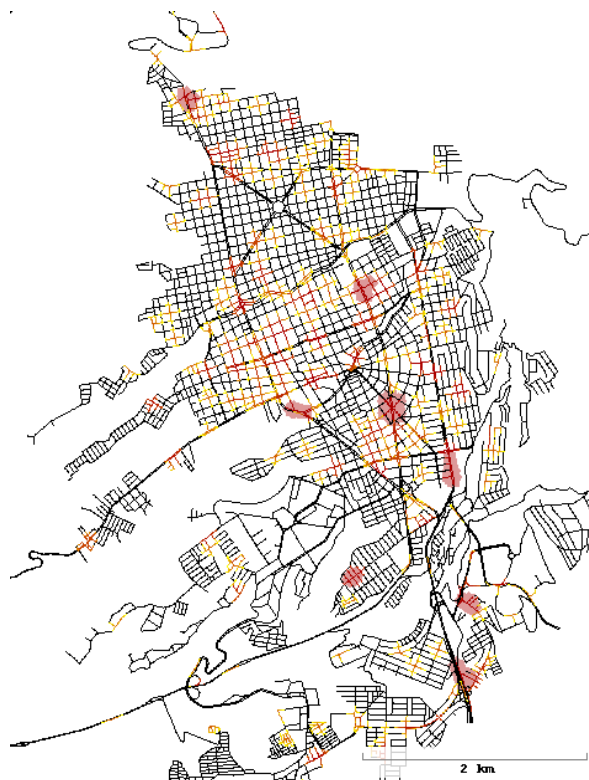


Fig. 6 Resultados de accidentes ocurridos durante el mes de diciembre

La comparación entre la vista general de resultados y la vista de accidentes que involucran motocicletas se presenta en la figura 5, 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) a concentrarse principalmente en la

denominada zona de meseta y no presentando puntos calientes en las zonas periféricas.

Por último, la comparación de la vista general con la vista de accidentes ocurridos durante el mes de diciembre también presentó diferencias sustanciales. En la figura 6 se presenta la distribución de puntos calientes de esta vista. Solamente se observan 8 puntos calientes, 6 de los cuales no están presentes en la vista general. 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.

VII. CONCLUSIONES Y 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 2016 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 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 [18] 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.

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 también 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 incluidos. Es importante resaltar que esta recomendación 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 incluso 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.

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 [8]. 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.

REFERENCIAS

- [1] P. J. Clark and F. C. Evans, "Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations," *Ecology*, vol. 35, no. 4, pp. 445–453, 1954.
- [2] D. W. Harvey, "Geographical Processes and the Analysis of Point Patterns: Testing Models of Diffusion by Quadrat Sampling," *Trans. Inst. Br. Geogr.*, no. 40, pp. 81–95, 1966.
- [3] W. R. Black, "Network Autocorrelation in Transport Network and Flow Systems," *Geogr. Anal.*, vol. 24, no. 3, pp. 207–222, 1992.
- [4] T. C. Bailey and A. C. Gatrell, *Interactive spatial data analysis*. 1995.
- [5] "ArcGIS Spatial Analyst." [Online]. Available: <http://www.esri.com/software/arcgis/extensions/spatialanalyst>. [Accessed: 31-Aug-2016].
- [6] "CrimeStat: Spatial Statistics Program for the Analysis of Crime Incident Locations." [Online]. Available: <http://nij.gov/topics/technology/maps/pages/crimestat.aspx>. [Accessed: 31-Aug-2016].
- [7] P. M. Dixon, A. H. El-Shaarawi, and W. W. Piegorisch, "Ripley's K function," *Encycl. Environmetrics*, vol. 3, pp. 1796–1803, 2002.
- [8] "SANET - Spatial analysis along networks." [Online]. Available: <http://sanet.csis.u-tokyo.ac.jp/>. [Accessed: 31-Aug-2016].
- [9] H. J. Miller, "Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T)," *Geogr. Anal.*, vol. 31, no. 4, pp. 373–399, 1999.
- [10] A. Okabe and I. Yamada, "The K-function method on a network and its computational implementation," *Geogr. Anal.*, no. 33, pp. 271–290, 2000.
- [11] Z. Xie and J. Yan, "Kernel Density Estimation of traffic accidents in a network space," *Comput. Environ. Urban Syst.*, vol. 32, no. 5, pp. 396–406, 2008.
- [12] A. Cockburn, *Crystal Clear: A Human-Powered Methodology for Small Teams*. Addison-Wesley Professional, 2004.
- [13] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber, J. Sutherland, and D. Thomas, "Principles behind the Agile Manifesto," *Manifesto for Agile Software Development*, 2001. [Online]. Available: <http://agilemanifesto.org/principles.html>. [Accessed: 31-Jul-2015].
- [14] "OGC® Approves KML as Open Standard." [Online]. Available: <http://www.opengeospatial.org/pressroom/pressreleases/857>. [Accessed: 15-Oct-2016].
- [15] "ArcMap - ArcGIS for Desktop." [Online]. Available: <http://desktop.arcgis.com/es/desktop/>. [Accessed: 15-Oct-2016].
- [16] "Google Maps." [Online]. Available: <http://maps.google.com>. [Accessed: 15-Oct-2016].
- [17] "MyGeoData - Free Online GIS Data Converter." [Online]. Available: <http://converter.mygeodata.eu/>. [Accessed: 15-Oct-2016].
- [18] P. Osborne, *The Mercator Projection*, 1st ed. Edinburgh, 2013.
- [19] O. Schabenberger and C. A. Gotway, *Statistical methods for spatial data analysis*. 2005.