



Análisis exploratorio de comportamientos de ciclistas voluntarios mediante minería de patrones espacio-temporales en Cuenca, Ecuador



Lorena Abad , Daniel Orellana 

Grupo de Investigación LactaLAB - Ciudades Sustentables, Departamento de Espacio y Población, Universidad de Cuenca, Cuenca, Ecuador.

Autor para correspondencia: lore.abad6@gmail.com

Fecha de recepción: 7 de octubre 2017 - Fecha de aceptación: 15 de febrero 2018

RESUMEN

El uso de técnicas de minería de datos combinadas con metodologías como los cubos espacio-temporales potencian en gran medida el análisis de bases de datos multidimensionales. Una de las áreas de aplicación más importantes es el estudio de la movilidad no motorizada, la cual constituye una de las prioridades más importantes de la mayor parte de ciudades alrededor del mundo. Este estudio busca extraer patrones en una base de datos de movilidad de ciclistas voluntarios en Cuenca (Ecuador) para identificar lugares y momentos relevantes para el diseño y aplicación de estrategias para mejorar la movilidad en bicicleta. La metodología está basada en la caja de herramientas para la minería de patrones espacio-temporales de ArcGIS. Los resultados demuestran la viabilidad de la metodología propuesta para la caracterización de la movilidad no motorizada y su potencial uso en otros conjuntos de datos de movilidad.

Palabras clave: Movilidad no motorizada, ciclismo urbano, análisis exploratorio, minería de patrones espacio-temporales, cubos espacio-temporales.

ABSTRACT

Data mining techniques combined with space-time cubes empower the analysis of multidimensional data. A study area in which these advanced analysis techniques can be applied pre-eminently is urban mobility, where investigation of non-motorized mobility patterns is a main priority for several cities around the world. The presented work aimed to extract spatio-temporal patterns from a human movement database containing volunteer-generated cycling data in Cuenca (Ecuador) with the objective to detect places and times where strategies can be applied that promote urban cycling. The methodology takes advantage of the capabilities of the space-time pattern mining toolbox in ArcGIS. The results demonstrate the viability of the proposed methodology for the characterization of non-motorized mobility patterns and its potential for analyzing other mobility datasets.

Keywords: Non-motorized mobility, urban cycling, exploratory analysis, space-time pattern mining, space-time cubes.

1. INTRODUCCIÓN

La minería de datos como una herramienta para alcanzar el descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases - KDD) (Fayyad, Piatetsky-Shapiro, & Smyth, 1996), ha revolucionado la manera en la que se analizan las grandes bases de datos, dejando de lado la estadística convencional para adaptarse a los nuevos retos que presentan (Hand, 1998). A partir de ello, se han explorado diversas fuentes de información que podrían beneficiarse de estas metodologías, incluyendo las bases de datos geoespaciales que comparten varias características con las bases de datos secundarias,



normalmente utilizadas en la minería de datos (Baçã, Lobo, & Painho, 2005) pero que deben ser abarcadas de manera especial debido a sus propiedades como autocorrelación, heterogeneidad, entre otras (Openshaw, 1999).

El desarrollo de estas metodologías ha permitido desarrollar técnicas que hoy en día permiten generar minería de patrones dentro de bases de datos no solo espacialmente relacionadas, sino también incluyendo variables como tiempo, y una base multidimensional de información a ser analizada. De ahí es de donde se combinan estas metodologías con los cubos espacio-temporales. Estos fueron inicialmente conceptualizados por Hägerstrand (1970), donde el espacio es representado en un plano de dos dimensiones, mientras que el tiempo es representado en la tercera dimensión, correspondiente al eje-Z. Su función inicial fue identificar trayectorias entre puntos origen/destino, donde podían identificarse patrones referentes a la velocidad y sitios de parada (Gatalsky, Andrienko, & Andrienko, 2004). Por tanto, siguiendo su propósito inicial de análisis de patrones de movimiento, pueden identificarse tendencias dentro de sets de datos de movilidad para caracterizar en cierta medida los problemas de transporte en una ciudad.

El tráfico vehicular es por lo general una de las fuentes principales de contaminación ambiental en varias ciudades en países en vías de desarrollo y países desarrollados, acompañado además por problemas de congestión vehicular, accidentes, contaminación acústica, y una falta de seguridad (Molina *et al.*, 2004). Sin embargo, los tipos de soluciones que se da a estos problemas generalmente se basan en la expansión de carreteras (Pojani & Stead, 2015), dando una alta prioridad al transporte vehicular, incurriendo en una expansión urbana que se acompaña de un desarrollo económico y social que consecuentemente incrementa nuevamente el parque automotor, regresado al punto de partida.

Según la Cuarta Encuesta de Percepción Ciudadana sobre la Calidad de Vida en la ciudad de Cuenca para el año 2016, la ciudad presenta un elevado uso del automóvil, donde en promedio se movilizan 1.1 personas por vehículo particular (CCCV, 2016), adquiriendo el segundo lugar como medio de transporte prioritario (35.9%) después del transporte público (bus: 43.1%) (GAD Municipal de Cuenca, 2015) y con un crecimiento del parque automotor del 10% anual (CCCV, 2016). La movilidad no motorizada toma un papel menor, ocupando el tercer lugar el traslado a pie con 15.9%, pero ubicando a la bicicleta en el último lugar, después de taxis y motos (3.6% en conjunto), con solo un 1.6% de la población dando prioridad a este medio de transporte (GAD Municipal de Cuenca, 2015). En este contexto, se han generado distintos esfuerzos para el análisis de los patrones de movilidad de ciclistas y peatones en la ciudad, destacando la investigación “Pies y Pedales”, la cual analiza desde un punto de vista metodológico, comportamental y perceptual la movilidad no motorizada (Orellana, Hermida, & Osorio, 2017).

El presente estudio propone alimentar el eje metodológico del proyecto “Pies y Pedales”, el cual tiene como objetivo entender los patrones de comportamiento de ciclistas voluntarios mediante cubos espacio-temporales en la ciudad de Cuenca. El presente artículo se encuentra estructurado de la siguiente manera: a) descripción de la metodología de cubos espacio temporales utilizada para el área de estudio, b) presentación de resultados y una corta discusión de estos en el contexto de la ciudad y c) conclusiones, limitaciones y recomendaciones del estudio.

2. METODOLOGÍA

El análisis se enfoca principalmente en el uso de la caja de herramientas para la Minería de Patrones Espacio Temporales de ArcGIS, donde se permite la creación de cubos espacio-temporales para el análisis aglomerado de una base de datos geoespacial tipo 7 según Openshaw (1999), es decir, que contenga coordenadas geográficas, datos de tiempo y atributos múltiples. En la presente sección se realiza una caracterización del área de estudio, además de una explicación de la recolección de los datos de ciclistas voluntarios, así como el pre-procesamiento llevado a cabo para preparar la base de datos primaria para la creación de los cubos espacio-temporales para la agregación de puntos registrados y de las velocidades medias de circulación.

2.1. Área de estudio

El área de estudio corresponde a la ciudad de Cuenca, ciudad categorizada como intermedia por su superficie (70.59 km²) y población (331,888 habitantes (INEC, 2010)), además de emergente por su desarrollo económico estable. La ciudad fue seleccionada por la Iniciativa para Ciudades Emergentes Sustentables del Banco Interamericano de Desarrollo (BID) en el año 2013 para encaminar el desarrollo de la ciudad hacia un modelo de sustentabilidad en tres aspectos: cambio climático y ambiente, urbano, y sustentabilidad fiscal y gubernamental (Interamerican Development Bank, 2014). Dentro de esta iniciativa, se encuentra el eje de transporte, en el cual se han visto reflejados varios esfuerzos por parte de planificadores urbanos para fortalecer la infraestructura existente para el fomento de la movilidad no motorizada como un reto dentro del plan de movilidad local (GAD Municipal de Cuenca, 2015).

2.2. Recolección de datos y preprocesamiento

La iniciativa “Científicos en Pedales” del proyecto “Pies y Pedales” convocó a ciclistas urbanos voluntarios interesados en participar activamente para el mejoramiento de la movilidad dentro del área de estudio (Grupo de Investigación LactaLAB - Ciudades Sustentables, 2015). Uno de los experimentos de la iniciativa permitió recolectar información de viajes diarios origen/destino de 20 ciclistas voluntarios mediante un GPS (QSTARZ GPS Sports Recorder (Model: BT-Q1300ST)) durante dos periodos de dos semanas entre 19/05/2016 y 14/11/2016 (Orellana, 2016).

La base de datos fue pre-procesada inicialmente con una depuración de información inconsistente debido a errores de precisión del GPS (Thierry, Chaix, & Kestens, 2013), para luego derivar variables como velocidad y rodamiento, en función de la resolución temporal del GPS (10 segundos), para lo cual se utilizó el software R (R Core Team, 2017). El análisis se enfocó en la hora del día, para lo cual se agregó la información en una sola fecha, conservando los registros originales de tiempo para posteriormente hacer análisis de días laborables y fines de semana, debido a distintos comportamientos de los ciclistas (Jiang, Ferreira, & González, 2015).

2.3. Análisis de datos

Dentro de la caja de herramientas de Minería de Patrones Espacio Temporales, se encuentra la opción para crear cubos espacio temporales a partir de agregaciones de puntos. Un cubo espacio temporal es una poderosa manera de visualizar y analizar información espacio temporal al agregar información en celdas espacio-temporales. El tamaño de la celda depende en el nivel de análisis a realizarse, el cual fue establecido en una distancia de 100 metros y un intervalo temporal de 1 hora. Un conteo de los puntos GPS de los ciclistas dentro de cada celda nos dará el valor de agregación de las distintas celdas dentro del cubo. Además, se genera un segundo cubo en donde los valores de las celdas corresponden a las velocidades medias de los ciclistas. El formato de la información es NetCDF (formulario de datos comunes en red), el cual fue presentado por Rew & Davis (1990), constituye un formato de condensación de información que permite el almacenamiento y recuperación de información multidimensional.

Dentro de la herramienta se encuentra el subconjunto de Utilidades, a partir del cual se realizó la visualización tridimensional de los datos; y el análisis de puntos calientes emergentes, así como el análisis de valores atípicos locales en dos dimensiones. Estos últimos requieren de una distancia y periodo de tiempo de vecindad, los cuales fueron establecidos en 500 metros y 4 horas respectivamente. Al aplicar el estadístico G_i^* de Getis-Ord, el análisis de puntos calientes emergentes puede llegar a determinar 17 patrones de: puntos calientes, los cuales se refieren a celdas con altos valores, y puntos fríos que corresponden a celdas con valores bajos; discriminando además entre celdas esporádicas, oscilantes, consecutivas y otras categorías (Environmental Systems Research Institute, 2015a). Con el I Anselin local de Moran pueden realizarse el análisis de valores atípicos, donde se calculan puntuaciones Z y valores p para cada celda, de acuerdo con sus vecindades espaciales y temporales (Environmental Systems Research Institute, 2015b). Todos los análisis se realizaron en ArcGIS 10.5 (Environmental Systems Research Institute, 2016).

3. RESULTADOS Y DISCUSIÓN

El resultado de la agregación de puntos GPS y velocidades medias se presentan en dos cubos espacio-temporales distintos, los cuales son analizados en función de la hora del día, en un análisis global, pero también realizando una distinción entre días laborables y fines de semana que permiten comparar las diferencias entre los patrones de movimiento. La herramienta utilizada permite hacer una minería de una base de datos cuyo tamaño es difícilmente manejable en términos estadísticos convencionales (Hand, 1998), los cuales superan los 140 mil registros para el presente estudio, identificando los patrones de movimientos y cuyos resultados son visualmente poderosos ya que permiten al analista reconocer sitios y momentos de interés que permitirán un análisis más detallado de los mismos dentro de la planificación de la ciudad. El estudio se enfoca en aquellos sitios donde la acumulación de puntos GPS es alta y las velocidades bajas.

Las siguientes secciones presentan los resultados de la aplicación de la herramienta en el set de datos de ciclistas voluntarios en la ciudad de Cuenca, además de una corta discusión para cada análisis, los cuales se abordarán en el siguiente orden: 1) visualización tridimensional de los cubos espacio-temporales de puntos GPS y velocidades medias, 2) análisis bidimensional de valores atípicos, y 3) análisis de puntos calientes emergentes en dos dimensiones.

3.1. Aglomeración de información en cubos espacio-temporales

La visualización tridimensional de la información es una herramienta particularmente poderosa al momento de analizar los resultados en softwares que permiten alternar entre distintos puntos de vista, de manera que puedan identificarse patrones, generalmente subjetivos y muy ligados a la interpretación del analista, pero que forman parte de un análisis inicial de la información generada con la herramienta en estudio.

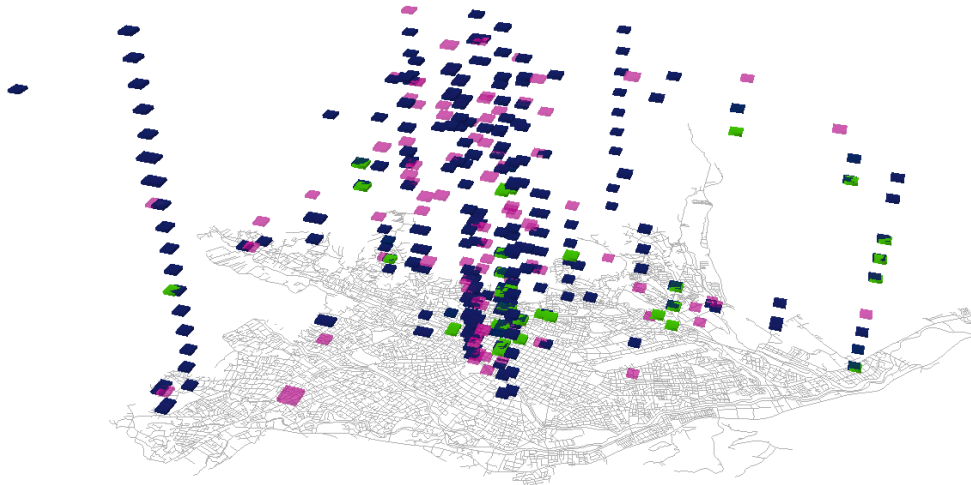


Figura 1. Visualización tridimensional del cubo espacio-temporal correspondiente a la aglomeración de puntos GPS. En azul se observa el análisis de la semana completa, en rosa para días laborables y en verde para fines de semana.

Al representar las celdas del cubo de puntos GPS que contienen el 90% de la información analizada, en el espacio y el tiempo, podemos apreciar que las aglomeraciones de puntos generalmente se dan hacia el centro de la ciudad y además tienden a aglomerarse en momentos del día cercanos al medio día (Fig. 1). El objetivo de la representación es identificar aquellos lugares en donde existe una gran afluencia de ciclistas, ya sea por la presencia de un punto de interés o un lugar de origen o destino común a los sujetos de estudio, que permita identificar sitios para la implementación de infraestructura necesaria para el

fomento del uso de la bicicleta como medio de transporte en la ciudad. En la misma ilustración pueden identificarse los valores correspondientes al análisis completo, así como al correspondiente a los días laborables y a los fines de semana. Salta a la vista que la representación de la semana completa y días laborables son bastante similares, inclusive superponiéndose unos a otros y dificultando en cierta medida su identificación, mientras que las celdas correspondientes a los fines de semana son menos densas y más esporádicas a lo largo del área y periodo de estudio.

Las mismas tendencias respecto al análisis por días de la semana pueden identificarse en el cubo correspondiente a las velocidades medias (Fig. 2), en donde se han representado aquellas velocidades por debajo de la media, que tienden a agruparse con celdas vecinas con valores también por debajo de la media, con el objetivo de identificar aquellos lugares donde los ciclistas disminuyen sus velocidades ya sea por la presencia de obstáculos en sus carriles de circulación, intersecciones peligrosas, presencia de sitios de interés, entre otros. En este caso se observa una figura mucho más desafiante a la vista, debido al efecto de oclusión (Wood, Dykes, & Slingsby, 2010) que se genera al superponer varias capas en un espacio tridimensional. Una exploración en un entorno SIG permite, sin embargo, identificar patrones no solo en el espacio sino también en el tiempo, identificando así que una mayor cantidad de celdas correspondientes a una disminución de la velocidad se ubican en horas coincidentes con alto tráfico vehicular, lo cual permitiría tomar medidas respecto a la seguridad de los ciclistas en el área de estudio.



Figura 2. Visualización tridimensional del cubo espacio-temporal correspondiente a la velocidad media de los ciclistas voluntarios. En azul se observa el análisis de la semana completa, en rosa para días laborables y en verde para fines de semana.

Aún si un entorno tridimensional permita explorar exhaustivamente los resultados generados, proyectarlos a un entorno bidimensional permite obtener una nueva perspectiva de análisis, en cierta medida disminuyendo el sesgo introducido por el analista al representar, por ejemplo, agrupaciones de información a manera de patrones de densidad, como se observa en la siguiente sección.

3.2. *Análisis de agrupación y detección de valores atípicos espaciales*

El análisis efectuado en base al I Anselin local de Moran permite identificar grupos y valores atípicos dentro de las celdas correspondientes a los cubos espacio-temporales. Siguiendo la tendencia de la sección pasada, se analizaron los grupos de valores altos, cuya nomenclatura en el algoritmo es HH y los valores atípicos espaciales elevados (HL) para el cubo de puntos GPS (Fig. 3); y el análisis contrario, para las velocidades medias (Fig. 4), es decir grupos de valores bajos (LL) y valores atípicos bajos (LH). La representación de estos resultados indica el porcentaje de los distintos casos, a lo largo del periodo

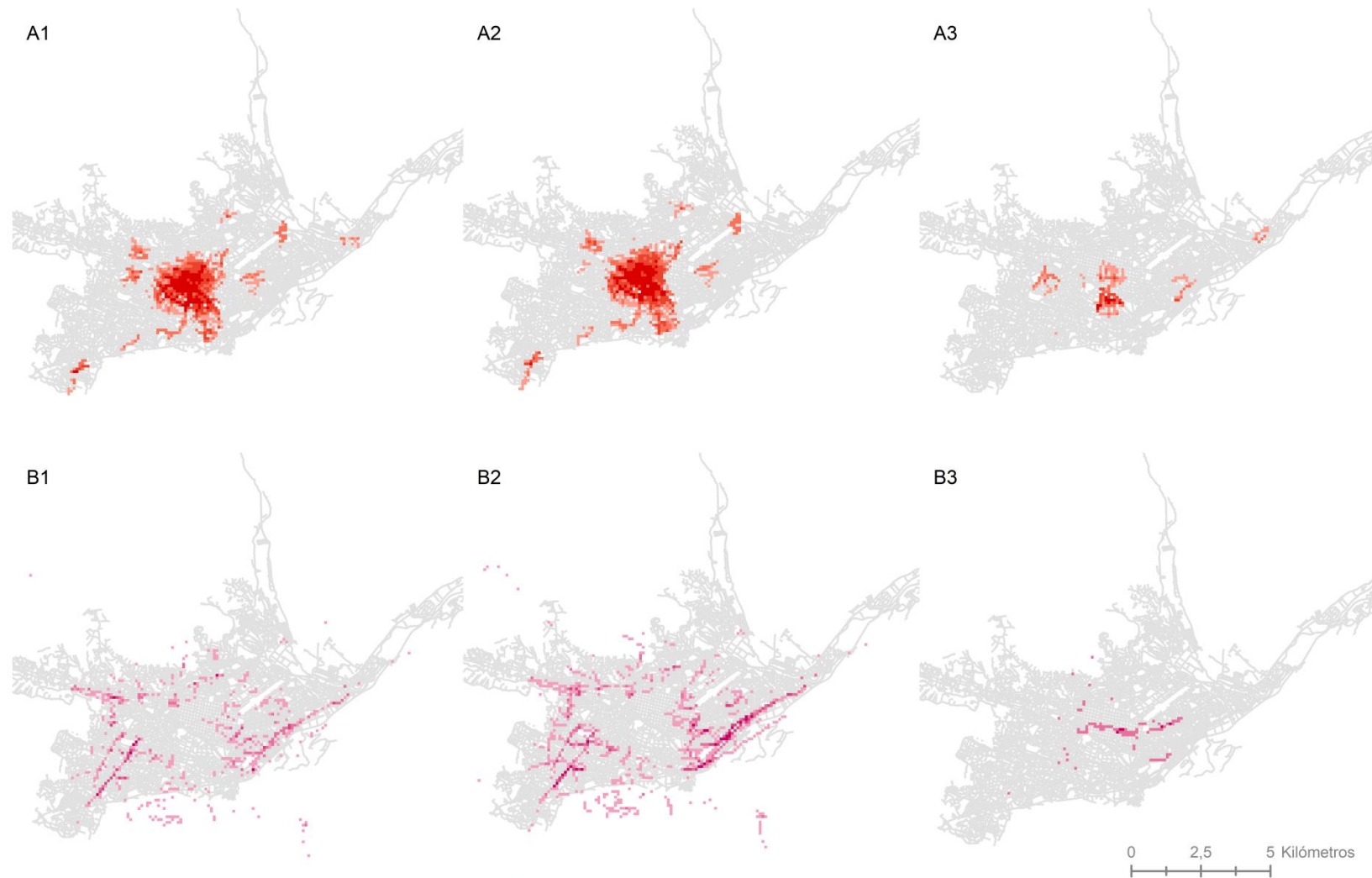


Figura 3. Identificación de grupos de valores altos (fila A) y valores atípicos espaciales altos entre vecinos de valores bajos (fila B) para el cubo de conteo de puntos GPS, diferenciados en el análisis de la semana completa (1), días laborables (2) y fines de semana (3). Colores más intensos indican mayor porcentaje de agrupación y valores atípicos a lo largo del día.

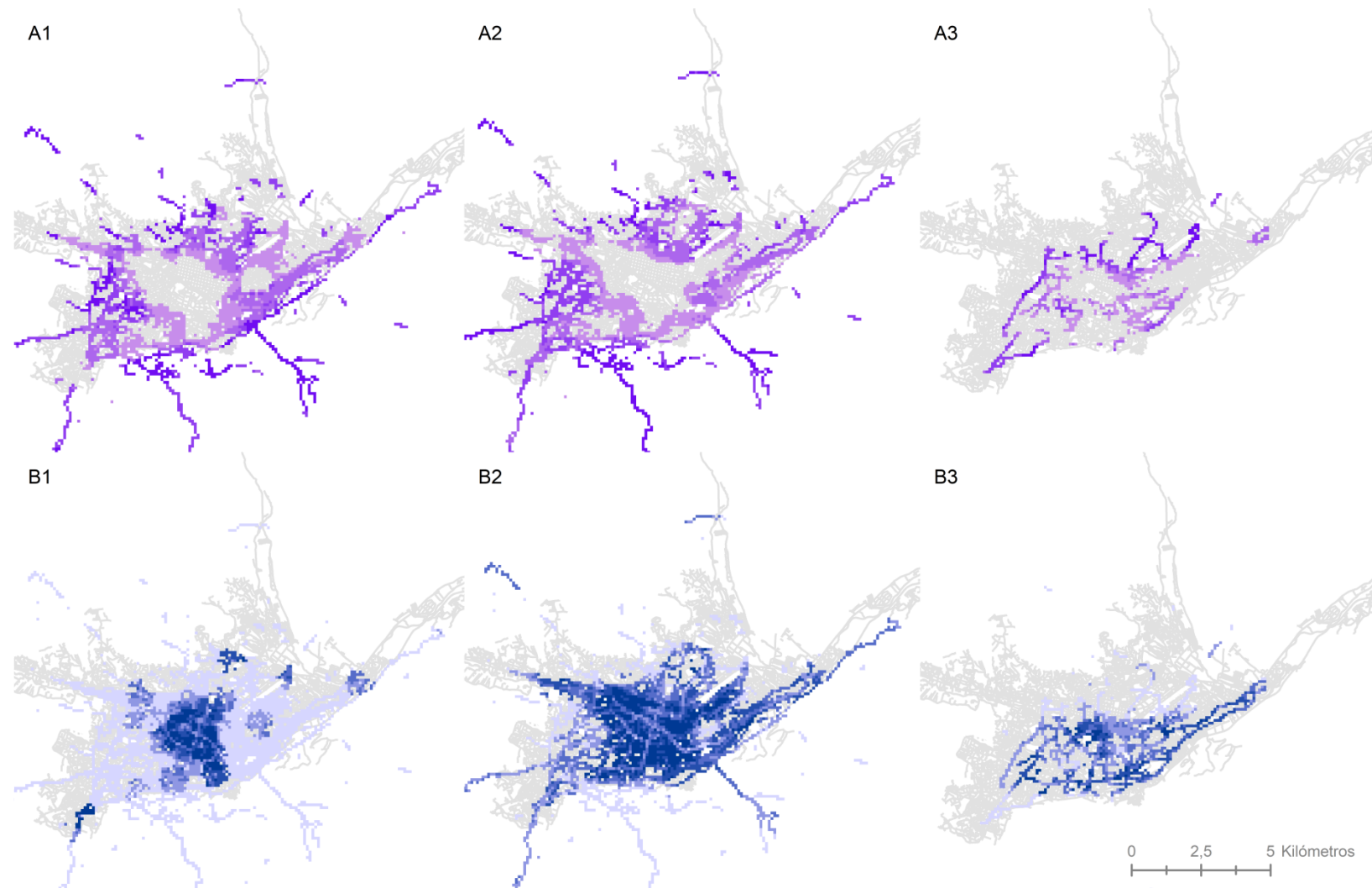


Figura 4. Identificación de grupos de valores bajos (fila A) y valores atípicos espaciales bajos entre vecinos de valores altos (fila B) para el cubo de velocidades medias, diferenciados en el análisis de la semana completa (1), días laborables (2) y fines de semana (3). Colores más intensos indican mayor porcentaje de agrupación y valores atípicos a lo largo del día.

completo de estudio, es decir en un solo día, de tal manera que se puede observar los lugares donde los ciclistas realizan paradas durante todo el día, y aquellos que corresponden únicamente a algunos momentos del día.

En la Figura 3, observamos la aglomeración de puntos en el caso A de agrupación, donde se observan tendencias similares para días laborables y la semana completa, mientras que los registros correspondientes a fines de semana son menores. De todas maneras, se observa que la concentración de información se da generalmente en el centro de la ciudad, en zonas como la Universidad de Cuenca, tanto en su campus Central como su campus Yanuncay, la Universidad del Azuay, Feria Libre y sectores cercanos al Control Sur y Parque Industrial en los extremos. Por su parte, el caso de valores atípicos presenta aquellos puntos que en el espacio son distintos a sus vecinos pero que de todas maneras son interesantes objetos de estudio al considerar que varios ciclistas coinciden en los mismos. De este análisis podemos principalmente identificar que calles, como la Avenida de las Américas y el margen del Río Paute, son sectores en donde se da una clara aparición de valores atípicos y en donde se podría considerar la inserción o el mejoramiento de infraestructura para el ciclista, como parqueaderos, ciclovías, entre otros.

Esta última tendencia analizada se identifica también en la Figura 4 para el caso de las agrupaciones de velocidades bajas, en donde se nota que hacia los sectores que rodean el centro de la ciudad, pueden encontrarse trayectos donde se puede intervenir a través de planes de movilidad. Para el caso de valores atípicos, por su parte, se puede apreciar que un análisis global identifica velocidades bajas en zonas centrales de la ciudad, mientras que para los días laborables se expande hacia otras zonas, identificando otros trayectos utilizados por los ciclistas, siendo, sin embargo, el fin de semana un ítem interesante al presentar valores atípicos bajos en las riberas de los ríos Tomebamba y Yanuncay y su unión con el río Paute.

3.3. Análisis de puntos calientes y fríos emergentes

Por último, se aplica la herramienta para identificación de patrones de puntos calientes y fríos en el área de estudio. Una gran cantidad de patrones son detectados, siendo más relevantes, para el análisis, aquellos correspondientes al cubo de velocidades medias. En la Figura 5 puede observarse nuevamente tendencias similares para el análisis global y los días laborables mientras que los fines de semana presentan principalmente patrones de puntos fríos oscilantes, significando que las velocidades medias fueron bajas hacia las últimas horas del día, pero que durante el transcurso de este se alternaron entre puntos calientes y fríos. Por tanto, se puede decir que las zonas identificadas en este caso (5c) presentan facilidades para la circulación de los ciclistas, pero que probablemente a lo largo del trayecto existan obstáculos o puntos de interés que obliguen a la persona a disminuir la velocidad. Se identifica además una zona cercana al parque industrial donde se observa una secuencia de puntos fríos consecutivos, pudiendo tratarse de intersecciones peligrosas en esta zona que dificultan la circulación. Finalmente, sectores cercanos a la ciudad clasificados como puntos fríos nuevos sugieren que las velocidades disminuyen al terminar el día.

Para el caso del análisis global (5a) y los días laborables (5b) se puede notar claramente que los patrones son mucho más variados y abarcan la totalidad del área de estudio. Se puede observar, por ejemplo, que en contraste con los resultados de la sección 3.2, el centro de la ciudad presenta generalmente puntos cálidos, indicando velocidades elevadas en estas zonas, con ciertas oscilaciones con puntos fríos en el transcurso del día. Sectores como el Hospital Regional, Feria Libre y calles del centro histórico como la Sangurima, presentan puntos fríos al terminar el día, pudiendo coincidir con horas de alta circulación vehicular que dificultan la circulación del ciclista.

Considerando que los participantes del experimento registraron únicamente sus viajes origen/destino (según lo solicitado) sin propósitos recreativos ni de deporte, podemos entender las claras diferencias que se presentan en los análisis de días laborables y fines de semana, en donde la circulación disminuye significativamente. Aun así, ciertos registros fueron identificados como puntos lejanos a la ciudad, ocurriendo en un fin de semana, los cuales son descartados del análisis, pudiendo considerarlos como errores de registro por parte de los voluntarios. Sin embargo, la tendencia es clara e indica que la bicicleta es un medio de transporte utilizado para moverse entre semana por la gran mayoría de zonas

de la ciudad de Cuenca, indicando una necesidad importante de infraestructura a tomar en cuenta en los planes de movilidad.

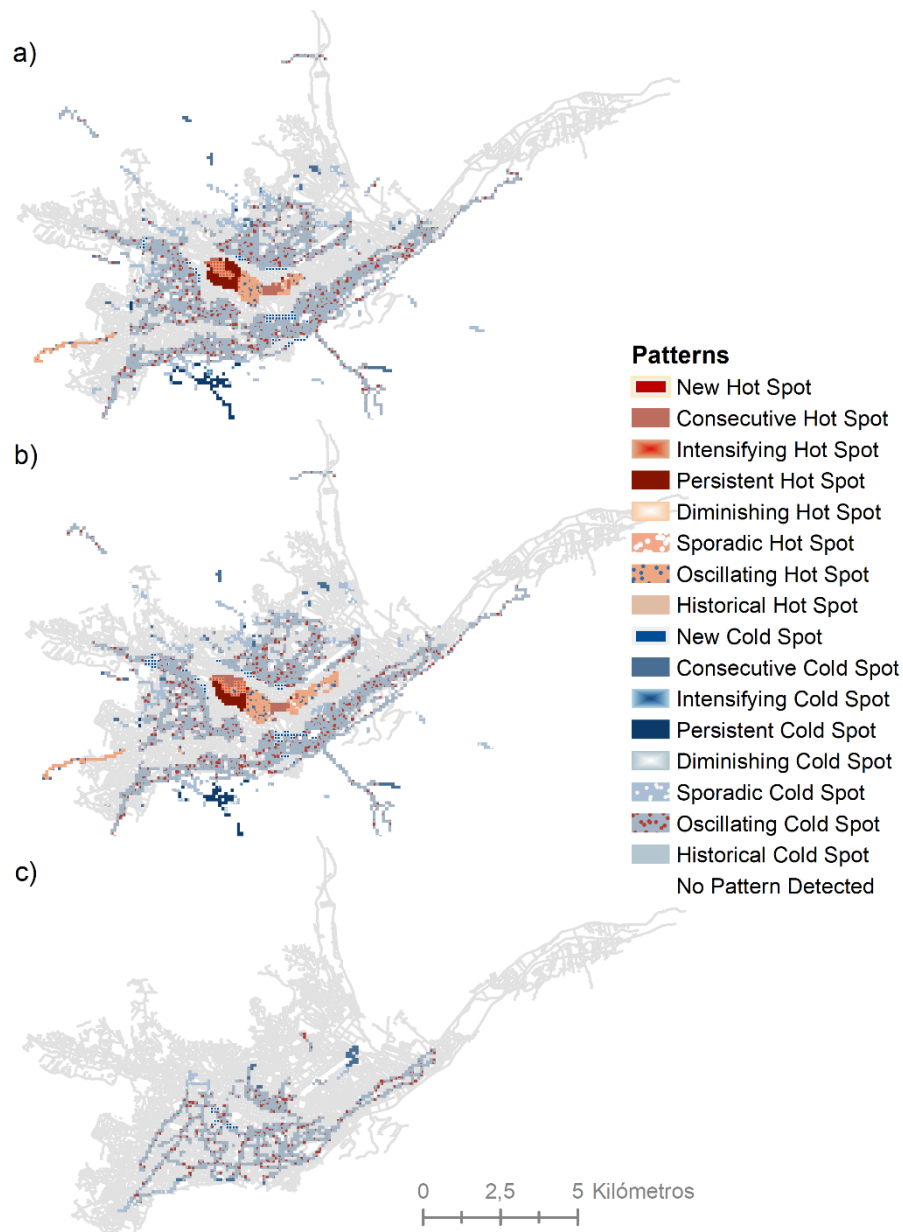


Figura 5. Tipificación de patrones de puntos calientes y fríos para el cubo de velocidades medias, presentando los resultados de la semana completa (a), días laborables (b) y fines de semana (c).

4. CONCLUSIONES Y RECOMENDACIONES

En el presente estudio, se han aplicado exitosamente técnicas de minería de datos para descubrir conocimiento (KDD) sobre los patrones de comportamiento espacial y temporal de ciclistas en la ciudad de Cuenca utilizando un set de datos de GPS recolectados voluntariamente por un grupo de participantes durante 15 días, considerando que la ciudad mantiene aún un tamaño manejable en términos de superficie, para la planificación de movilidad, sin necesidad de una sectorización para su evaluación. En la ciudad se han llevado a cabo varios esfuerzos para el mejoramiento de las redes de transporte

motorizado y no-motorizado, sin embargo, aún existen ciertos puntos que pueden beneficiarse de este tipo de estudios.

Las diferentes utilidades dentro de la caja de herramientas ha permitido entender los patrones de circulación de los participantes en el experimento en una manera global, pudiendo identificar sectores, no solo en el espacio sino también en el tiempo, en donde se puedan incluir medidas a los planes de movilidad como, por ejemplo, la implementación o mejoramiento de infraestructura para bicicletas (ciclovías, parqueaderos), introducción de mecanismos para aplacar intersecciones peligrosas, o inclusive la dirección de esfuerzos por parte de los agentes de tránsito para que en horas pico permita que los ciclistas puedan circular de manera segura.

Las limitaciones del estudio se centran en la baja precisión de los GPS utilizados, los cuales pueden introducir errores en el análisis si no se invierten exhaustivas horas de preprocesamiento a la base de datos inicial. Otras fuentes de errores se caracterizan por la inexperiencia de los voluntarios al momento de levantar la información o simplemente descuidos al no apagar el GPS cuando los viajes hayan terminado. Además, al ser un experimento con participantes voluntarios, no se puede garantizar una representatividad de la población que utiliza la bicicleta como medio de transporte, ya que sesgos socioeconómicos entre otros factores pueden introducirse en la identificación de patrones. Una futura investigación en esta área debería centrarse en la identificación de patrones de fuentes de datos colaborativos, conocidos como “*crowdsourced data*”, para la evaluación del rendimiento de la herramienta y comparación de sus resultados.

AGRADECIMIENTOS

El presente trabajo es parte del proyecto “Estudio de los patrones de movilidad de ciclistas y peatones para una movilidad sustentable - Pies y Pedales”, financiado por la Dirección de Investigación de la Universidad de Cuenca (DIUC).

REFERENCIAS

- Baço, F., Lobo, V., Painho, M. (2005). *On the particular characteristics of spatial data and its similarities to secondary data used in data mining*. Retrieved from http://www.novaims.unl.pt/docentes/vlobo/Publicacoes/2_13_lobo05_secondary_data.pdf
- CCCV. (2016). *Calidad de vida en Cuenca. Cuarta encuesta de percepción ciudadana*. Retrieved from [http://www.cccv.ec/download/20170316_EncuestaCCCV2016\(Sitio\).pdf](http://www.cccv.ec/download/20170316_EncuestaCCCV2016(Sitio).pdf)
- Environmental Systems Research Institute. (2015a). *How emerging hot spot analysis works*. Retrieved from <http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/learnmoreemerging.htm>
- Environmental Systems Research Institute. (2015b). *How local outlier analysis works*. Retrieved from <http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/learnmorelocaloutlier.htm>
- Environmental Systems Research Institute. (2016). *ArcGIS Desktop 10.5*. ESRI. Retrieved from <http://www.esri.com/arcgis/about-arcgis>
- Fayyad, U. U., Piatetsky-Shapiro, G., Smyth, P. (1996). *Knowledge discovery and data mining: Towards a unifying framework*. Int. Conf. on Knowledge Discovery and Data Mining, pp. 82-88. <https://doi.org/10.1.1.27.363>
- GAD Municipal de Cuenca. (2015). *Plan de movilidad y espacios publicos (Tomo I)*. Cuenca, Ecuador.
- Gatalsky, P., Andrienko, N., Andrienko, G. (2004). *Interactive analysis of event data using space-time cube*. Proc. 8th Int. Conf. on Information Visualisation, IV 2004, 145-152. <https://doi.org/10.1109/IV.2004.1320137>

- Grupo de Investigación LactaLAB - Ciudades Sustentables. (2015). *Científicos en pedales - Científicos a pies*. Retrieved from <https://lactalab.ucuenca.edu.ec/investigacion/pies-y-pedales/cientificos-en-pedales/>
- Hägerstraand, T. (1970). What about people in regional science? *Papers in Regional Science*. <https://doi.org/10.1111/j.1435-5597.1970.tb01464.x>
- Hand, D. J. (1998). General data mining: Statistics and more? *The American Statistician*, 52(2), 112-118.
- INEC. (2010). Información Censal.
- Interamerican Development Bank. (2014). *Methodological guide: Emerging and sustainable cities initiative* (2nd ed.). IDB, 161 p.
- Jiang, S., Ferreira, J., González, M. C. (2015). *Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore*. *IEEE Transactions on Big Data*, 1-13. <https://doi.org/10.1109/TBDDATA.2016.2631141>
- Molina, L. T., Molina, M. J., Slott, R. S., Kolb, C. E., Gbor, P. K., Meng, F., ... Gertler, A. W. (2004). Air quality in selected megacities. *Journal of the Air & Waste Management Association*, 54, 1-73. <https://doi.org/10.1080/10473289.2004.10471015>
- Openshaw, S. (1999). *Geographical data mining: key design issues*. 4th International Conference on GeoComputation, Virginia, pp. 1-9.
- Orellana, D. (2016). Métodos para el análisis de patrones de movilidad no motorizada. In: *Comunidades Urbanas Energéticamente Eficientes* (1st ed., pp. 141-147). EDUFES.
- Orellana, D., Hermida, C., Osorio, P. (2017). Comprendiendo los patrones de movilidad de ciclistas y peatones: Una síntesis de literatura. *Revista Transporte Y Territorio*, 16, 168-184.
- Pojani, D., Stead, D. (2015). Sustainable urban transport in the developing world: Beyond megacities. *Sustainability*, 7(6), 7784-7805. <https://doi.org/10.3390/su7067784>
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from <https://www.r-project.org/>
- Rew, R., Davis, G. (1990). NetCDF: an interface for scientific data access. *IEEE Computer Graphics and Applications*, 10(4), 76-82. <https://doi.org/10.1109/38.56302>
- Thierry, B., Chaix, B., Kestens, Y. (2013). Detecting activity locations from raw GPS data: a novel kernel-based algorithm. *International Journal of Health Geographics*, 12, 14. <https://doi.org/10.1186/1476-072X-12-14>
- Wood, J., Dykes, J., Slingsby, A. (2010). Visualisation of origins, destinations and flows with OD maps. *The Cartographic Journal*, 47(2), 117-129. <https://doi.org/10.1179/000870410X12658023467367>