

FUSIÓN DE DATOS PARA LA PLANIFICACIÓN DE SERVICIOS URBANOS SOSTENIBLES – CASOS DE USO EN TRANSPORTE, MEDIO AMBIENTE Y TURISMO

**Miguel Picornell
Ricardo Herranz
Manuel Álvarez
Iris Galloso
*Kineo Mobility Analytics, S.L***

Artículo Recibido: 28/04/2018

Artículo Aceptado: 14/09/2018

Resumen

La planificación y gestión de ciudades sostenibles requiere entender el comportamiento de la población: la planificación y operación del transporte necesita información precisa, fiable y actualizada sobre la demanda de viajes; la definición de estrategias efectivas que mitiguen la exposición a la contaminación requiere conocer la distribución espacio-temporal de la población a lo largo del día; el diseño de una oferta turística sostenible necesita conocer los patrones de actividad de los turistas. Las fuentes de datos tradicionales, basadas fundamentalmente en encuestas, proporcionan información muy valiosa, pero no están exentas de inconvenientes. En general, las encuestas resultan caras y lentas de realizar, lo que limita el tamaño de la muestra y su frecuencia de actualización. Además, están sujetas a respuestas incorrectas e imprecisas y dependen de la disposición a responder de los entrevistados.

La posibilidad de recoger datos geolocalizados procedentes de dispositivos móviles personales, de manera dinámica y a un coste inferior al de los métodos tradicionales, abre nuevas oportunidades para superar estos problemas. En particular, los datos procedentes de las redes de telefonía móvil permiten obtener un tamaño de muestra uno o dos órdenes de magnitud superior al de las encuestas convencionales. Asimismo, el alto nivel de penetración de los servicios móviles en prácticamente todos los estratos de la sociedad garantiza muestras muy bien distribuidas. La elevada granularidad temporal de los datos permite determinar con detalle la localización del móvil a lo largo del día y su resolución espacial resulta en general adecuada para estudios a escala urbana y metropolitana.

WPSReview International on Sustainable Housing and Urban Renewal (RI-SHUR)

La presente comunicación aborda el análisis de registros anonimizados de telefonía móvil y su fusión con otras fuentes de datos para proporcionar información sobre actividad y movilidad de la población. Se discute la necesidad de enriquecer la información obtenida mediante telefonía móvil con diversas fuentes de datos, se expone la metodología y se presentan tres casos de uso: el estudio de la movilidad general y en transporte público en Málaga mediante la fusión de datos de telefonía móvil y datos de tarjeta inteligente de transporte; el estudio de la exposición a la contaminación en Madrid a partir de mapas dinámicos de población y de un modelo de dispersión de contaminantes; y el análisis de los visitantes al Parque Nacional de la Sierra de Guadarrama a partir de datos de telefonía móvil y datos de aforo.

Abstract

The planning and management of sustainable cities requires understanding the behavioural patterns of the population: transport planning and operation requires accurate, reliable and updated travel demand information; the definition of effective strategies to mitigate the exposure to air pollutants needs information on the spatio-temporal distribution of the population along the day; tourist's activity and mobility patterns are essential to design a sustainable touristic offer. Traditional data collection methods, such as household travel surveys, provide rich travel and demographic data, but they also suffer from shortcomings: they depend on users' willingness to answer, people may provide incorrect or imprecise answers, and they are expensive and require months to complete, which limits the size of the sample and the frequency with which information is updated.

New digital data sources make it possible to complement and/or replace traditional travel surveys, overcoming some of their main limitations. In particular, mobile phone records can give access to a sample that is usually at least one order of magnitude higher than the one provided by traditional sources, and that is also well distributed across the different socioeconomic segments, given the high penetration of mobile phone services. Additionally, the high temporal granularity of mobile phone data allows us to determine in detail the location of the device along the day and their spatial resolution is in general suitable to study population's behaviour at urban and metropolitan scale.

This contribution focuses on the analysis of anonymised mobile phone data and their fusion with other data sources to provide information on population's activity and mobility. We discuss the need to enrich the information obtained from mobile telephone data with different data sources and we described the methodology followed in three use cases: the study of mobility, including public transport mobility, in the city of Málaga using mobile phone data and intelligent transport card data; the study of population exposure to air pollutants in the city of Madrid using dynamic population maps and an air pollutants dispersion model; and the characterisation of the visitors to the Parque Nacional de la Sierra de Guadarrama using mobile phone data and people counts.

WPSReview International on Sustainable Housing and Urban Renewal (RI-SHUR)

Palabras clave

fusión de datos; datos de telefonía móvil; patrones de movilidad y actividad; transporte; calidad del aire; turismo

Key Words

mix of data, mobile phone data; pathers of mobility and activity; transport; air quality; tourism.

1.- Introducción

El conocimiento de los patrones de conducta de la población es esencial para el diseño de políticas sostenibles. Esta información ha sido tradicionalmente obtenida a través de encuestas. Las encuestas proporcionan información muy completa, pero adolecen de algunas limitaciones. La realización de encuestas es costosa en términos de tiempo y dinero, lo que conlleva que el tamaño de la muestra sea reducido y en muchos casos la información esté desactualizada. Además, la dependencia de la disponibilidad a contestar del entrevistado y las imprecisiones de las respuestas pueden producir muestras sesgadas (por ejemplo, capturando sólo usuarios con mayor disponibilidad de tiempo) o información incompleta (muchas veces los entrevistados olvidan reportar algunas actividades, estiman de manera muy aproximada tiempos de viaje, etc.).

La gran cantidad de datos producidos por el uso generalizado de dispositivos móviles geolocalizados permite obtener información de forma continua sobre la movilidad y actividad de las personas, así como a observar, en tiempo real, su reacción a medidas o fenómenos puntuales. A lo largo de la última década, distintos estudios han abordado el análisis de la movilidad a partir de las trazas geolocalizadas procedentes de teléfonos móviles [1,2], redes sociales de Internet [3], registros de uso de tarjetas de crédito [4] o tarjetas inteligentes de transporte [5], entre otros ejemplos. Particularmente interesantes resultan los datos procedentes de las redes de telefonía móvil, gracias al tamaño muestral que proporcionan y a su elevada resolución temporal. Diferentes ejercicios de validación llevados a cabo en los últimos años han demostrado que los datos de telefonía móvil, adecuadamente tratados, permiten extraer matrices origen-destino y otras estadísticas de viajes coherentes con las obtenidas mediante encuestas [6]. Por otro lado, los datos de telefonía móvil presentan también algunas limitaciones. En particular, la resolución espacial de los registros de telefonía, en general localizados al nivel de las celdas de la red, hace que no siempre sea posible la identificación del modo de transporte utilizado y la ruta elegida en los desplazamientos identificados, especialmente en el caso de desplazamientos de corta distancia en entornos urbanos. Una alternativa para superar esta limitación es el uso de los datos procedentes de las tarjetas inteligentes de transporte público para extraer las matrices de viajes en transporte público [7,8,9,10].

WPSReview International on Sustainable Housing and Urban Renewal (RI-SHUR)

En la presente comunicación se discuten tres ejemplos de casos prácticos donde los datos de telefonía móvil han sido fusionados con distintas fuentes de datos para obtener información sobre la movilidad y actividad de la población y se han empleado en distintas aplicaciones relacionadas con la planificación urbana: transporte y movilidad, medioambiente y turismo. En la Sección 2 se detallan el tipo de datos empleados y se describen y discuten los tres casos de uso. En la Sección 3 se presentan las conclusiones extraídas de los casos de uso sobre el potencial del uso de datos provenientes de telefonía móvil para la planificación y gestión de ciudades sostenibles.

2.-Datos y descripción de casos de uso

2.1.-Fuentes de datos

A continuación se describen las características de los distintos datos empleados en los casos de uso. Algunas fuentes de datos, como los datos de telefonía móvil y los datos del Padrón Municipal de Habitantes, son comunes a todos los casos, mientras que otras fuentes son específicas para cada caso.

A) Datos de telefonía móvil. Se han empleado datos de registros de telefonía móvil anonimizados proporcionados por Orange España. Los datos incluyen los registros generados cada vez que se produce un evento de interacción de un terminal móvil con la red, ya sea al efectuar o recibir una llamada, enviar un mensaje de texto o iniciar una sesión de datos. Los datos incluyen también determinados eventos pasivos, tales como cambios de áreas de cobertura. Cada registro contiene, entre otros, los siguientes campos:

- Id anonimizado del usuario
- Hora del evento
- Antena a la que se conecta el dispositivo en el momento del evento
- Tipo de evento: llamada, sms, datos.

La muestra proporcionada corresponde aproximadamente a un 27% de la población española. Para la mayoría de los usuarios la granularidad temporal es de entre 20 y 30 minutos, mientras que la granularidad espacial depende de la densidad de celdas de telefonía, variando desde entre unas decenas o cientos de metros en entornos urbanos hasta varios kilómetros en zonas rurales con menor densidad de población.

B) Datos del Padrón Municipal de Habitantes. Los datos de número de habitantes a nivel sección censal se utilizan como marco muestral, para expandir la muestra obtenida con los datos de telefonía al total de la población.

C) Datos de tarjeta inteligente de transporte. Estos datos fueron utilizados para el estudio de la movilidad en Málaga. Los datos de la tarjeta de transporte incluyen la siguiente información:

- Id anonimizado de la tarjeta.

WPSReview International on Sustainable Housing and Urban Renewal (RI-SHUR)

- Información sobre la hora de validación de la tarjeta.
- Localización de la validación, que puede corresponder a la localización de la estación o a la posición del vehículo en el que se produjo la validación. En este último caso, la localización se obtiene a través del SAE¹
- Tipo de validación. Esta puede ser entrada o salida. Para los autobuses sólo se dispone de la información de entrada.

Otra información sobre la red de transporte público, no asociada a la tarjeta de transporte, pero relevante para el estudio, incluye datos de la red y de la oferta de servicios (paradas, incluyendo posición, y líneas existentes), datos del SAE, y valores totales de demanda (incluyendo tanto a los viajeros usuarios de la tarjeta de transporte como a los que utilizan otros títulos de transporte).

D) Datos de contaminantes del aire en Madrid. Se ha utilizado la información de concentración de NOx medida en las distintas estaciones del Ayuntamiento de Madrid, que se ha empleado para obtener los niveles de contaminantes en cada zona de la ciudad.

E) Conteos de visitantes al Parque Nacional de la Sierra de Guadarrama. Estos datos incluyen datos de aforos de vehículos en los aparcamientos del parque, contadores de pisada y encuestas presenciales en los centros de visitantes del parque.

F) Mapas de usos del suelo. Mapas de uso de suelo provenientes del Sistema de Información Sobre Ocupación del Suelo Español (SIOSE), que se han utilizado para refinar la localización de algunas actividades.

2.2.- Metodología y casos de uso

2.2.1.- Obtención de patrones de movilidad y actividad a partir de registros anonimizados de telefonía móvil

El análisis de los datos de telefonía móvil comprende los siguientes subprocesos:

A) Selección de la muestra. Se seleccionan aquellos usuarios que presenten una frecuencia de registros suficiente para obtener patrones de movilidad y actividad con fiabilidad. La muestra resultante representa entre un 15 y un 20% de la población total de España. En función de las necesidades del estudio y de las propiedades de los distintos usuarios, de los usuarios potenciales seleccionan aquellos usuarios que formarán parte de la muestra efectiva.

B) Identificación de lugar de residencia y trabajo Para cada usuario válido se utilizan datos históricos para construir un perfil de actividad. Este perfil recoge las horas de mayor y menor actividad, así como las zonas más visitadas y el tiempo promedio de estancia en cada una de ellas. A partir del perfil de actividad, y cruzando las

¹ SAE sistema de ayuda a la explotación. El SAE contiene la información GPS de cada vehículo.

WPSReview International on Sustainable Housing and Urban Renewal (RI-SHUR)

localizaciones con los mapas de uso de suelo se obtiene la residencia y el lugar de trabajo del usuario.

C) Estimación de movilidad y actividad. Para el o los días de estudio, se reconstruyen la posición de las actividades y los viajes realizados por cada uno de los usuarios incluidos en la muestra de cada día. Se entiende como 'viaje' el desplazamiento entre las localizaciones de dos actividades consecutivas, y como 'actividad' una interacción o conjunto de interacciones con el entorno que tienen lugar en una misma localización y que motivan que el individuo se desplace hasta allí. La información obtenida del perfil de actividades del usuario se combina con distintos criterios de tiempos de estancia para identificar actividades y viajes a partir de los registros del día de estudio, filtrando las estancias intermedias subordinadas al viaje y realizadas entre etapas del mismo (por ejemplo, una parada intermedia en un viaje por carretera).

D) Elevación de la muestra al total de la población. Utilizando información sociodemográfica, la información se expande al total de la población. La opción por defecto es la utilización de factores de elevación basados en el hogar a nivel de sección censal, pero la solución permite otras aproximaciones en función de las características del estudio y la información disponible.

2.2.2.- Movilidad en la ciudad de Málaga

A) Movilidad general. La movilidad general en la ciudad de Málaga fue obtenida a partir de los registros de telefonía móvil:

1. Se obtuvo la residencia de todos los usuarios potenciales de la muestra y se filtraron los residentes en Málaga.
2. Se obtuvieron los viajes a nivel muestral para un día laborable promedio. Una vez obtenidos los viajes, se elevó la muestra al total de la población de la ciudad.
3. Los viajes se agregaron en matrices origen destino según una zonificación acordada con el cliente. Estas matrices se segmentaron en tres franjas horarias: hora punta de la mañana, hora punta de la tarde y hora valle.

B) Movilidad en transporte público. Para este estudio se combinaron datos de la tarjeta inteligente de transporte y datos de telefonía móvil. Para la obtención de la movilidad en transporte público se siguieron los siguientes pasos:

1. Obtención de parada de subida. Ésta se obtiene directamente de los datos proporcionados por la tarjeta, ya sea de forma explícita o a partir del SAE, en los casos en que sólo se dispone de la información del coche en que se ha validado.
2. Estimación de parada de bajada. En los casos en los que la información de la parada de bajada no está disponible (por ejemplo, los autobuses), para la estimación de la parada de bajada se usa información longitudinal de varios días (por ejemplo, la posición de la parada de subida en un tiempo subsecuente es usada para estimar la posición de bajada del viaje anterior). Se emplean criterios de tiempos de espera y de estancia para distinguir entre bajada en parada de destino y bajada en parada de transbordo.

WPSReview International on Sustainable Housing and Urban Renewal (RI-SHUR)

3. Elevación de la muestra. La muestra de viajes de transporte público se expandió usando la información de viajes totales.
4. Estimación de origen real y destino final. Las matrices obtenidas contienen orígenes y destinos a nivel de parada. Para la estimación del origen y destino real del viaje se reconstruyen las áreas de atracción de cada parada a partir de los datos de telefonía móvil y criterios de accesibilidad (distancia a la parada desde las distintas zonas). Finalmente, el origen y destino final de los usuarios en cada parada es asignado en las distintas zonas pertenecientes al área de atracción en función del potencial de atracción de cada zona, estimado a partir de la información de uso de suelo.

2.2.3.- Exposición a la contaminación en la ciudad de Madrid

En este caso de uso se estudió la exposición real de los habitantes de la ciudad de Madrid a las emisiones de NOx. Normalmente la exposición de los habitantes a los contaminantes se estima en base a los niveles de contaminantes en el lugar de residencia. Esta medida no refleja la exposición real, ya que en muchas ocasiones pasamos la mayor parte del día fuera de nuestro lugar de residencia. En este estudio se hizo una estimación dinámica de la exposición a contaminantes considerando la posición real de la población a distintas horas del día.

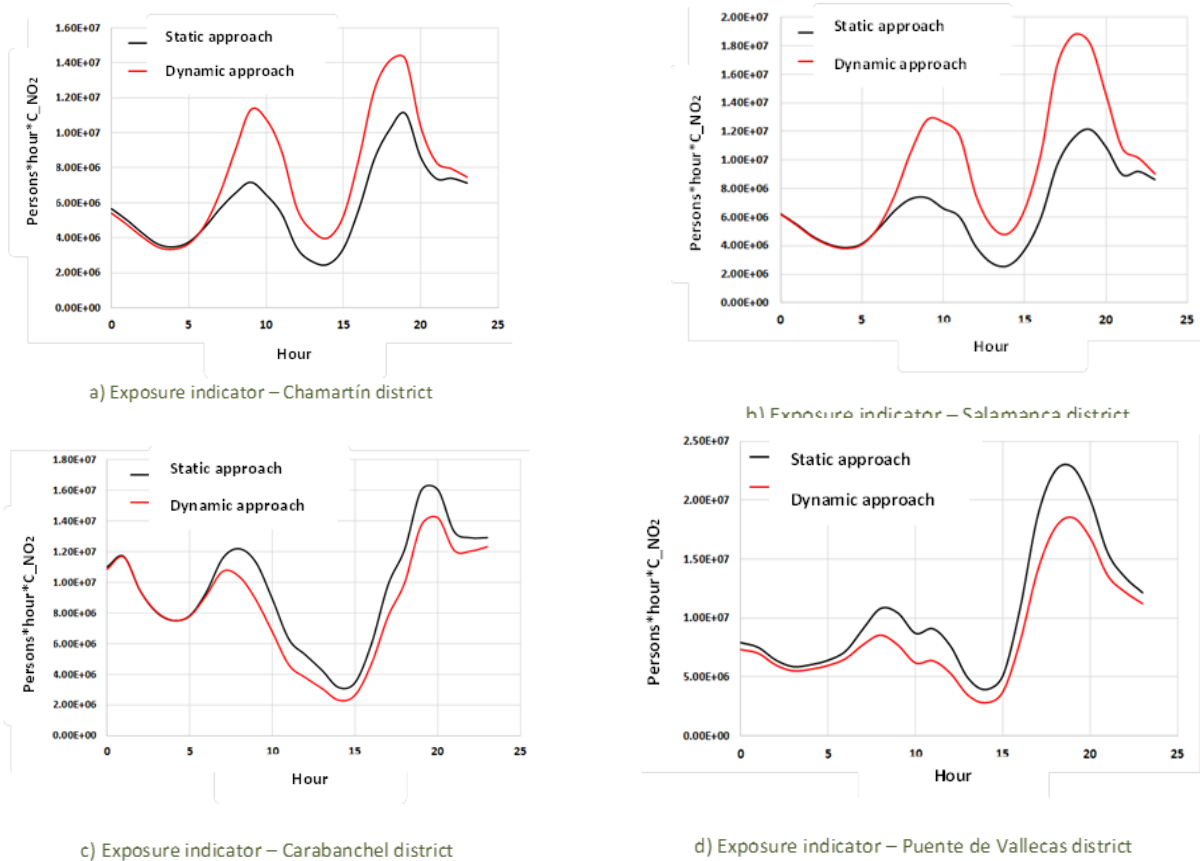
Los niveles de NOx en las distintas zonas de la ciudad fueron estimados con un modelo de dispersión de partículas desarrollado por el Departamento de Ingeniería Química y Medioambiental de la UPM con datos obtenidos de las estaciones de medición de contaminantes del ayuntamiento de Madrid.

La presencia de personas en las distintas zonas de la ciudad a diferentes horas del día se obtuvo a partir de los datos de telefonía móvil, según la metodología descrita en la Sección 2.2.1. Con esta información se obtuvo un índice de exposición, definido como número de personas por hora multiplicado por la concentración de contaminantes. En la Figura 1 se puede observar una comparativa del valor de este indicador obtenido usando el lugar de residencia y la posición real de las personas a las distintas horas del día en distintas zonas de la ciudad.

Como podemos observar en la figura, estas dos curvas coinciden por la mañana y por la noche. Sin embargo, para la mayor parte de las horas del día son muy distintas. Así, hay zonas donde se está subestimando la exposición real a contaminantes, principalmente zonas con mucha actividad económica, mientras que para zonas de carácter residencial se está sobreestimando.

WPSReview International on Sustainable Housing and Urban Renewal (RI-SHUR)

Figura 1. Comparativa del indicador de exposición para distintas horas del día calculado en base al lugar de residencia y a la posición real de la población.



2.2.4 Análisis de los visitantes al Parque Nacional de la Sierra de Guadarrama

Tradicionalmente, la afluencia de visitantes al parque se ha calculado mediante una combinación de datos de aforo en los estacionamientos y contadores de pisadas. Así, gracias a los aforos de parking se puede estimar el número total de visitantes, mientras que los contadores de pisadas dan información sobre las rutas seguidas. Adicionalmente, mediante encuestas de tipificación se pueden obtener la procedencia, edad, género y opiniones de los visitantes al parque. Aunque sean complementarios y se puedan combinar, estos datos no siempre son capaces de capturar la imagen completa sobre los visitantes y su movilidad dentro del parque. Así, no todos los visitantes llegan al parque en coche, algunos aparcan fuera del estacionamiento, y los contadores de pisadas no son capaces de capturar todas las rutas visitadas o a los visitantes que no siguen ninguna ruta. Por otro lado, el tiempo de estancia en el parque no se puede detectar de forma fiable con ninguna de las fuentes mencionadas anteriormente. En este caso de uso, estudiamos los flujos de visitantes al parque usando datos de telefonía móvil. Empleando la metodología descrita en la Sección 2.2.1, se obtuvo la procedencia de los visitantes al igual que la presencia de estos en el parque y la duración de su estancia. Los datos de encuestas, aforos y conteos de

WPSReview International on Sustainable Housing and Urban Renewal (RI-SHUR)

pisadas se utilizaron para corregir la muestra en los casos en que la granularidad espacial de los datos móviles no permitía distinguir correctamente entre usuarios ubicados dentro y fuera del parque. Los datos obtenidos de la telefonía móvil permitieron tener una información completa sobre el perfil (procedencia, edad y género), el tiempo de estancia y el número de visitantes, así como la recurrencia de las visitas, información que resulta complicado obtener de forma fiable a partir de las fuentes tradicionales.

3.- Conclusiones

En esta ponencia hemos presentado tres ejemplos distintos del potencial que tiene la explotación de los datos de dispositivos móviles personales obtenidos de forma pasiva, en particular los datos provenientes de las redes de telefonía móvil, para la planificación y diseño de políticas de desarrollo urbano. Los casos de uso presentados muestran que, si bien los datos de telefonía proporcionan una información muy rica, por sí solos no son suficientes para obtener una imagen completa de la movilidad. Es necesario por tanto desarrollar metodologías y algoritmos de fusión de datos, para integrar los datos de telefonía con otras fuentes relevantes para el problema a analizar. Por último, cabe resaltar que la utilización de datos de telefonía móvil abre interesantes oportunidades para la planificación y la gestión de servicios urbanos, tales como la posibilidad de obtener información permanentemente actualizada con un coste reducido o la oportunidad de analizar periodos especiales (Navidad, Semana Santa, etc.) y eventos concretos que influyen significativamente en las dinámicas de la ciudad.

4.- Referencias bibliográficas

- [1] S. Çolak, L.P. Alexander, B.G. Alvim, S. R. Mehndiratta, M.C. González (2015), "Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities", *Transportation Research Record: Journal of the Transportation Research Board* 2526, pp. 126-135.
- [2] M. Picornell, T. Ruiz, M. Lenormand, J. J. Ramasco, T. Dubernet, E. Frías-Martínez, "Exploring the potential of phone call data to characterize the relationship between social network and travel behavior", *Transportation*, vol. 42, nº 4, pp. 647-668, 2015.
- [3] M. Lenormand, A. Tugores, P. Colet, J. J. Ramasco, "Tweets on the Road", *PLoS ONE* 9(8): e105407, 2014.
- [4] S. Sobolevsky, I. Sitjo, R. Tachet Des Combes, B. Hawelka, J. Murillo Arias, C. Ratti, "Money on the move: Big Data of bank card transactions as the new proxy for human mobility patterns and regional delineation. The case of residents and foreign visitors in Spain", *IEEE International Congress on Big Data*, 2014.

WPSReview International on Sustainable
Housing and Urban Renewal (RI-SHUR)

- [5] H. Samiul, C. M. Schneider, S. V. Ukkusuri, M.C. González, "Spatiotemporal patterns of human mobility", *Journal of Statistical Physics*, vol. 151, n° 1-2, pp. 304-318, 2013.
- [6] L. Alexander, S. Jiang, M. Murga, and M.C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data", *Transportation Research Part C: Emerging Technologies* vol. 58 pp. 240-250, 2015.
- [7] Trépanier, M., Tranchant, N., & Chapleau, R. (2007). Individual trip destination estimation in a transit smartcard automated fare collection system. *Journal of Intelligent Transportation Systems*, 11(1), 1-14.
- [8] Munizaga, M. A., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.
- [9] Cui, A. (2006). Bus passenger origin-destination matrix estimation using automated data collection systems (Doctoral dissertation, Massachusetts Institute of Technology).
- [10] Alsger, Azalden A., et al. "Use of smartcard Fare Data to Estimate Public Transport Origin–Destination Matrix." *Transportation Research Record: Journal of the Transportation Research Board* 2535 (2015): 88-96.