

# Estimación de matriz origen-destino a partir de datos de tarjetas inteligentes: una revisión bibliográfica

Gabriel Tenenbaum de Oliveira\*, Carlos David Nassi

Universidad Federal de Rio de Janeiro / Programa de Engenharia de Transportes / COPPE, Brasil

\* Autor para correspondencia:  
gto@pet.coppe.ufrj.br

## RESUMEN

Este artículo tiene como objetivo revisar la bibliografía de la aplicación de datos de tarjetas inteligentes de sistemas de recaudo electrónico – SRE – para la obtención de matrices origen-destino de viajes en transporte público. Los SREs generan datos de forma continua e íntegra y pueden substituir la toma de datos en terreno, que son mucho más dispendiosas y sujetas a errores. Para contextualizar y facilitar la comprensión de los trabajos, inicialmente son abordados aspectos relacionados a las tecnologías y procesos de los SER y las ventajas de utilizar sus datos. El proceso de estimación del origen y destino de los viajes, ya estudiado por varios autores, es abordado en detalle, explicando las reglas lógicas que han sido exploradas, los resultados obtenidos, sus limitaciones y las etapas de validación utilizadas. El texto concluye con visiones más amplias sobre el papel de esta tecnología en el futuro del planeamiento de los sistemas de transporte.

Palabras clave: tarjeta inteligente, SRE, origen-destino, caracterización de la demanda

## ABSTRACT

*This article aims to review the literature about the application of smartcard data from Automated Fare Collection – AFC systems to characterize the demand of transit systems. AFC systems generate data in a continuously and comprehensively manner and are able to replace field surveys, which are more costly and error-prone. To contextualize and facilitate understanding of the literature, aspects related to technologies and processes of this type of system and the advantages of using such data are first addressed. The trip origin-destination estimation process, undertaken by several authors, is then discussed in detail, examining aspects such that the logical rules introduced, the results obtained, its limitations and validation procedures. The paper concludes with broader views on the role of these systems in the future of transportation planning.*

*Keywords: smart card, AFC, origin-destination, demand characterization*

## 1. INTRODUCCIÓN

Los sistemas de recaudo electrónico – SRE – son usados en la actualidad en un gran número de ciudades en todo el mundo, y gracias a sus beneficios, la tendencia actual indica su futura aplicación en más ciudades que aún no los tienen. Estos sistemas, aplicados con el objetivo de facilitar el recaudo de la tarifa, representan una fuente continua, precisa e íntegra de datos sobre la demanda. De hecho, estos datos, disponibles en forma abundante y con el nivel de detalle de cada transacción, permiten una amplia gama de estudios y aplicaciones.

Entre las aplicaciones posibles, se destaca la posibilidad de determinar matrices origen-destino (OD) en sistemas de transporte público urbano, las cuales constituyen insumos básicos para su propia planeación. Por medio del procesamiento de los datos con base en reglas lógicas, es posible obtener matrices similares a las obtenidas como resultado de grandes y dispendiosas encuestas de campo. En la literatura internacional, una gran cantidad de autores han estudiado este tema en diversos países de América Latina como Chile (Devilleine *et al.*, 2013; Munizaga y Palma, 2012; Munizaga *et al.*, 2014) y Brasil (Farzin, 2008; Guerra *et al.*, 2014). También se destacan estudios en Norteamérica, en Canadá (Chu y Chapleau, 2013; Morency *et al.*, 2007; Pelletier *et al.*, 2011) y Estados Unidos (Cui, 2006; Dumas, 2015; Zhao *et al.*, 2007); en Europa, en Inglaterra (Gordon, 2012; Wang *et al.*, 2011); y en Asia, en China (Lianfu *et al.*, 2007).

El presente artículo científico tiene como objetivo presentar una revisión bibliográfica sobre la aplicación de datos provenientes de SRE para la estimación de matrices OD en sistemas de transporte público, destacando las reglas lógicas utilizadas en el procesamiento, los resultados encontrados, sus limitaciones y etapas de validación posteriores.

El trabajo está dividido de la siguiente manera: la Sección 2 presenta una revisión bibliográfica, sistemática y teórica, sobre los componentes de un SRE buscando la mejor comprensión de los aspectos básicos relacionados a la utilización de este tipo de datos. En la Sección 3 son revisadas las ventajas de emplear este tipo de sistemas para la caracterización de la demanda, en comparación con sistemas de recolección de datos manuales. En la Sección 4 se trata detalladamente el proceso de estimación de origen y destino en el sistema de transportes. Finalmente, la Sección 5 presenta las conclusiones y esboza recomendaciones para investigaciones y aplicaciones en sistemas reales.

## 2. SISTEMAS DE RECAUDO ELECTRÓNICO

Según una de las primeras investigaciones acerca de la utilización de datos de sistemas de recaudo electrónico en el planeamiento del transporte (Zhao *et al.*, 2007), estos sistemas abarcan una amplia gama de tecnologías empleadas en el procesamiento de datos de transacciones en el sistema de transporte público para fines de recaudación tarifaria. En ese contexto, las transacciones se refieren al acto de utilización de un modo de transporte público por un usuario en específico, mediante la validación de la tarjeta inteligente y el registro de ingreso al vehículo (y salida del mismo, en el caso de “sistemas cerrados”). Los SRE son, generalmente, categorizados como sistemas de recolección de datos automático – ADC<sup>1</sup> por su sigla en inglés.

La tarjeta inteligente con microchip sin contacto (llamada en inglés smart card) es la tecnología de tarjetas más utilizada para la realización de transacciones en un sistema SRE, tanto en Brasil (Carvalho, 2003; Guerra *et al.*, 2014), como en otros países alrededor del mundo, debido a que presenta una serie de ventajas en relación a otras tecnologías (Pelletier *et al.*, 2011). La primera implementación de estas tarjetas en sistemas de transporte público fue realizada finalizando los años 90 en las ciudades de Washington y Tokio (Devilleine *et al.*, 2013).

Para recibir la tarjeta inteligente, el usuario debe registrarse y recargar la misma en los puntos de distribución del sistema. Las transacciones son realizadas en el momento en que el usuario aproxima la tarjeta a un validador, dispositivo físico presente en los modos de transporte. Para el éxito de la transacción, no se requiere el contacto físico entre la tarjeta y el validador, gracias a que ambos están dotados de antenas, lo que vuelve más conveniente y ágil su aplicación en el contexto de sistemas de transporte urbano (Pelletier *et al.*, 2011). Una transacción almacena básicamente datos sobre la validación (fecha y horario de inicio del viaje o transbordo), sobre la tarjeta (usuario, número consecutivo de viaje de la tarjeta), sobre la tarifa aplicada (tarifa vigente, tipo de tarifa pagada, subsidio) y sobre el servicio (estación, línea, sentido, vehículo, conductor, validador), entre otros.

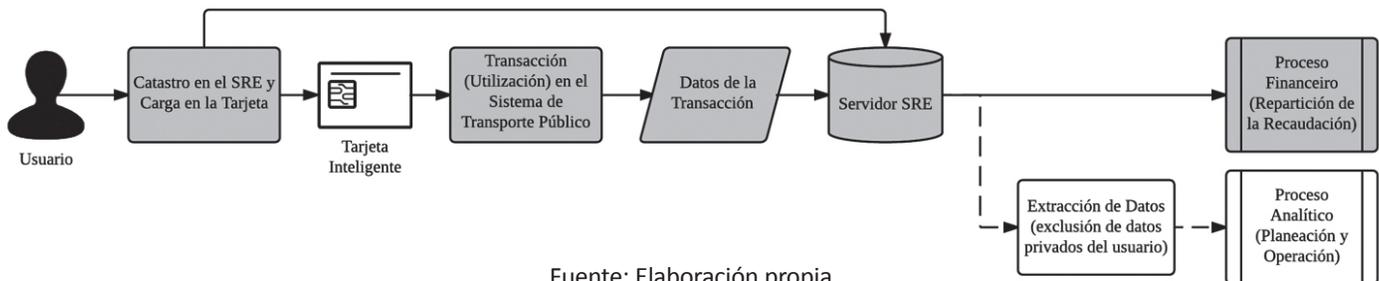
Los validadores pueden estar localizados en la red de transporte de forma fija –como por ejemplo en las estaciones de tren, metro, barcos, BRT, y demás modos denominados estructurantes– o de forma móvil –como en el caso de validadores instalados al interior de buses e colectivos, denominados modos capilares. Existen casos intermediarios en los que el sistema de transporte conserva tanto estaciones como vehículos con validadores –como puede ocurrir en sistemas de bus con terminales–, casos en los cuales el sistema es denominado como de uso mixto.

<sup>1</sup> Otros sistemas de recolección de datos automáticos incluyen la localización vehicular automática por GPS y de conteo automático de pasajeros, los cuales son indicados generalmente por sus siglas en inglés, respectivamente AVL (automatic vehicle location) y APC (automatic passenger counting).

El validador puede realizar la transmisión en tiempo real de los datos de transacciones para un servidor único del SRE, o almacenar los datos a lo largo del día, realizando la transmisión de datos por lotes al final del día. En este último caso se nota que hay una falta de sincronía entre la realización de la transacción y la transmisión de los datos al servidor. Esta diferencia temporal varía y es de interés para los agentes de control del SRE que ésta sea minimizada, buscando el procesamiento efectivo en tiempo hábil. El servidor del SRE puede almacenar también información relativa a los servicios previstos (tabla de horario, vehículo despachado, conductor, etc.) La presencia de estos datos puede ser recomendada para recolectar información no sólo de la demanda, sino también sobre la oferta de la red de transporte.

Posteriormente al almacenaje de los datos en el servidor, se prosigue con su procesamiento para fines de repartición de la recaudación, esto es, la distribución de las tarifas relativas a las transacciones entre los diferentes operadores del sistema de transporte público. Zhao (2007) nota que la concepción del SRE para este fin específico dificulta, o limita, su utilización para análisis operativos o de planeamiento, bien sea porque los datos vienen en un formato poco amigable para su uso, o por no contener otros campos importantes para esos fines. Diversos autores destacan que, antes de su utilización en el proceso de planeación, los registros de transacciones son separados de los datos sobre la identidad de los usuarios, para garantizar la privacidad personal (Pelletier *et al.*, 2011). Un esquema general y simplificado de las etapas presentes en este sistema, desde la transacción por parte del usuario hasta su utilización en el proceso de planeación, es presentado en la Figura 1.

**Figura 1: Etapas entre a la realización de una transacción hasta su procesamiento analítico**



Fuente: Elaboración propia

### ¿Qué define al usuario?

En el sistema de recaudo electrónico, usuario puede ser considerado todo aquel ciudadano que realizó por lo menos una transacción en el período de un año (Bagchi y White, 2005). En el caso de que el sistema cuente con información de desafiliación del usuario, usuario es todo aquel que cuenta con una tarjeta activa.

### ¿Qué es considerado como un viaje?

Un viaje es un movimiento desde un punto de origen a un punto de destino (Ortúzar y Willumsen, 2011). El origen y el destino son lugares en que las necesidades de los usuarios son satisfechas por la realización de actividades (Munizaga *et al.*, 2014). Cada viaje es compuesto por un grupo de uno o más movimientos en un modo en particular, siendo que cada movimiento caracteriza una etapa del viaje. Los puntos de abordaje y descenso son las posiciones en que la etapa comienza y termina, respectivamente (Munizaga y Palma, 2012).

Transbordos designan la continuidad entre las etapas en el contexto de un mismo viaje. Estos son consecuencia de la interacción entre la red de transporte y la satisfacción de la necesidad de los usuarios (Munizaga *et al.*, 2014). El primer abordaje de un viaje puede ser llamado abordaje inicial; los abordajes siguientes pueden ser llamados abordajes intermedios. Similarmente, el descenso más cercano al destino del viaje será llamado descenso final y las bajadas anteriores, descensos intermedios.

Durante el procesamiento, la detección acertada a partir de los datos de las transacciones entre abordajes iniciales o transbordos es altamente relevante para agrupar las etapas de los viajes y para una caracterización de la demanda que sea representativa de la realidad.

### 3. APLICACIÓN DE DATOS DE SRE PARA CARACTERIZACIÓN DE DEMANDA

Una de las primeras fases del proceso de planeación de sistemas de transporte es la recolección de datos e información sobre las condiciones existentes de oferta y demanda. Según Vuchic, la fase de recolección de datos es la parte más costosa del proceso de planeamiento. Dentro de los tipos de datos recolectados y necesarios para la planeación, se encuentran datos sobre el origen y destino con detalles sobre el motivo del viaje, generalmente obtenidos de la realización de encuestas de campo (Vuchic, 2005, p. 486).

En general, las encuestas de campo recopilan una cantidad significativa de datos, aunque el tamaño de la muestra sea generalmente limitado por la razón entre calidad (profundidad de las preguntas) y cantidad (tamaño de la muestra) (Morency *et al.*, 2007). Al ser realizadas de forma manual, las encuestas consumen tiempo para ser digitalizadas, adicionando otro factor de error (Attanucci y Wilson, 1981; en Zhao *et al.*, 2007).

Específicamente, es en esta fase de recolección de datos y caracterización de la demanda que los SRE pueden ser de gran ayuda para el proceso general de planeación, afirmación concordante con Vuchic (2005). Los SRE no pueden ser considerados, de forma bruta, una encuesta de transporte per se (Chu y Chapleau, 2013) y en ningún caso se recomiendan para substituir por completo la existencia de otras fuentes de datos (Bagchi y White, 2005). Sin embargo, numerosos trabajos han demostrado tanto la capacidad como la tendencia de explorar información sobre viajes y sobre el comportamiento de los usuarios con base en el procesamiento de gran escala de datos disponibles en los SRE.

Hasta el inicio de los años 2000, los estudios relacionados a SREs en la literatura científica se limitaban a análisis de la tecnología empleada (Morency *et al.*, 2007). A partir de autores como Buneman (1984; en Wilson *et al.*, 2009), Barry *et al.* (2002; en ZHAO, 2004), Bagchi (2003; en Bagchi y White, 2005) y Zhao (2004) se comenzó a desarrollar el conocimiento sobre la prospección de información proveniente de la gran cantidad de datos suministrados por los SRE para la caracterización de la demanda y para la planeación de transportes.

En relación al uso de estos sistemas para la prospección de información, la literatura permite elaborar comparaciones con metodologías manuales para la recolección de datos, según se muestra en la Tabla 1. Las ventajas, aunque son menos numerosas, son importantes para hacer investigación ya que los datos están disponibles en gran cantidad, cobertura territorial y exactitud. Los errores y desventajas que surgen son relativamente pequeños frente al volumen de transacciones tratadas.

Según Chu y Chapleau (2013), los SREs presentan propiedades alineadas con tres aspectos recomendados por la literatura actual para investigaciones de gran escala sobre movilidad y caracterización de la demanda, principalmente en lo relacionado al entendimiento del comportamiento a escala individual. En primer lugar, por medio de los datos de este tipo de sistema es posible realizar una investigación de varios días (multi-day survey) con el mismo usuario, comprendiendo su comportamiento y sus posibles variaciones, lo cual es imposible de determinar mediante levantamientos de información de un único día. En segundo lugar, estos sistemas permiten levantar información de tipo panel (panel surveys) –levantamiento de información del mismo usuario en puntos espaciados de tiempo– para detectar variaciones históricas –a lo largo del tiempo– del comportamiento. Finalmente, estos sistemas entregan datos de forma continua, permitiendo monitorear cambios en los patrones de comportamiento incrementalmente.

Pelletier *et al.* (2011) destacan tres categorías de aplicaciones de SRE para la planeación de transporte: estudios de ámbito estratégico, táctico y operativo. La Tabla 2 detalla las etapas correspondientes en la aplicación de los datos para las estimaciones de origen y destino en sistemas de transporte público.

**Tabla 1: Ventajas y desventajas de la prospección de información a partir de SER en comparación con metodologías manuales de recolección de datos**

Ventajas
<p><b>Alcance:</b> cobertura espacial y temporal extensiva del sistema de transporte público, en contraste con el sistema de levantamientos de campo, que son concentrados generalmente en períodos o recorridos específicos. Días de fin de semana, de grandes eventos y de tiempo meteorológico atípico, así como líneas de transporte poco utilizadas son monitoreadas de forma total y efectiva <sup>(1) y (2)</sup>;</p> <p><b>Exactitud:</b> los datos son recolectados automáticamente, siendo posible, por ejemplo, entrar en el detalle de hora-minuto-segundo exacto de la realización de la transacción. En los métodos tradicionales, en el caso de transbordo entre etapas, puede haber contabilización de dos viajes o de dos usuarios diferentes <sup>(2)</sup>;</p> <p><b>Tiempo de procesamiento:</b> con el soporte de herramientas de procesamiento y análisis, los datos pueden ser procesados rápidamente <sup>(1)</sup>;</p> <p><b>Interferencia Humana:</b> dependencia menor del encuestador en campo o del usuario, lo que puede minimizar errores de declaración;</p> <p><b>Escala Individual:</b> métodos tradicionales realizan conteos agregados de pasajeros<sup>(2)</sup>;</p>
<p>Fuentes: <sup>(1)</sup> Zhao <i>et al.</i> (2007); <sup>(2)</sup> Bagchi y White (2005); <sup>(3)</sup> Morency <i>et al.</i> (2007); <sup>(4)</sup> Pelletier <i>et al.</i> (2011); <sup>(5)</sup> Munizaga y Palma (2012); <sup>(6)</sup> Munizaga <i>et al.</i> (2014); <sup>(7)</sup> Devillaine <i>et al.</i> (2013); <sup>(8)</sup> Chu y Chapleau (2013).</p>
Desventajas
<p><b>Estructura de datos:</b> organizar los datos para su uso requiere tiempo dentro de una investigación. Necesidad de agregación de las etapas para definir el viaje completo <sup>(1) y (2)</sup>;</p> <p><b>Dato no georreferenciado:</b> comúnmente el SRE no está vinculado directamente con el sistema de localización vehicular (AVL), requiriendo su sincronización para localización espacial de la transacción <sup>(1) y (3)</sup>;</p> <p><b>Ausencia de datos extra-transacción:</b> no existe información socioeconómica sobre el usuario <sup>(4)</sup>. No hay información sobre el motivo del viaje <sup>(2) y (5)</sup>, aunque existan métodos para realizar su estimación <sup>(6)</sup>. Transbordos entre las etapas del viaje y el destino final del viaje no son informados, aunque existen métodos para deducirlos <sup>(2) e (3)</sup>, pudiendo ser evitado en “sistemas cerrados”, con control de entrada y salida <sup>(7)</sup>;</p> <p><b>Validación exógena:</b> existe la necesidad de realizar encuestas para confirmar los análisis y el uso de las premisas adoptadas <sup>(2) e (6)</sup>;</p> <p><b>Equidad en su uso:</b> no hay estudios indicando los diferentes grupos de la población que son de hecho beneficiarios de este tipo de sistemas, lo que puede generar una distorsión en los resultados. Además, es deseable que el sistema tenga una alta penetración en el mercado para disponer de una muestra adecuada de la población. En Santiago de Chile, por ejemplo, debido a que la tarjeta inteligente es la única opción para acceder al sistema de buses y es ampliamente utilizada en el metro, la penetración alcanza el 97% de la demanda <sup>(5)</sup>;</p> <p><b>Posibles sesgos:</b> la evasión tarifaria no es captada por el SRE. Una misma persona utilizando diferentes tarjetas, más de una persona utilizando la misma tarjeta (viaje en grupo) y tarjetas transferibles pueden inducir sesgos en los datos. El poseedor de una tarjeta puede, ocasionalmente, utilizar otro medio de pago <sup>(7) e (8)</sup>;</p> <p><b>Error sistemático:</b> mal funcionamiento de un validador o de una transmisión puede generar problemas de captación y procesamiento de los datos;</p> <p><b>Eventos aleatorios y problemas de seguridad pública:</b> estos tipos de eventos causan desvíos operativos que podrían confundir los análisis realizados <sup>(7)</sup>.</p>
<p>Fuentes: <sup>(1)</sup> Zhao <i>et al.</i> (2007); <sup>(2)</sup> Bagchi y White (2005); <sup>(3)</sup> Morency <i>et al.</i> (2007); <sup>(4)</sup> Pelletier <i>et al.</i> (2011); <sup>(5)</sup> Munizaga y Palma (2012); <sup>(6)</sup> Munizaga <i>et al.</i> (2014); <sup>(7)</sup> Devillaine <i>et al.</i> (2013); <sup>(8)</sup> Chu y Chapleau (2013).</p>

**Tabla 2: Diferentes categorías de aplicación de datos de SRE para planeación de transportes**

Ámbito Estratégico
Aplicación para planeación a largo plazo de la red de transporte, agregando por tipo de usuario, de comportamiento de viaje y comportamientos habituales en diferentes escalas temporales (Bagchi y White, 2005; Chu y Chapleau, 2013; Morency <i>et al.</i> , 2007), incluso antes y después un cambio en la prestación del servicio (Mojica, 2008; en Chu y Chapleau, 2013; Gordon, 2012). La aplicación en esta escala es limitada debido a que generalmente los datos no están relacionados con un perfil socioeconómico de los usuarios. Una estimación del lugar de domicilio a partir de datos de SRE hecha por Dumas (2015) permitió, sin embargo, un análisis de equidad entre etnias de los pasajeros de la red de transporte público de Boston.
Ámbito Táctico
Aplicación para ajustes en la oferta del servicio. Es posible entrar en el detalle de la demanda para diferentes días de la semana, líneas, ofertas de viaje y tramos de diferentes modos de transporte. De esta manera es viable detectar, por ejemplo, la sección más cargada. El estudio de los patrones de transbordo se destaca como un paso para la adecuación geométrica de la red y de los horarios programados según la demanda (Bagchi y White, 2005). La determinación de la matriz OD también puede ser definida como una aplicación en el ámbito táctico (Pelletier <i>et al.</i> , 2011), siendo que es posible, incluso, inferir el tipo de actividad realizada (Devillaine <i>et al.</i> , 2012; Ortega-Tong, 2013; Goulet-Langlois, 2015; Kusakabe y Asakura, 2014).
Ámbito Operativo
Aplicación para monitorear indicadores de desempeño de la red de transporte como velocidad y tiempo de viajes (Nuñez <i>et al.</i> , 2013), cumplimiento de horarios programados y kilometraje recorrido por vehículo; reducción de tiempos de parada de buses (Milkovits, 2008) así como para la detección de fraude y errores sistemáticos (Farzin, 2007).

#### 4. ESTIMACIÓN DE LA MATRIZ OD A PARTIR DE DATOS DE SRE

Las transacciones registran, según lo explicado anteriormente, el abordaje realizado por los usuarios, siendo ese el modo en que se conoce la información relativa al origen del viaje, es decir, a la entrada del usuario en un determinado vehículo o modo de transporte o, inclusive, parada o estación. Sin embargo, para fines de planeamiento de transporte, es de vital importancia que se conozca el punto de descenso del usuario, tanto dentro de una etapa del viaje, como del viaje por completo, indicando su destino final. Esta información, en caso de ser estimada con éxito, permite al analista llegar, a partir de los datos del SRE a su disposición, a una matriz de lugares de abordaje y descenso de usuarios, la cual se torna útil para obtener una matriz completa de OD en la red de transporte.

##### 4.1 Autores y casos analizados, métodos y premisas adoptadas

La derivación de una matriz OD con base en reglas lógicas a partir de los datos de transacciones ha sido estudiada desde los primeros trabajos que tratan sobre la aplicación de sus datos en el planeamiento de transporte.

Varios autores realizaron trabajos de investigación en sistemas de distinta cobertura y complejidad (combinación de modos de transporte). Diferentes series de datos han sido utilizadas para inferir la localización del punto de descenso en el caso de modos capilares, como el posicionamiento vehicular automático (AVL) y el horario programado de inicio de la oferta de viaje en el sistema. Algunos tipos de lenguaje de programación han sido empleados para tratar los datos, siempre constandingo la necesidad de entender el lenguaje de consulta a bancos de datos, SQL (*Structured Query Language*). Las metodologías propuestas fueron aplicadas en diversos escenarios, lo que ha contribuido al entendimiento lógico del problema. Algunas han llegado incluso a ser aplicadas para la caracterización de la demanda de sistemas reales en la industria. En la Tabla 3 se presenta un resumen de los tipos de información según los autores consultados.

Tabla 3: Relación de autores que estudiaron el problema de estimación de abordaje y descenso a partir de datos del SRE

Año	Autor(es)	Ciudad, País (Tasa de Penetración del SRE)	Modo(s) Estudiado(s)	Datos para Localización en MC	Lenguaje de programación, Software utilizado	Cantidad días (d.), transacciones (t.), viajes (v.) analizados	Estimativa de Descenso (porcentual del total analizado)	Expansión para Matriz OD	Contribución	Aplicación en Sistema Real
1984	Buneman (en Farzín, 2008)	San Francisco, EUA	Tren (ME) (control de entrada y salida)	No necesario	ND	ND	ND	ND	Primer uso de SRE para OD.	ND
2002	Barry et al. (en Zhao, 2004)	Nueva York, EUA	Metro (ME)	No necesario	ND	ND	ND	ND	Primer uso en sistema abierto.	ND
2004	Zhao	Chicago, EUA (88%)	Metro (ME) (+Bus para correr regla)	AVL	C++, ArcGIS	6 d., 2,5 mi t.	71%	Sí, Red de metro	Estimativa en los transbordos ME-MC.	Sí
2006	Cui	Chicago, EUA	Bus (MC)	AVL	SQL	5 d., 2,7 mi t.	60%	Sí, Muestra de la red	Estimativa de abordaje y descenso en MC.	No
2007	Trépanier (en Munitzaga, Palma, 2012)	Gatineau, Canadá	Bus (MC)	ND	ND	ND	66%	ND	ND	ND
2007	Lianfu	Changchu, China	Bus (MC)	Seguimiento en campo	MNA	MNA	MNA	MNA	Método sin AVL	MNA
2008	Farzín	São Paulo, Brasil	Bus (MC)	AVL	SQL, GIS	1 d., 658k v. (10mi t. / 6,3mi v. no total SRE)	77% (6% / 8% del total de t. y de v.)	Sí, Zona de tráfico	Primer caso brasileño.	No
2010	Guerra	Alagoas, Brasil	Bus (MC)	Tabla de horarios	TransCAD	7 d., 24,9k t. PM (172 k t. no total SRE)	56% (8% en total)	Sí, Zona de tráfico	Uso de tabla de horarios, SIG e distribución de intervalo de viajes.	No
2011	Wang	Londres, Inglaterra (> 80% Bus)	Bus (MC)	AVL	Java, SQL	72.4k t.	66% (promedio)	No	Perfil de carga, Variabilidad, Análisis tiempo de transbordo.	No
2012	Gordon	Londres, Inglaterra	Bus (MC) (en Sistema Multimodal)	AVL	Java	10 d., 6,3 mi t.	77%	Sí, Red de Londres	Múltiples condiciones para definir transbordo.	Sí
2012	Munitzaga y Palma	Santiago, Chile (97%)	Sistema Multimodal (ME+MM+MC)	AVL	SQL, C++, GoogleEarth API	7 d., 38 mi t.	83%	Sí, Zona de tráfico	Uso del tiempo para inferir el destino.	Sí, después de validación.
2012	Nasiboglu et al.	Izmir, Turquía	Metro, Bus, Ferry y Tranvía (ME + MC)	No especificado		1,5mi t.	No especificado	No especificado	Investigación poco clara.	No especificado
2015	Dumas	Boston, EUA	Sistema Multimodal (ME+MM+MC)	AVL	Java	21 d., 18,3 mi t.	56% Bus 77% Tranvía 75% Tren	No	Método de Gordon (2012). Contribución estimación domicilio.	No

Abreviaciones: ND = dato no disponible en la bibliografía consultada / MNA = método no aplicado / ME = modos estructurantes / MC = modos capilares / MM = modos mixtos / k = 103 / mi = 106

## 4.2 Reglas lógicas para determinación del punto de abordaje y descenso

En términos de las reglas lógicas estudiadas para estimar el punto de descenso, algunas premisas básicas han sido desarrolladas por los investigadores. En primer lugar, la regla principal para la determinación del destino de la etapa del viaje, se basa en la idea de que un número significativo de usuarios permanece o retorna al punto de destino de la etapa de viaje anterior para comenzar la siguiente etapa del viaje (o el viaje siguiente) (Barry *et al.*, 2002; en Zhao, 2004). Esta inferencia es válida si se cumple lo siguiente:

- No hay modos de transporte privado (carro, motocicleta, bicicleta, etc.) entre dos transacciones consecutivas.
- Los pasajeros no caminan una distancia grande para abordar en un punto diferente al punto de descenso. Esta condición es especialmente válida cuando la distancia entre puntos es superior a distancias aceptables para la caminata, e igualmente para modos estructurantes con un espaciamiento considerable entre estaciones.
- Los pasajeros, en el último viaje del día, retornan al punto donde comenzaron el primer viaje de ese mismo día, concepto que puede ser extendido para considerar el primer viaje del día siguiente.

Un punto de vista de extrema importancia es la detección de cuáles transacciones constituyen abordajes de inicio de viaje, y cuáles constituyen transbordos entre etapas del mismo viaje. Este tema ha sido estudiado desde Bagchi y White (2005) como forma de caracterización de los patrones de demanda. Solamente con esta diferenciación es posible salir de una perspectiva de etapa del viaje y llegar a una matriz OD, en la cual está incluido el concepto de destino de viaje como el lugar de realización de alguna actividad (Munizaga y Palma, 2012).

El intervalo entre transacciones ha sido un parámetro utilizado para esta clasificación. Una opción es establecer un límite fijo, por ejemplo, de 30 minutos o 1 hora, dentro de las que dos transacciones indican que hubo un transbordo. Guerra indica que es importante que ese intervalo sea compatible con el porte de la red estudiada (Guerra *et al.*, 2014). Esta elección arbitraria, aunque de simple implementación, falla en la identificación de actividades cortas y en el caso de amplios intervalos entre ofertas de viaje. De forma paralela, se puede suponer que, independientemente del intervalo de tiempo entre etapas, en el caso de transacciones consecutivas en una misma línea hubo un destino entre las dos, por ser poco probable que alguien baje del modo y luego retome el viaje sin algún motivo que no sea la realización de alguna actividad (Munizaga y Palma, 2012).

Para la detección del lugar de descenso en el caso de modos capilares, en que los validadores están presentes dentro de los vehículos, todos los autores en este caso utilizaron una segunda fuente primaria de datos. En general, fueron usados datos provenientes del sistema AVL con base en el uso de GPS a bordo de los vehículos (Cui, 2006; Farzin, 2008; Munizaga y Palma, 2012; Wang *et al.*, 2011; Zhao, 2004). Fuentes alternativas de información sobre la localización han sido el acompañamiento del recorrido a bordo del vehículo (Lianfu *et al.*, 2007) y la utilización de la tabla de horarios programados junto con la distribución normal del tiempo de viaje (Guerra, 2011). En el análisis de estos casos se consideró que: (i) la mayor parte de los validadores permanece instalada en un mismo vehículo por un tiempo razonable (Zhao, 2004); (ii) que el pasajero ingresó al vehículo aproximadamente en el mismo momento en que es realizada la transacción, y por tanto, que ese dato corresponde al punto y hora de inicio del viaje (Munizaga *et al.*, 2014); aunque está claro que en muchos casos existen asientos antes de pasar la registradora del vehículo y que, en algunos casos, usuarios pueden validar su viaje solo en el momento de bajada (Farzin, 2008).

En todos los casos se requirió una etapa compleja e importante de emparejamiento de las bases de datos, es decir, entre los datos de las transacciones de la tarjeta inteligente y los datos de localización del vehículo. La limitación de datos primarios de localización, a escala de vehículo, puede reducir significativamente la muestra tratada, como en el caso de Farzin (2008) en el cual, debido a la inexistencia de un sistema AVL en los buses de São Paulo, solo pudo emplear datos de 658 mil viajes entre un total de 6,3 millones de viajes diarios (10% del total).

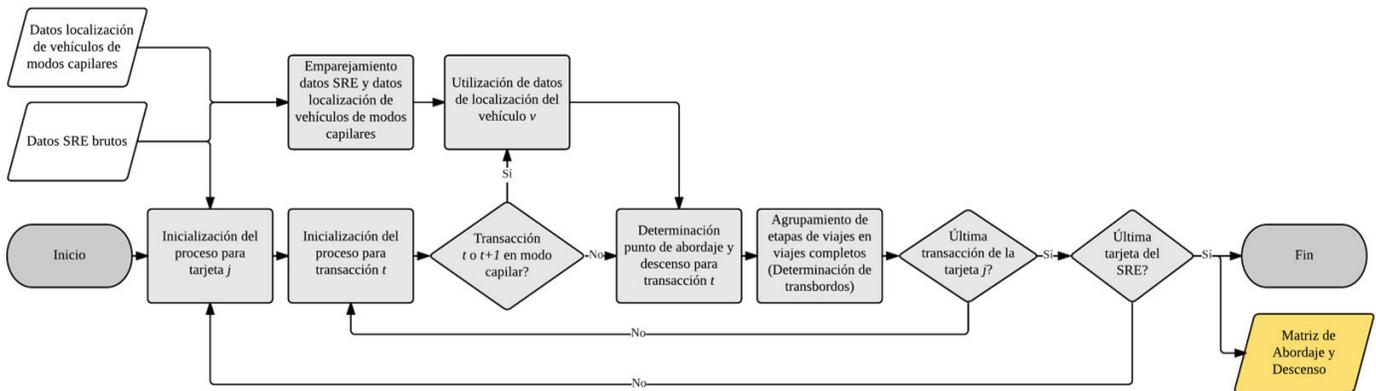
La escala de determinación de los lugares de abordaje y descenso varía en cada trabajo. En los modos estructurantes es posible localizar esos puntos a nivel de la estación. En los modos capilares, esos lugares pueden ser localizados a nivel de coordenadas geográficas provenientes de AVL, a nivel de paraderos físicamente existentes o a nivel agregado en la escala de zona de tráfico, opción que minimiza la imprecisión del método originada por la diferencia de tiempo entre abordaje y descenso del pasajero en el vehículo y la validación de la tarjeta (Guerra, 2011).

Los transbordos que involucran un modo capilar en una etapa del viaje también presentan peculiaridades, y la metodología puede variar de acuerdo con el modo en que fue realizada la etapa original (Munizaga y Palma, 2012). Saliendo de una línea de un modo estructurante, la intersección entre ésta y la línea del modo capilar constituye el destino del modo estructurante (Zhao, 2004), es decir, el transbordo ocurre en la estación más cercana, espacial y temporalmente, del punto de abordaje en la siguiente etapa (Munizaga y Palma, 2012). Saliendo de una línea de un modo capilar, Munizaga y Palma (2012) establecen que el transbordo ocurre en un punto de espacio-tiempo que minimiza la función ponderada del tiempo de viaje en el vehículo y del tiempo estimado de caminata hasta el lugar donde se realiza el próximo abordaje, denominando ésta función de tiempo generalizado. En todos los casos, puede ser adoptada una distancia máxima de caminata por encima de la cual la transacción realizada no correspondería a un transbordo. Esta distancia fue determinada por Zhao (2004) como 400 m y por Munizaga y Palma (2012) como 1.000 m, siendo verificado posteriormente que esta variable podría ser tratada de forma más flexible (Munizaga *et al.*, 2014). Gordon (2012) explora otras condiciones para inferir transbordos como el concepto de tortuosidad (circuitry) entre segmentos de viaje y extensión total del viaje completo.

Otras premisas deben ser consideradas verdaderas. En el caso de análisis con datos de varios días, es importante la consideración de que cada usuario posee apenas una tarjeta y viceversa (Munizaga y Palma, 2012). Sin embargo, cuando son detectados viajes en grupo con la misma tarjeta, cada etapa del viaje puede tener un factor de expansión equivalente al tamaño del grupo (Devillaine *et al.*, 2013).

En general, las reglas lógicas son realizadas de forma secuencial dentro de determinados ciclos. El ciclo más grande corresponde a la interacción de todas las tarjetas del sistema. Dentro de este, es realizado un ciclo para una tarjeta determinada, un ciclo de interacción para determinar los puntos de abordaje y descenso de todas las transacciones. En caso de que en esta determinación haya modos capilares, se emplea una fuente de datos que reúne la localización espacial de los vehículos en relación a los paraderos. Puede ocurrir, todavía dentro del ciclo de una misma tarjeta, la determinación de las etapas del viaje. Un modelo general de estimación del punto de abordaje y descenso de las transacciones y del origen y destino de los viajes, considerando un bajo nivel de detalle, es representado en la Figura 2.

Figura 2: Estructura de ciclo para estimar los lugares de abordaje y descenso



Fuente: Elaboración propia

#### 4.3 Resultados obtenidos y expansión para matriz origen-destino

En la literatura explorada, considerando la muestra de transacciones utilizada, que varió entre 24,9 mil y 38 millones – y que puede ser diferente de la cantidad total de información disponible sobre el sistema, como en Farzin (2008) –, se llegó a un mínimo de 56% y un máximo de 83% de éxito en la estimación del punto de descenso.

Con una estimación exitosa del lugar y momento del descenso de cada etapa del viaje, es posible calcular el tiempo de viaje, el de transbordo, perfiles de carga y otras variables de importancia relacionadas a la operación, uno de los objetivos del trabajo de Wang *et al.* (2011).

Si la agrupación de las etapas de viaje dentro del viaje completo es exitosa, es posible realizar la expansión para una matriz OD completa, input esencial para la caracterización de la demanda y para el planeamiento operativo, táctico y estratégico de sistemas de transporte. Para eso se hace necesario comparar la matriz de abordaje y descenso obtenida (matriz semilla) con una matriz de conteo de pasajeros en diversos puntos de la red. Algunos autores lograron realizar la expansión de la matriz al nivel de la red estudiada (Cui, 2006; Zhao *et al.*, 2007; Gordon, 2012), mientras que otros llegaron al nivel de ciudad, utilizando para ello zonas de tráfico (Farzin, 2008; Guerra, 2011; Munizaga y Palma, 2012).

Es importante resaltar que la obtención de esas matrices puede ser realizada para cualquier subconjunto en el espacio-tiempo necesarios (aunque considerando la matriz expandida, este nivel se podría limitar a zonas de tráfico utilizadas para construir factores de expansión) (Munizaga y Palma, 2012).

#### 4.4 Limitaciones, validaciones y despliegue de los resultados

La limitación más básica encontrada en este tipo de aplicación para los datos de SRE reside en la necesidad de una segunda fuente de información primaria para la localización espacial de las transacciones en modos capilares. De acuerdo con Cui (2006), conocer la localización del vehículo es imprescindible para estimar los pares OD. Cuando hay sistemas automáticos, como el AVL, interesa que estos estén presentes en la mayor parte de la red y que sean capaces de ofrecer información precisa. En este sentido, el caso de Santiago de Chile es referente: existen incentivos financieros a los operadores por la obtención de datos de GPS, aumentando el interés de las mismas en instalar sistemas AVL eficaces (Munizaga y Palma, 2012).

Las estimaciones realizadas pueden presentar sesgos y limitaciones. En el procesamiento de algunas reglas, en caso de que haya un error en la estimación, se puede suponer que esto ocurre de forma similar en la totalidad de la extensión de la red, lo que convierte a la matriz de abordaje y descenso inferida semejante a la real (Zhao *et al.*, 2007). A pesar de eso, se reconoce la necesidad de iniciar procesos de validación de los resultados hallados con el procesamiento de transacciones con base en reglas (Chu y Chapleau, 2013; Guerra, 2011, p. 112). Esta importante fase fue realizada en la investigación de Munizaga *et al.* (2014), posteriormente a la confección de la matriz OD multimodal de Santiago, y fue dividida de la siguiente forma: (i) validación interna de la información, es decir, a través de análisis estadísticos de los propios datos utilizados para la confección de la matriz OD; (ii) validación externa de la información con base a una fuente de datos externa. En el caso de Santiago, la validación externa fue realizada de dos formas: con base a una encuesta OD en que usuarios del metro indicaron, además de información sobre el viaje, datos de su número de tarjeta inteligente; y con base a viajes realizados por voluntarios.

Una vez estimados los embarques, descensos y transbordos, las matrices de viaje pueden servir también de base para otros algoritmos de post-procesamiento. Los trabajos de Devillaine en Santiago (2012) y Ortega-Tong (2013) y de Goulet-Langois (2015) en Londres, por ejemplo, estiman el motivo del viaje realizado al nivel del usuario, habiendo estos dos últimos analizado también la regularidad de viajes. Dumas (2015) utiliza el origen del primer viaje de los usuarios para estimar su lugar de domicilio y analizar la equidad de viajes entre etnias en Boston. En sistemas de transporte “cerrados”, donde la validación en la salida es obligatoria y el destino de los viajes es automáticamente conocido, otros también se interesaron por la estimación del motivo de viaje, como Kusakabe y Asakura en Osaka (2014), y por el análisis de la variabilidad de viajes, como Zhong *et al.* en Singapur (2015).

Entre los casos estudiados, tres de ellos especificaron la aplicación del método en el monitoreo de sistemas reales. En Chicago, el método de Zhao (2004) fue desarrollado en una plataforma de fácil acceso y los resultados obtenidos fueron comparados con una gran encuesta de campo con usuarios de metro. En este caso, hubo validación del método, el cual pasó a ser empleado para reportar el total de pasajeros-kilómetros transportados por el sistema, datos enviados para la base de datos de transporte público del organismo federal de control (Zhao *et al.*, 2007). En Santiago, posteriormente a la fase de validación, el método de Munizaga y Palma (2014) ha sido utilizado para fines de planeación tanto por los operadores como por la Secretaría de Transportes. En Londres, la agencia Transport for London utiliza el método de Gordon (2012), denominado ODX (origin-destination-interchange inference process) en algunos análisis e incluso para remplazar algunas pesquisas anteriores.

## 5. CONCLUSIÓN

En el presente artículo se mostró el potencial rol que los sistemas de recaudo electrónico (SRE) tienen en la caracterización de la demanda y en la planeación de sistemas de transporte, al ser posible inferir el origen y el destino de los viajes por medio del procesamiento de sus datos con base en reglas lógicas.

Los SRE, consagrados como medios de pago seguros, prácticos y versátiles, generan datos de forma constante e integral, que pueden servir de base para investigaciones similares a las realizadas en campo, las cuales son por naturaleza más dispendiosas y sujetas a errores. Para su utilización, no obstante, es indispensable el conocimiento de aspectos básicos relacionados a las tecnologías y los procesos que componen este tipo de sistemas, trazando un cuadro teórico sólido sobre la relación entre los datos digitales y los viajes que son realizados en el mundo real.

El nivel de desagregación y la diversidad de elementos que componen estos datos permiten aplicaciones en diversas escalas. Lógicamente, quien realiza el estudio debe estar atento no solo a las ventajas de estas metodologías en relación a las tradicionales, sino también a las desventajas, que pueden dificultar o desorientar los análisis o distorsionar los resultados obtenidos. En particular, fueron abordados en profundidad los trabajos ya realizados sobre estimación de los lugares de abordaje y descenso, se exploraron las premisas y reglas lógicas adoptadas y los resultados obtenidos y sus limitaciones. En términos de investigaciones futuras, se sugiere que sean adicionados análisis de resultados desde el punto de vista del comportamiento de los usuarios, es decir, cómo se distribuyen los pares origen-destino en el caso de usuarios regulares.

La aplicación de SRE para la caracterización de la demanda constituye una oportunidad para que entidades de planeación, gestión y fiscalización del transporte público urbano tengan acceso a información en tiempo real de la demanda. Para la industria de transportes es posible construir una mejor relación con el usuario en específico a partir de la información encontrada, pudiendo influenciar sus patrones de uso de transporte (Bagchi y White, 2005). En caso de la existencia de datos de registro, como por ejemplo de dirección y teléfono, puede ser ofrecida información al usuario, de acuerdo con los servicios que este utiliza. La privacidad del usuario y el sigilo en el manejo de su información personal continúa siendo, sin embargo, uno de los puntos cruciales en la utilización de datos de SRE (Pelletier *et al.*, 2011).

De acuerdo con lo afirmado por Farzin (2008), aunque las aplicaciones realizadas puedan ser consideradas meramente una entrada de información para el proceso de planeación, éstas representan un cambio de paradigma en el transporte público, pasando éste de ser una fuente rica en datos a una fuente rica en información.

## AGRADECIMIENTOS

Los autores agradecen a los revisores y los editores por sus importantes comentarios y contribuciones para profundizar la investigación.

## REFERENCIAS

- Attanucci, J., Wilson, N.H.M. (1981) Bus Transit Monitoring Manual: Volume 1: Data Collection Program Design, US Department of Transportation.
- Bagchi, M., White, P.R. (2005) The potential of public transport smart card data. *Transport Policy* 12, 464–474.
- Buneman, K. (1984) Automated and Passenger-Based Transit Performance Measures. *Transportation Research Record* 992, 23-28.
- Carvalho, C.H.R. (2003) Panorama da Bilhetagem no Brasil, in: *Bilhetagem Automática e Gestão nos Transportes Públicos - Série Cadernos Técnicos Volume 1* 8–20.
- Nuñez, C., Munizaga, M., Gschwender, A. (2013) Cálculo de indicadores de calidad de servicio del sistema de transporte público de Santiago a partir de datos pasivos. XVI Congreso Chileno de Ingeniería de Transportes.

- Chu, A., Chapleau, R. (2013) Smart Card Validation Data as a Multi-Day Transit Panel Survey to Investigate Individual and Aggregate Variation in Travel Behavior. En Zmud, J., Lee-Gosselin, M., Munizaga, M.A. y Carrasco, J.A. (Eds.) *Transport Survey Methods: Best Practice for Decision Making*. Emerald Goup Publishing Limited 649–671.
- Cui, A. (2006) *Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems*. Tesis de Maestría. Massachusetts Institute of Technology, USA.
- Devillaine, F. (2012) *Estimación de viajes y actividades en base a sistemas tecnológicos de transporte público*. Tesis de Maestría. Universidad de Chile, Chile.
- Devillaine, F., Munizaga, M., Palma, C., Zúñiga, M. (2013) Towards a Reliable Origin-Destination Matrix from Massive Amounts of Smart Card and GPS Data: Application to Santiago. En Zmud, J., Lee-Gosselin, M., Munizaga, M.A. y Carrasco, J.A. (Eds.) *Transport Survey Methods: Best Practice for Decision Making*. Emerald Goup Publishing Limited 695–709.
- Dumas, R. (2015) *Analyzing Transit Equity Using Automatically Collected Data*. Tesis de Maestría. Massachusetts Institute of Technology, USA.
- Farzin, J.M. (2007) Summary of fraudulent farecard use patterns in São Paulo’s public transit network.
- Farzin, J.M. (2008) Constructing an Automated Bus Origin – Destination Matrix Using Farecard and Global Positioning System Data in São Paulo, Brazil. *Transportation Research Record* 2072, 30–37.
- Gordon, J.B. (2012) *Intermodal Passenger Flows on London’s Public Transport Network*. Tesis de Maestría. Massachusetts Institute of Technology, USA.
- Goulet-Langlois, G. (2015) *Exploring Regularity and Structure in Travel Behavior Using Smartcard Data*. Tesis de Maestría. Massachusetts Institute of Technology, USA.
- Guerra, A.L. (2011) *Determinação de Matriz Origem/Destino Utilizando Dados do Sistema de Bilhetagem Eletrônica*. Tesis de Maestría. Universidade Federal de Minas Gerais, Brasil.
- Guerra, L., Barbosa, H.M., Kelli, L. (2014) Estimativa de matriz origem/destino utilizando dados do sistema de bilhetagem eletrônica: proposta metodológica. *Revista Transportes* 22, 26–38.
- Kusakabe, T., Asakura, Y. (2014) Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C* 45, 179-191.
- Lianfu, Z., Shuzhi, Z., Yonggang, Z., Ziyin, Z. (2007) Study on the method of constructing bus stops OD matrix based on IC card data. 2007 International Conference on Wireless Communications, Network and Mobile Computing, WiCOM, 3142–3145.
- Milkovits, M.N. (2008) Modeling the factors affecting bus stop dwell time: Use of automatic passenger counting, automatic fare counting, and automatic vehicle location data. *Transportation Research Record* 2072, 125-130.
- Mojica, C. (2008) *Examining changes in transit passenger travel behavior through a smart card activity analysis*. Tesis de Maestría. Massachusetts Institute of Technology, MA, USA.
- Morency, C., Trépanier, M., Agard, B. (2007) Measuring transit use variability with smart-card data. *Transport Policy* 14, 193–203.
- Munizaga, M., 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.
- Munizaga, M., Devillaine, F., Navarrete, C., Silva, D. (2014) Validating travel behavior estimated from smartcard data. *Transportation Research Part C Emerging Technologies* 44, 70–79.

- Ortega-Tong, M.A. (2013) Classification of London's Public Transport Users Using Smart Card Data. Tesis de Maestría. Massachusetts Institute of Technology, USA.
- Ortúzar, J. de D., Willumsen, L. G. (2011) Modelling Transport. Wiley, Chichester. 2011.
- Pelletier, M.-P., Trépanier, M., Morency, C. (2011) Smart card data use in public transit: A literature review. *Transportation Research Part C Emerging Technologies* 19, 557–568.
- Trépanier, M., Tranchant, N., Chapleau, R. (2007) Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems* 11, 1-14.
- Vuchic, V.R. (2005) *Urban Transit: Operations, Planning and Economics*. John Wiley & Sons.
- Wang, W., Attanucci, J.P., Wilson, N.H.M. (2011) Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems. *Journal of Public Transportation* 14, 131–150.
- Wilson, N.H.M., Zhao, J., Rahbee, A., (2009) The potential impact of automated data collection systems on urban public transport planning. En: Wilson, N.H.M., Nuzzolo, A. (Eds.), *Schedule-Based Modeling of Transportation Networks: Theory and Applications*. Springer.
- Zhao, J. (2004) The Planning and Analysis Implications of Automated Data Collection Systems : Rail Transit OD Matrix Inference and Path Choice Modeling Examples. Tesis de Maestría. Massachusetts Institute of Technology, USA.
- Zhao, J., Rahbee, A., Wilson, N.H.M. (2007) Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering* 22, 376–387.
- Zhong, C., Manley, E., Arisona, S.M., Batty, M., Schmitt, G. (2015) Measuring variability of mobility patterns from multiday smart-card data. *Journal of Computational Science* 9, 125-130.

