



**UNIVERSIDAD DE CHILE**

**FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS**

**DEPARTAMENTO DE INGENIERÍA CIVIL**

**ESTIMACIÓN DE VIAJES Y ACTIVIDADES EN BASE A SISTEMAS  
TECNOLÓGICOS DE TRANSPORTE PÚBLICO**

**TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA  
INGENIERÍA, MENCIÓN TRANSPORTE**

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL**

**FLAVIO MARIO DEVILLAINÉ GÓMEZ**

**PROFESORA GUÍA:  
MARCELA MUNIZAGA MUÑOZ**

**MIEMBROS DE LA COMISIÓN:  
CRISTIÁN CORTÉS CARRILLO  
ANTONIO GSCHWENDER KRAUSE  
CAROLINA PALMA ALVARADO  
MARTIN TRÉPANIER**

**SANTIAGO DE CHILE  
JUNIO 2012**

## RESUMEN

El objetivo de la tesis es desarrollar y aplicar una metodología de estimación de viajes y actividades en base a datos pasivos de transacciones y de posicionamiento global de buses recolectados a través de sistemas tecnológicos de transporte público. El aprovechamiento de estos datos constituye un nuevo paso en la revolución de la información, que mediante la construcción de aplicaciones tienen potencial para transformarse en valiosas herramientas de planificación, más representativas y con menos errores y sesgos que las existentes en la actualidad.

Para tal fin se desarrolla una metodología que consta de dos pasos principales: un primer módulo realiza un análisis secuencial de las transacciones, y apoyado en los estudios previos de estimación de puntos de bajada para transacciones efectuadas en transporte público, reconoce y diferencia actividades y viajes realizados por usuarios de transporte público en las ciudades de Santiago, Chile y Gatineau, Canadá. En forma posterior, un segundo módulo procesa las actividades estimadas previamente y con información adicional les asigna un propósito determinado.

Una vez que los viajes y actividades se encuentran estimados, se lleva a cabo un análisis de las potencialidades de la metodología y sus resultados. Sobre estos, se puede observar el comportamiento social, en términos de asignación de tiempo y dinámica de viajes, en ambas ciudades; se establece una comparación entre ellas, encontrándose diferencias, y se mapean los hallazgos en términos de comportamiento temporal para cada tipo de actividad y propósito, concluyéndose que la metodología presenta potencialidades no sólo como herramienta para describir la demanda de transporte, sino también como instrumento para apoyar decisiones de planificación.

*“La ciencia se compone de errores,  
que a su vez, son los pasos hacia la verdad.”*

*-Julio Verne*

## *Agradecimientos*

En primer lugar, quiero agradecer a mi familia, y en forma muy especial a mis padres, Mario y Ana, por haber sido un apoyo permanente en este largo proceso de crecimiento académico, profesional y personal que ha significado mi instrucción en la Universidad de Chile, como también por haberme inculcado a temprana edad la rectitud, el civismo y la educación como nortes de vida. En segundo lugar, deseo agradecer a mi profesora guía, la Dra. Marcela Munizaga, por haberme dado la oportunidad de trabajar en un proyecto tan interesante, haber colaborado conmigo en ya casi innumerables investigaciones, y especialmente por brindarme la chance de vivir una experiencia tan inolvidable como es realizar una pasantía en el extranjero. A mi polola, Valentina, que constituye un apoyo emocional incondicional cada vez que lo necesito, y alegra mis días, mis lugares, mis instantes, mis segundos. A mis amigos transportistas y no transportistas; a los primeros, por hacer del día a día en la universidad más liviano y permitirme el honor de ser colega con un grupo tan excepcional de personas; a los segundos, por brindarme su amistad desde tiempos inmemoriales y soportar mis certeras y oportunas “acotaciones técnicas” cada vez que surge un tema asociado al transporte. A mis perros, Lupo y León, dos de los amigos más fieles que yo pudiese llegar a encontrar. Finalmente, a las instituciones y fondos que contribuyeron al financiamiento de esta investigación, FONDECYT y el Canadian Bureau of Education.

# Índice General

<b>Resumen</b> .....	i
<b>Agradecimientos</b> .....	iii
<b>Índice General</b> .....	iv
<b>Lista de definición de símbolos</b> .....	vi
<b>1. Introducción</b> .....	1
1.1 Motivación.....	1
1.2 Objetivos.....	3
1.3 Estructura de la tesis.....	3
<b>2. Revisión bibliográfica</b> .....	5
2.1 Etapas de Viaje.....	5
2.2 Tipos de operación y transacciones.....	6
2.3 Modelos de estimación de bajadas y generación de matices Origen/Destino.....	7
2.3.1 Modelo Hoffman & O’Mahony (2005).....	7
2.3.2 Modelo Trépanier et al. (2007).....	8
2.3.3 Modelo Zhao et al. (2007).....	10
2.3.4 Modelo Lianfu et al. (2007).....	11
2.3.5 Modelo Munizaga et al. (2011).....	14
2.4 Utilización de la información para describir a la población.....	17
2.5 Discusión del estado del arte en estimación de etapas y viajes.....	21
<b>3. Análisis de las bases de datos de transacciones</b> .....	22
3.1 Base de datos de Santiago, Chile.....	23
3.1.1 Transantiago: sobre el origen y otros atributos interesantes de las bases de datos de transporte público de Santiago.....	23
3.1.2 Descripción de las características generales de la base de datos de Transantiago e información contenida.....	25
3.1.3 Cuantificación de usuarios y transacciones en la dimensión temporal.....	27

3.1.4	Cuantificación de usuarios y transacciones en la dimensión espacial.....	30
3.2	Base de datos de Gatineau, Canadá.....	33
3.2.1	Breve reseña demográfica de la ciudad de Gatineau y sobre el funcionamiento de su sistema de transporte público.....	33
3.2.2	Descripción de la base de datos de Gatineau.....	35
<b>4.</b>	<b>Metodología de estimación de viajes y actividades.....</b>	<b>37</b>
4.1	Descripción general de la metodología.....	37
4.2	Módulo de Identificación de Actividades y Viajes (AIM).....	38
4.2.1	Patrón de los 30 minutos.....	39
4.2.2	Patrón dos transacciones consecutivas en Metro.....	39
4.2.3	Patrón dos transacciones consecutivas en el mismo recorrido de bus.....	40
4.2.4	Patrón comparación entre distancia euclidiana y distancia en ruta de un viaje.....	40
4.2.5	Patrón bus no abordado.....	49
4.3	Módulo de Asignación de Propósito (PAM).....	51
4.3.1	Definición de propósitos.....	51
4.3.2	Criterios de asignación de propósitos.....	52
<b>5.</b>	<b>Viajes y actividades en Santiago y Gatineau.....</b>	<b>56</b>
5.1	Estadísticas generales sobre viajes y actividades.....	56
5.2	Aplicación a Gatineau: evolución temporal de actividades.....	64
5.3	Aplicación a Santiago: análisis cruzado día-actividad.....	67
5.4	Aplicación a Santiago: perfiles de uso de tiempo.....	70
<b>6.</b>	<b>Validación.....</b>	<b>73</b>
6.1	Validación endógena.....	73
6.2	Validación exógena.....	77
<b>7.</b>	<b>Conclusiones y líneas de investigación futura.....</b>	<b>79</b>
	<b>Referencias Bibliográficas.....</b>	<b>82</b>
	Anexo A: CD con software Python para la implementación de la metodología.....	85

# Lista de definición de símbolos

## 1. Variables

$R$	Recorrido (secuencia de paradas) []
$s_{rik}^j$	Parada $j$ -ésima de la $r$ -ésima ruta del usuario $i$ , para el día $k$ []
$V_{rik}$	Ruta auxiliar de la $r$ -ésima ruta del usuario $i$ , para el día $k$ []
$J_{ik}$	Secuencia de rutas auxiliares del usuario $i$ , para el día $k$ []
$N_k$	Número de rutas auxiliares tomadas en el día $k$ []
$M$	Distancia máxima de caminata [m]
$O_L$	Número de viajes en transporte público generados en zona $L$ []
$O_{CARD\ i}$	Número de viajes en transporte público generados en zona $L$ asociados a tarjetas []
$D_L$	Número de viajes en transporte público con destino en zona $L$ []
$D_{CARD\ i}$	Número de viajes en transporte público con destino en zona $L$ asociados a tarjetas []
$x_i$	Coordenada UTM $x$ asociada a la bajada de la transacción $i$ [m]
$y_i$	Coordenada UTM $y$ asociada a la bajada de la transacción $i$ [m]
$t_i$	Tiempo asociado a la bajada de la transacción $i$ [s]
$f_w$	Peso asociado a la caminata []
$d_{i-post}$	Distancia euclidiana hasta lugar de transacción siguiente a $i$ [m]
$s_w$	Velocidad de caminata promedio [m/s]
$d$	Distancia máxima de caminata [m]
$etapas_{ijt}$	Etapas de viaje asociadas al origen $i$ , destino $j$ y período $t$ []

$d_{ruta}$	Distancia en ruta [m]
$d_{euclidiana}$	Distancia euclidiana [m]
$N_j$	Conjunto de viajes con igual cantidad de etapas que el viaje $j$ y que no están en la zona crítica []
$t_{subida}^x$	Tiempo de subida asociado a la transacción $x$ [s]
$t_{bajada}^x$	Tiempo de bajada asociado a la transacción $x$ [s]
$d_{x-1;x}$	Distancia euclidiana entre la bajada de la transacción $x-1$ y la subida de la transacción $x$ [m]
$v_{cam}$	Velocidad de caminata [m/s]
$t_{bus}^x$	Tiempo de pasada del último bus asociado a la transacción $x$ en el mismo paradero en el cual fue realizada [s]
$n_{bus}^x$	Número de buses del servicio abordado por el usuario en la transacción $x$ que pasaron por el paradero durante la estadía del usuario en éste []



## 2. Funciones

$d_{rik}$	Parada de bajada asociada a $V_{rik}$ []
$R_L^O$	Coefficiente de expansión para generación []
$R_L^D$	Coefficiente de expansión para atracción []
$O_{Pi}'$	Generación de viajes zona $i$ []
$D_{Pi}'$	Atracción de viajes zona $i$ []
$O_{Pi}$	Generación de viajes zona $i$ balanceada []
$D_{Pi}$	Atracción de viajes zona $i$ balanceada []
$Tg_i$	Tiempo generalizado asociado a la transacción $i$ [s]
$f_{it}$	Factor de expansión para corrección por viajes sin destino, para origen $i$ y período $t$ []
$f_t$	Factor de expansión para corrección por viajes sin origen ni destino, para período $t$ []
$f_{d_j}$	Cociente de distancias asociado al viaje $j$ []
$f_{d_j}^{real}$	Cociente de distancias calculado en base a viajes no pertenecientes a zona crítica, con igual cantidad de etapas que viaje $j$ []
$V_{sol_i}$	Verosimilitud de la solución $i$ []
$t_e^{x-1;x}$	Tiempo estimado de espera entre bajada de transacción $x-1$ y subida de transacción $x$ []
$D_{bus}^x$	Función binaria para pasadas de buses por paradero asociadas a la transacción $x$ []

# Capítulo 1

## Introducción

### 1.1 Motivación

El auge y desarrollo de las tecnologías de la información representa hoy en día una herramienta invaluable en múltiples aspectos de la vida cotidiana. La automatización de procesos que antes requerían largas esperas y expresa intervención humana ha resultado por un lado en un mejor aprovechamiento de recursos tecnológicos, capitales, humanos y de tiempo, y por otra parte ha constituido una revolución en cuanto al nivel de información manejada. Es en este contexto que el transporte público urbano se ve doblemente beneficiado gracias a la penetración de sistemas de cobro vía tarjetas inteligentes y a la utilización de sistemas de posicionamiento global; se tiene como resultado de esto un sistema ágil, eficiente e integrado, dotado además de una amplia base de datos, que representa y describe el comportamiento de millones de usuarios de transporte público día a día. El aprovechamiento de esta información es la base que da pie a la presente tesis.

Mirando en forma específica el caso de Santiago de Chile y su sistema de transporte público, Transantiago, se observa que los datos recabados a través de las tarjetas inteligentes representan a la gran mayoría de los usuarios de transporte público, debido a la casi total penetración de las tarjetas de pago; si además se considera la gran cantidad de información almacenada por transacción, y las posibilidades que existen al cruzar esta información con aquella proveniente de las emisiones GPS de los buses, es posible colegir que un análisis riguroso de estas fuentes de información posee un gran potencial en cuanto a estimación de comportamientos de viaje, patrones de conducta y actividades, y planificación de sistemas de transporte público. El análisis anterior se torna aún más interesante si se cuenta con una segunda base de datos, de otro sistema de transporte público, como lo es el de la ciudad canadiense de Gatineau; la inclusión de estos datos en el análisis y las potenciales comparaciones y complementos posibles al trabajar con estas bases de datos puede sólo enriquecer un estudio de este tipo.

No es hasta el surgimiento de las tarjetas inteligentes de información y su irrupción en los sistemas de transporte público urbano, hacia el año 2000, que este tipo de información es

considerada en los estudios en materia de transporte. Un primer acercamiento a su uso práctico lo constituye la descripción estadística del comportamiento del usuario de transporte público y categorización de la población a través de sus patrones de viaje. Ejemplos de lo anterior se hallan en Morency et al. (2007) y en Agard et al. (2009), en donde se utilizan técnicas de minería de datos para realizar una descripción de la población e identificar patrones por grupos; tales estudios revisten particular interés, dado que los resultados encontrados en este tipo de aplicaciones son comparables a los resultados que podrían encontrarse en estudios de comportamiento a través de datos obtenidos en encuestas de viajes, como el de Olguín et al. (2009). Otra línea de trabajo en el procesamiento y utilización de estos datos la constituye la estimación de matrices Origen/Destino de viajes en transporte público; Trépanier et al. (2007) describe una metodología a través de la cual es posible estimar bajadas en transporte público y estimar destinos de cada una de las etapas de un viaje, para el caso de Montreal; Munizaga & Palma (2012) aplican esta metodología a la ciudad de Santiago, introduciendo un sistema de cálculo y estimación de hora y lugar de bajadas para el modo metro. Similar experiencia se observa en Lianfu et al. (2007) y Zhao & Rahbee (2007) que desarrollan metodologías muy similares aplicadas a los sistemas de transporte público de Changchun y Chicago, respectivamente. En todos estos casos, el enfoque de determinación es a nivel “totalmente desagregado”, esto es, se analiza cada viajero en forma particular, un concepto originalmente introducido por Chapleau (2000).

El estado del arte es, por tanto, la estimación de matrices O/D a nivel de etapas de viaje, entendiendo por una etapa el intervalo espacio/temporal de viaje que transcurre entre una transacción y el instante en el cual el usuario desciende del vehículo (o modo) en el cual realizó dicho tramo. Sin embargo, dada la estructura de muchas ciudades y de sus sistemas de transporte público, un viaje frecuentemente se encuentra constituido por más de una etapa; en consecuencia, la metodología disponible podrá describir exhaustivamente cada una de las etapas en forma individual, no obstante no podrá discriminar cuándo comienza o finaliza un viaje. La actual metodología de estimación de los puntos de origen y destino de cada viajero consiste en definir ventanas de tiempo arbitrarias, lo cual introduce importantes sesgos; para mejorar esta metodología, se propone incorporar información adicional al sistema, como por ejemplo frecuencias de transporte público y atraktividad y accesibilidad a una zona vía uso de suelos, además de establecer criterios de acuerdo al comportamiento encadenado de cada uno de los usuarios, diferenciando y distinguiendo el rol preponderante y dinámico de las dimensiones espacial y temporal.

## 1.2 Objetivos

El principal objetivo del presente trabajo es desarrollar una metodología, partiendo del estado del arte en modelos de estimación de etapas de viaje en un sistema de transporte público y de los análisis conjuntos de las bases de datos de transporte público de las ciudades de Santiago y Gatineau, que permita estimar la ocurrencia de actividades entre las etapas previamente determinadas por modelos anteriores. A su vez, esto posibilitará la identificación de viajes, actividades y sus propósitos en forma precisa, posibilitando la generación de matrices Origen/Destino y perfiles de uso de tiempo, en base a sistemas tecnológicos de transporte público. Finalmente, llevar a cabo una caracterización y un mapeo geográfico-temporal de las actividades estimadas de esta forma, y aplicar la metodología a las bases de datos de Santiago y Gatineau y comparar los resultados obtenidos.

Los objetivos más específicos son:

- Identificar patrones de conducta y casos recurrentes que permitan discriminar actividades de trasbordos en la población, en base a literatura preexistente, formas de funcionamiento del sistema de transporte público y a los datos asociados a uso de suelo y a las transacciones Bip!
- Determinar y caracterizar temporalmente los viajes y actividades llevadas a cabo por los usuarios en una base de datos de transporte público.
- Desarrollar una subrutina que permita determinar los propósitos de las actividades estimadas.
- Aplicar la metodología en las bases de datos de Santiago y Gatineau.
- Llevar a cabo una validación de la metodología general en base a la Encuesta Origen/Destino de Metro de Santiago 2010.
- Comparar los resultados obtenidos en ambas ciudades y explicar las posibles diferencias en un contexto de distintos ámbitos sociales, geopolíticos, operacionales, etc.
- Concluir sobre los alcances de la metodología y su validez como herramienta de evaluación y planificación.

## 1.3 Estructura de la tesis

La primera parte de la tesis corresponde a una revisión bibliográfica del estado del arte en metodologías de estimación de etapas, viajes y matrices Origen/Destino en base a información obtenida de sistemas de transporte público inteligentes; este estudio se presenta en el Capítulo 2.

El Capítulo 3 corresponde a un análisis de las bases de datos de transporte público de las ciudades de Santiago y Gatineau, que apunta a entender las características asociadas a cada una de ellas.

En el Capítulo 4 se presenta la metodología, para lo cual se explica ésta de manera general y en forma posterior se explican separadamente sus principales módulos constituyentes: El Módulo de Estimación de Actividades y Viajes (AIM, *Activity Identification Module*) y el Módulo de Asignación de Propósito (PAM, *Purpose Assignment Module*).

El Capítulo 5 corresponde a un compendio de resultados obtenidos en base a la metodología, tanto de viajes como de actividades. Se incluyen diversas comparaciones y aplicaciones específicas para cada ciudad.

En el capítulo 6 se trata la temática de la validación de la metodología, para lo cual se presentan dos herramientas: la validación endógena y la validación exógena.

Finalmente, el Capítulo 7 corresponde a los alcances de la metodología, las líneas futuras de investigación y las conclusiones finales del trabajo.

# Capítulo 2

## Revisión bibliográfica

El aprovechamiento de la información proporcionada por los sistemas tecnológicos implementados en el transporte público, tales como las tarjetas inteligentes para usuarios y los sistemas de posicionamiento global en buses, posibilita la estimación de matrices de etapas en transporte público, siendo posible describirlas en una dimensión tanto espacial como temporal.

El desafío de estimar una matriz de etapas de viaje (elemento fundamental que será definido con posterioridad en este mismo capítulo) tiene como componente fundamental la determinación de las dimensiones espaciales y temporales de cada uno de los abordajes y bajadas en los diversos modos y/o vehículos de transporte público, para cada uno de los usuarios; esto es, estimar la posición geográfica y la hora en la cual estos eventos ocurren.

En el presente capítulo se realizará primeramente una revisión de los modelos existentes para estimación de abordajes y bajadas utilizando datos proporcionados por sistemas tecnológicos implementados en servicios de transporte público, partiendo por una descripción de los conceptos relevantes del contexto; posteriormente se revisará literatura concerniente a análisis estadísticos y de clasificación que es factible realizar en base a esta misma información.

### 2.1 Etapas de viaje

Uno de los conceptos fundamentales en la presente temática es el de “etapa de viaje”. Por ésta se entiende el tramo de viaje comprendido entre el abordaje de un usuario a un modo o vehículo de transporte público y su posterior bajada de éste.

El concepto de etapa varía según los modos de transporte; mientras para el modo bus se define una etapa como el lapso temporal y geográfico transcurrido entre el abordaje y descenso de un vehículo, en los modos metro y tren se entenderá una etapa como el lapso temporal/ geográfico durante el cual el usuario hará uso del modo en cuestión, sin

representar una interrupción de etapa los transbordos de vehículo entre líneas que no involucren una nueva transacción de pago. La anterior diferencia se efectúa debido a que en el caso del modo metro no se registra una nueva transacción al efectuarse trasbordos intra-modales.

Una etapa de viaje quedará espacial y temporalmente definida por las coordenadas geográficas y tiempos, tanto del abordaje que dio inicio a la etapa como de la bajada que constituye su fin. Adicionalmente, y dependiendo de la tecnología disponible en los sistemas de transporte público, una etapa puede poseer información complementaria relativa al ID de la tarjeta que efectúa la transacción, el modo, recorrido y vehículo específico en que fue realizada la etapa, tarifa recaudada, y otros; suplementariamente, el procesamiento de esta información y su cruce con información manejada por las centrales de servicios de transporte público permite agregar datos adicionales, como operadores de los vehículos involucrados en la etapa, etc.

Entendiendo un viaje como el desplazamiento y tiempo conllevado por un usuario entre que parte desde su origen y llega a su destino (lugar en donde pretende realizar la actividad que motivó su viaje en primer lugar), un viaje podrá en términos generales estar constituido por una o más etapas, dependiendo del número de transbordos vehiculares y/o cambios de modo que el usuario deba efectuar en su recorrido.

## **2.2 Tipos de operación y transacciones**

Uno de los factores clave en el contexto de las aplicaciones tecnológicas para transporte público, lo constituye la utilización de tarjetas inteligentes; a través de estas, los usuarios pueden realizar transacciones sin necesidad de hacer ningún tipo de contacto, acercando las tarjetas a validadores que emiten campos electromagnéticos en su proximidad. El conjunto constituye una operación de transacción monetaria fácil y rápida (Finkenzeller, 2003), en la cual como ya se ha dicho se recaba además una gran cantidad de información.

Dependiendo de las características particulares de cada sistema de transporte público, un usuario puede ser requerido de realizar una validación sólo al abordar un modo o vehículo (caso Santiago, Chile), o alternativamente realizar validaciones tanto al subir como al bajar de un modo o vehículo (caso metro de Londres, Inglaterra). En el contexto de la identificación de etapas de viaje, el segundo caso no reporta ninguna dificultad para determinar cada una de las etapas que constituyen los viajes de un usuario, ya que existe un registro de tanto el inicio como del fin de cada una de las etapas, correspondientes respectivamente a la validación de entrada y la de salida del vehículo o modo. Es en el primer caso, en el cual sólo se requieren validaciones en el abordaje, en el cual es necesario

un modelo de estimación de bajadas para describir en detalle cada una de las etapas asociadas a los viajes de un usuario.

## **2.3 Modelos de Estimación de Bajadas y Generación de Matices Origen/Destino**

Se presentan a continuación una serie de modelos que tienen como finalidad resolver el problema de estimación de bajadas para etapas, introducido en el acápite anterior. Es importante mencionar que cada uno de estos modelos ha sido desarrollado en el contexto de un sistema de transporte público en particular, y por tanto incluyen elementos singulares que responden al nivel de tecnología presente en cada uno de estos sistemas, como también a sus características de operación.

### **2.3.1 Modelo Hofmann y O'Mahony (2005)**

El Modelo de Hofman y O'Mahony (2005) constituye un primer acercamiento al uso de tarjetas inteligentes para estimar dinámicas de movimiento de los usuarios en un sistema de transporte público y matrices Origen/Destino; si bien la fuente de los datos, así como la ciudad de aplicación no es reportada por los autores por cuestiones de confidencialidad, se da a conocer que se trata de un sistema de transporte público en el cual sólo se dispone del modo bus, con dos posibles formas de pago: tarjeta de transporte público o efectivo, la que sólo se realiza al abordar los vehículos sin necesidad de registrar la salida del sistema por parte del usuario. El sistema cuenta además con una plataforma de recolección de datos automatizada, que almacena para cada transacción realizada con tarjetas el ID asociado a la tarjeta (único para cada dispositivo), la hora de la transacción y la ruta de transporte público en la cual se realizó la transacción; ordenando esta información primero por usuario y posteriormente por orden cronológico de transacciones, se puede realizar un seguimiento ordenado de los lugares en los cuales un usuario validó su tarjeta; cruzando además esta información con el sistema que recolecta información sobre la hora en que un bus llega a un paradero, es posible además estimar las coordenadas geográficas de la transacción asociándolas a la posición del paradero que se registró anterior a la hora de la transacción.

El problema que surge naturalmente es, como ya se ha mencionado, la identificación de trasbordos y actividades; el modelo de Hofmann y O'Mahony (2005) lo afronta estableciendo una serie de reglas bajo las cuales se consideraría que un determinado patrón de transacciones serían trasbordos. Para que dos transacciones consecutivas de un mismo usuario (mismo ID) puedan ser consideradas dos etapas de un viaje (es decir que entre ellas se efectuó un trasbordo y no una actividad), deben cumplir con los siguientes requisitos:



- Ambas transacciones se registraron en el mismo día cronológico.
- La segunda transacción fue efectuada entre 1 y 90 minutos posteriormente a la primera.
- El recorrido de transporte público asociado a la segunda transacción debe ser distinto del recorrido asociado a la primera.

El criterio de los 90 minutos como límite entre transacciones es fundamentado por los autores en base a estudios y encuestas cuyas conclusiones apuntan a que el 99% de los usuarios del sistema en estudio demoraría menos de 90 minutos en concluir una etapa de su viaje y realizar un trasbordo para una etapa subsecuente; sin embargo, esto no asegura que una actividad no se haya realizado dentro de este intervalo, lo cual constituye una de las limitaciones del modelo.

Procesando todas las transacciones aplicando el criterio anterior, se puede obtener un mapa espacio-temporal de los viajes, a partir del cual obtener una matriz Origen/Destino es directo.

### 2.3.2 Modelo Trépanier et al. (2007)

Este modelo fue diseñado por Trépanier et al. (2007) para el sistema de transporte público de la localidad de Gatineau, Québec; contempla un único modo, bus, que consta en total de una flota de 302 buses, y cuya modalidad de operación en cuanto a transacciones es de validación sólo al inicio de las etapas, llevándose éstas a cabo a bordo del bus. Los buses se encuentran equipados con Sistemas de Posicionamiento Global y emisores de localización, información en base a la cual es posible obtener las coordenadas en que se registra cada transacción.

Definiendo los índices  $i$  como usuario (dado por el ID único de la tarjeta),  $j$  como indicador de la secuencia de paradas en una ruta de transporte público,  $r$  como el indicador de número de ruta de transporte público que el usuario toma en el día,  $k$  como el indicador de día; una ruta de transporte público estará entonces definida, en un sentido único, por una secuencia ordenada de paradas  $s$ :

$$R = S^j \quad (2.1)$$

Luego la parada  $j$ -ésima de la  $r$ -ésima ruta del usuario  $i$ , para el día  $k$ , estará dada por la expresión  $S_{rik}^j$ . Definiendo ahora el concepto de “ruta auxiliar”,  $V_{rik}$ , esta corresponde a la

secuencia de paradas donde el usuario puede descender posterior a haber abordado una ruta dada en una parada  $j=B$ ;

$$V_{rik} = S_{rik}^j, \forall j > B \quad (2.2)$$

La jornada de etapas  $J_{ik}$  corresponde a la secuencia de rutas auxiliares tomadas por el individuo  $i$  en el día  $k$ ; siendo por tanto  $N_k$  el número de rutas auxiliares que el individuo toma en el día  $k$ :

$$J_{ik} = V_{rik}, r = 1, \dots, N_k \quad (2.3)$$

La distancia es un elemento clave a la hora de estimar el punto de bajada, esto bajo el supuesto lógico que las personas caminarán para acceder de los paraderos de una ruta a los de la siguiente que tomará, pero sólo hasta un cierto límite. Definiendo  $d(a,b)$  como la distancia euclidiana entre los puntos  $a$  y  $b$ , el modelo debe estimar la localización óptima de bajada, considerando los valores de  $r$  asociados a cada usuario; se debe por tanto estimar la parada de bajada  $d_{rik}$  de la ruta auxiliar donde la distancia a la parada de abordaje de la siguiente ruta  $S_{(r+1)ik}^B$  sea mínima:

$$d_{rik} = z \rightarrow \underset{z}{\text{Min}} d\left(S_{(r+1)ik}^B, z\right) \quad z \in V_{rik} \quad (2.4)$$

$$r < N_k \quad d\left(S_{(r+1)ik}^B, z\right) < M$$

Se introduce un valor de tolerancia para la máxima distancia de caminata  $M$ . La ecuación (2.4) se utiliza para todos los casos en que existe una transacción siguiente en el día, denominados “casos normales”. Para la última etapa del día ( $r=N_k$ ), se deben realizar supuestos de cara a estimar la bajada, dado que no existe una siguiente transacción. Un primer supuesto corresponde a asumir que el individuo retornará a la parada en la que realizó la primera transacción del día, si es que ésta se relaciona con la ruta tomada ( $r=N_k$ ):

$$d_{rik}' = z \rightarrow \underset{z}{\text{Min}} d\left(S_{lik}^B, z\right) \quad z \in V_{rik} \quad (2.5)$$

$$r = N_k \quad d\left(S_{lik}^B, z\right) < M$$

Alternativamente, y de fallar la búsqueda de bajadas bajo el criterio de la ecuación (2.5), se asume que el individuo terminará su día en la parada que corresponda a aquella donde se efectúe la primera transacción del día siguiente:

$$d_{rik}'' = z \rightarrow \underset{z}{\text{Min}} d\left(s_{li(k+1)}^B, z\right) \quad z \in V_{rik} \quad (2.6)$$

$$r = N_k \quad d\left(s_{li(k+1)}^B, z\right) < M$$

Si existe alguna etapa a la cual no pueda serle estimada una bajada a través de la batería de ecuaciones (2.4), (2.5) y (2.6), se procede a comparar la transacción correspondiente a esa etapa con otras transacciones efectuadas en el mismo día y que procedan de la misma ruta, y se asignará como bajada aquella de la transacción que registre una menor diferencia de tiempo con la transacción cuya bajada no podía ser determinada.

La metodología general se resume en la siguiente figura:

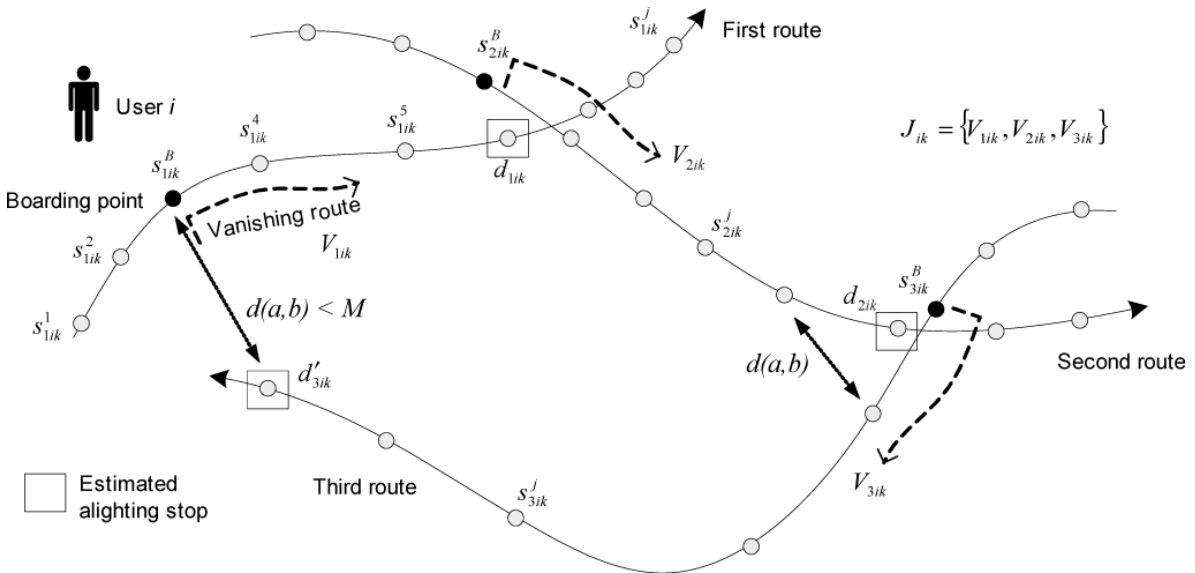


Figura 2.1: Modelo de estimación de bajada para etapas  
Fuente: Trépanier et al. (2007)

### 2.3.3 Modelo Zhao et al. (2007)

El modelo de Zhao et al. (2007) fue diseñado para inferir una matriz de etapas de viaje en el sistema de transporte público de la ciudad de Chicago. Al igual que en los modelos anteriores, la validación se lleva a cabo sólo al inicio de las etapas, sin embargo este modelo contempla integración modal, al existir tanto bus como tren (3 líneas de tren); las

validaciones se efectúan a bordo de los buses y en las estaciones en las cuales el usuario aborda el tren. También en este sistema de transporte público se cuenta con información geo-referenciada de un alto porcentaje de los buses, en tiempo real, y además se cuenta con información de las transacciones realizadas en las estaciones de tren.

La metodología general de inferencia de bajadas en este caso es igual al modelo de Trépanier et al. (2007). La riqueza adicional de este estudio corresponde a la integración modal. En términos generales, y debido a la posibilidad de integrar la información de ambos modos, la línea de tren es tratada como una línea de bus para términos de inferencia metodológica de etapas. La única diferencia radica en la estimación del posicionamiento de cada transacción para el tren, que es más directa; dado que se posee el registro de la estación en la cual se llevó a cabo, obtener las coordenadas de dicha transacción es directo. La siguiente figura ejemplifica el caso de una estimación de bajada para el caso de combinación intermodal bus-tren-bus:

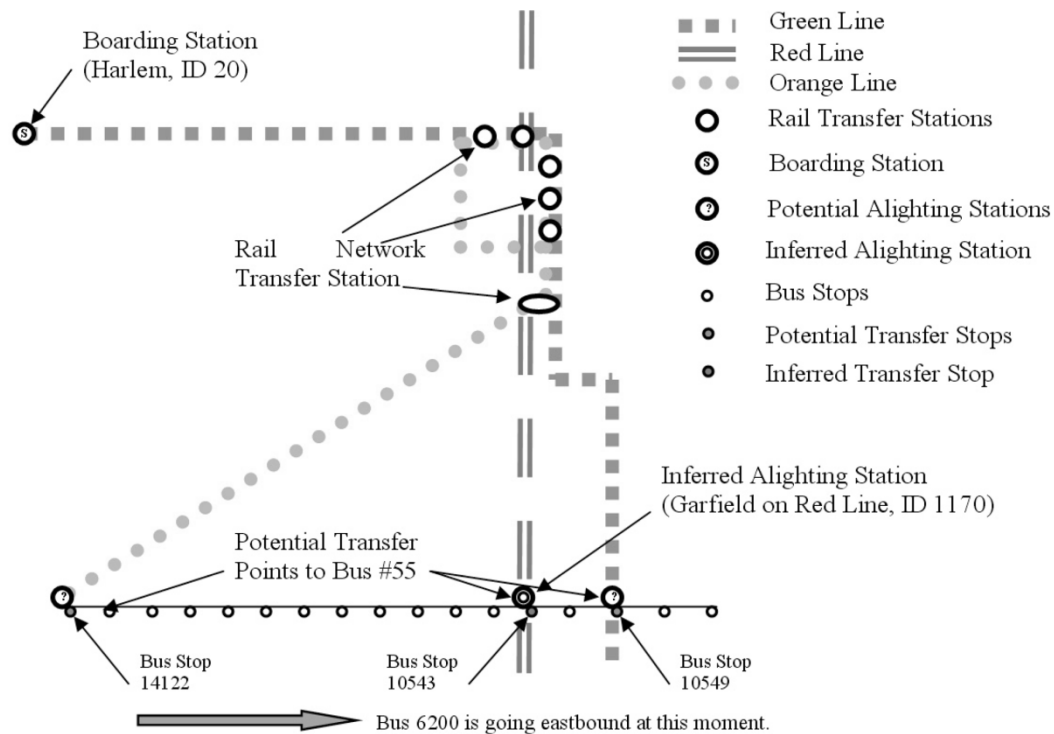


Figura 2.2: Estimación de bajadas intermodales para etapas

Fuente: Zhao et al. (2011)

### 2.3.4 Modelo Lianfu et al. (2007)

Lianfu et al. (2007) propone una metodología de recolección de datos y de estimación de matrices Origen/Destino de viajes para el sistema de transporte público de la ciudad de Changchun. Un aspecto importante a recalcar sobre este sistema es que estimaciones indican que sólo un 20-25% de los usuarios utilizan tarjetas inteligentes, no obstante

encuestas revelan que el comportamiento en términos de distribución de viajes de este tipo de usuarios es similar al de los usuarios de boleto unitario. Este hecho será relevante a la hora de estimar las matrices de viajes, ya que será necesaria la aplicación de factores de corrección de cara a una correcta representación de la población.

En el sistema de transporte público de Changchun, China, se observa un solo modo, bus; la validación se lleva a cabo a bordo de los buses, quedando registrada la hora de la transacción; paralelamente, cada chofer debe a lo largo del recorrido llenar una hoja itinerario, informando sobre las horas exactas de llegada a cada parada. Un cruce de información es luego realizado, permitiendo conocer la parada de subida asociada a cada transacción, comparando tiempos del itinerario con los de la transacción.

Se procede posteriormente a analizar, para una misma transacción, el período punta mañana y punta tarde. Un fuerte supuesto del modelo es que el destino del viaje (asumido único) realizado en punta mañana será el paradero en el cual se genere, para ese usuario, la primera transacción en punta tarde; viceversa, el destino del viaje realizado en punta tarde (también asumido único) será el paradero que haya generado la primera transacción en el período punta mañana.

Obtenida esta información, se dispone mediante un reordenamiento de datos de una matriz Origen/Destino de viajes de las transacciones efectuadas mediante tarjetas inteligentes a nivel de paraderos. Se procede luego a estimar la matriz Origen/Destino representativa de todos los usuarios de transporte público a través de la siguiente metodología. El primer paso consiste en establecer una zonificación adecuada de acuerdo a la conformación urbana y distribución de usuarios de transporte público. Cada una de estas zonas podrá contener varias paradas de buses, de distintos recorridos. A continuación, se calculan coeficientes de expansión para generación  $R_L^O$  y para atracción  $R_L^D$  a través de las siguientes ecuaciones:

$$R_L^O = \frac{O_L}{\sum_{i=1}^n O_{CARD\ i}} \quad (2.7)$$

Donde  $O_L$  corresponde a la generación de viajes de transporte público de la zona  $L$ ,  $O_{CARD\ i}$  corresponde a la generación de etapas asociadas a transacciones de tarjetas inteligentes en la zona  $i$ ,  $n$  corresponde al número de paradas de bus en la zona  $i$ .

$$R_L^D = \frac{D_L}{\sum_{i=1}^n D_{CARD\ i}} \quad (2.8)$$

Donde  $D_L$  corresponde a la atracción de viajes de transporte público de la zona  $L$  (Igual a  $O_L$  del otro período para el mismo día),  $D_{CARD\ i}$  corresponde a la atracción de etapas asociadas a transacciones de tarjetas inteligentes de todas las paradas en la zona  $i$ ,  $n$  corresponde al número de paradas de bus en la zona  $L$ .

El siguiente paso corresponde a calcular la generación  $O_{Pi}$  y atracción  $D_{Pi}$  de viajes a nivel zonal. Definiendo:

$$O_{Pi}' = O_{CARD\ i} \cdot R_L^O \quad (2.9)$$

$$D_{Pi}' = D_{CARD\ i} \cdot R_L^D \quad (2.10)$$

Con un alto grado de probabilidad,  $\sum O_{Pi}' \neq \sum D_{Pi}'$ , y por tanto se hace necesario balancear las cantidades; para esto se mantienen constantes los  $O_{Pi}'$  y se recalculan los  $D_{Pi}'$ . Para esto sean:

$$\begin{aligned} O_{Pi} &= O_{Pi}' \\ D_{Pi} &= D_{Pi}' \cdot \left( \frac{\sum_{i=1}^n O_{Pi}'}{\sum_{i=1}^n D_{Pi}'} \right) \end{aligned} \quad (2.11)$$

Donde  $O_{Pi}$  y  $D_{Pi}$  son la generación y atracción de viajes de transporte público, respectivamente, posterior al balanceo.

El paso final corresponde a distribuir los vectores de generación y atracción en la matriz Origen/Destino. Esto se realiza a través del método de Furness, utilizando como matriz a priori la matriz de viaje correspondiente a las transacciones realizadas con tarjetas inteligentes, previamente determinada en la metodología, e imponiendo generaciones y atracciones totales sobre zonas iguales a  $O_{Pi}$  y  $D_{Pi}$ , respectivamente.

### **2.3.5 Modelo Munizaga y Palma (2012)**

En este trabajo se utiliza la información proveniente de la base de datos de transacciones realizadas en el sistema de transporte público de Santiago, Transantiago. Este sistema posee muchas particularidades que lo hacen idóneo de cara a utilizar los datos recolectados. En primer lugar, la penetración de tarjetas inteligentes en la población de usuarios es cercana al 100%, debido en gran medida a que la integración tarifaria intermodal es sólo posible con este modo de pago, y al hecho de que es la forma de pago única para el caso del modo bus; sólo una pequeña porción de usuarios recurre al boleto unitario del metro como forma de pago, correspondiente a quienes sólo realizan una única etapa por viaje en transporte público y específicamente en el modo metro. Es por tanto posible deducir, sometiendo estos datos a análisis, un comportamiento representativo de la totalidad de los usuarios. No obstante lo anterior, una desventaja de este método de pago corresponde al alto nivel de evasión registrado en el modo bus.

Otro factor relevante corresponde al avanzado sistema tecnológico que contempla Transantiago; la totalidad de los buses se encuentra dotado de un sistema de posicionamiento global, que emite las coordenadas del vehículo cada 30 segundos; adicionalmente, por cada transacción registrada se almacena gran cantidad de datos: vehículo o estación de metro en la cual se realizó, hora exacta, tarifa cancelada, etc. El post-procesamiento de esta información permite, con una metodología similar a la implementada en Trépanier et al. (2007), estimar coordenadas y horas de subidas y bajadas de etapas; a su vez, la utilización de esta información da pie a interesantes análisis de patrones de viajes y actividades; este es el foco principal del trabajo.

Se describe a continuación la metodología empleada en este trabajo para determinar las etapas; ésta es una extensión al trabajo de Trépanier et al. (2007), que contempla la inclusión de un costo generalizado en lugar de una distancia geográfica para estimar paraderos de bajada, y además adiciona el modo metro y el elemento de zonas pagas para abordar buses. Estos últimos corresponden a paraderos de recinto cerrado en los cuales el usuario paga su pasaje al acceder a éste, en validadores ubicados en la entrada. De esta forma, los buses que paran en estas zonas abren todas las puertas para que los pasajeros puedan abordar, dado que no es necesario que entren por la puerta frontal para cancelar su pasaje. Este sistema de paraderos prepagos contribuye a agilizar los tiempos marginales de abordado de pasajeros y por ende tienen un impacto en los tiempos de ciclo. Es importante destacar también que más de una línea de bus puede pasar por cada zona paga

Munizaga y Palma (2012) proponen estimar la bajada de una etapa como la posición de la trayectoria del bus abordado que minimiza el tiempo generalizado al punto de la siguiente

transacción. Cada bajada estará caracterizada por una terna  $(x_i, y_i, t_i)$ , que describe las coordenadas geográficas y la hora de la bajada para la posible bajada  $i$ , respectivamente. Se define el tiempo generalizado  $Tg_i$  asociado a la transacción (o etapa)  $i$  como

$$Tg_i = t_i + f_w \cdot \frac{d_{i-post}}{S_w} \quad (2.12)$$

Donde  $f_w$  corresponde al peso de la desutilidad asociada a caminar,  $d_{i-post}$  corresponde a la distancia geográfica euclidiana desde la posición  $i$  hasta el lugar de la siguiente transacción, y  $S_w$  es la velocidad de caminata promedio.

El problema de optimización asociado a encontrar el punto óptimo de bajada es el siguiente:

$$\begin{aligned} & \text{Min}_i Tg_i \\ & \text{s.a. } d_{i-post} < d \end{aligned} \quad (2.13)$$

Siendo  $d$  la distancia euclidiana tolerable como límite de caminata.

Para el caso del Metro, es conocida la estación de subida de cada etapa; para estimar la estación de bajada, se utiliza un procedimiento similar al del caso del bus; se busca, en torno a la siguiente transacción y dentro de un radio  $d$ , la estación de bajada más cercana a la posición de la siguiente transacción; para estimar la hora de bajada de las etapas realizadas en metro<sup>1</sup> se utiliza una estimación de tiempo de viaje que contempla los tiempos en vehículo, tiempos de caminata desde los accesos de la estación hasta el andén, tiempo de espera en andén y tiempo de trasbordos. Para estimar la ruta utilizada dentro del metro, se emplea un algoritmo de rutas mínimas Dijkstra, y se asume que el individuo utiliza la ruta de menor costo.

Las zonas pagas representan un problema particular, ya que en éstas la validación se efectúa a la entrada del recinto; por ende, es a priori desconocido el vehículo y el recorrido abordado por el usuario, siendo el universo factible todas aquellas líneas que se detienen en el paradero asociado a la zona paga. Por ende, se hace necesaria la implementación de un módulo que estime el vehículo abordado por el usuario, para posteriormente, en base a las emisiones de ese vehículo en particular y a la siguiente transacción asociada, estimar las coordenadas y hora de bajada. En este módulo, se aplica la metodología propuesta por

---

<sup>1</sup> Para el caso de las etapas realizadas en bus no es necesario aplicar una metodología elaborada para estimar la hora de bajada; simplemente se utiliza una interpolación espacio-temporal tomando como referencia las emisiones GPS del bus abordado geográfica y secuencialmente anterior y posterior a la bajada.



Chriqui & Robillard (1975) para definir líneas comunes de transporte público, considerando como destino del tramo de línea común las coordenadas de la siguiente transacción; a través de este procedimiento, se “asigna” una línea y específicamente un vehículo al usuario.

En base a este procedimiento es posible describir las etapas de viaje de la población de usuarios. Lo siguiente corresponde a identificar, en base a las etapas, los viajes. Para esto, se toma el supuesto de que una actividad sucederá cada vez que se produzca una ventana de 30 minutos o más entre el tiempo estimado de bajada y el tiempo de la siguiente transacción<sup>2</sup>. En virtud de la definición dada de viaje, una actividad marcará el fin de un viaje, al considerarse la localización de ocurrencia de esta actividad como destino; subsecuentemente, la transacción posterior corresponderá al comienzo de un nuevo viaje, siendo la localización anteriormente mencionada el origen de este nuevo viaje.

Una vez que los viajes han sido estimados y la matriz procedente de los datos ha sido conformada, corresponde aplicar factores de balance o de corrección a dicha matriz, dado que no todos los viajes son reconocibles con el procedimiento explicado, o bien algunas etapas (y consecuentemente viajes) han arrojado una estimación errática; particular interés revisten aquellos viajes no detectables debido a la evasión, efecto no homogéneo en la ciudad.

Para subsanar aquellos casos en los cuales no se detectó destino, se construye la siguiente familia de factores de expansión:

$$f_{it} = \frac{\sum_j etapas_{ijt}}{\sum_{j \neq nulo} etapas_{ijt}} \quad (2.14)$$

Donde  $i$  corresponde al origen,  $j$  corresponde al destino y  $t$  corresponde al período. Aquellos casos en los cuales no se pudo estimar ni el origen ni el destino de la transacción son incluidos a través de la siguiente familia de factores:

$$f_t = \frac{\sum_{ij} etapas_{ijt}}{\sum_{ij \neq nulo} etapas_{ijt} \cdot f_{it}} \quad (2.15)$$

La evasión es un problema que podría ser abordado con una familia de factores similares a la ecuación (2.15), sin embargo la información disponible respecto a esta problemática no permite fundamentar supuestos en virtud de los cuales se pueda corregir la matriz.

---

<sup>2</sup> Las implicancias y limitaciones de este supuesto revisten particular importancia en el presente trabajo y son discutidas posteriormente en este capítulo.

## **2.4 Utilización de la información para describir a la población**

Uno de los usos más interesantes que posee la información recabada tanto a través de encuestas como también a través de sistemas tecnológicos corresponde a la clasificación de los patrones de conducta de la población, e identificación de actividades. La descripción en clústeres (o categorías) de la conducta de los usuarios de transporte público y de sus patrones de actividades aporta información muy relevante de cara a realizar una modelación que permita diferenciar los transbordos de la generación de actividades. Por este motivo, se procede a continuación a realizar una breve revisión sobre literatura relacionada.

Una primera revisión interesante para detectar patrones conductuales útiles a la hora de modelar comportamientos y sucesos de actividades, la constituye el trabajo de Olguín et al. (2009), que en base a los resultados de la Encuesta Origen/Destino 2001 realiza un análisis de comportamientos y preferencias de la población de Santiago de Chile, distinguiendo diferentes tipos de conductas para distintas categorías de usuarios; todo este análisis conductual se apoya en la teoría microeconómica de actividades y viajes descrita en su forma más general en Jara-Díaz et al. (2008).

Uno de los resultados más interesantes de la investigación de Olguín et al. (2009) de cara al presente trabajo lo constituye la caracterización del patrón de actividades por tipo de día para cada propósito; conocer a qué hora la mayor parte de la población está desarrollando sus actividades, en función del propósito de esta última, sin duda ayuda a construir un modelo explicativo de generación de actividades.

En la Figura 2.3 se presenta una caracterización del uso de tiempo de los individuos a lo largo del día. Cada uno de los propósitos considerados en el estudio está representado por una curva, que describe para las distintas horas del día qué porcentaje de la población se encuentra realizando cada tipo de actividad. Las horas punta de viajes son claramente identificables en el día laboral, como también las horas recurrentes de trabajo entre las 8:00 y las 19:00 para gran parte de la muestra; en dicho tipo de día es además observable que muy poca gente se encuentra realizando otro tipo de actividades que no sea trabajo o estar en el hogar. Similares análisis se pueden realizar para el día sábado, en el cual alrededor de un 35% de los individuos trabaja en horas cercanas al medio día, y para el día domingo, en el cual la actividad totalmente preponderante es estar en el hogar, y la segunda actividad más importante pasa a ser la recreación, concentrándose esta en horas posteriores al almuerzo (15:00- 21:00).

Otra caracterización interesante de este trabajo es el análisis de los patrones de actividades según el sector geográfico; este estudio evidencia que en sectores diferentes de la ciudad,

los patrones de conducta de los individuos son distintos, en base a que las actividades que se desarrollan son asimismo diferentes.

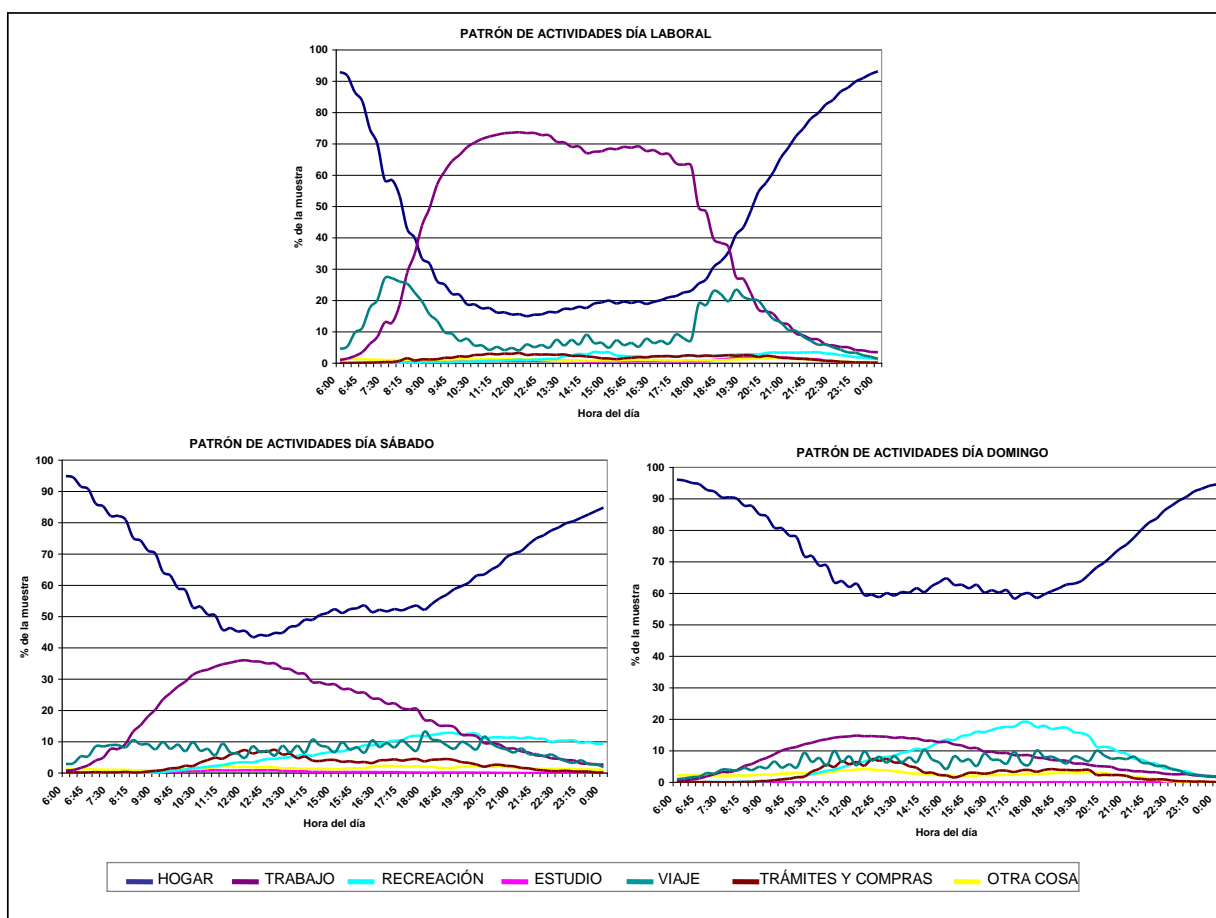


Figura 2.3: Patrones de Actividad según Día

Fuente: Olgún et al. (2009)

La duración de las actividades es otro tema abordado por el trabajo de Olgún et al. (2009) que reviste importancia en la temática de la presente tesis. En la Figura 2.4 se presenta una serie de histogramas que representan las horas al día que en promedio un individuo dedica a las distintas actividades según propósito, tipo de día y sector de residencia. Se observa que en general la duración de las actividades dependería fuertemente de estos factores al registrarse marcadas diferencias; por ejemplo, los fines de semana existe una mayor cantidad de actividades relacionadas al hogar, y también mayor cantidad de horas asignadas a la recreación.

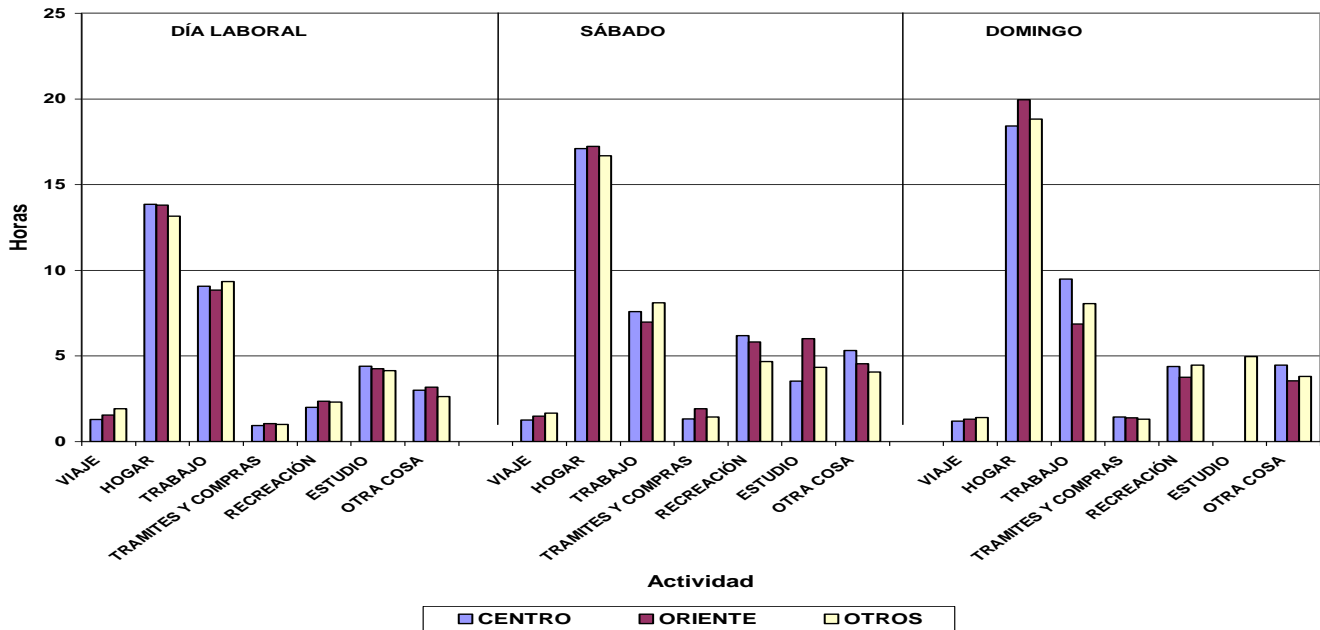


Figura 2.4: Duración Promedio de Actividades según sector

Fuente: Olgún et al. (2009)

En otro trabajo relacionado con la utilización de datos para efectos de comprender el comportamiento del usuario, Bagchi y White (2005) desarrollan un análisis más bien valorativo de los datos obtenidos a partir del sistema de recolección automático de datos de transporte público de la ciudad de Bradford, Inglaterra. Se reconoce la relevancia de las dimensiones espaciales y temporales de cara a identificar comportamiento de usuarios, y se obtienen estadísticas de las tasas de uso de transporte público para distintos tipos de usuarios (identificables en base al tipo de tarjeta utilizado); se realiza también un interesante análisis sobre el abandono del uso del transporte público, y consecuentemente fluctuaciones en la partición modal, en base a la inactividad de tarjetas inteligentes.

Morency et al. (2007) presenta un interesante análisis en el cual se utilizan los datos provenientes del sistema de transporte público de Gatineau para catalogar la variabilidad de comportamientos de los distintos tipos de usuarios de transporte público; empleando metodologías de minería de datos, se extraen importantes conclusiones sobre los patrones de variabilidad tanto espacial como temporal de viajes, encontrándose en general marcados tipos de patrones de comportamiento que presentan conductas muy distintas en términos de viajes realizados a lo largo del día y a la hora en que se efectúan; mientras algunos usuarios poseen una estructura de viajes más bien simétrica, viajando en la mañana y tarde en horarios similares para un período prolongado de tiempo de observación, otros son más bien viajeros irregulares; analizar la dimensión temporal de los viajes y los patrones conductuales de los viajeros en este aspecto resulta útil para comprender las preferencias

de la población, siempre entendiendo que detrás de cada viaje hay una necesidad que la motiva.

La Figura 2.5 describe uno de los principales hallazgos de Morency et al. (2007), que dice relación con la determinación de 4 patrones recurrentes de viaje, presentados según las categorías de tarjeta inteligente; para cada tipo de tarjeta, esto es, distintos tipos de usuario, se distinguen 4 grupos de comportamiento, reflejados en el eje y. Los puntos o centroides reflejan a qué horas realiza una mayor cantidad de transacciones cada categoría de tarjeta y sub-categoría o grupo de comportamiento. Se distinguen diferencias no solo entre los distintos tipos de tarjetas (de esta forma se observa que en general los adultos mayores se comportan diferente en términos de las horas del día en que realizan sus viajes), sino también para un mismo tipo de tarjeta. Los viajeros “simétricos”, que viajan una vez en la mañana y otra vez en la tarde, son identificables en varios sub-grupos (A-I 1, A-I 2, A-I 3, A-E 1, A-E 2, A-E 3, A-E 4, A-R 1, A-R 3), diferenciándose principalmente en las horas en las que efectúan sus viajes. Otros subgrupos tienden a realizar viajes sólo en el mediodía (A-I 4, A-R 4, S 1).

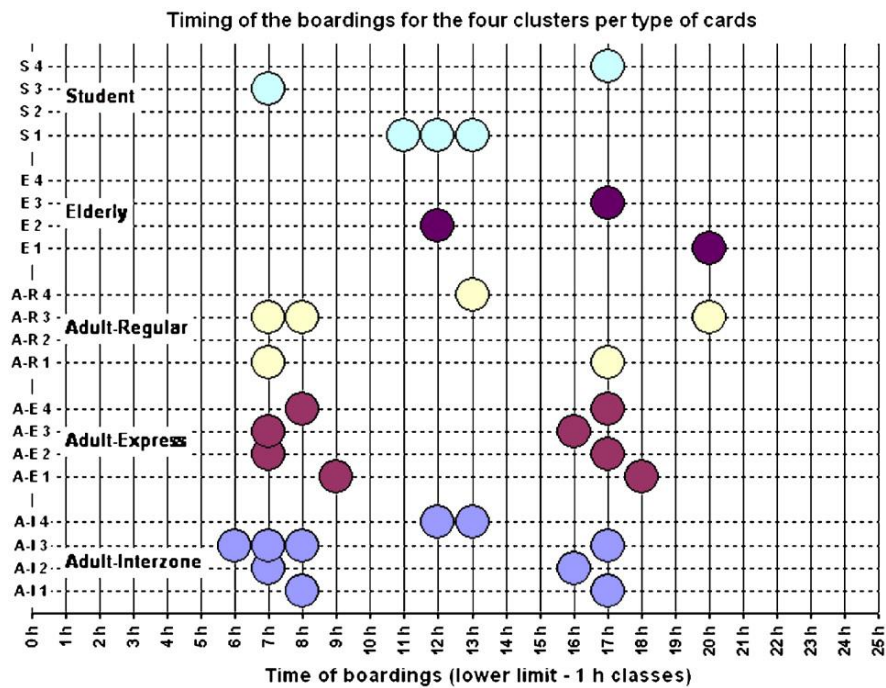


Figura 2.5: Horas de viaje para los distintos tipos de tarjeta y sus subcategorías.

Fuente: Morency et al. (2007)

Agard et al. (2009) emplea técnicas similares con la misma base de datos y amplía el trabajo de Morency et al. (2007), profundizando el análisis de los patrones de comportamiento en un contexto temporal más amplio, analizando variaciones en el uso de transporte público a nivel semanal y mensual, para los distintos tipos de usuarios.

## **2.5 Discusión del estado del arte en estimación de etapas y viajes**

Se han presentado en este capítulo diversas metodologías que tienen como finalidad la estimación de etapas y viajes en transporte público. Una primera observación importante es que la metodología más actualizada de identificación de viajes (Munizaga et al. 2011) realiza el supuesto de que toda ventana de tiempo, entre bajada de una etapa y la siguiente validación, correspondiente a 30 minutos o más, constituye una actividad. Esto implica que aquellas actividades cuya duración sea en la realidad menor a 30 minutos, como podría ser pagar una cuenta o ir a dejar o a buscar a alguien, no están siendo recogidas, a la vez que los viajes que motivan estas actividades no son detectados. Por otro lado, si eventualmente hubiese trasbordos con un tiempo de espera mayor a 30 minutos, estas ventanas de tiempo serían incorrectamente identificadas como actividades. No se ha abordado tampoco, hasta el momento, el problema de identificar propósitos de las actividades.

Además de los problemas de identificación presentados, el supuesto induce sesgos importantes; tal y como se reporta en el trabajo de Olguín et al. (2009), es natural que la duración de una actividad se vea directamente relacionada a múltiples factores, como el sector en el cual ocurre (características de uso de suelo), la hora del día y el propósito de la actividad misma; paralelamente, el tiempo de trasbordo dependerá directamente de las condiciones locales de operación de transporte público. La inclusión de estos elementos en un modelo que tenga como propósito predecir el suceso de actividades y, subsecuentemente, los viajes que estas actividades motivan, es el foco fundamental del presente trabajo.

# Capítulo 3

## **Análisis de las bases de datos de transacciones**

Entender las características asociadas a las dos bases de datos que se utilizan en el presente estudio, y ser capaces de extraer la mayor cantidad de información descriptiva posible de las poblaciones involucradas, es un requisito fundamental para poder entender los fenómenos que rigen el comportamiento de los usuarios de transporte público. En este contexto, las bases de datos asociadas a transacciones en transporte público contienen una gran cantidad de información sobre los patrones de conducta de los usuarios en múltiples aspectos: veces que el usuario hace uso del servicio al día, preferencias de modos y líneas, horarios y día de la semana en que se usa el servicio, y en general numerosas variables que permiten llevar a cabo una categorización según categorías de usuarios, para poder finalmente entender de mejor forma el comportamiento general de la población en el sistema. Más aún, resulta de interés para el presente estudio analizar las diferencias y similitudes entre ambas bases de datos, tanto las explicables tecnológicamente como también socialmente.

La estructura del presente capítulo es la siguiente: en primera instancia se revisará la base de datos correspondiente a la ciudad de Santiago, Chile; se repasará primero los orígenes de esta base de datos a partir de la instauración del sistema actual de transporte público; posteriormente se describirán las características fundamentales de esta base de datos y se analizará la información y variables contenidas en ella; luego se analizará el comportamiento de viaje del usuario de transporte público de Santiago tanto en la dimensión temporal como en la dimensión espacial. En forma consecutiva, se procederá a revisar la base de datos de la ciudad de Gatineau, Canadá.

## **3.1 Base de datos de Santiago, Chile**

### **3.1.1 Transantiago: sobre el origen y otros atributos interesantes de las bases de datos de transporte público de Santiago**

Las bases de datos de uso de transporte público para el caso de Santiago surgen como una consecuencia indirecta de la reestructuración del sistema concretada en Febrero de 2007, tras la implementación de Transantiago. Esta transición marcó un antes y un después en la estructura y funcionamiento del sistema de transporte público de la ciudad de Santiago, llevándose a cabo modificaciones y reestructuraciones medulares; atrás quedaba el sistema de las “micros amarillas”, cuyo origen se remontaba al sistema de licitaciones implementado por el primer gobierno de la concertación en 1990, heredero a su vez de la liberalización del sistema de transporte llevada a cabo durante la dictadura militar entre los años 1979 y 1983 (Díaz et al. 2004), y que permitía un virtual libre albedrío a cualquier empresario deseoso de probar suerte en el área del transporte público urbano, con escasa o nula intervención y fiscalización por parte del gobierno en lo que respecta al sistema tarifario, al diseño de recorridos y a la calidad del servicio. En su lugar, se concibe un nuevo sistema en el cual el Estado sería parte activa en la fiscalización y diseño ordenado de recorridos y tarifas, siendo el otro actor relevante los empresarios del sistema de transporte; estos empresarios serían ahora determinados mediante el otorgamiento de concesiones, una por cada una de las 10 zonas de operación establecidas para el funcionamiento del sistema, además de la concesión de operación para una serie de servicios troncales.

La puesta en marcha de este sistema no estuvo exenta de una acalorada polémica debido a una serie de errores de diseño y de implementación, que significarían un detrimento dramático en la calidad del servicio percibido por la gran mayoría de la población de Santiago. Profundizar sin embargo en los detalles de la historia de Transantiago no es un objetivo específico de este trabajo. Una apreciación documentada y estructurada de esto puede encontrarse en Muñoz y Gschwender (2008). No obstante, es relevante para establecer los antecedentes del presente trabajo que dos de los cambios más profundos acarreados por la implementación de Transantiago fueron la integración tarifaria, por una parte, y la penetración de las tarjetas inteligentes para llevar a cabo las transacciones (indispensable para implementar la susodicha integración). Esta última inclusión significó un salto tecnológico en términos de la operación del sistema y su interacción con los usuarios; reemplazando a los antiguos cobradores automáticos (prácticamente no operacionales), a los cobradores humanos, y más importante, a los conductores cobradores,



se brindó al servicio una mayor eficiencia y mayor seguridad. Por otra parte, estos avances tecnológicos pueden ser mirados como necesarios desde otra perspectiva: al tomar el Estado un rol activo en la planificación (y especialmente fiscalización) del sistema, se hizo necesario contar con tales sistemas tecnológicos como apoyo al rol fiscalizador. Los impactos de esta inyección tecnológica al sistema para los usuarios fueron también considerables: se vieron obligados, a partir de la puesta en marcha de Transantiago, a utilizar la tarjeta Bip! para abordar cualquier recorrido de bus y para poder hacer efectiva la integración tarifaria entre servicios de buses y metro.

Es precisamente este hecho particular el que otorga a las bases de datos de Transantiago una característica adicional, que hace que su análisis sea aún más interesante: la penetración de las tarjetas inteligentes<sup>3</sup> en el sistema es del orden de un 97% de los usuarios, con lo cual es posible afirmar que cualquier conclusión obtenida en base a la susodicha base de datos será representativa para virtualmente la totalidad de la población que utilice el sistema. Esta característica es prácticamente única al caso de Santiago; estudios de conducta del pasajero de transporte público en otras ciudades del mundo se ven obligados a obtener conclusiones sobre el comportamiento del usuario en base a datos con muy poco nivel de significancia, debido a la escasa penetración de las tarjetas inteligentes como forma de pago, viéndose forzados a corregir algebraicamente y a construir factores de expansión para obtener resultados y conclusiones representativas de toda la población. Ejemplo de esto lo constituye el trabajo de Hofman y O'Mahony (2005) que estima viajes y trasbordos en transporte público en base a un sistema con una tasa de penetración de tarjetas inteligentes de un 12,5%. Similar caso se observa en Lianfu et al. (2007), estudio que presenta un sencillo modelo para estimar patrones de viaje y matrices Origen/Destino en la ciudad de Chagchun, cuyo sistema de transporte público posee una tasa de penetración de tarjetas de un 25%. No obstante lo anterior, un serio problema acarreado tras la implementación de Transantiago corresponde al alza sustancial en las tasas de evasión tarifaria, en parte debidas a la facilidad que un sistema de tarjetas inteligentes sin torniquete brinda para ejecutar este tipo de malas prácticas, sobre todo para una población de usuarios disconforme con la deficiente calidad inicial del sistema, un estigma que hasta el día de hoy persiste no tanto en el ámbito operacional de Transantiago como sí en la opinión pública.

---

<sup>3</sup>La penetración de tarjetas inteligentes corresponde al porcentaje de transacciones realizadas a través de estos dispositivos con respecto al total de transacciones en el sistema de transporte público, considerando los modos bus y metro.

### **3.1.2 Descripción de las características generales de la base de datos de Transantiago e información contenida**

El sistema de registro de datos de transacciones opera almacenando en forma automática la totalidad de las transacciones registradas en un determinado período de tiempo en cualquier validador ubicado ya sea a bordo de un bus, en una estación de metro o una zona paga. La base de datos que se utiliza en el presente inciso del capítulo para describir el comportamiento y las preferencias de los usuarios corresponde al registro de todas las transacciones Bip! realizadas entre los días 20 y 26 de Marzo del año 2009; este registro consta de un total de 35 millones de transacciones. Para cada una de estas transacciones, se almacena una gran cantidad de datos y variables; se presenta a continuación una breve enumeración y descripción de las principales variables, algunas de las cuales son obtenidas directamente de los registros proporcionados por Transantiago y otras han sido calculadas en base a post-procesamiento de las variables originales, a través de la metodología descrita en Munizaga y Palma (2012).

*ID:* corresponde al número de identificación asociado a la tarjeta Bip! que realiza la transacción. Es único para cada tarjeta.

*Hora transacción:* corresponde a la hora en que se produce la validación de la tarjeta.

*Pago:* es el importe pagado por el usuario en la validación en cuestión.

*Sitio:* lugar donde se realizó la transacción; si esta se efectúa a bordo de un bus corresponde a la patente del vehículo, si se realiza en el metro corresponde al nombre de la estación, y si la transacción se efectúa en una zona paga es el código del validador.

*Latitud y longitud:* ubicación de la transacción. Este dato se obtiene cruzando la información de transacciones con información de los buses, metro y paraderos.

*Coordenadas X,Y subida:* corresponde a las coordenadas UTM de la transacción, calculada en base a la latitud y longitud. Todas se encuentran en la zona 19H, WGS84, correspondiente a Santiago.

*Servicio*: servicio del bus donde se realizó la transacción, obtenido en base a la patente del vehículo de las tablas de asignación de servicio proporcionadas por Transantiago.

*Sentido (ida o retorno)*: sentido del bus donde se realizó la transacción, obtenido en base a la patente del vehículo de las tablas de asignación de servicio proporcionadas por Transantiago.

*Período*: período en el cual se realizó la transacción, según definición de períodos de Transantiago.

*Tipo transporte*: tipo de transporte abordado en la transacción; puede ser bus o metro. Si la transacción se realiza en zona paga, también es registrado.

*Tipo de operador*: define si las transacciones en bus son en servicios locales o troncales.

*Escolar*: variable booleana que define si la transacción es escolar o no.

*Tiempo transacción anterior*: corresponde a la hora de la transacción anterior del mismo ID.

*Tiempo transacción posterior*: corresponde a la hora de la transacción posterior del mismo ID.

*Distancia euclid transacción anterior*: corresponde a la distancia euclidiana hasta las coordenadas de la transacción anterior del mismo ID.

*Distancia euclid transacción posterior*: corresponde a la distancia euclidiana hasta las coordenadas de la siguiente transacción del mismo ID.

*Coordenadas X,Y bajada*: coordenadas UTM estimadas para la bajada de la etapa a que da lugar la transacción en cuestión. El procedimiento general de cálculo corresponde al desarrollado originalmente por Trépanier et al. (2007), elaborado y aplicado para el caso de Santiago por Munizaga et al. (2010) para la misma base de datos. En caso de que la transacción sea efectuada en una zona paga, se ejecuta una subrutina basada en líneas comunes para identificar el bus abordado por el usuario; en caso de que la transacción sea en metro, se hace una estimación en base a velocidades comerciales del metro y tiempos de caminata y transbordos al interior de las estaciones.

*Tiempo bajada:* tiempo estimado de bajada, para la etapa que da a lugar la transacción en cuestión. El procedimiento de cálculo corresponde a la misma metodología que el de las coordenadas de bajada.

*Paradero subida:* código del paradero de subida, obtenido de cruzar información de las coordenadas de subida con información de distribución de paraderos y recorridos proporcionada por Transantiago.

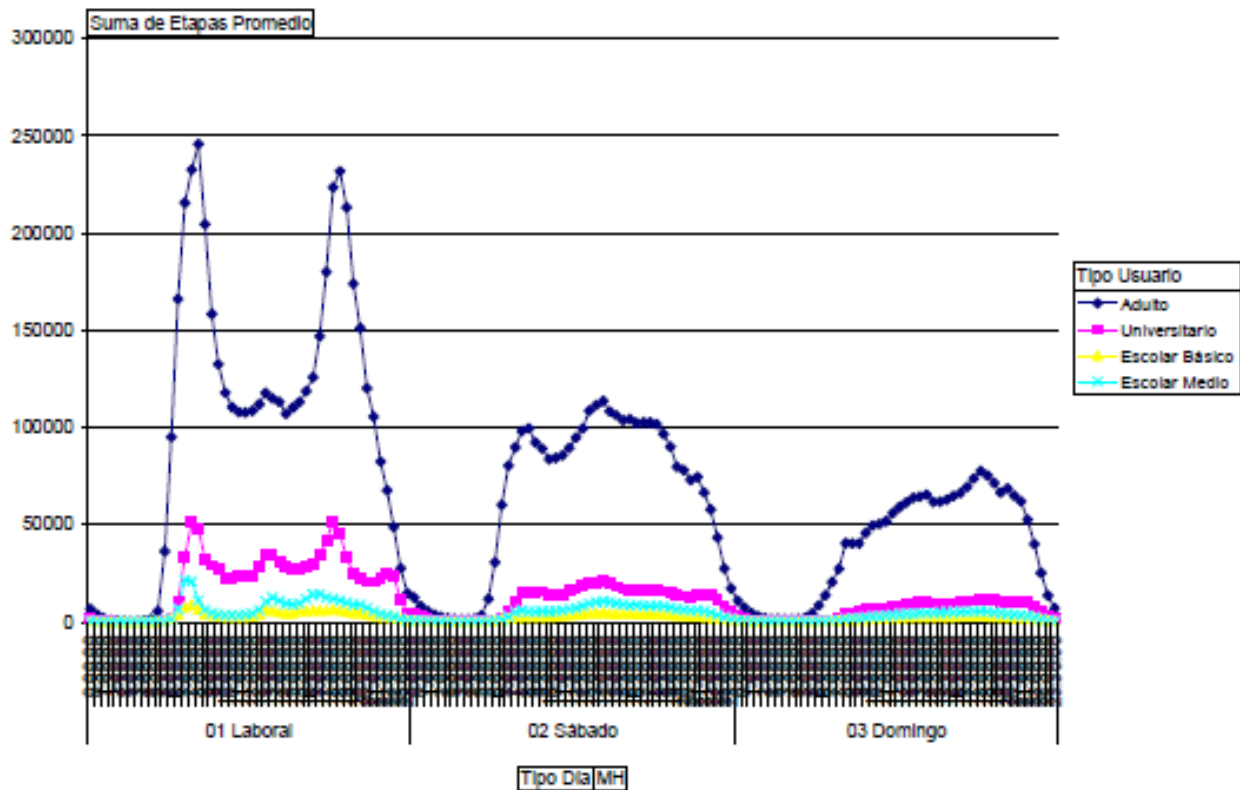
*Paradero bajada:* código del paradero de bajada, obtenido de cruzar información de las coordenadas de bajada con información de distribución de paraderos y recorridos proporcionada por Transantiago.

*Primera del día:* variable de carácter booleano que identifica aquellas transacciones que corresponden a la primera del día para el ID que la efectúa.

*Última del día:* variable de carácter booleano que identifica aquellas transacciones que corresponden a la última del día para el ID que la efectúa.

### **3.1.3 Cuantificación de usuarios y transacciones en la dimensión temporal**

Un primer factor interesante a analizar corresponde a la cantidad de transacciones registradas por cada tipo de usuario en cada tipo de día. Se observan en la Figura 3.1 las distribuciones diarias de transacciones para las tarjetas tipo adulto, universitario, escolar básico y escolar medio. Las puntas mañana y tarde son claramente distinguibles en los días laborales entre las 7:00 y las 9:00, y entre las 17:30 y las 19:30, respectivamente, tanto en los adultos como en los estudiantes universitarios. Los días sábado en cambio se registran menos transacciones, y se observa una distribución más bien uniforme a lo largo del día para todos los tipos de tarjetas. Los días domingo también se observa una distribución más bien uniforme, aunque ligeramente más concentrada en las horas de la tarde (en torno a las 18:00-19:00).



*Figura 3.1: Distribución de transacciones según tipo de día y usuario*  
*Fuente: Transantiago, 2011*

Un resumen de las transacciones por cada día se puede ver en la Tabla 3.1. A nivel de día de la semana, es observable un nivel más o menos constante de transacciones en los 5 días laborales, con una cantidad ligeramente mayor de transacciones los días lunes y viernes; una menor cantidad se registra el día sábado, y una menor cantidad aún el día domingo. Similares patrones se observan para la utilización de boletos de metro correspondientes a viajes efectuados en metro y cancelados a través de este medio de pago (y por tanto sin posibilidad de integración tarifaria inter-modal). Esta disminución dice relación con la menor cantidad de actividades laborales y de estudio externas al hogar que se desarrollan los fines de semana, y en particular los días domingo en los cuales un alto porcentaje de la población permanece en su hogar. No obstante lo anterior, una mayor cantidad de actividades con otros propósitos se desarrollan los fines de semana (recreativas, compras, culto religioso, etc.), sin embargo el aumento en este tipo de actividades es menor en comparación con la disminución de actividades laborales y de estudio (Sectra, 2001)

Fecha	Tipo Día	Total Etapas Bips	Total Etapas Boletos	Total Etapas
31-05-2010	Laboral	6.484.378	185.676	6.670.054
01-06-2010	Laboral	6.303.518	177.973	6.481.491
02-06-2010	Laboral	6.225.939	169.709	6.395.648
03-06-2010	Laboral	6.259.172	175.297	6.434.469
04-06-2010	Laboral	6.478.077	192.656	6.670.733
05-06-2010	Sábado	3.990.844	127.126	4.117.970
06-06-2010	Domingo	2.377.022	73.255	2.450.277
<b>Promedio Laboral</b>		<b>6.350.217</b>	<b>180.262</b>	<b>6.530.479</b>
<b>Total Semanal</b>		<b>38.118.950</b>	<b>1.101.692</b>	<b>39.220.642</b>

*Tabla 3.1: Número de transacciones por día de la semana*

*Fuente: Transantiago, 2011*

La figura 3.2 muestra la partición modal desagregada por pago para los 7 días de la semana. Tres modos son reportados: bus, metro y zona paga; no obstante las transacciones efectuadas en zonas pagas corresponden también al modo bus, son tratadas separadamente debido a que procesamientos adicionales son requeridas en estos casos para llevar a cabo la metodología de estimación de bajadas. Se observan particiones muy similares para los cinco días de la semana, y también estructuras de pago semejantes; alrededor de un 60% de las transacciones realizadas de lunes a viernes se efectúan en validadores a bordo de un bus, un 34% se realizan en estaciones de Metro, y un 6% en zonas pagas. Los fines de semana, el modo bus toma más partición de mercado: los sábados, un 69% de las transacciones se registran en buses, un 30% en estaciones de Metro, y sólo un 1% en zonas pagas. Los domingos, la diferencia es mayor aún; un 73% de las transacciones se registran en buses, un 26% en Metro, y un 1% en zonas pagas. Algunos de los fenómenos que pueden explicar esta diferencia entre días de la semana y fines de semana son el hecho de que los horarios de funcionamiento de los buses son más flexibles, en particular en horas de la noche los fines de semana, que la red de metro cubre principalmente viajes hacia los centros laborales (no tan relevantes los fines de semana) y también a que la gran mayoría de las zonas pagas no se encuentran operativas los días sábado y domingo.

Tipo día	Fecha	BUS	METRO	ZONA PAGA	Total general
Laboral	31-05-2010	3.891.423 <b>60,0%</b>	2.175.577 <b>33,6%</b>	417.378 <b>6,4%</b>	6.484.378
	01-06-2010	3.799.099 <b>60,3%</b>	2.103.506 <b>33,4%</b>	400.913 <b>6,4%</b>	6.303.518
	02-06-2010	3.748.590 <b>60,2%</b>	2.080.924 <b>33,4%</b>	396.424 <b>6,4%</b>	6.225.938
	03-06-2010	3.777.219 <b>60,3%</b>	2.087.578 <b>33,4%</b>	394.375 <b>6,3%</b>	6.259.172
	04-06-2010	3.892.937 <b>60,1%</b>	2.194.687 <b>33,9%</b>	390.453 <b>6,0%</b>	6.478.077
Sábado	05-06-2010	2.761.187 <b>69,2%</b>	1.206.334 <b>30,2%</b>	23.323 <b>0,6%</b>	3.990.844
Domingo	06-06-2010	1.743.557 <b>73,4%</b>	621.414 <b>26,1%</b>	12.051 <b>0,5%</b>	2.377.022
<i>Total general</i>		23.614.012 <b>61,9%</b>	12.470.020 <b>32,7%</b>	2.034.917 <b>5,3%</b>	38.118.949

*Tabla 3.2: Partición modal de transacciones según día de la semana y pago*

*Fuente: Transantiago, 2011*

### 3.1.4 Cuantificación de usuarios y transacciones en la dimensión espacial

La dimensión espacial es un aspecto importante a analizar de cara a estimar actividades. Conocida la localización de las posibles actividades/trasbordos, y reconociendo que los puntos de atracción de actividades, según cuales sean éstas, se encuentran más bien sectorizados, es posible inferir (aun cuando no se conozca de forma directa el propósito del viaje) cuándo un cierto patrón de conducta por parte de un usuario está enmascarando el desarrollo de una actividad.

La información de la que se dispone en la base de datos en términos de localización de transacciones es bastante específica, dado que se dispone de las coordenadas de origen de la transacción. Para el caso de las transacciones llevadas a cabo en una estación de metro o zona paga, la obtención de estas coordenadas es directa cruzando el código de la estación de metro o paradero con sus coordenadas geográficas<sup>4</sup>; para el caso de transacciones realizadas en validadores a bordo de un bus, las coordenadas de subida se obtienen mediante el post-proceso explicado en el capítulo 2, a través de la metodología de

<sup>4</sup> Transantiago dispone de una base de datos con las coordenadas geográficas de todas las estaciones de metro y paraderos de bus, la cual fue proporcionada para la realización de este estudio.







La Figura 3.3 resume un análisis exploratorio, realizado con una sub-muestra de 50.000 transacciones (aproximadamente un 0,2% de la muestra total), describiendo la cantidad de validaciones totales por comuna según el tipo de día<sup>5</sup>; se observa en general que la mayor cantidad de transacciones se encuentra en la comuna de Santiago y en algunas comunas de la zona oriente de la ciudad, como Las Condes, Providencia y La Reina, además de algunas comunas de otros sectores con alta población, como Maipú y Lo Prado. Estos resultados se condicen con el hecho de que las comunas que presentan la mayor cantidad de transacciones poseen al menos uno de los siguientes dos factores determinantes:

- Comunas densamente pobladas, y por consiguiente en las cuales es natural esperar una mayor cantidad de transacciones
- Comunas con una alta actividad comercial y laboral.

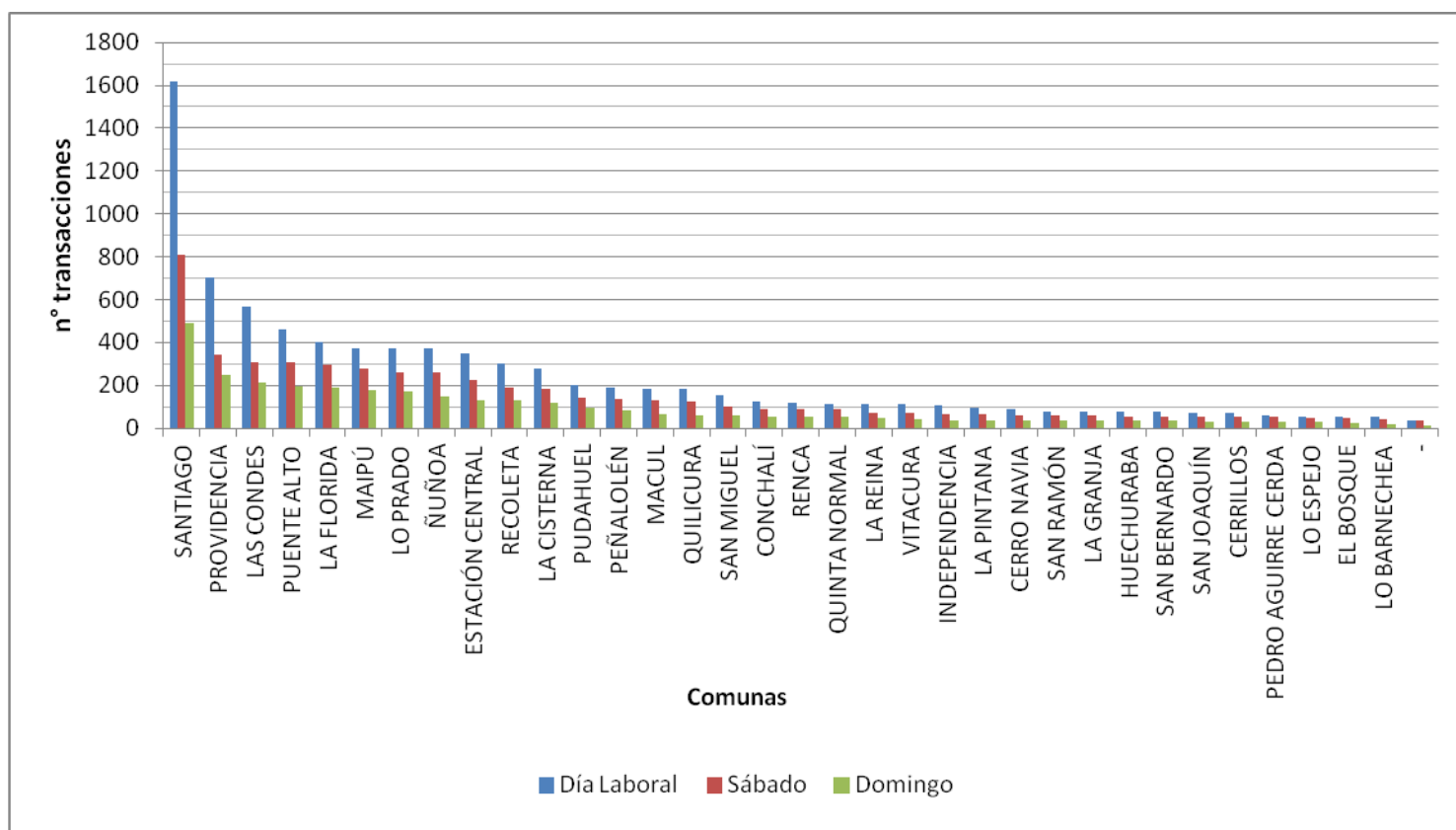


Figura 3.3: Transacciones generadas por comuna según tipo de día

Fuente: elaboración propia

<sup>5</sup> Día Laboral se encuentra normalizado, es decir corresponde a la suma de todos los días desde lunes a viernes dividida por 5; esto es realizado con el propósito de comparar este tipo de día con las transacciones en días sábado y domingo.

Un punto importante a tener en consideración es que este un análisis a nivel de transacciones, y no viajes, dado que en el presente capítulo se estudia sólo información directamente obtenible a partir de la base de datos. Por lo tanto, no se debe confundir con un análisis enfocado en base a origen y destino de viajes; este último se lleva a cabo en el capítulo 5. Similares análisis al realizado, en términos de información obtenible en forma directa desde la base de datos de transacciones y de GPS, puede ser visto en Beltrán et al. (2011).

## **3.2 Base de datos de Gatineau, Canadá**

### **3.2.1 Breve reseña demográfica de la ciudad de Gatineau y sobre el funcionamiento de su sistema de transporte público**

La ciudad de Gatineau se ubica en la provincia de Quebec, Canadá; está conformada por los distritos municipales de Gatineau, Hull, Aylmer, Chelsey, Cantley, Masson-Angers y Buckingham. Su población en el año 2010 se estima en 300.000 habitantes, en base al censo nacional efectuado por el gobierno canadiense el año 2006. En la actualidad, se encuentra conurbada con la ciudad de Ottawa, capital de Canadá, de la cual se encuentra separada por el río Ottawa; no obstante, todos los servicios públicos y divisiones administrativas son independientes para ambas ciudades. En términos económicos, la ciudad de Gatineau es sede de una gran cantidad de agencias e instituciones gubernamentales y federales canadienses, debido a su proximidad con la capital del país; la prestación de servicios es también un aspecto relevante de la actividad económica de la ciudad.

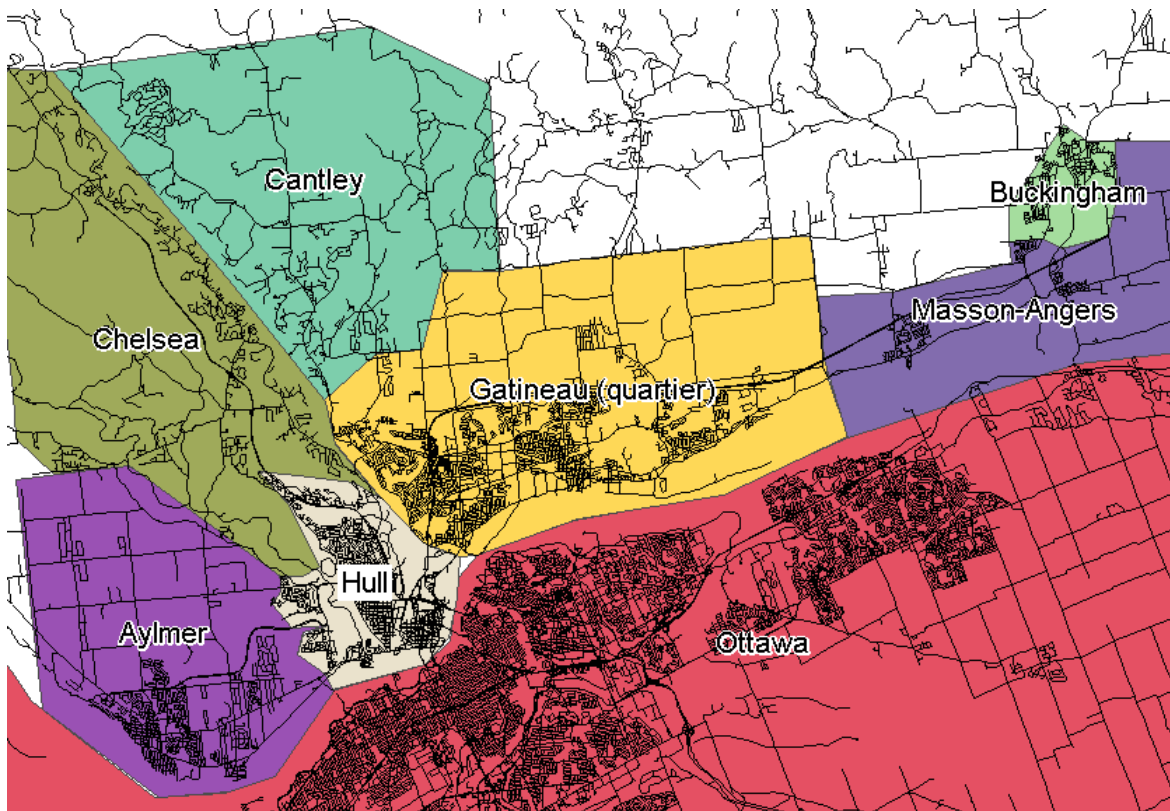


Figura 3.4: Distribución geopolítica de la ciudad de Gatineau

Fuente: Piriou (2008)

El sistema de transporte público de Gatineau se compone en su totalidad de buses, y es operado por dos entidades privadas independientes; por una parte *OC Transpo* opera aquellos recorridos que tienen origen en la ciudad de Ottawa y destino en la ciudad de Gatineau, o vice-versa; por otro lado la *Société de transport de l'Outaouais (STO)* opera aquellos recorridos que sólo circulan en la ciudad de Gatineau. Los tickets individuales de pasaje no son intercambiables, sin embargo todas las formas de prepago por una determinada cantidad de tiempo, asociables a las tarjetas magnéticas de una u otra institución, son utilizables en cualquier recorrido de uno u otro operador, indistintamente.

El sistema de pago a través de tarjetas inteligentes fue instaurado por la STO en el año 2001, pudiendo los usuarios que utilizan esta forma de pago acceder a distintos tipos de tarifas: cargos unitarios por pasaje, abonos semanales y abonos mensuales, estas últimas dos modalidades correspondientes al pago de una cantidad fija por el derecho a pasajes ilimitados en un período de una semana y un mes, respectivamente. Existen además diferentes tipos de tarjetas según sean los atributos y condiciones del usuario; las tarjetas tipo adulto-normal pueden ser adquiridas en múltiples puntos de la ciudad, y los estudiantes, menores de 12 años y adultos mayores pueden optar tarjetas que les permiten

acceder a tarifas rebajadas; además existen diferentes tipos de tarjetas tipo adulto que permiten acceder con tarifas rebajadas a ciertos recorridos (expresos e inter-zonales). Cada bus posee a bordo uno o más validadores, un cobrador automático que posibilita el pago del pasaje mediante efectivo, un sistema GPS y una unidad almacenadora de datos, que guarda en forma interna todos los datos asociados a las transacciones efectuadas en el vehículo. Una vez que termina el día de servicio de la máquina, el último chofer en operarla es responsable de conectar esta unidad y descargar su contenido a una base de datos central presente en el terminal, la cual almacena la totalidad de las transacciones. El marcaje de la hora y posición al inicio de su recorrido de servicio se registran en forma infrarroja tras la salida del vehículo a la expedición, parámetros que son tomados como referencia en todos los registros ulteriores.

Dentro del sistema de transporte público de la ciudad de Gatineau, la penetración de tarjetas validadoras como forma de pago del pasaje alcanza aproximadamente un 80%; no obstante, según señala Bagchi y White (2005), existe un sesgo asociado al hecho observable de que son aquellos usuarios más recurrentes del sistema de transporte público quienes adscriben a este sistema de pago, mientras que los usuarios ocasionales utilizan mayoritariamente el pago en efectivo a bordo de los vehículos o los tickets unitarios adquiridos en forma previa; esto es explicable debido a que para un viajero recurrente (2 o más viajes en transporte público al día) es más económicamente conveniente adquirir abonos.

### **3.2.2 Descripción de la base de datos de Gatineau**

La base de datos de transporte público de Gatineau con la cual se trabaja en el presente estudio corresponde a los datos recabados por la STO en todos sus recorridos en el período comprendido entre los años 2002 y 2009, en forma temporalmente continuada<sup>6</sup>; en su totalidad la base de datos contiene 45.087.964 registros, a la vez que la red de transporte público de Gatineau por parte de la STO está compuesta por aproximadamente 1500 paradas y 107 recorridos. Por cada transacción se almacenan una serie de variables, que se listan a continuación.

- ID de la tarjeta
- Tipo de tarjeta (adulto, menor, estudiante)
- Fecha de validación
- Hora de validación

---

<sup>6</sup> Quedan excluidas de la base de datos por tanto todas las transacciones efectuadas en recorridos operados por *OC Transpo*.

- Vehículo en que se efectúa la validación y chofer que lo opera
- Parada, recorrido, y sentido en que se efectúa la validación (Obtenidos en base a la variable anterior e información de recorridos de la STO)
- Hora en que el vehículo inicia el recorrido durante el cual se efectúa la transacción

Tres son los atributos más valorables de la base de datos de Gatineau. Por un lado al tratarse de una ciudad de tamaño más bien reducido, el sistema es en general manejable desde el punto de vista estadístico y de manipulación de las bases de datos. Adicionalmente, la base de datos canadiense describe un espacio temporal continuo y muy extenso, cubriendo un total de nueve años. Por último, la base de datos de Gatineau es de antigua data, puesto que recaba datos desde el año 2001; esto permite que potenciales cambios en las costumbres, usos o gustos en los usuarios de transporte público (normalmente observables en espacios de tiempo más bien extensos, excepto en casos de cambios abruptos en las condiciones de servicio) sean más fácilmente observables en la base de datos de Gatineau.

Todos estos atributos en la base de datos de Gatineau han permitido a diversos grupos de investigación desarrollar en base a ella muy interesantes aplicaciones y estudios académicos, como también la elaboración de herramientas de planificación para la STO y las autoridades reguladoras del transporte público. Trépanier et al. (2004) introduce la programación orientada a objetos para abordar el problema de procesar la información contenida en la base de datos y elaborar criterios para el diseño de recorridos de la STO; Chu & Chapleau (2008) y Chu & Chapleau (2010) utilizan esta información para reconstruir el itinerario de los usuarios de transporte público y para “realizar un retrato espacio-temporal de la red de transporte público”; Piriou (2008) caracteriza los lugares de ocurrencia de actividades en Gatineau empleando técnicas de minería de datos sobre la información contenida en la base de datos de transporte público; Morenc et al. (2007)<sup>7</sup> aborda por primera vez la estimación espacio-temporal de bajadas, introduciendo la noción de etapa de viaje y posibilitando el estudio de los destinos de las transacciones. Finalmente, Trépanier y Morency (2010) utilizan los datos contenidos en la base de datos para abordar la problemática de la fidelidad del usuario de transporte público.

---

<sup>7</sup> Esta metodología es explicada con detalle en el capítulo 2.

# Capítulo 4

## Metodología de estimación de viajes y actividades

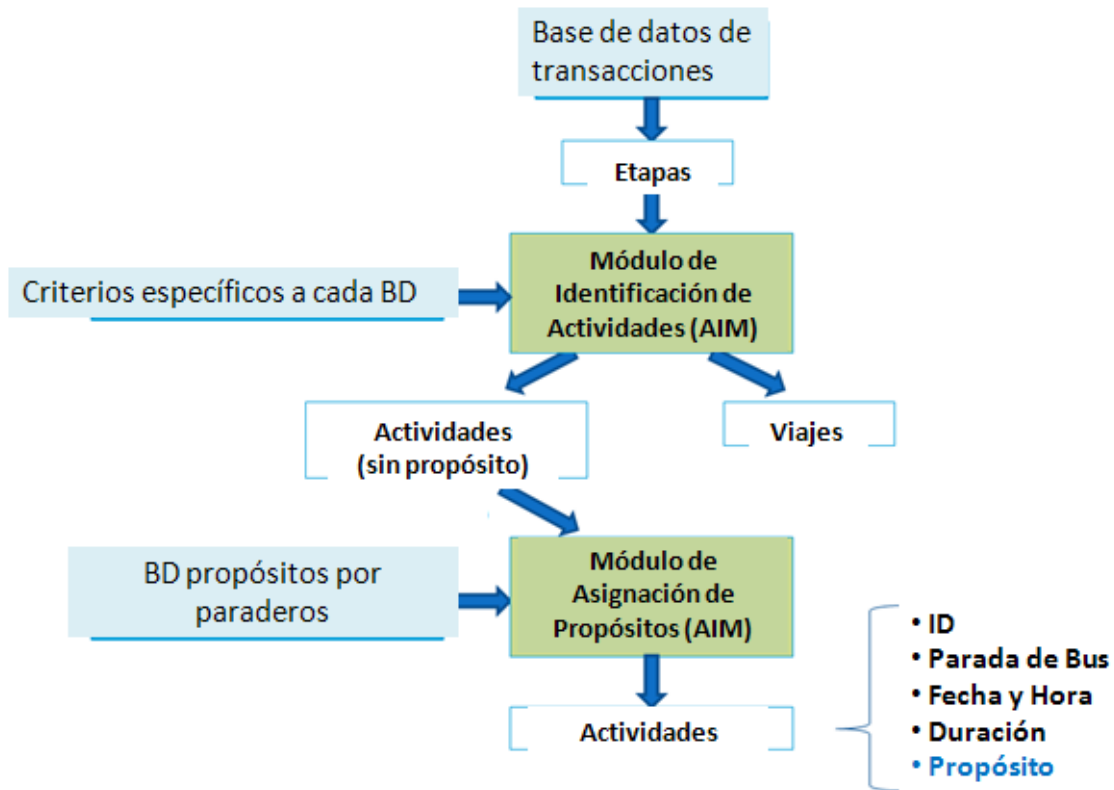
El presente capítulo tiene como objetivo la descripción de la metodología desarrollada para estimar viajes y actividades en las dos bases de datos disponibles. Con tal fin, se procederá en primera instancia a explicar en forma general la metodología, y posteriormente se presentarán con mayor detalle cada uno de los pasos y módulos de que esta consta.

### 4.1 Descripción general de la metodología

La metodología del presente trabajo toma como punto de partida los trabajos de Munizaga y Palma (2012) y Trépanier et al. (2007), utilizando en particular sus algoritmos de estimación de paradero de bajada por transacción para las bases de datos de transporte público de Santiago y Gatineau, respectivamente. En forma posterior, se construye en base a cada transacción lo que se denomina una *etapa* de viaje, esto es, una cierta franja espacio-temporal de viaje comprendida entre el origen, un trasbordo, o el destino. Hecho esto, es necesario identificar en base a estas etapas los viajes propiamente tales: puesto que un viaje, definido entre un origen y un destino, puede contener una o más etapas (definiéndose éstas por los sucesivos trasbordos), se debe estimar en definitiva la conformación de estos viajes, y en forma análoga y complementaria, cuáles ventanas temporales entre bajadas y abordajes son trasbordos, y cuáles son actividades. El problema de estimar viajes y actividades puede, por lo tanto, ser visto como las dos caras de una misma moneda: al estimar que una ventana temporal es una actividad, esta inmediatamente definirá el fin de un viaje y el comienzo de otro. Esta problemática es abordada en la metodología a través del denominado *Módulo de Identificación de Actividades (AIM, Activity Identification Module)*, que toma como entrada las etapas de viaje, y otros criterios específicos para cada base de datos, y retorna como salida las actividades (de las cuales aun no se conoce el propósito) y los viajes. El siguiente paso constituye en la estimación de un atributo fundamental para describir actividades: el propósito. A través del *Módulo de Asignación de*

*Propósito (PAM)*, se estimará para cada una de las actividades previamente detectadas un cierto propósito, quedando por tanto éstas totalmente descritas.

La metodología general se resume en la figura 4.1.



*Figura 4.1: Diagrama de la metodología general*  
*Fuente: Devillaine et al. (2012a)*

## 4.2 Módulo de Identificación de Actividades y Viajes (AIM)

La estimación de bajadas a través de las metodologías Munizaga y Palma (2012) y Trépanier et al. (2007) han sido explicadas con detalle en el Capítulo 2 y por tanto serán omitidas en el presente acápite.

La funcionalidad del Módulo de Estimación de Actividades está definida por el establecimiento de patrones de conducta a través de los cuales sea posible identificar cuándo una determinada ventana temporal entre transacciones constituye un trasbordo y cuándo corresponde a la ejecución de una actividad en destino, definiendo por tanto el fin

de un viaje para el usuario de transporte público involucrado. Tales criterios deben ser identificados tomando en consideración características asociadas al comportamiento de cada individuo, características intrínsecas del transporte público y una combinación entre ambos aspectos. De esta forma es posible identificar, bajo el supuesto de un comportamiento racional por parte de cada usuario, cánones conductuales bajo los cuales es inferible si el individuo en cuestión realizó un trasbordo o una actividad.

En base a las variables presentadas en el Capítulo 3, se procederá a continuación a enumerar y describir los distintos criterios a través de los cuales se llevará a cabo la identificación de cada una de las ventanas temporales entre transacciones como una actividad o un trasbordo.

#### **4.2.1 Patrón de los 30 minutos**

La primera regla o patrón es la utilizada en el trabajo de Munizaga y Palma (2012) y corresponde a la categorización de una ventana temporal entre una bajada y la siguiente transacción, cuando el tiempo estimado de espera (descontando el tiempo de acceso) es mayor a 30 minutos. Aunque este número puede parecer en un principio arbitrario, ha sido utilizado en el mencionado trabajo, además de otros estudios relacionados al tema (Hoffman y O'Mahony 2005, Park y Kim 2008). Las justificaciones para este número en cada trabajo, cuando se mencionan, apuntan a razones similares: principalmente se basan en la experiencia técnica y social respecto de los usuarios de los modeladores, y en estudios focalizados al transporte público, según los cuales un bajo porcentaje de la población realizaría trasbordos de más de 30 minutos.

#### **4.2.2 Patrón dos transacciones consecutivas en metro**

El segundo patrón de identificación tiene relación con el registro de dos transacciones consecutivas, para un mismo ID y en el mismo día, en estaciones de metro; si tal circunstancia llegara a producirse, inmediatamente la ventana registrada entre esas dos transacciones será catalogada como una actividad.

La razón que fundamenta este criterio es la forma de operación del sistema de metro y su metodología de cobro. Considerando que la circulación en la red de metro de Santiago requiere por parte del usuario una única validación en la estación de abordaje, siendo innecesario revalidar para efectuar trasbordos, la única razón por la que un individuo se vería en la necesidad de validar nuevamente en una estación (pagando por tanto una nueva tarifa) es para volver a entrar al sistema en forma posterior a una salida. Bajo el supuesto de comportamiento racional y maximización de utilidades y beneficios, una salida de la red de



metro implica por tanto al usuario un cobro adicional, y es por tanto directo deducir que el individuo efectuó en forma previa a su segunda validación en el metro una actividad que le reportó una cierta utilidad, asumiendo en forma posterior el costo de volver a entrar en el sistema. Un ejemplo de lo anteriormente descrito sería para un usuario de metro en camino a su trabajo, bajarse y salir de la estación de Metro en un sector con amplia cobertura de servicios como por ejemplo bancos con la finalidad de realizar un trámite; una vez hecho esto, volver a ingresar al metro en la misma estación (o alguna adyacente) para continuar su viaje hacia el trabajo.

#### **4.2.3 Patrón dos transacciones consecutivas en el mismo recorrido de bus**

El tercer patrón de identificación de actividades corresponde al caso en que se registran dos transacciones seguidas, para igual ID y en el mismo día, en dos buses que sirvan el mismo recorrido, incluyendo tanto el caso en que ambos buses sirvan el mismo sentido del recorrido, como también el caso en que los buses sirvan el mismo recorrido en sentidos opuestos.

El fundamento de este criterio es muy similar al del patrón anterior. Dada la estructura operativa y tarifaria de Transantiago al momento de la realización del presente trabajo, un usuario puede realizar hasta tres trasbordos en el modo bus sin incurrir en cargos adicionales; no obstante, si se observa que dos transacciones consecutivas se registran en buses que sirven el mismo recorrido y en el mismo sentido, es razonable suponer que el individuo realizó un encadenamiento de viajes, y en algún punto cercano a la segunda transacción bajó del primer vehículo para llevar a cabo una actividad, tras la cual abordó nuevamente un bus del mismo recorrido para iniciar un viaje hacia su siguiente destino. Por otro lado, si se observa que las transacciones consecutivas se registran en buses que sirven el mismo recorrido pero en sentidos opuestos, un trasbordo en estas circunstancias no tiene sentido pues implicaría una ineficiencia fehaciente en la ruta seguida por el usuario, el cual recorrería en el mismo viaje un tramo de ruta dos veces. En virtud de lo anterior, la situación puede interpretarse como un caso en el cual el usuario llega a su punto de destino, desarrolla su actividad, y una vez concluida ésta, utiliza el mismo recorrido para cubrir la primera etapa de su siguiente viaje.

#### **4.2.4 Patrón comparación entre distancia euclidiana y distancia en ruta de un viaje**

Dado que se posee información sobre el paradero de subida, y estimación sobre el paradero de bajada asociadas a una transacción (ambas obtenidas a través de post-procesamiento de la información en la base de datos de transacciones y emisiones GPS de buses), en

particular es interesante analizar la distancia euclidiana ( $d_{euclidiana}$ ) entre paradero de subida y bajada, y la distancia recorrida ( $d_{ruta}$ ) a lo largo del viaje. La distancia euclidiana de un viaje corresponde a la distancia entre el paradero de subida de la primera etapa y el paradero de bajada de la última etapa; la distancia recorrida, o distancia en ruta de un viaje, corresponde a la distancia total recorrida por el usuario en su viaje, obtenida sumando la distancia recorrida en cada modo utilizado en cada etapa. La figura 4.2 muestra una representación de ambas distancias para un viaje genérico de tres etapas, en donde los paraderos de subidas están representados por cruces y los paraderos de bajada están representados por círculos. El espacio entre un punto de bajada y el siguiente punto de subida representa las distancias de acceso o caminata entre etapas<sup>8</sup>, y no se incluyen en la distancia total en ruta por considerarse despreciables respecto a los tramos recorridos en cada etapa

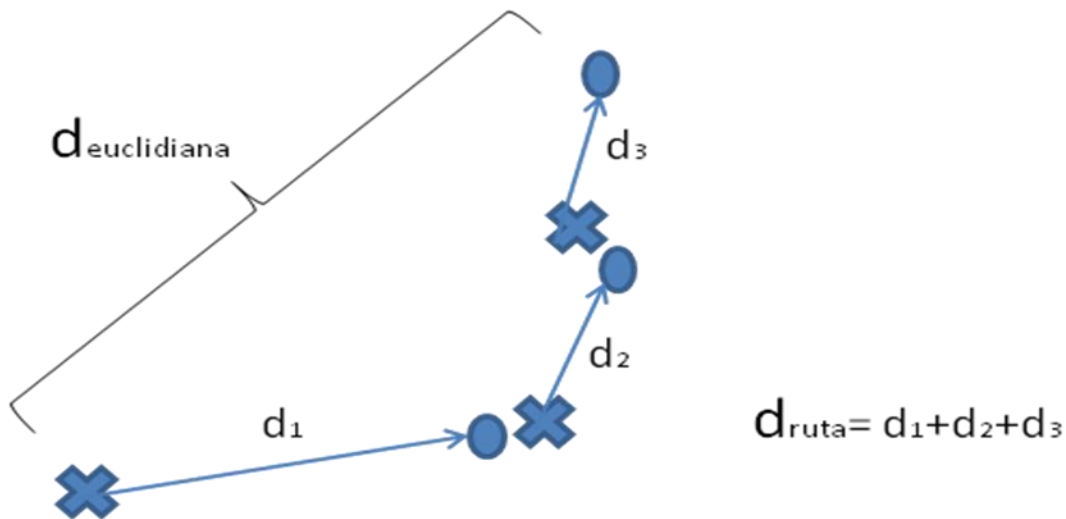


Figura 4.2: Distancia euclidiana y distancia en ruta de un viaje de tres etapas  
Fuente: elaboración propia

Para el análisis de ambas distancias, un indicador adecuado y de relativamente sencilla interpretación es el cociente entre ambas, denominado  $f_d$  :

$$f_d = \frac{d_{ruta}}{d_{euclidiana}} \quad (4.1)$$

Un viaje será factible en la medida que el cociente  $f_d$  sea mayor o igual a 1. Valores menores a 1 implican una distancia euclidiana mayor que la distancia en ruta, lo cual es una

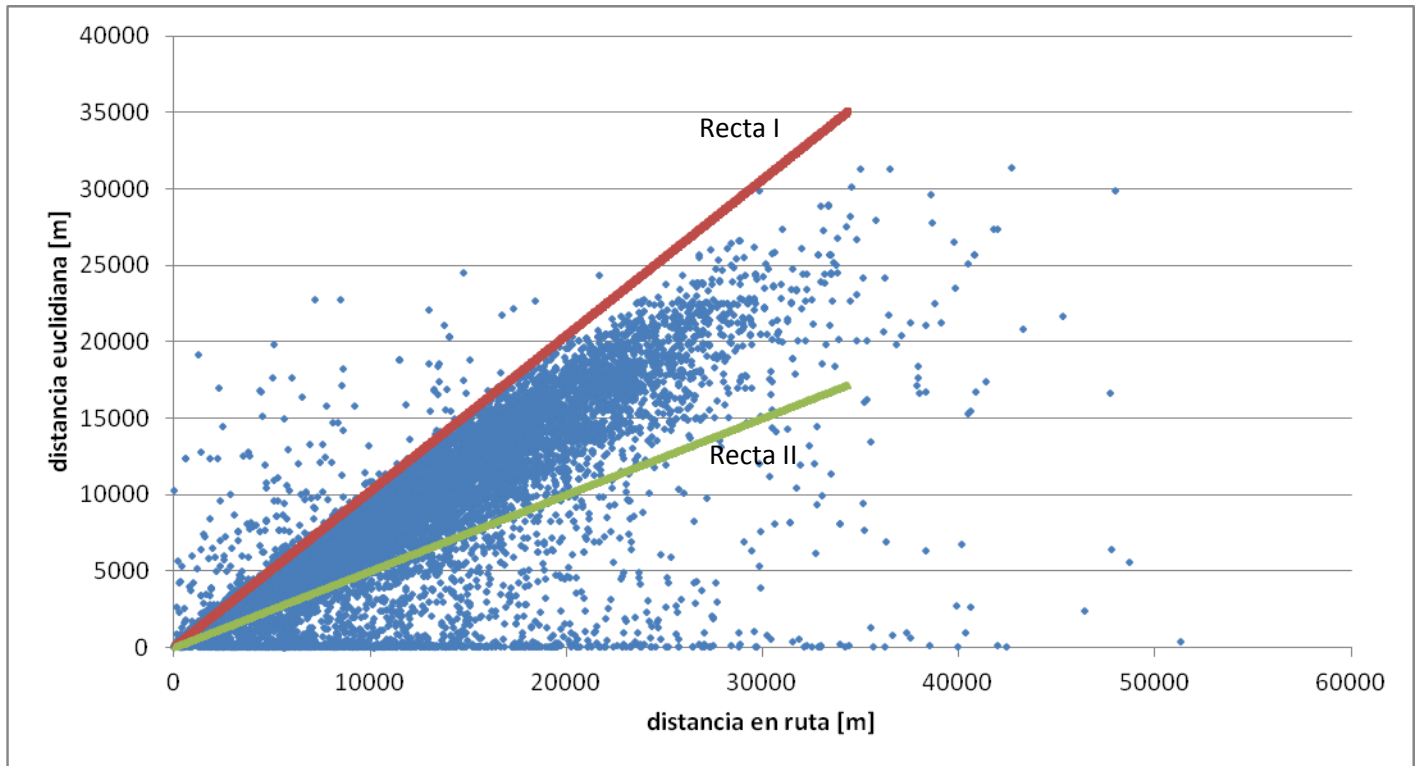
<sup>8</sup> Estas distancias nunca superan los 500 metros según la metodología de estimación de bajadas de Munizaga y Palma (2012). Si no se encuentra un paradero correspondiente al próximo servicio abordado por el usuario dentro de este radio, se considera que el punto de bajada de esta etapa no puede ser estimado.

clara imposibilidad geométrica. Cualquier viaje que posea esta característica presenta, por tanto, errores de estimación de distancia, atribuibles a falta de información en paraderos ya sea de subida o bajada, errores en las coordenadas geográficas asociadas, o bien errores asociados al post-procesamiento de datos.

Sobre el universo de viajes factibles, un análisis interesante constituye ver el comportamiento de  $f_d$  al cruzar su valor con el número de etapas estimadas en el viaje. Valores altos de  $f_d$  en viajes de una etapa caracterizan viajes realizados en recorridos cuyo trazado es intrincado y mucho más largo respecto a la distancia euclidiana de los pares origen/destino que sirven; valores de  $f_d$  cercanos a 1 en viajes de 1 etapa por el contrario caracterizan viajes realizados en recorridos cuyo trazado de servicio es más bien directo en términos de distancias.

Sin embargo, la aplicación más interesante de este factor está en aquellos viajes con altos valores de  $f_d$  y más de una etapa estimada (2, 3 ó 4 etapas). En términos prácticos, esta situación revela casos en los cuales la distancia recorrida en el viaje ha sido mucho mayor que la distancia euclidiana, y en donde se registró más de una etapa. Sin embargo, se dan en la realidad muchas situaciones que bajo la metodología utilizada por Munizaga y Palma (2012) serían procesadas como un viaje de más de una etapa, pero en realidad esconden más de un viaje, bajo circunstancias no observables o procesables directamente por dicha metodología pero sí identificables a través del factor  $f_d$ ; por ejemplo, los viajes “ida y retorno” con la finalidad de realizar actividades cortas podrían verse revelados al considerar este factor, ya que el individuo tiende a comenzar y terminar su viaje en paraderos relativamente cercanos (presumiblemente en la vecindad de la zona de residencia o trabajo), lo que genera una distancia euclidiana muy baja y consecuentemente un  $f_d$  alto; tal patrón de transacciones sería detectado por la metodología estándar como un solo viaje con uno o más trasbordos, sin embargo queda revelado en virtud del análisis anterior que en estricto rigor se registran dos viajes mediados por una actividad, tomando en consideración que realizar un trasbordo en un área alejada para acceder a un paradero de destino relativamente cercano, y al cual con alta probabilidad se podría acceder a través de un viaje con menor cantidad de etapas, es un comportamiento irracional desde el punto de vista de un usuario maximizador de utilidad. Otro caso de estudio que cabe en el contexto de la situación anteriormente descrita es el encadenamiento de viajes y actividades. Es posible que un individuo encadene numerosas actividades de corta duración a realizarse en diversos destinos en el marco de un viaje más largo; un ejemplo de lo anterior es realizar trámites bancarios antes de ir al trabajo en la mañana, o pasar de compras en el viaje de regreso del trabajo al hogar. En la medida que estas detenciones impliquen una mayor cantidad de etapas y desvíos considerables con respecto a la ruta que el individuo utilizaría para cubrir el par origen/destino preponderante (viaje del hogar al trabajo y del trabajo al hogar respectivamente en los ejemplos), esto se verá revelado en una mayor distancia de viaje para igual distancia euclidiana, y consecuentemente aumentará el factor  $f_d$ .

Se exhibe a continuación un gráfico de dispersión detallando el comportamiento, para una sub-muestra de 50.000 transacciones, de la distancia euclidiana con respecto a la distancia en ruta a nivel de viajes estimados con la metodología de Munizaga y Palma. (2012).



*Figura 4.3: Gráfica de distancia en ruta vs. distancia euclidiana*  
*Fuente: elaboración propia*

Como se observa, la Figura 4.3 revela dos tipos de tendencia claramente identificables. Existe por una parte una nube de observaciones que tiende a la recta identidad, esto es, observaciones en las cuales la distancia en ruta es ligeramente superior a la distancia euclidiana (o equivalentemente, poseen un valor de  $f_d$  cercano a 1). Por otro lado, existen observaciones que tienden al eje de la distancia en ruta, es decir poseen una distancia euclidiana entre paradero de subida y paradero de bajada estimados para el viaje en cuestión cercana a cero; estos serían los casos con valores elevados para  $f_d$  y que por tanto revelarían comportamientos sospechosos de encadenamiento de viajes y actividades o viajes de “ida y retorno”. Ambos comportamientos han sido divididos en el gráfico a través de rectas trazadas arbitrariamente, que se explican a continuación. La recta I está definida por la ecuación

$$d_{euclidiana} = d_{ruta} \cdot 1,02 \quad (4.2)$$

que define la identidad, en la cual la distancia en ruta es igual a la distancia euclidiana, considerando un 2% de holgura sobre la distancia euclidiana. Esta holgura es considerada

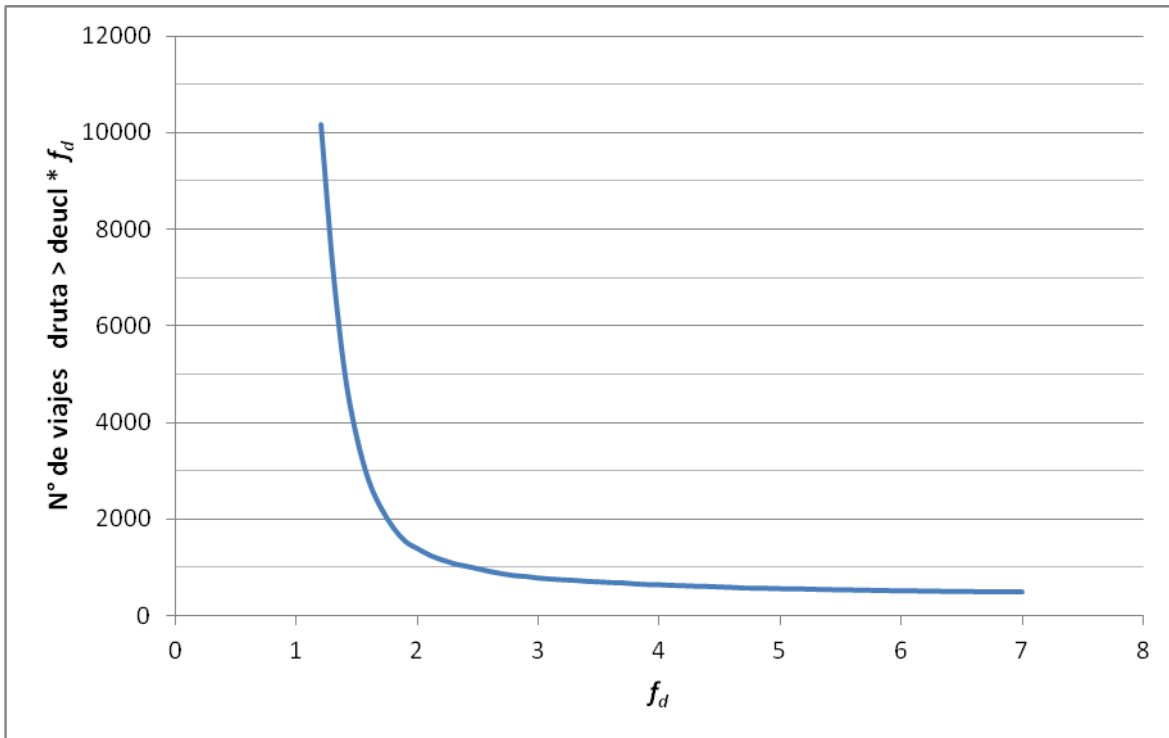
en virtud de los errores asociados a la medición de la posición geográfica de los buses por parte de los aparatos de GPS y su propagación en los diversos post-procesos que involucran esta variable. Aquellas observaciones que se encuentran en la región ubicada sobre esta recta (1,5% del total de observaciones) corresponden a viajes cuya distancia en ruta estimada es inferior a la distancia euclidiana entre los paraderos de inicio y término del viaje, lo cual es una situación físicamente imposible; estas observaciones por tanto presentan algún error ya sea en la medición de datos o en alguno de sus post-procesamientos.

La recta II está definida por la ecuación

$$d_{euclidiana} = \frac{d_{ruta}}{2} \quad (4.3)$$

con lo cual todas las observaciones que se encuentran en la región ubicada bajo esta recta (denominada “región crítica”) representan aquellos viajes estimados en donde la distancia en ruta es mayor al doble de la distancia euclidiana. Estos casos son a priori aquellos sospechosos de esconder encadenamientos o viajes “ida y retorno”. El factor 2 utilizado como límite de  $f_d$  para identificar viajes sospechosos no es arbitrario, sino que ha sido obtenido en base a un análisis de sensibilidad que se explica a continuación.

Si se deja la pendiente de la recta II, que delimita la región de viajes sospechosos, como variable (es decir el límite de  $f_d$  para identificar viajes sospechosos no es conocido) y se grafica en función de este valor límite el número de observaciones en la región sospechosa, se obtiene el gráfico de la Figura 4.4. En este, se denota que el número de observaciones en la región crítica posee un punto de inflexión alrededor de un valor límite de  $f_d$  de 2, a partir del cual el número de casos sospechosos tiende a estabilizarse. Es presumible que estos casos que se mantienen estables albergan viajes adicionales escondidos bajo la estimación de un solo viaje; aunque puede que algunos encadenamientos se encuentren escondidos en valores de  $f_d$  inferiores a 2, no existe una forma clara de distinguirlos y por tanto serán considerados como viajes únicos correctamente estimados por la metodología preexistente (Munizaga et al. 2012).

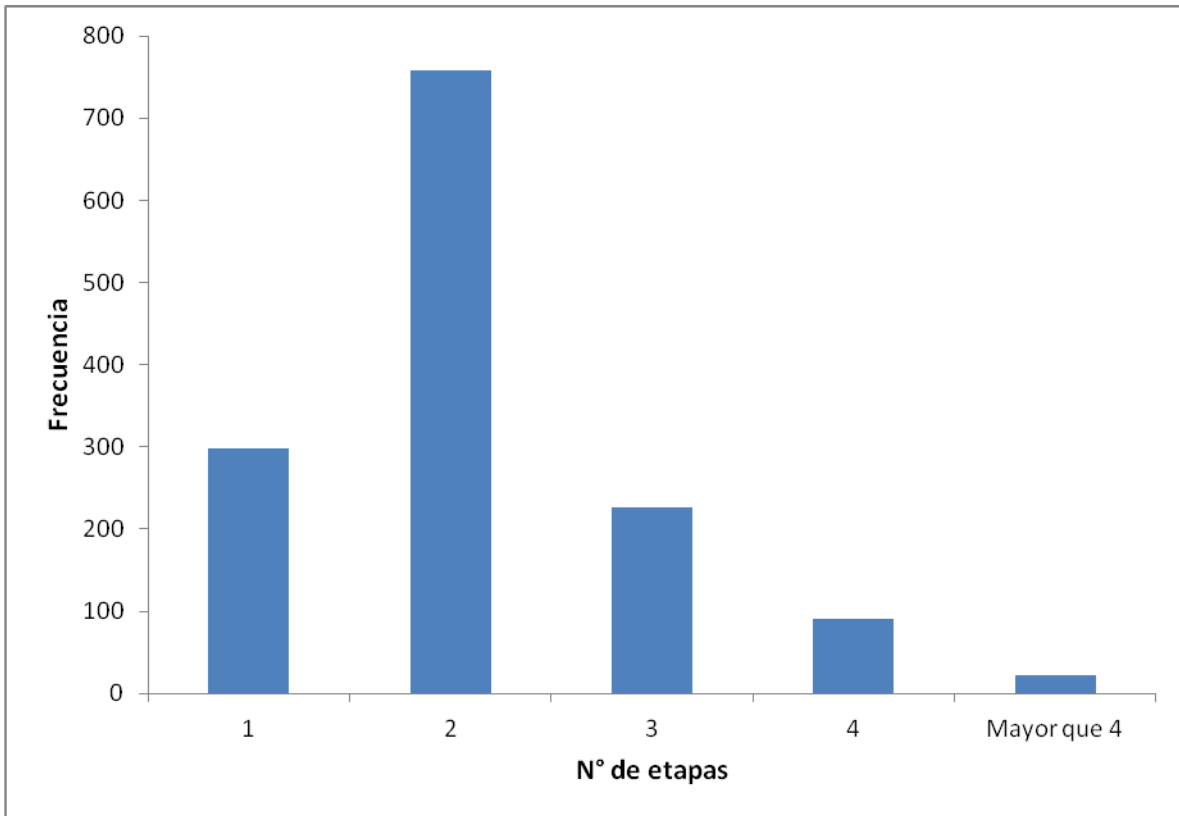


*Figura 4.4: Número de viajes en región crítica en función del valor límite de  $f_d$*

*Fuente: elaboración propia*

Estableciendo por tanto en 2 el límite de  $f_d$ , se obtiene un total de 2017 observaciones (viajes) en la región crítica, equivalentes al 12,1% de los viajes totales analizados; considerando este antecedente, se esperaría por tanto, al aplicar un criterio de corte basado en la razón  $f_d$ , un aumento de igual orden en los viajes totales estimados.

En forma concreta, el patrón de distancia en ruta/distancia euclidiana contempla separar todo viaje de más de una etapa que cumpla con la condición  $f_d > 2$ , en dos o más viajes. Un aspecto importante a observar es la composición en términos de etapas de los viajes que se encuentran en la región crítica. En la Figura 4.4 se observa que existe una cantidad no despreciable de viajes que poseen sólo una etapa. Claramente estos 298 viajes, correspondientes al 21,3%, no pueden ser divididos puesto que no poseen un trasbordo, y es probable que tal registro resulte de un usuario que efectivamente tomó un recorrido con numerosos desvíos o geográficamente tortuoso. La mayoría de los viajes en la región crítica sin embargo poseen dos etapas; en estos 758 viajes, correspondientes al 54,3% del total de viajes en la región crítica, llevar a cabo la separación del viaje es trivial, puesto que éste sólo posee un trasbordo.



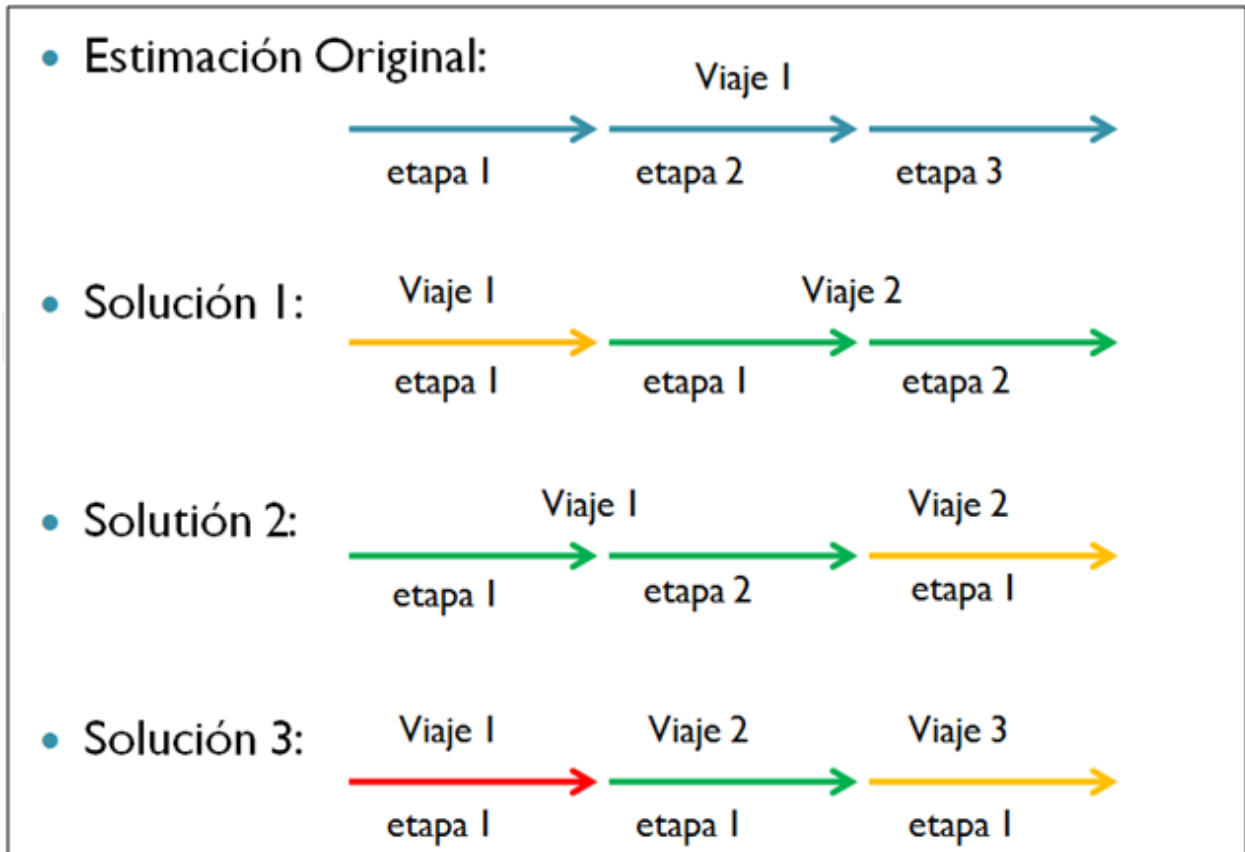
*Figura 4.5: Histograma de número de etapas para viajes en la región crítica*  
*Fuente: elaboración propia*

Para los casos de viajes en la región crítica con tres o más etapas, correspondientes al 24,4%, surge el problema de dónde realizar el (o los) corte(s) del viaje, es decir, cuál (o cuáles) de los trasbordos previamente estimados son realmente actividades. Para resolver este problema, se introduce el concepto de *verosimilitud de  $f_a$* , denotada  $V$ . Para definir formalmente esta verosimilitud, es necesario primero entender que dado un viaje de 3 o más etapas en la región crítica, existe más de una posible solución sobre cómo proceder para dividir el susodicho viaje; a su vez, cada una de esas soluciones puede contener dos o más viajes. Para entender esta idea en forma más clara, se presenta a continuación un ejemplo con un viaje de tres etapas, esquematizado en la Figura 4.6.

Sea un viaje en la zona crítica con 3 etapas, en el cual consecuentemente se estimaron 2 trasbordos (trasbordo 1 y trasbordo 2). Para dividir este viaje, existen 3 posibles soluciones, enumeradas a continuación:

1. La etapa 1 del viaje original es un (nuevo) viaje, y las etapas 2 y 3 forman parte de un mismo (nuevo) viaje. Esto es, el trasbordo 1 es una actividad, mientras que el trasbordo 2 es en efecto un trasbordo.

2. Las etapas 1 y 2 del viaje original son un solo (nuevo) viaje, y la etapa 3 del viaje original es otro (nuevo) viaje. Esto es, el trasbordo 1 era en efecto un trasbordo, mientras que el trasbordo 2 es en realidad una actividad.
3. Las etapas 1, 2 y 3 son cada una un (nuevo) viaje. Esto es, tanto el trasbordo 1 como el trasbordo 2 son actividades.



*Figura 4.6: Ejemplo: posibles formas de división de un viaje de tres etapas  
Fuente: elaboración propia*

Luego, es necesario definir un criterio para seleccionar en forma fundada alguna de las 3 posibles soluciones.

*Proposición:* Generalizando el ejemplo anterior, se llegará a que cualquier viaje en la zona crítica con  $n$  etapas, siendo  $n \geq 2$ , poseerá una cantidad  $s$  de formas de ser dividido (o soluciones), donde

$$s = 2^{n-1} - 1 \quad (4.3)$$

*Demostración:* considérese el conjunto ordenado y discreto  $\{1, 2, \dots, n\}$  correspondiente a las transacciones de un viaje de  $n$  etapas en la zona crítica. El número  $s$  de soluciones que



tendrá el problema de “dividir el susodicho viaje en un subconjunto de nuevos viajes de cardinalidad entre 2 y  $n$ ”, posee un análogo si se caracteriza como elementos en un nuevo conjunto, digamos  $C_t$ , los posibles trasbordos o actividades; como en general todo viaje posee un número de trasbordos igual al número de etapas menos 1, el conjunto  $C_t$  tendrá cardinalidad  $n-1$ . Luego,  $s$  será igual al número de subconjuntos que sea posible formar a partir del conjunto  $C_t$  de entre 1 y  $n-1$  elementos; cada uno de estos subconjuntos puede ser considerado en este contexto como el subconjunto de las actividades dentro de la solución a la que pertenecen. Por tanto, descontando la solución correspondiente al conjunto vacío (cero divisiones en el viaje original), que no ofrece en este caso una solución al problema sino que deja el viaje sin cortar, se tendrá de acuerdo al teorema de Newton que

$$s = \sum_{k=1}^{n-1} \binom{n-1}{k} = \sum_{k=0}^{n-1} \binom{n-1}{k} - \binom{n-1}{0} = 2^{n-1} - 1 \quad (4.4)$$

*Fin demostración.*

Se define, para un determinado viaje en la zona crítica con un conjunto  $C$  de formas de ser dividido (soluciones), la verosimilitud de la solución  $i \in C$  como

$$V_{sol_i} = - \sum_{\text{viajes } j \in sol_i} |f_{d_j} - f_{d_j}^{real}| \quad (4.5)$$

Donde  $f_{d_j}$  corresponde al coeficiente  $f_d$  del viaje  $j$ , y  $f_{d_j}^{real}$  corresponde al promedio de los coeficientes  $f_d$  de todos los viajes que no están en la zona crítica y que poseen igual número de etapas que el viaje  $j$ . Formalmente:

$$f_{d_j}^{real} = \frac{1}{\#N_j} \cdot \sum_{\text{viajes } k \in N_j} f_{d_k} \quad (4.6)$$

Siendo  $N_j$  el conjunto de todos los viajes con igual cantidad de etapas que el viaje  $j$  y que no están en la zona crítica y son viajes físicamente factibles, es decir que cumplen con la restricción

$$\frac{1}{1,02} \leq f_{d_k} \leq 2 \quad (4.7)$$

Y  $\#N_j$  la cardinalidad del conjunto  $N_j$ . La cota inferior de la restricción en la ecuación (4.7) viene dada por la holgura del 2% aplicada a la distancia en ruta para efectos de considerar un viaje como físicamente posible, explicada anteriormente en este capítulo.

Conceptualmente, la verosimilitud de una determinada solución representa el grado de semejanza que esta presenta con los viajes que se definen como “correctamente estimados” (esto es, que no se encuentran en la región crítica) en términos de cercanía entre sus coeficientes  $f_d$ . Constituye una función más completa a minimizar que  $f_d$ , ya que recoge casos en que la solución con el mínimo cociente puede no ser la real (por ejemplo casos de recorridos entre un par origen/ destino con rutas que no son euclidianamente directas, como por ejemplo las que deben dar rodeos por accidentes geográficos).

Por construcción, la verosimilitud de una solución cualquiera será negativa; consecuentemente dado un determinado viaje en zona crítica, la solución cuya verosimilitud sea mayor, será aquella solución que en el global contiene los viajes que más se aproximan a los valores promedio de  $f_d$  de los viajes considerados como correctamente calculados según el criterio de distancia y que poseen igual cantidad de etapas, es decir aquellos viajes que pertenecen al conjunto  $N_j$ . En conclusión, se adopta la solución que posea una verosimilitud más alta.

Cabe notar que esta metodología de verosimilitud puede refinarse a través de una especificación más detallada de los casos en base a los cuales se construyen los coeficientes  $f_d^{real}$ ; por ejemplo, este coeficiente podría ser calculado específicamente para cada viaje en la zona crítica, sus posibles soluciones, y más aún, para cada viaje asociado a las distintas soluciones, considerando sólo el universo de viajes que además de haber pasado el criterio de distancia y tener igual cantidad de etapas que el viaje en cuestión, tengan igual origen y destino; este procedimiento entregaría una estimación mucho más exacta ya que contrastaría las posibles soluciones y sus viajes contra coeficientes específicos considerados como correctos, y que además poseen características similares tanto en términos geográficos como de número de etapas (y por ende quedarían recogidos efectos como servicios que cubren rutas en forma no tan directa o singularidades geográficas que obliguen a tomar desvíos entre el par origen/destino en cuestión). Sin embargo, implementar un algoritmo de esta naturaleza tiene un costo computacional muy elevado, razón por la cual en el presente estudio no será llevado a cabo.

#### **4.2.5 Patrón bus no abordado**

La dimensión temporal juega en el presente problema, al igual que la dimensión espacial, un rol absolutamente preponderante. Es necesario por ende establecer un criterio que tenga como pivote conceptual la susodicha dimensión. En términos específicos, existen muchas variables en la base de datos de transacciones que representan tiempos, ya sea obtenidas directamente como información o bien como post-proceso de ésta; no obstante, una variable temporal sencilla de obtener y muy relevante de cara a identificar actividades

corresponde al tiempo entre una bajada y la siguiente transacción registrada para el mismo usuario. Esta ventana temporal corresponde al período en el cual el usuario podría potencialmente realizar una actividad en la localización estimada de bajada, o bien efectuar un trasbordo; por lo tanto, correspondería al tiempo estimado de espera en caso de que sea un trasbordo, o bien al tiempo de duración de la(s) actividad(es) en caso de que se haya estimado el fin del viaje en la etapa previa. Se define por tanto el *tiempo estimado de espera* entre la bajada de la transacción  $x-1$  y la transacción  $x$  como

$$t_s^{x-1;x} = t_{subida}^x - \left( t_{bajada}^{x-1} + \frac{d_{x-1;x}}{v_{cam}} \right) \quad (4.8)$$

Donde  $t_{subida}^x$  corresponde al tiempo en que se efectuó la transacción asociada a la etapa  $x$ , y  $t_{bajada}^{x-1}$  corresponde al tiempo estimado de bajada para la etapa respectiva a la transacción  $x-1$ ,  $d_{x-1;x}$  corresponde a la distancia euclidiana entre la bajada de la transacción  $x-1$  y la subida de la transacción  $x$ , y  $v_{cam}$  corresponde a la velocidad promedio de caminata, asumida igual a 4 [km/h] por ser este el valor tomado por el Ministerio de Vivienda y Urbanismo y por el Ministerio de Obras Públicas para efectos de cálculo y diseño (Ministerio de Vivienda y Urbanismo, 2009).

Es posible obtener cuántos buses del mismo recorrido<sup>9</sup> en el cual se efectuó una transacción pasaron por el paradero de abordaje del último bus, digamos nuevamente, la transacción  $x$ . Con esta información se puede reconocer si el usuario dejó pasar una cantidad significativa de buses. De ser esto afirmativo, se podría sostener que el usuario no se encontraba efectuando un trasbordo en ese momento, sino desarrollando una actividad fuera del paradero.

Formalmente, el criterio de corte se define a continuación:

Sea  $t_{bus}^x$  el tiempo en que pasó el bus del recorrido tomado por el usuario en la transacción  $x$  anterior a la misma, en el mismo paradero donde esta ocurrió. Sea  $D_{bus}^x$  tal que

$$D_{bus}^x = \begin{cases} 1 & \text{si } t_{bus}^x > t_{bajada}^{x-1} + \frac{d_{x-1;x}}{v_{cam}} \\ 0 & \text{si } t_{bus}^x \leq t_{bajada}^{x-1} + \frac{d_{x-1;x}}{v_{cam}} \end{cases} \quad (4.9)$$

Sea  $n_{bus}^x$  el número de buses del servicio abordado por el usuario en la transacción  $x$  que pasaron por el paradero durante la estadía del usuario en éste.

<sup>9</sup> Este criterio se podría mejorar si se incluyeran en el análisis las líneas comunes, esto es, considerar los buses de todas las líneas que conectan el par origen/destino de la etapa en cuestión recorrida por el usuario. Esto, sin embargo, podría aumentar los tiempos de procesamiento computacional de este criterio considerablemente.

Definido lo anterior, el criterio de corte por frecuencia se activará (cortando el viaje y definiendo la ventana temporal como una actividad) cuando se cumplan las siguientes condiciones:

$$D_{bus}^x = 1 \quad (4.10)$$

$$n_{bus}^x \geq \mu \quad (4.11)$$

Donde  $\mu$  corresponde al umbral de tolerancia de buses no abordados. Para el caso de este trabajo, el valor de  $\mu$  tomado corresponde a 3, en base a la experiencia de los modeladores al considerarse que en Santiago es infrecuente que un pasajero deje pasar más de esta cantidad de buses en el paradero por razones de que este venga a capacidad. Finalizado este criterio, y el módulo de estimación de actividades (y, complementariamente, de viajes), se procede a ejecutar el siguiente módulo de la metodología, correspondiente al Módulo de Asignación de Propósito.

### 4.3 Módulo de Asignación de Propósito (PAM)

Una vez terminada la ejecución del Módulo de Identificación de Actividades, cada una de las actividades estimadas posee como atributo el ID asociado, la parada en la periferia de la cual la actividad se habría desarrollado, la fecha y hora, y la duración. El PAM tiene como finalidad la estimación del atributo *propósito* para cada una de las actividades determinadas en el paso anterior de la metodología.

#### 4.3.1 Definición de Propósitos

El PAM toma como inputs los atributos pre-estimados de las actividades, y también información de uso de suelo en el caso de Gatineau, para asignar, de acuerdo a criterios previamente diseñados, uno de cuatro propósitos a cada actividad:

- **Trabajo:** constituye la actividad laboral principal del día
- **Estudio:** actividad de carácter estudiantil-académica principal del día
- **Hogar:** es el conjunto de actividades desarrolladas en el hogar después de haber llegado a éste y hasta el momento en que el usuario de transporte público vuelve a salir de él.
- **Otros:** cualquier actividad cuyo propósito no esté comprendido en una de las categorías anteriores.

Se presentan en el siguiente subinciso los criterios de asignación de propósitos.

### **4.3.2 Criterios de Asignación de Propósitos**

Los criterios de asignación de propósitos son diferentes para ambas ciudades debido a que en el caso de Gatineau se contaba, al momento de realizar el presente trabajo, con información adicional referente a uso de suelos. Esta información constó de una base de datos en donde para cada paradero de bus de la red de transporte público de la ciudad de Gatineau, se detallaba un vector de uso de suelos, que contenía en orden de predominancia los principales propósitos de las inmediaciones. Para recoger esta data en la metodología, se optó por diferenciar el módulo PAM en cada ciudad. También se diferenciaron los criterios para asignación en un día laboral y en fines de semana, para reconocer el hecho de que la naturaleza de las actividades cambia en uno u otro tipo de día; principalmente, el propósito otros tiende a ser más corto los días de semana con respecto a las actividades de estudio y trabajo, mientras que los fines de semana, las actividades de otros propósitos tienen una predominancia mayor tanto en duración como en frecuencia (Olguín et al. 2009).

Se presentan a continuación los criterios de asignación de propósitos para ambas ciudades en un día laboral (Tabla 4.1) y para un fin de semana (Tabla 4.2).

Base de datos Propósito	Gatineau (día laboral)	Santiago (día laboral)
Trabajo	<ul style="list-style-type: none"> <li>• Tipo de tarjeta es adulto</li> <li>• Se cumple una de las siguientes:               <ul style="list-style-type: none"> <li>- La actividad dura más de 5 horas</li> <li>- La actividad dura entre 2 y 5 horas, y la zona del paradero no es hospital ni recreacional</li> </ul> </li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>	<ul style="list-style-type: none"> <li>• Tipo de tarjeta es adulto</li> <li>• La actividad dura más de 5 horas</li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>
Estudio	<ul style="list-style-type: none"> <li>• Tipo de tarjeta es estudiante o menor</li> <li>• Se cumple una de las siguientes:               <ul style="list-style-type: none"> <li>- La actividad dura más de 5 horas</li> <li>- La actividad dura entre 2 y 5 horas, y la zona del paradero es educacional</li> </ul> </li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>	<ul style="list-style-type: none"> <li>• Tipo de tarjeta es estudiante</li> <li>• La actividad dura más de 5 horas</li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>
Hogar	<ul style="list-style-type: none"> <li>• El viaje anterior a la actividad fue el último del día</li> <li>• El viaje antes de la actividad no es el único del día</li> </ul>	<ul style="list-style-type: none"> <li>• El viaje anterior a la actividad fue el último del día</li> </ul>
Otros	<ul style="list-style-type: none"> <li>• Si el tipo de tarjeta es adulto, se cumple una de las siguientes:               <ul style="list-style-type: none"> <li>- La actividad dura entre 30 minutos y 2 horas</li> <li>- La actividad dura entre 2 y 5 horas, y la zona del paradero es hospital o recreacional</li> </ul> </li> <li>• Si el tipo de tarjeta es estudiante o menor, se cumple una de las siguientes:               <ul style="list-style-type: none"> <li>- La actividad dura entre 30 minutos y 2 horas</li> <li>- La actividad dura entre 2 y 5 horas, y la zona del paradero no es educacional</li> </ul> </li> <li>• El viaje antes de la actividad no es el único del día</li> </ul>	<ul style="list-style-type: none"> <li>• La actividad dura entre 1 minuto y 2 horas</li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>

Tabla 4.1: Criterios de asignación de propósitos para día laboral

Fuente: elaboración propia

Base de datos Propósito	Gatineau (fin de semana)	Santiago (fin de semana)
Trabajo	<ul style="list-style-type: none"> <li>• Tipo de tarjeta es adulto</li> <li>• Se cumple una de las siguientes: <ul style="list-style-type: none"> <li>- La actividad dura más de 5 horas, y la zona del paradero no es hospital ni recreacional</li> <li>- La actividad dura entre 2 y 5 horas, y la zona del paradero es industrial</li> </ul> </li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>	<ul style="list-style-type: none"> <li>• Tipo de tarjeta es adulto</li> <li>• La actividad dura más de 5 horas</li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>
Estudio	<ul style="list-style-type: none"> <li>• Tipo de tarjeta es estudiante o menor</li> <li>• La actividad dura más de 5 horas</li> <li>• La zona del paradero es educacional</li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>	<ul style="list-style-type: none"> <li>• Tipo de tarjeta es estudiante</li> <li>• La actividad dura más de 5 horas</li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>
Hogar	<ul style="list-style-type: none"> <li>• El viaje anterior a la actividad fue el último del día</li> <li>• El viaje antes de la actividad no es el único del día</li> </ul>	<ul style="list-style-type: none"> <li>• El viaje anterior a la actividad fue el último del día</li> </ul>
Otros	<ul style="list-style-type: none"> <li>• Si el tipo de tarjeta es adulto, se cumple una de las siguientes: <ul style="list-style-type: none"> <li>- La actividad dura entre 30 minutos y 2 horas</li> <li>- La actividad dura entre 2 y 5 horas, y la zona del paradero no es industrial</li> <li>- La actividad dura más de 5 horas, y la zona del paradero es hospital o recreacional</li> </ul> </li> <li>• Si el tipo de tarjeta es estudiante o menor, se cumple una de las siguientes: <ul style="list-style-type: none"> <li>- La actividad dura entre 30 minutos y 5 horas</li> <li>- La actividad dura más de 5 horas, y la zona del paradero no es educacional</li> </ul> </li> <li>• El viaje antes de la actividad no es el único del día</li> </ul>	<ul style="list-style-type: none"> <li>• La actividad dura entre 1 minuto y 5 horas</li> <li>• El viaje anterior a la actividad no fue el último del día</li> </ul>

Tabla 4.2: Criterios de asignación de propósitos para fines de semana (sábado y domingo)

Fuente: elaboración propia

De esta forma, todas las actividades reciben una asignación de propósito en base a los criterios presentados. Cada actividad tiene por tanto, a la salida del PAM, los atributos ID, paradero, fecha y hora, duración y propósito.

La justificación de los criterios presentados en las tablas 4.1 y 4.2 se centra en el estudio previamente realizado por Olgún et al. (2009). Los límites temporales para el establecimiento y asignación de los diversos propósitos fueron tomados de los promedios de duración de actividades de la Figura 2.4, puesto que constituye al momento de la realización del presente trabajo una de las fuentes más completas y recientes sobre uso de tiempo. No obstante lo anterior, el mencionado trabajo se encuentra desarrollado como ya se ha dicho en base a la Encuesta Origen/Destino 2001, que incluye tanto a usuarios de transporte privado como de transporte público, por lo que una limitación (y posible fuente de sesgo) del presente trabajo lo constituye la suposición de que ambos usuarios poseen comportamientos similares en términos de duración de sus actividades según propósito.

Concluida la fase PAM, las actividades, etapas y viajes se encuentran procesados y listos para análisis.



# Capítulo 5

## Viajes y actividades en Santiago y Gatineau

En este capítulo se presentan los resultados obtenidos a partir de la aplicación de la metodología a las bases de datos de Santiago y Gatineau, como también los análisis realizados tanto para la dinámica de viajes como para actividades. Si bien ambas estimaciones son complementarias en la metodología, la información proporcionada por cada una es analizable desde distintas aristas.

Las gráficas y los análisis mostrados en el presente capítulo son un compendio de las aplicaciones realizables a través de la metodología.

### 5.1 Estadísticas generales sobre viajes y actividades - análisis de impacto de mejoras metodológicas

Se presentan en este inciso cifras generales referentes a viajes y actividades tanto en Santiago como en Gatineau.

Se presentan en la Tabla 5.1 los resultados relativos a la cantidad de viajes estimados según cantidad de etapas para cada ciudad, y los tiempos de viaje promedio para cada una de las categorías. Se exhiben además los tiempos promedios ponderados por cantidad de viajes según número de etapas para un viaje promedio en cada ciudad. Se observa que en general los viajes en Santiago presentan una mayor cantidad de etapas comparativamente respecto a Gatineau; sin embargo estos tienden también a ser ligeramente más largos. Mientras en Santiago los viajes de una etapa son aproximadamente la mitad del total, en Gatineau estos corresponden a un considerable 86%. Este fenómeno es probablemente debido al tamaño y forma de la ciudad, como también a la estructura de los servicios de transporte público;

Santiago es una ciudad radial geográficamente y espacialmente más grande y habitada que Gatineau, que a su vez es una ciudad de forma más bien lineal. Adicionalmente, la estructura del transporte público en Santiago es de tipo troncal-alimentador desde la implementación de Transantiago, mientras que en Gatineau se observa un sistema de líneas directas. De esta forma, en Santiago se estima un tiempo promedio por viaje ponderado de 41,9 minutos, y un valor bastante menor de 26,7 minutos para Gatineau. En términos de etapas por viaje, se obtiene para Santiago un promedio de 1,68, mientras que en Gatineau se estima que cada viaje posee en promedio 1,15 etapas.

	Santiago		Gatineau	
	Número de Viajes	Tiempo promedio [min]	Número de Viajes	Tiempo promedio [min]
1 etapa	7.959.600 <b>(52,05%)</b>	26,6	17.411.237 <b>(86,05%)</b>	23,1
2 etapas	4.937.547 <b>(32,29%)</b>	42,5	2.537.057 <b>(12,5%)</b>	46,3
3 etapas	1.655.519 <b>(10,83%)</b>	75,7	241.430 <b>(1,19%)</b>	71,1
4 o más etapas	737.811 <b>(4,83%)</b>	126,8	42.827 <b>(0,21%)</b>	100,6
	Tiempo promedio ponderado [min]	41,9	Tiempo promedio ponderado [min]	26,7
	<b>Etapas por viaje promedio []</b>	<b>1,68</b>	<b>Etapas por viaje promedio []</b>	<b>1,15</b>

*Tabla 5.1: Número de viajes según etapas y tiempos promedio de viaje para Santiago y Gatineau*

*Fuente: elaboración propia*

En términos del impacto de las mejoras metodológicas, se muestran en la Tabla 5.2 los resultados obtenidos del procesamiento de datos para Santiago utilizando la metodología estándar (Munizaga y Palma, 2012), y aplicando las mejoras metodológicas. Se observa que tras la aplicación de estas, aumenta tanto el número de viajes estimados, desde 18.712.103 a 20.871.496 (correspondiente a un aumento de un 14%), y también se produce un incremento del 1,9% de la proporción de viajes que posee estimación de tiempo de viaje. Respecto a la estructura de los viajes en términos de cantidad de etapas, la introducción de las mejoras metodológicas produce notorios cambios: se observa en contraste con la metodología basal una tendencia, por un lado, a reducir el número de etapas de los viajes,

produciéndose un aumento del 11% global en los viajes con una etapa estimados, un aumento del 6% global de los viajes de 2 etapas estimados, mientras que los viajes de 3 etapas bajan alrededor de un 2% global, y los viajes de 4 o más etapas reducen una drástica caída de un 3% global, correspondiente a una disminución del 30% de este tipo de viajes, mayormente debida a la mejor identificación del encadenamiento de viajes y actividades. Todo esto se traduce en una disminución del promedio de etapas por viaje desde 1,87 a 1,68. Por otra parte, los tiempos de viaje estimados aumentan considerablemente para el caso de los viajes de una etapa en 10 minutos, correspondiente a un incremento del 60%; sin embargo, los viajes de 2, 3, y 4 o más etapas disminuyen sus tiempos de viaje estimados, lo que en un ponderado global resulta en una disminución del tiempo promedio de viaje desde 49,7 minutos a 41,9 minutos, correspondiente a un decremento del 16%.

	<b>Metodología Estándar</b>		<b>Metodología incluyendo mejoras</b>	
	Etapas analizadas:	38.118.949	Etapas analizadas:	38.118.949
	Viajes estimados:	18.712.103	Viajes estimados:	20.871.496
	Viajes con estimación de tiempo de viaje:	13.381.784 <b>(71,51%)</b>	Viajes con estimación de tiempo de viaje:	15.290.477 <b>(73,41%)</b>
	<b>SOBRE EL TOTAL DE VIAJES CON ESTIMACIÓN DE TIEMPO DE VIAJE:</b>		<b>SOBRE EL TOTAL DE VIAJES CON ESTIMACIÓN DE TIEMPO DE VIAJE:</b>	
	<b>Viajes</b>	<b>Tiempo de viaje promedio [min]</b>	<b>Viajes</b>	<b>Tiempo de viaje promedio [min]</b>
<b>1 Etapa</b>	5.493.357 <b>(41,05%)</b>	16,60	7.959.600 <b>(52,05%)</b>	26,55
<b>2 Etapas</b>	5.114.300 <b>(38,2%)</b>	51,29	4.937.547 <b>(32,29%)</b>	42,48
<b>3 Etapas</b>	1.722.699 <b>(12,8%)</b>	85,15	1.655.519 <b>(10,83%)</b>	75,71
<b>4 o más Etapas</b>	1.051.428 <b>(7,85%)</b>	153,02	737.811 <b>(4,83%)</b>	126,78
	<b>Tiempo de viaje promedio ponderado:</b>	<b>49,67</b>	<b>Tiempo de viaje promedio ponderado:</b>	<b>41,86</b>
	<b>Etapas por Viaje promedio:</b>	<b>1,87</b>	<b>Etapas por Viaje promedio:</b>	<b>1,68</b>

*Tabla 5.2: Resultados obtenidos aplicando metodología estándar (Munizaga y Palma, 2012) e implementando mejoras metodológicas.*

*Fuente: elaboración propia*

Las cifras generales obtenidas para actividades se exhiben en la Tabla 5.3. Se observan en esta tanto en Santiago como en Gatineau resultados similares en cuanto a tipo de

actividades llevadas a cabo en un día laboral; alrededor de un 50% de las actividades llevadas a cabo corresponden a la categoría de actividades asociadas al hogar, y un 30% a actividades de tipo laboral. Otros propósitos corresponde a alrededor de un 10-12% de las actividades en ambas ciudades, porcentaje similar al del propósito Estudio. Cabe resaltar que este último propósito sólo puede ser ejecutado por usuarios con tarjetas de tipo estudiante. Por otro lado, los fines de semana la estructura de desarrollo de actividades difiere en ambas ciudades; en Gatineau se observa una clara predominancia de Otros propósitos como la principal actividad a desarrollarse fuera del hogar, con un 43,6% de las actividades totales, mientras que la actividad trabajo solo ocupa un 6,1%. En Santiago en cambio se observa una tendencia similar pero con diferencias menos exacerbadas: un 26% del total de actividades en un fin de semana corresponde a Otros propósitos, mientras que un 15,6% del total de actividades corresponde a actividades de tipo Trabajo.

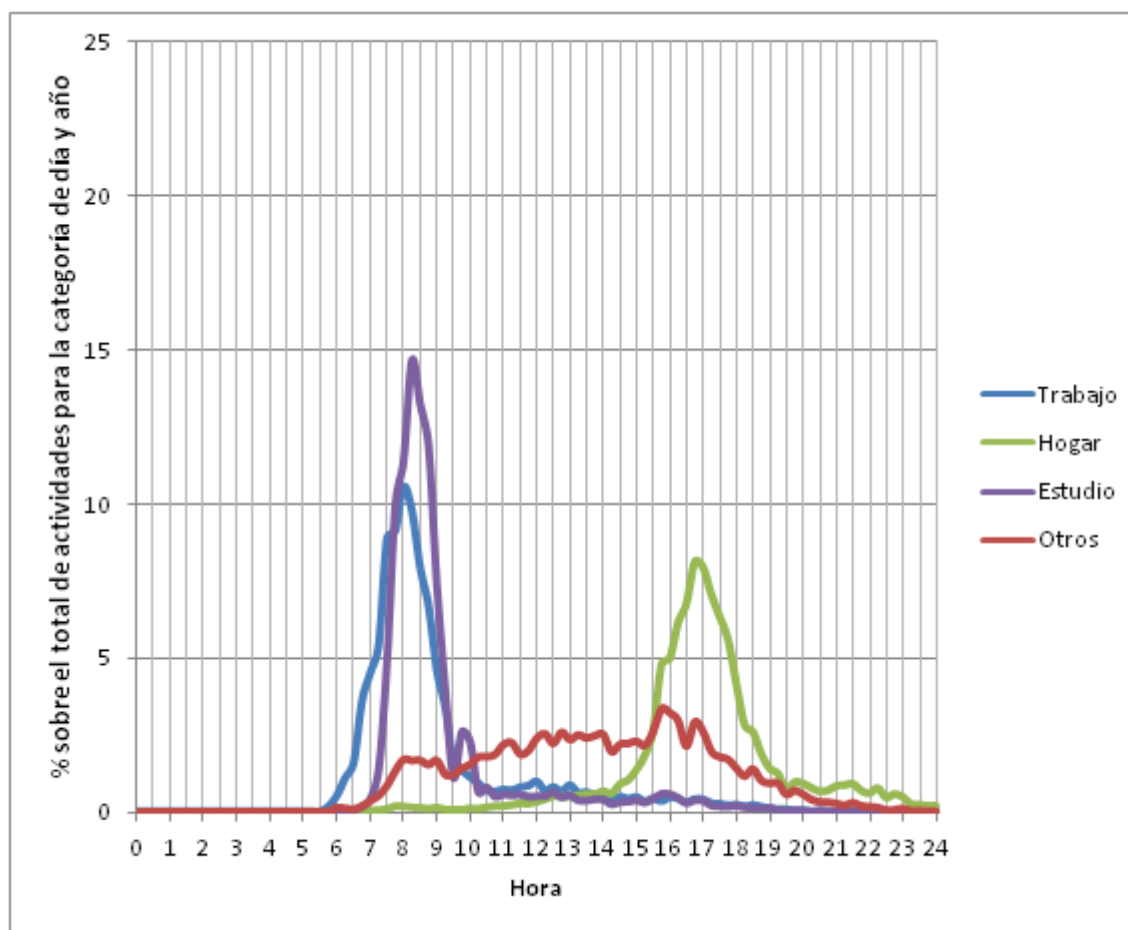
	Gatineau		Santiago	
	<i>Día Laboral</i>	<i>Fin de Semana</i>	<i>Día Laboral</i>	<i>Fin de Semana</i>
<b>Trabajo</b>	4.723.959 <b>(29,9%)</b>	100.320 <b>(6,1%)</b>	3.507.083 <b>(30,6%)</b>	300.733 <b>(15,6%)</b>
<b>Estudio</b>	1.178.042 <b>(7,5%)</b>	15.057 <b>(0,9%)</b>	1.269.941 <b>(11,1%)</b>	73.376 <b>(3,8%)</b>
<b>Hogar</b>	7.872.839 <b>(49,8%)</b>	813.513 <b>(49,4%)</b>	5.431.202 <b>(47,4%)</b>	1.055.687 <b>(54,7%)</b>
<b>Otros</b>	2.034.262 <b>(12,9%)</b>	717.894 <b>(43,6%)</b>	1.242.310 <b>(10,8%)</b>	501.455 <b>(26,0%)</b>

*Tabla 5.3: Actividades según día de la semana en Santiago y Gatineau*

*Fuente: elaboración propia*

La hora de inicio de los distintos tipos de actividades es un interesante caso de análisis que puede revelar profundos patrones y diferencias socio-culturales. Se muestra a continuación una serie de gráficas a propósito de este tópico. De la Figura 5.1, Se observa en Gatineau un peak de comienzo de actividades de Trabajo alrededor de las 8:00 horas, y un peak de comienzo de actividades de Estudio alrededor de las 8:20. Las actividades de Otros propósitos siguen un patrón de comienzo más bien uniforme a lo largo del día, comenzando

a las 8:00, observándose un ligero peak alrededor de las 16:00, y decayendo progresivamente hasta las 23:00 horas. El propósito Hogar, cuya hora de inicio coincide con la llegada de los usuarios de transporte público a su lugar de residencia, presenta un peak alrededor de las 16:45 horas. Se analizará a continuación un gráfico similar para el caso de la ciudad de Santiago.



*Figura 5.1: Tiempos de comienzo de actividades para lunes-jueves en Gatineau, año 2009*  
*Fuente: elaboración propia*

En la Figura 5.2 se observa que Santiago posee tiempos de comienzo de actividades un tanto diferentes con respecto a Gatineau. Las actividades laborales tienen un peak a las 8:30, más tarde que el propósito estudio, que a su vez presenta el peak a las 8:00; además, se observa para el propósito estudio una distribución de 3 peaks a lo largo del día: adicional a la de la mañana se observa una alrededor de las 14:00, probablemente actividades académicas posteriores al almuerzo, y otro peak alrededor de las 18:30, atribuible a actividades académicas en horas de la tarde. Otros propósitos presentan una distribución similar al caso de Gatineau, distribuidos en forma uniforme a lo largo del día. La hora de inicio de las actividades en el hogar presentan la mayor diferencia en el caso de Santiago

con respecto a Gatineau; esto puede observarse en forma más clara en la Figura 5.3, donde se destaca la hora de inicio del propósito Hogar en ambas ciudades.

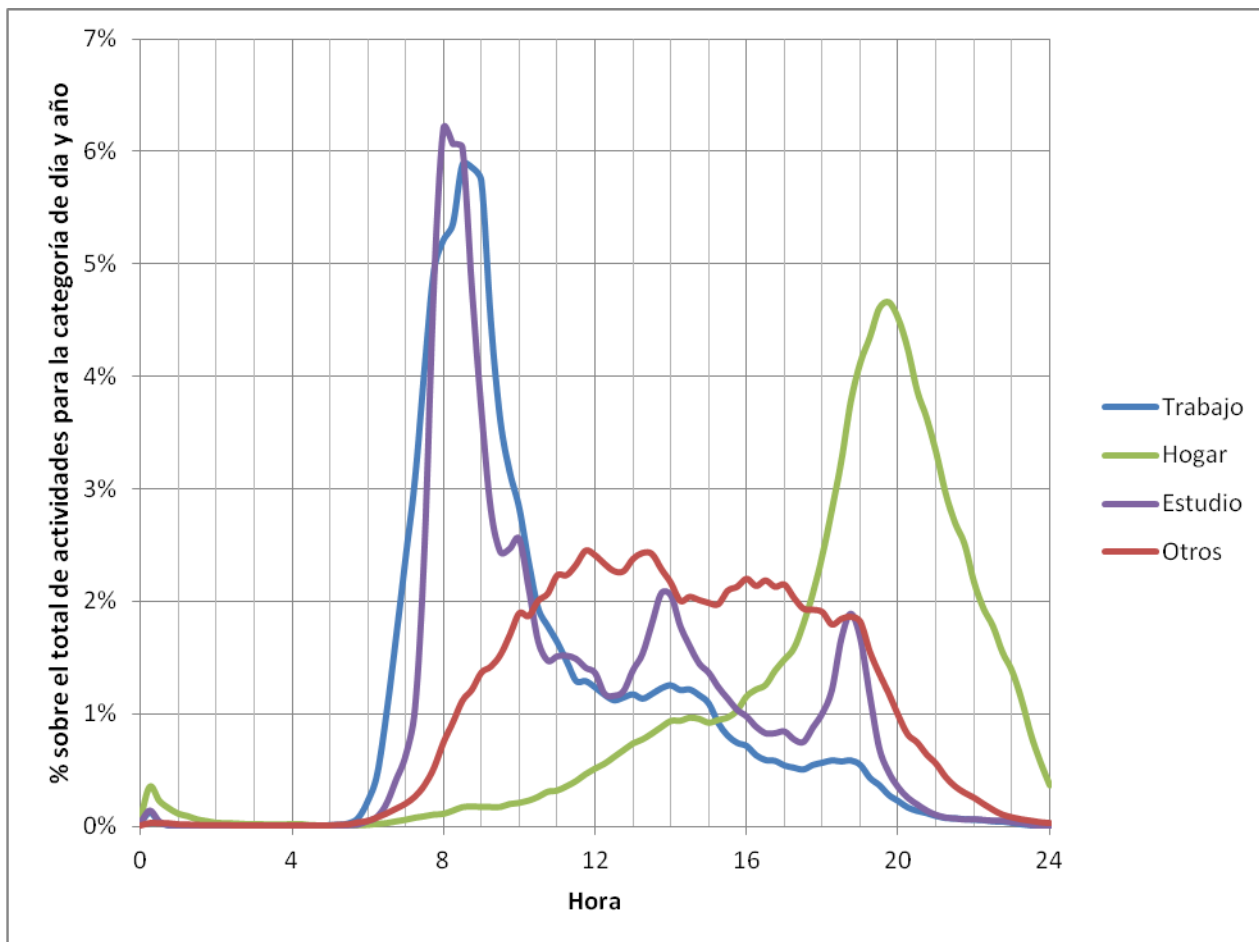
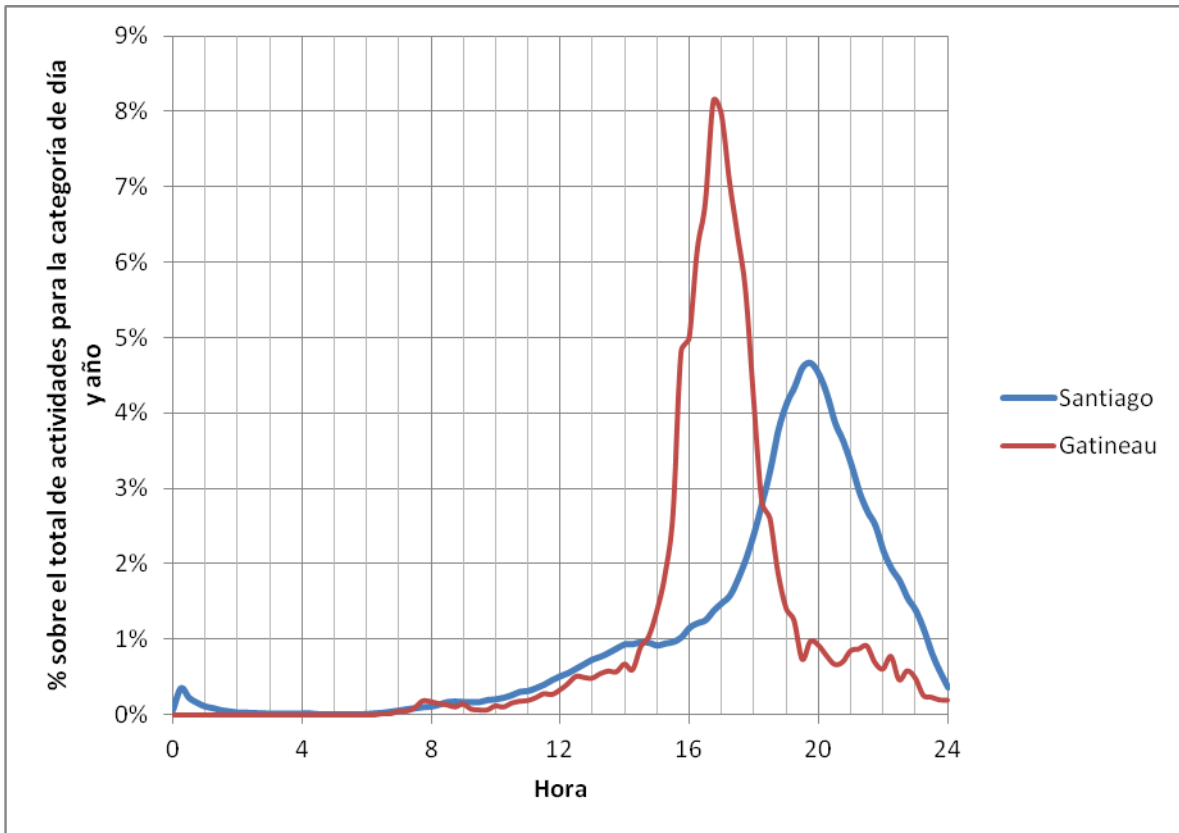


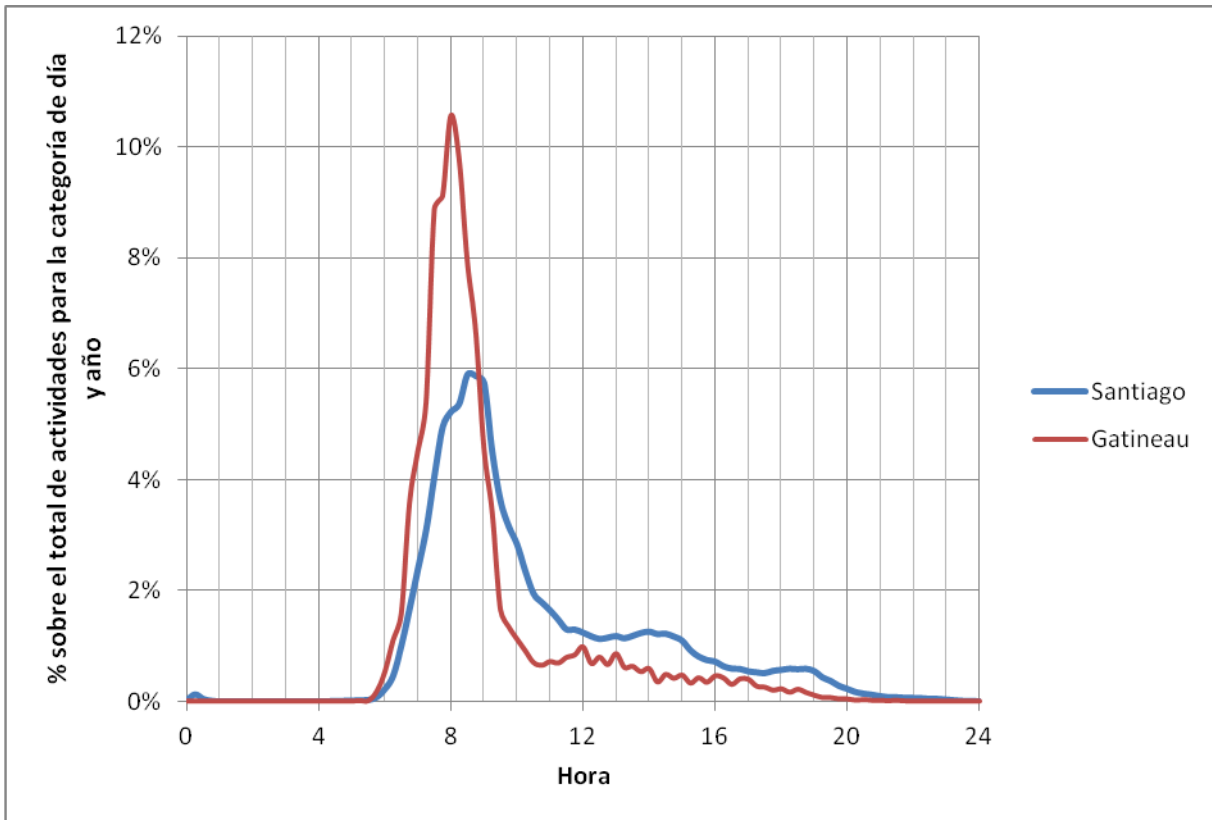
Figura 5.2: Tiempos de comienzo de actividades para lunes-jueves en Santiago, año 2010  
Fuente: elaboración propia



*Figura 5.3: Comparación entre horas de comienzo de actividades en el hogar durante lunes-jueves en Santiago y Gatineau*  
*Fuente: elaboración propia*

Se observa una marcada diferencia en los horarios de retorno al hogar en Santiago y Gatineau; la ciudad chilena presenta una distribución con una mayor desviación estándar centrada alrededor de las 19:45. La ciudad canadiense presenta una distribución con un peak más pronunciado, que sucede más temprano alrededor de las 16:45 y con una varianza menor. Este fenómeno de comportamiento social puede ser explicado por variaciones socio-culturales entre ambos países, contrastes climáticos, y también diferencias en la cantidad legal de horas laborales semanales, que es 5 horas menor en Canadá.

Las actividades con propósito Trabajo también presentan diferencias entre ambas ciudades. Se observa en la Figura 5.4 que el peak de comienzo de actividades laborales en Santiago es menos pronunciado y sucede más tarde que en la ciudad de Gatineau, siendo de un 6% de ocurrencias concentradas por cuarto de hora a las 08:30, y de un 11% alrededor de las 08:00 en la ciudad de Gatineau. Consecuentemente se observa una curva mayormente distribuida a lo largo del día en Santiago, presentando mayor cantidad de actividades laborales que comienzan en la tarde (12:00-19:00).

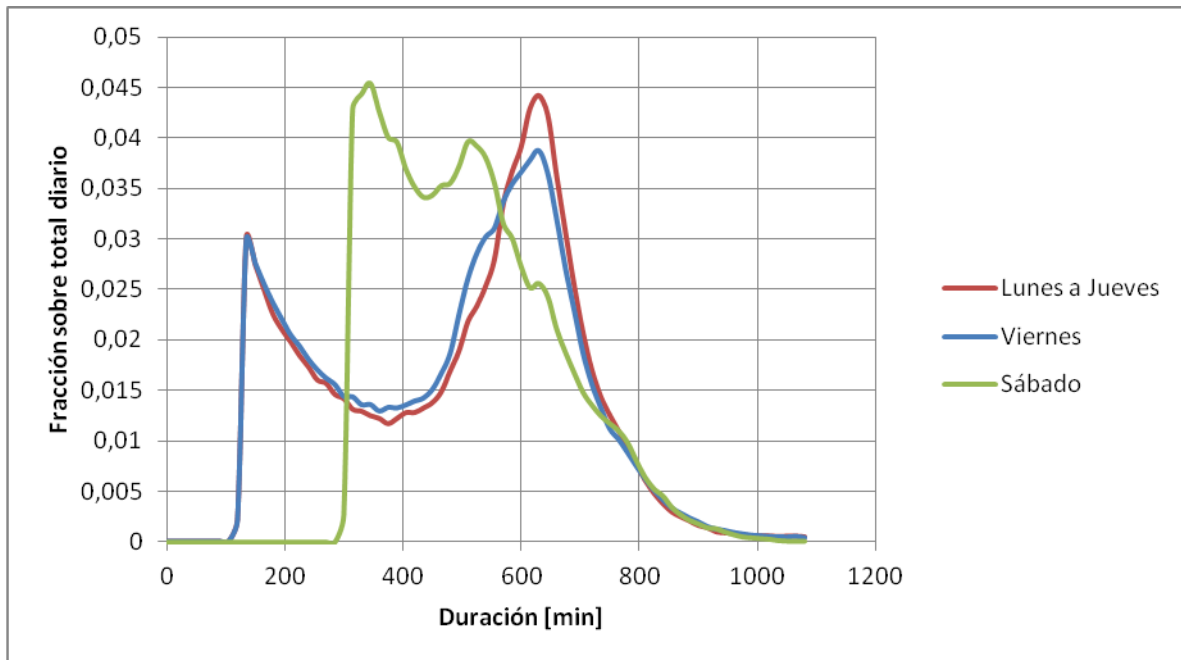


*Figura 5.4: Comparación entre horas de comienzo de actividades trabajo durante lunes-jueves en Santiago y Gatineau*  
*Fuente: elaboración propia*

Respecto a las actividades, no solo es relevante analizar el propósito, frecuencia y hora de ocurrencia, sino también la duración. Se presentan a continuación estas estimaciones para el caso de Santiago. En la Figura 5.5 se observa, en cuanto a duración de las actividades laborales, una ligera tendencia a la sucesión de actividades más cortas los días viernes en comparación con los días lunes a jueves, aunque el peak en ambos casos es entorno a los 600 minutos. Cabe destacar que esta cifra esconde las potenciales actividades no detectables debido a que el usuario no accede a ellas a través de modos asociables a tarjetas inteligentes, tales como trámites, almuerzos u otras actividades de otros propósitos comúnmente efectuadas en el transcurso de la jornada laboral. Se observa adicionalmente una cantidad considerable de actividades laborales entre los 120 minutos, progresivamente decayendo hasta los 400 minutos; este subconjunto de actividades corresponde al universo de actividades que potencialmente podrían ser de tipo “otros propósitos”, debido a que incuestionablemente existen actividades de estos propósitos con estas duraciones; sin embargo, para un mayor grado de identificación y refinamiento en este aspecto, se requiere de una mayor cantidad de información, por ejemplo a través de data referente a uso de



suelos. Con respecto al día sábado, las actividades laborales tienen una clara tendencia a ser más cortas, siendo el peak en torno a los 350 minutos y 500 minutos, decayendo progresivamente en forma posterior.



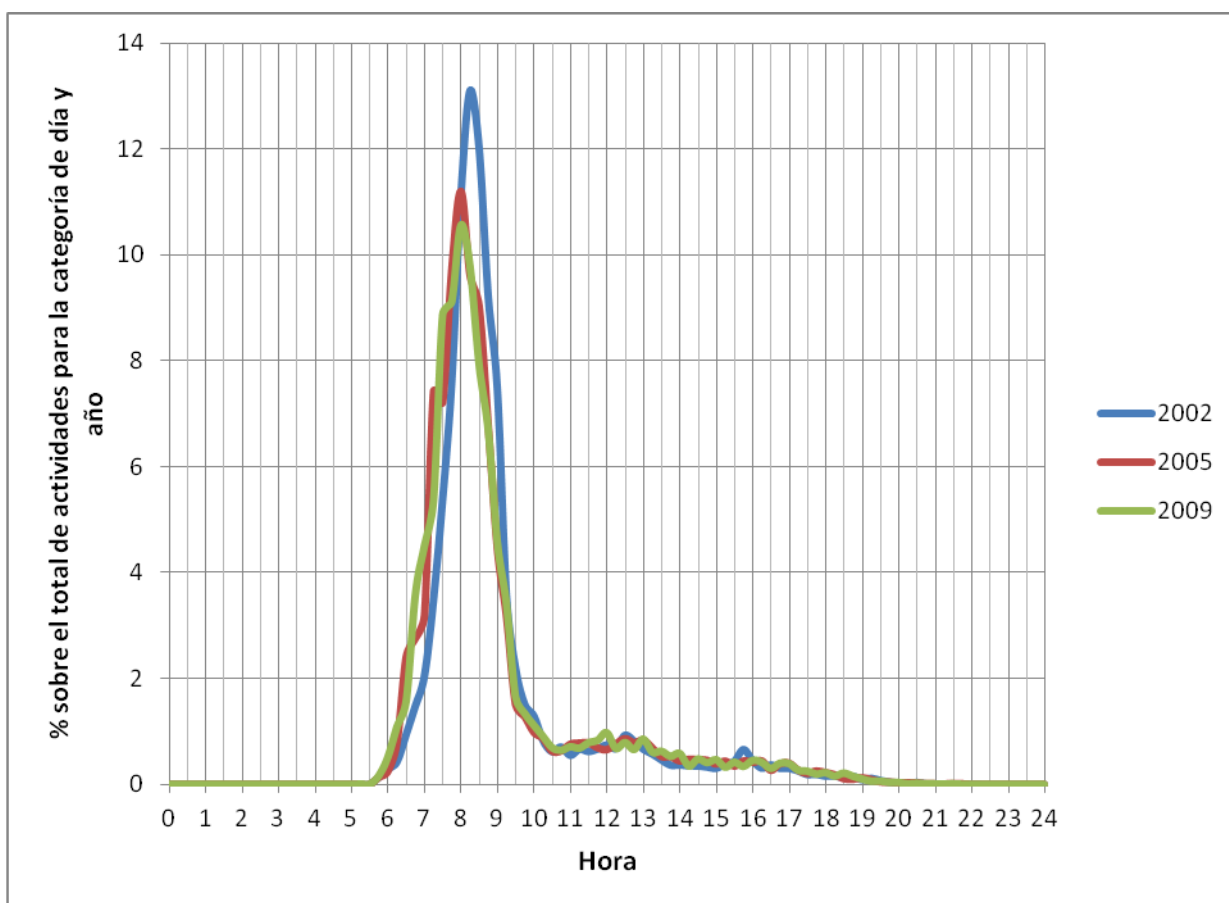
*Figura 5.5: Duración de actividades trabajo en Santiago para los distintos días de la semana*

*Fuente: elaboración propia*

## 5.2 Aplicación a Gatineau: evolución temporal de actividades

El amplio horizonte temporal de 9 años de la base de datos de transporte público asociada a la ciudad de Gatineau permite un análisis evolutivo en cuanto a la ocurrencia de actividades. Para tal propósito, se tomaron 3 años específicos dentro de la base de datos, uno al comienzo de esta (2002), otro en la mitad (2005) y otro al final (2009), y se llevó a cabo un análisis comparativo de la ocurrencia de actividades en cada uno de estos años, encontrándose interesantes resultados.

En una primera aplicación, se analizó la evolución de los tiempos de inicio de cada uno de los propósitos en los tres años. Se observan en la Figura 5.6 dos tendencias claras. Por una parte, los tiempos de inicio de las actividades laborales se han visto desplazados con una tendencia a ocurrir más temprano: mientras en el 2002 el peak de actividades sucedía a las 8:20, ya en el 2005 el peak sucedía a las 8:00, horario que se mantuvo en el 2009. Por otro lado, se observa una tendencia sostenida a la disminución de la concentración de actividades en los peaks, en pos de curvas de comienzo de actividades con una mayor desviación estándar, entendiéndose por ende que la distribución de horarios de entrada a trabajar presentaría una mayor holgura. De esta forma, mientras en el 2002 el peak registraba un 13% de las actividades, en 2005 el peak agrupaba un 11% de las actividades y en el 2009 sólo un 10,5% de las transacciones estaban incluidas en el cuarto de hora peak.

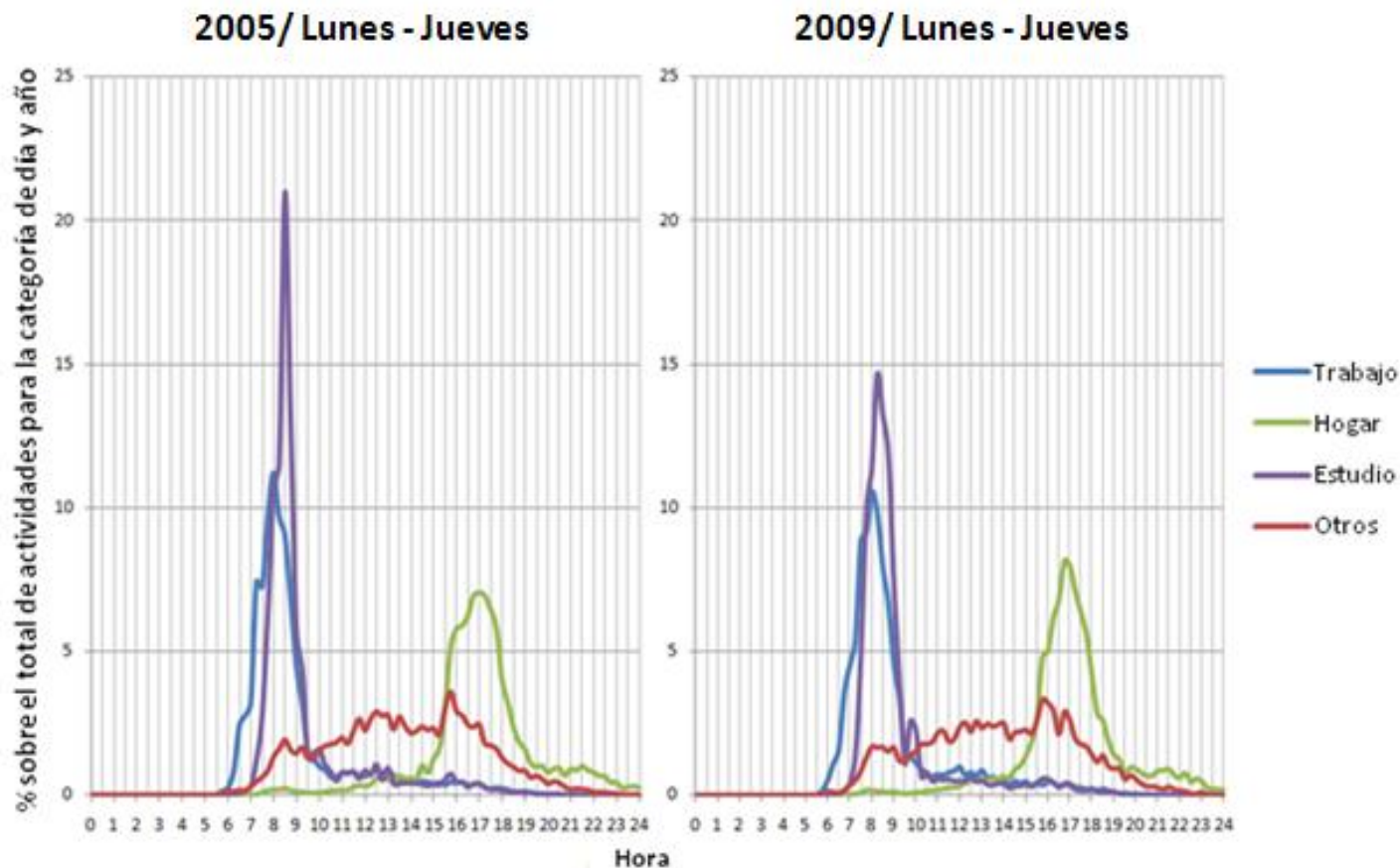


*Figura 5.6: Evolución de hora de comienzo de actividades laborales durante lunes-jueves en Gatineau para los años 2002, 2005 y 2009*  
*Fuente: elaboración propia*

Las causas de estos fenómenos observables pueden responder a distintas dinámicas sociales: cambios programados de horario, o bien sencillamente a la tendencia y necesidad

de un día laboral más largo, con miras a una mayor productividad. Se presenta a continuación un análisis similar, sólo considerando el año 2005 y 2009, para los 4 propósitos.

Se observan en la Figura 5.7 dinámicas marcadas en cuanto a horarios de comienzo de actividades. En el caso de actividades de estudio, se nota un descenso marcado en la magnitud del peak, desde un 21% el 2005 hasta un 15% en el 2009, además de un ligero adelantamiento en la hora de inicio desde las 8:30 en 2005 hasta las 8:15 en el 2009. Al igual que para el caso de las actividades laborales, se observa para el propósito estudio una tendencia a aumentar la desviación estándar de la distribución de comienzo de actividades entorno al peak, viéndose este mismo reducido. Se destaca además la aparición en el 2009 de un pequeño peak de estudio a las 10:00 correspondiente a la reprogramación de entrada a clases en un período particular del año. Las actividades laborales no sufren un cambio tan drástico en este par de años (la diferencia más observable se registra entre el 2002 y el 2005), al igual que las actividades de otros propósitos que mantienen su distribución más bien uniforme a lo largo del día. Las actividades del Hogar si observan una variación: un aumento en el peak, desde un 6,5% a un 7,5% manifestándose una tendencia por parte de los habitantes de Gatineau a llegar más tarde a sus hogares. Aunque esto recoge el promedio de todo el año, es muy probable que existan efectos estacionales debido a las fuertes diferencias climáticas observables en Gatineau. No obstante, la detección de este tipo de efectos requeriría una mayor desagregación de la información.



*Figura 5.7: Evolución de hora de comienzo de todas las actividades durante lunes-jueves en Gatineau para los años 2005 y 2009*  
*Fuente: elaboración propia*

### 5.3 Aplicación a Santiago: análisis cruzado día-actividad

La masividad de la base de datos de Santiago en término de número de transacciones por unidad de tiempo permite realizar desagregaciones y análisis que no son posibles con la base de datos de la ciudad de Gatineau. Una aplicación interesante en este aspecto es analizar cómo un determinado propósito sucede en los distintos días de la semana.

La Figura 5.8 muestra un detalle con las distribuciones de hora de inicio de cada propósito por día en la ciudad de Santiago. La Figura 5.8 (a) muestra el propósito estudio; este presenta una curva característica tanto en lunes-jueves como viernes de tres peaks: uno en la mañana, asociable a actividades estudiantiles principales y más frecuentes, que parten entre las 8:00 y las 10:00. Otro peak se observa alrededor de las 14:00, asociable a

actividades académicas posteriores al horario de almuerzo; finalmente, un tercer peak se registra entre las 18:30 y las 19:00 horas, correspondiente a actividades estudiantiles de la tarde y nocturnas. El día sábado se observa una distribución de actividades de estudio más concentrada en la mañana, con un peak más tardío alrededor de las 9:00 y menos pronunciado que los días de semana; el peak post- almuerzo se ve reemplazado por una distribución más uniforme entre las 11:00 y las 17:00, y el peak de la tarde no se registra.

La Figura 5.8 (b) presenta el propósito trabajo; en general se observan distribuciones de hora de inicio de actividades muy similares entre los días lunes-jueves y viernes, observándose en ambos casos un peak a las 8:30 horas con un 6% de las actividades de trabajo totales del día, el cual decae progresivamente hasta las 11:00 del día, momento en el cual se mantiene una generación de actividades laborales de un 1% sobre el total diario hasta las 15:00 horas. De 16:00 a 20:00 se observa otra distribución uniforme entorno a un 0,5% del total diario. Los días sábado difieren en términos de inicio de actividades laborales en que se observa un peak doble en la mañana: uno en torno a las 8:00 y otro alrededor de las 9:00; se observa también que no se genera una cantidad significativa de actividades laborales en forma posterior a las 17:00 horas, a diferencia de los días de semana. Los días domingo una distribución de inicio de actividades laborales más concentrada en la tarde: esta tiene su peak a las 9:00, y se ostenta una alta generación de actividades hasta alrededor de las 11:00. Entre 11:00 y 15:00 se genera proporcionalmente una mayor cantidad de actividades con respecto a los días de semana (alrededor de un 1,5% más por cuarto de hora), para finalmente decaer progresivamente hasta alrededor de las 17:00, en forma similar a la curva del día sábado.

La Figura 5.8 (c), que representa el propósito otros, revela que en general este tipo de actividades se producen siguiendo un patrón de ocurrencia uniforme a lo largo del día. Las curvas asociadas para este propósito en los días lunes-jueves, viernes y sábado son bastante similares, mostrando actividades desde las 7:00, llegando a su máximo a las 10:00 horas, en torno a un 2% de la generación diaria de actividades por cada cuarto de hora, cantidad que se mantiene uniforme hasta las 18:00 horas para el caso del día sábado y entorno a las 19:00-20:00 horas para el caso de los días de semana, decayendo progresivamente hasta las 24:00 horas. El día domingo presenta una distribución similar, pero más densa en horarios de la tarde: en la mañana se encuentra desplazada alrededor de una hora respecto al resto de los días de la semana, mientras que entre las 12:00 y 19:00 horas se observa una mayor generación de actividades con otros propósitos, en alrededor de un 0,75-1%.

Las actividades en el hogar, exhibidas en la Figura 5.8 (d), muestran distribuciones diferentes para cada tipo de día. Los días lunes-jueves presentan una cantidad considerable de actividades en el hogar a partir de las 14:00, en torno a un 1% del total diario; a partir de las 16:00 horas, la cantidad de actividades en el hogar por cuarto de hora (o llegadas al hogar) comienza a aumentar progresivamente, hasta alcanzar el máximo alrededor de las 19:30 horas con un 4,5%, para posteriormente declinar progresivamente hasta las 24:00

horas. El día viernes posee una distribución similar a la del día lunes jueves, pero con una concentración de actividades ligeramente superior entre las 14:00 y las 17:00 horas, y una concentración de peak menor en un 1%, esto entre las 18:00 y las 22:00 horas; no obstante se observan llegadas al hogar más tarde en el día viernes, desde las 22:00 horas en adelante. Respecto al día sábado, se observa una tendencia similar al día viernes, pero más acentuada: una cantidad aun mayor de llegadas al hogar entre las 14:00 y las 17:00 horas, una cantidad ligeramente menor pero constante (comparativamente con el día viernes) de llegadas entre las 17:00 y 22:00 horas, y una cantidad similar a las del día viernes entre 22:00 y 24:00 horas; adicionalmente se observa una cantidad considerable de llegadas entre las 00:00 y las 02:00 horas, correspondientes a retornos al hogar posteriores a actividades realizadas los días viernes en la noche. Los días domingo en general presentan una distribución similar a los días sábado, pero con una mayor concentración de llegadas al hogar entre las 14:00 y 17:00 horas y las 20:00 y 23:00 horas.

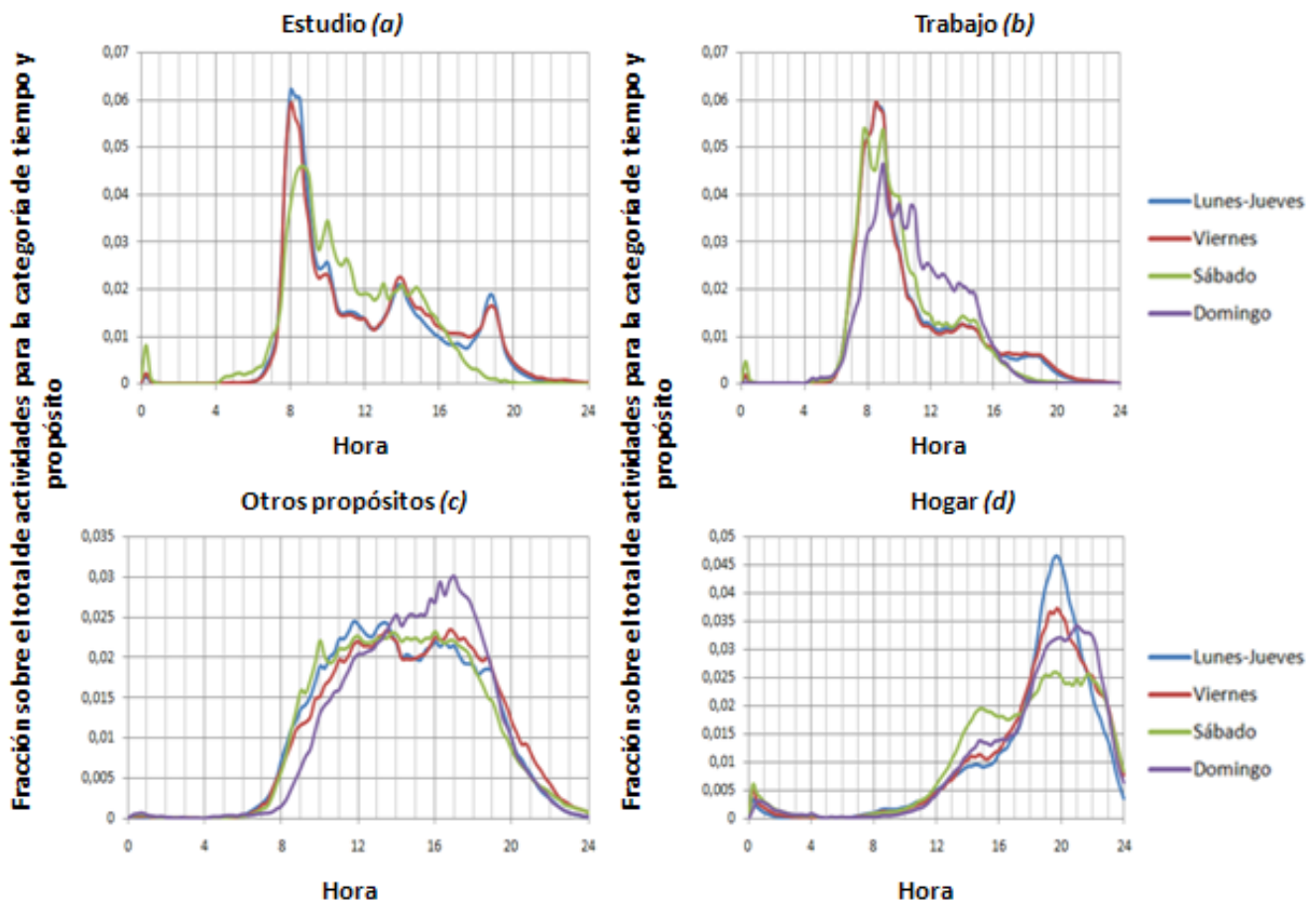


Figura 5.8: Hora de comienzo de actividades en los 4 tipos de días de la semana en Santiago, para los propósitos estudio (a), trabajo (b), otros (c) y hogar (d)

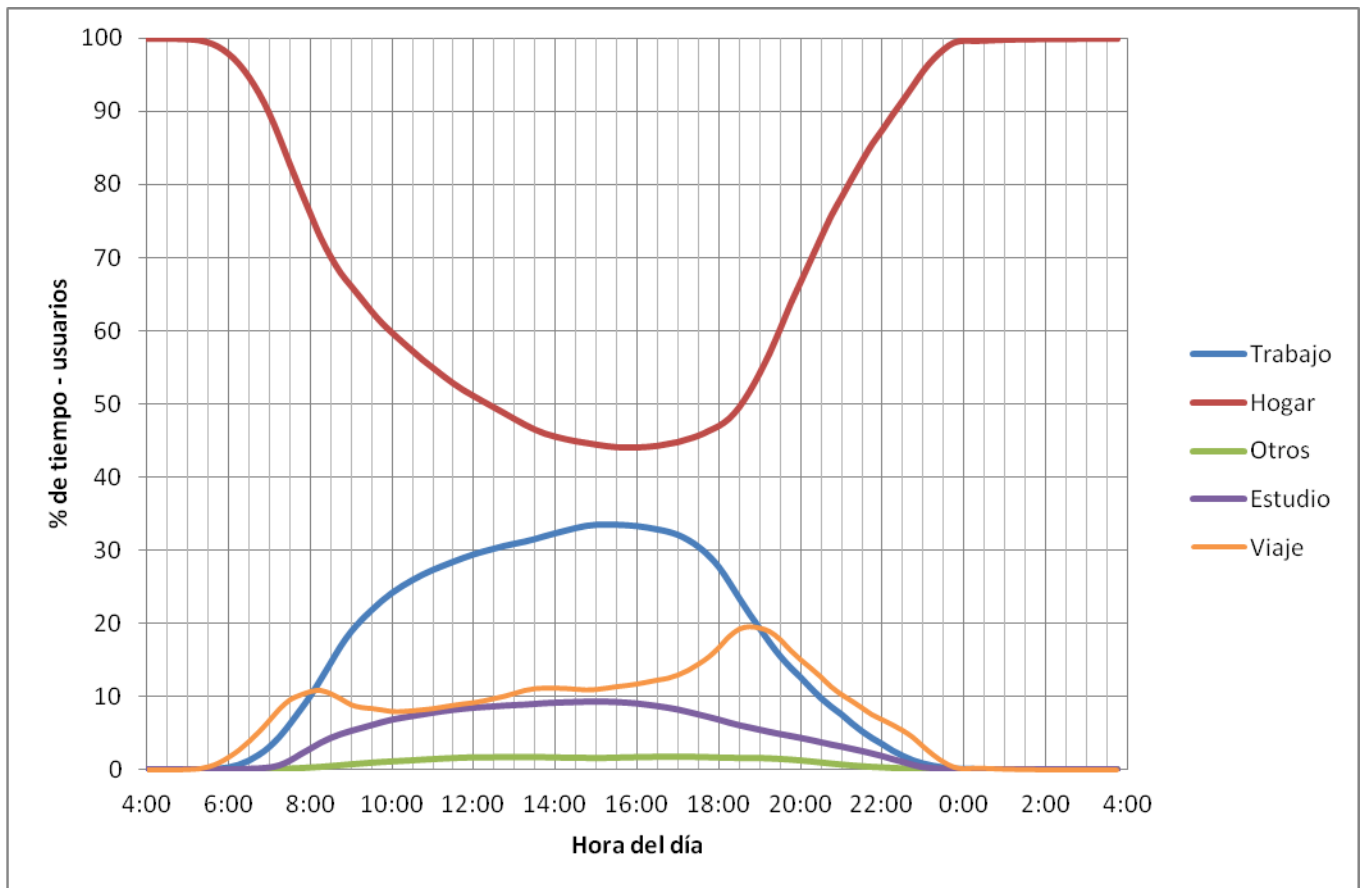
Fuente: elaboración propia

## 5.4 Aplicación a Santiago: perfiles de uso de tiempo

Resulta de gran interés para entender el comportamiento del usuario de transporte público, tanto como el mismo análisis sociológico a que da pie, estudiar la dinámica de uso de tiempo involucrada en el día a día. Tal investigación se hace posible desde el momento que en el presente trabajo se estima y caracteriza la ocurrencia de actividades y viajes en su dimensión temporal, incluyendo en particular propósito, hora de inicio y duración. Aparece además un nuevo uso de tiempo: Viaje.

La Figura 5.9 muestra un perfil de uso de tiempo bastante típico, que describe el porcentaje de asignación de uso de tiempo sobre el universo de usuarios de transporte público que utilizaron la tarjeta ese día, y poseen un día-id sin información faltante (esto es, una tarjeta que en un mismo día posee sólo viajes y etapas de viaje a los cuales se les ha podido estimar tanto subida como bajada). A las 4 se asume que todos los usuarios de transporte público se encuentran en su hogar. Es importante resaltar que la curva denominada hogar no sólo agrupa a tales actividades, sino también el conjunto de todas las actividades no detectables que el usuario pudiese haber desarrollado utilizando otros modos de transporte antes de haber efectuado su primera validación del día o en forma posterior a efectuar la última. Estas por supuesto no pueden ser detectadas debido a la falta de información.

A partir de las 5:00 se puede observar inflexiones en la curva de hogar, mientras que en forma simultánea comienzan a aparecer viajes. El mínimo asociado a la curva hogar, correspondiente con los máximos de las curvas trabajo y estudio (esto es, el momento del día en que la mayor cantidad de gente se encuentra fuera de su casa mayormente trabajando, viajando o realizando actividades de estudio) se encuentra contenido en el intervalo 15:00-16:00. A partir de este punto de inflexión, más gente comienza a salir del trabajo o estudio con respecto a las que inician actividades de estos tipos, y más retornan a sus casas con respecto a los que salen de ellas. Las horas punta mañana (7:00-9:00), punta medio día (12:30-14:30) y punta tarde (18:00-20:00) son claramente identificables, siendo esta última la más densa en términos de usuarios involucrados en viajes, con un peak cercano al 20%.



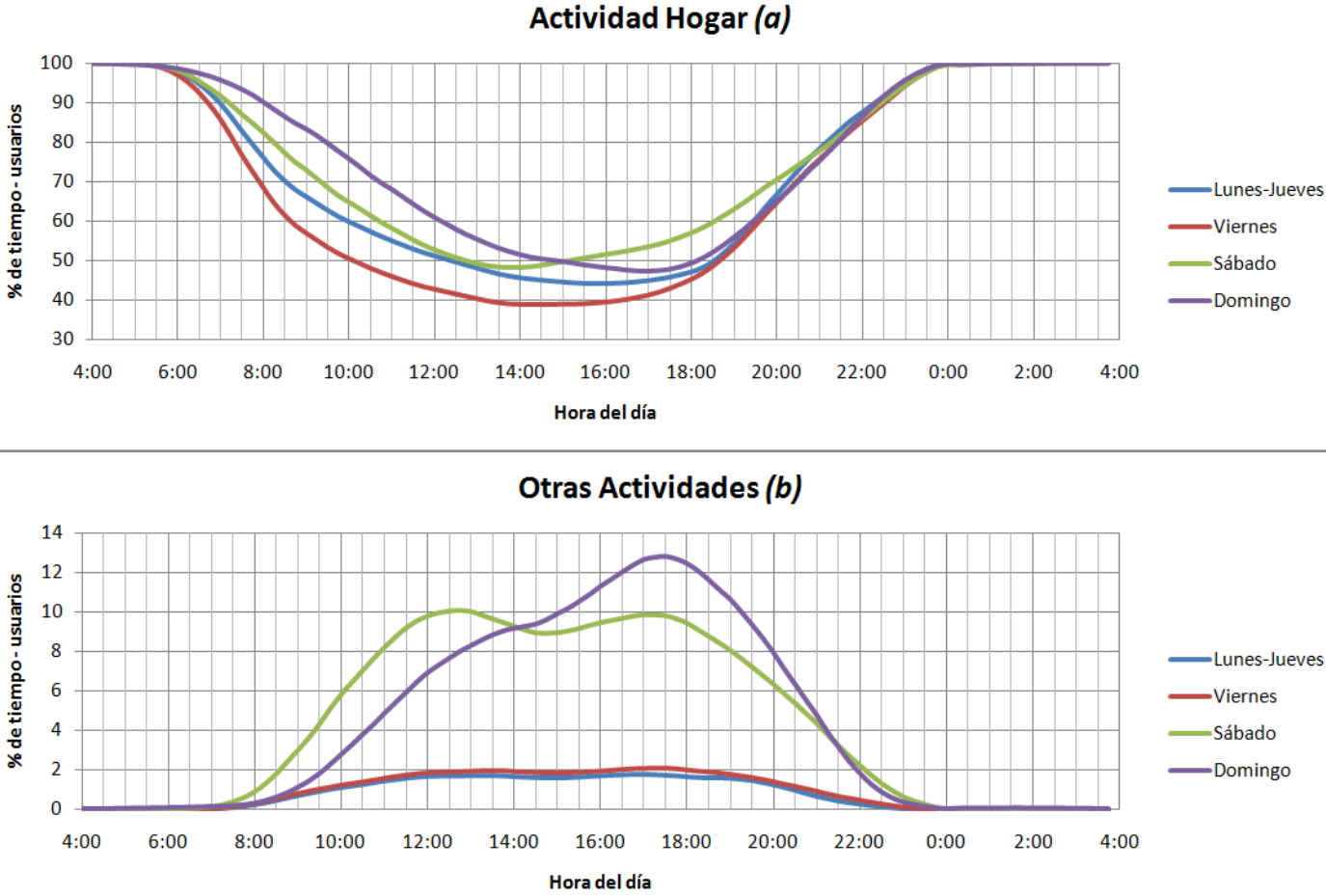
*Figura 5.9: Perfil de uso de tiempo para lunes-jueves en Santiago  
Fuente: elaboración propia*

Se muestra a continuación un detalle de cómo se comportan los perfiles de las actividades hogar (más actividades no detectables) y otros propósitos para los distintos días de la semana. En la Figura 5.10 (a) se observa que los perfiles de las actividades en el hogar siguen un patrón similar para los cuatro tipos de días, no obstante se observan algunas diferencias. Los días viernes el punto de inflexión de retorno al hogar sucede alrededor de las 14:30 horas, mientras que los días lunes-jueves este punto se encuentra en torno a las 16:00 horas, observándose por tanto una tendencia a regresar más temprano al hogar los días viernes. Los días sábado y domingo se caracterizan por tener menos salidas del hogar (los mínimos de los perfiles son un 10% mayores que en los días de semana); el punto de inflexión los días sábado es el que se sucede más temprano de todos los días de la semana, en torno a las 13:30-14:00 horas, mientras que la inflexión del día domingo es la más tardía, produciéndose alrededor de las 17:00-17:30 horas.

La Figura 5.10 (b) muestra los perfiles de otras actividades. En este caso, se observan claras diferencias tanto en la forma como en la magnitud de las distribuciones entre los días de



semana, sábados y domingos. Mientras los días de semana muestran distribuciones uniformes a lo largo del día y con magnitudes que no superan el 2%, el día sábado se observa una distribución semi-uniforme con dos peaks de un 10% a las 12:30 y a las 17:00; el día domingo por otro lado presenta una distribución más gaussiana, y con mayor concentración en horas de la tarde, registrándose un único peak de un 13% alrededor de las 17:30 horas.



*Figura 5.10: Análisis cruzado de uso de tiempo para cada día en Santiago, para los propósitos hogar (a) y otras actividades (b)*  
*Fuente: elaboración propia*

# Capítulo 6

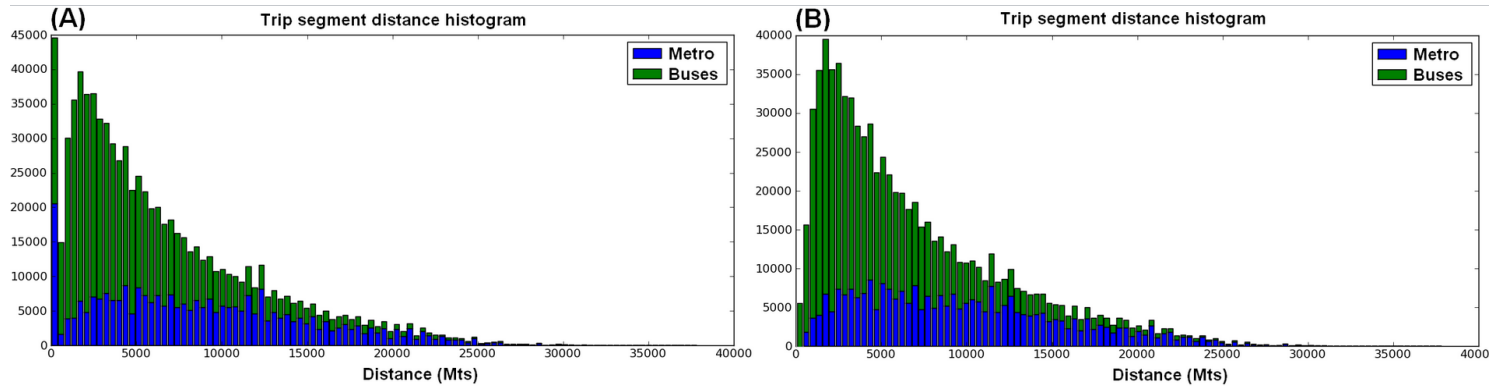
## Validación

La validación constituye un aspecto fundamental en el desarrollo de toda la metodología; sin esta etapa, no se puede tener credibilidad en cuanto a los resultados entregados. En el presente trabajo, se proponen dos formas de validación: endógena y exógena, cada una de las cuales apunta a detectar y también corregir, distintos aspectos de la metodología.

### 6.1 Validación endógena

Una forma de validar las estimaciones de bajada involucradas en la primera fase de la metodología (Munizaga et al. 2011, Trépanier et al 2007) corresponde a lo que se ha denominado en el presente trabajo como validación endógena. Este proceso implica realizar un análisis exploratorio de todas las variables derivadas de la estimación tanto de puntos de bajada como también de actividades, tales como tiempos de viaje, tiempos de etapas de viaje, distancias y velocidades. Se presentan a continuación ejemplos de la aplicación de este tipo de validación.

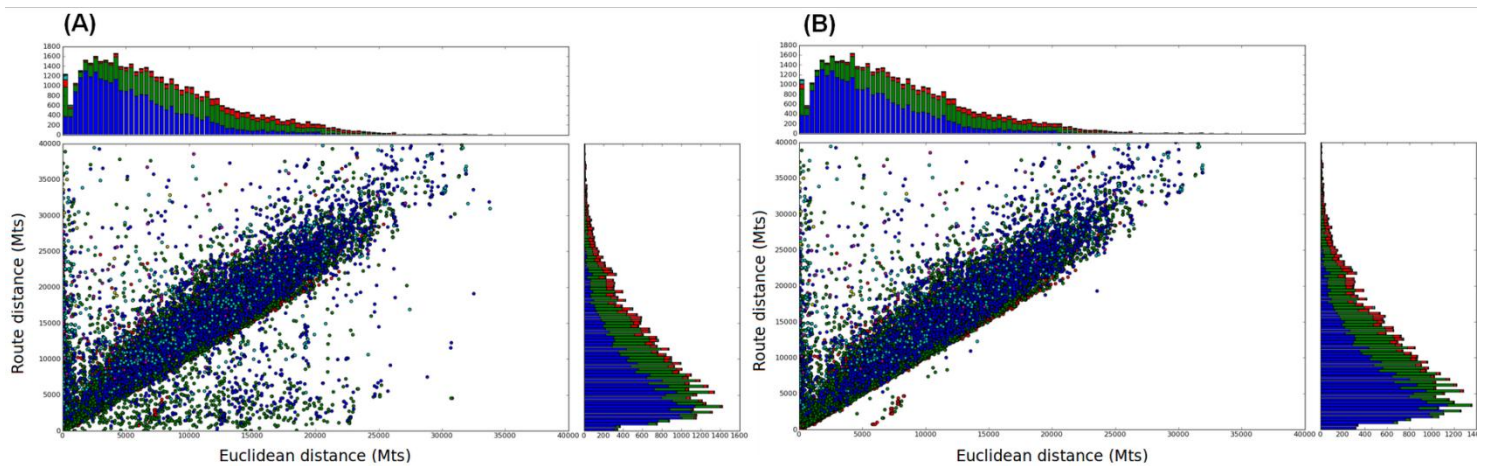
La Figura 6.1 (a) muestra el resultado de una primera versión del software donde se observó una cantidad anormal de etapas de viaje con distancias cero o muy bajas, que fueron analizadas con detalle y permitieron encontrar errores en la data de entrada (secuencias de paradas de bus, rutas de servicios de bus, etc.) y en la estimación de bajada. La Figura 6.1 (b) muestra los resultados después de llevar a cabo correcciones en el software respecto a los problema mencionados.



*Figura 6.1: Histogramas de distancias en etapas de viaje, antes (a) y después (b) de correcciones al software*  
*Fuente: Devillaine et al. (2012 b)*

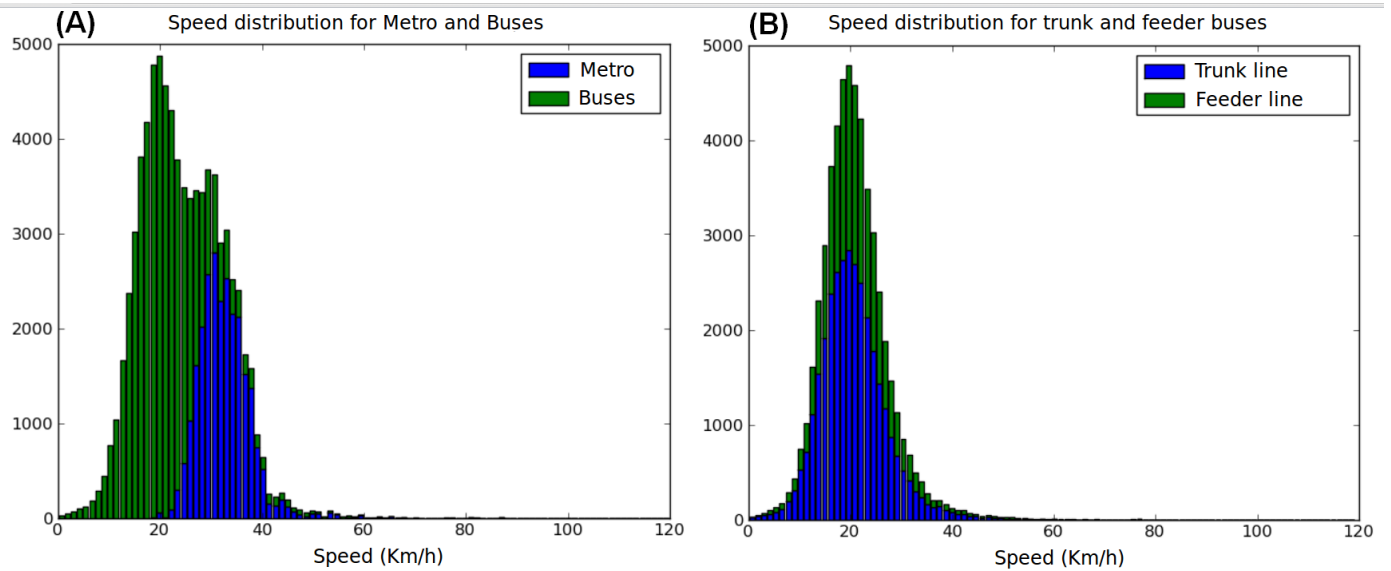
Otra variable interesante a considerar corresponde, como ya se ha mencionado en el capítulo 4, a las distancias euclidianas y en ruta de viaje. No solo sirven estas para mostrar dinámicas interesantes que metodológicamente pueden no estarse recogiendo, como fue el caso de emplearlas anteriormente, sino que también pueden ser utilizadas para detectar errores en los datos de entrada y pre-procesos asociados a las estimaciones de puntos de bajada.

La Figura 6.2 muestra nuevamente una situación de antes y después de correcciones al software y a los datos de entrada. En la figura 6.2 (a) se observa que una gran cantidad de estimaciones de viajes poseen una distancia euclidiana mayor que la distancia en ruta, lo cual es claramente una imposibilidad matemática. Analizando estos casos particulares, se encontró nuevamente una serie de problemas en la codificación de las rutas de transporte público que hacían contar ciertas etapas con distancias negativas, lo cual fue corregido. Posteriormente a estas correcciones, se obtienen los resultados expuestos en la Figura 6.2 (b), en el cual si bien todavía persisten ciertos casos de anormalidad geométrica, estos se redujeron en más de un 95%, y los restantes han sido identificados como un recorrido específico de buses que presenta problemas con las emisiones de GPS. Los colores representan viajes según cantidad de etapas (azul= 1, verde=2, rojo=3, cian=4 o más).



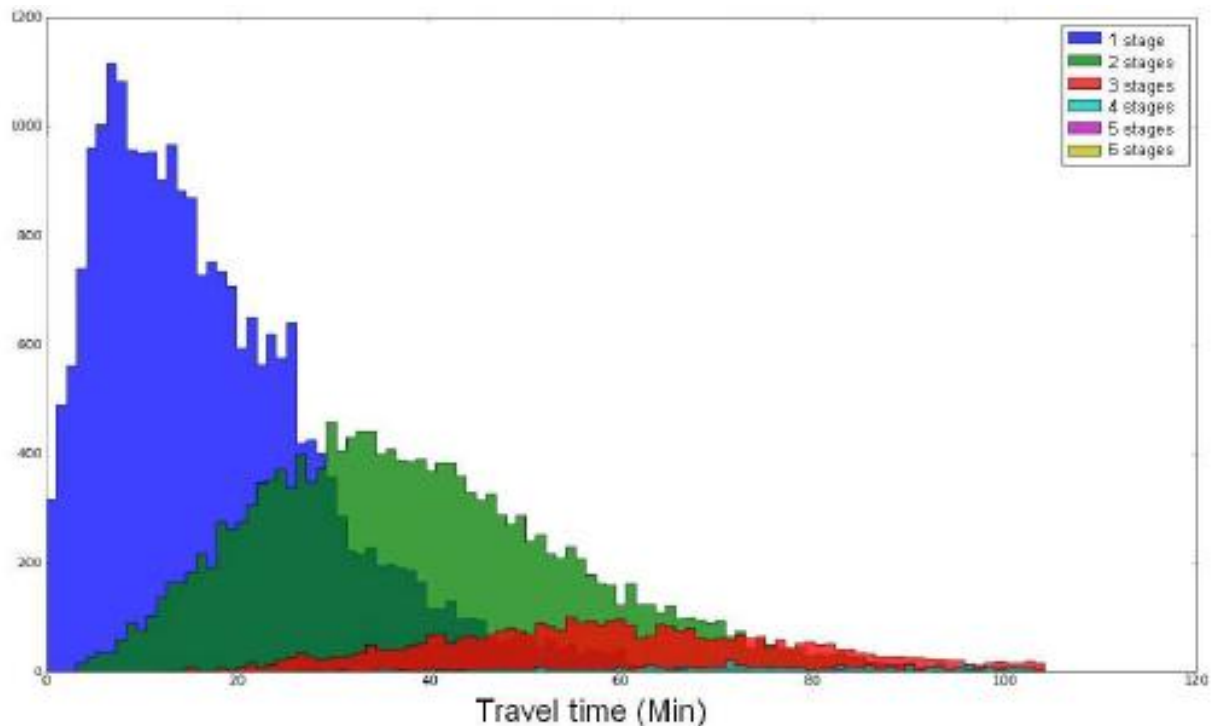
*Figura 6.2: Gráficos de distancia euclidiana vs. distancia en ruta de viajes antes (a) y después (b) de depuración de errores de datos y modificaciones al software*  
*Fuente: Devillaine et al. (2012 b)*

La distribución de velocidades es otra variable interesante a observar de cara a una validación endógena. Esta es directamente calculable en la medida de que para cada etapa se dispone de la distancia en ruta y también de las horas de subida y de bajada, y por tanto del tiempo de cada etapa. La Figura 6.3 muestra las distribuciones de velocidades en etapas desarrolladas en buses y metro (Figura 6.3 a) y un desglose de velocidades en buses troncales y alimentadores (Figura 6.3 b). Se observa en ambos casos velocidades con distribución normal, centradas en 20 km/h para el caso de buses y alrededor de 30 km/h para el caso del metro. No se observan anomalías en los resultados exhibidos (velocidades mayores a la correspondiente a la circulación normal de bus o metro).



*Figura 6.3: Distribuciones de velocidades comerciales para Metro y buses (a) y para el desglose entre buses troncales y alimentadores (b)  
Fuente: Devillaine et al. (2012 b)*

La Figura 6.4 permite observar que los viajes de una etapa son más cortos y tienen una menor dispersión que los de dos etapas, los cuales a su vez presentan menor duración y dispersión que los de 3 etapas. Este gráfico permite además observar un clúster de viajes de una etapa con tiempo de viaje cero o cercano a cero, que será corregido en el futuro.



*Figura 6.4: Histogramas de tiempos de viaje de acuerdo al número de etapas*  
*Fuente: Devillaine et al. (2012 b)*

## 6.2 Validación exógena

La validación exógena consiste en comparar los resultados metodológicos con otras fuentes de información que permitan un chequeo cruzado de las variables calculadas y post-procesos (principalmente estimación de paraderos de bajada y estimación de viajes y actividades). En forma específica, se utilizó una encuesta Origen/Destino tomada la semana anterior al horizonte temporal de la base de datos de Santiago estudiada, realizada a 7.266 usuarios de la red de Metro. En esta encuesta, cada uno de los pasajeros encuestados es requerido para responder detalles específicos respecto al viaje en curso: intersecciones para sus puntos de origen y destino, estación de metro en el cual hace ingreso a la red, número de etapas del viaje y los modos en que cada una se desarrolló, estación de metro de salida, además de otros datos asociados al ingreso familiar y el ID de la tarjeta, estas últimas para una parte de los encuestados (aproximadamente un 30%). Aunque una validación definitiva se lograría si se comparase la misma semana tanto en términos de la base de datos como de la encuesta, al momento de la realización de este trabajo esa opción no era posible; no obstante en la actualidad se trabaja en procesar a través de la metodología las semanas correspondientes a la franja temporal de la toma de encuesta con el mayor número de ID's, de forma tal de realizar finalmente la validación con una coincidencia perfecta.

No obstante como ya se ha explicado el match de días entre la encuesta y la semana procesada no es perfecto, la suposición de que el principal viaje diario (esto es, el viaje correspondiente a los propósitos trabajo o estudio, desde y hacia el hogar) permanece igual puede sostenerse, considerando que el desfase es de sólo una semana. Un primer resultado obtenible en base a esta comparación exploratoria es que la metodología en general tiende a predecir con precisión el destino declarado por el encuestado: cerca de un 70% de los datos analizados de esta forma arrojan coincidencias entre el destino declarado y el estimado por la metodología, utilizando un radio de error tolerable de 500 metros. Otros propósitos no puede ser comparado en esta ocasión debido a que la naturaleza de este tipo de actividades no es de periodicidad diaria y por tanto se requiere de un match perfecto. Otro interesante resultado apunta a que las etapas en las cuales los usuarios tomaron modos de transporte no rastreables a través de tarjetas inteligentes (como buses interurbanos o taxis), o bien sencillamente efectuaron una caminata larga (mayor a 1000 metros), son considerablemente relevantes: de los 7266 encuestados, el 21% de ellos declararon haber hecho al menos una etapa de su viaje en un modo no rastreable con tarjetas inteligentes. En términos de la metodología, una etapa efectuada en este tipo de modos presenta el mismo problema que una etapa evadida (Devilleine et al. 2012 b).

## Capítulo 7

# Conclusiones y líneas de investigación futura

Se ha propuesto en el presente trabajo una metodología para estimar viajes y actividades en base a sistemas tecnológicos de cobro automático de transporte público y a la literatura preexistente sobre estimación de paraderos de bajada para dichos sistemas. En forma posterior, se han mostrado diversas aplicaciones de la metodología para las ciudades de Santiago, Chile y Gatineau, Canadá.

La información contenida en estas bases de datos contiene valiosa información subyacente a los usuarios mismos, y presenta enormes potencialidades, proveyendo de una fuente continua y masiva de datos. La metodología elaborada en el presente trabajo ha sido desarrollada para incluir no sólo literatura referente a utilización de datos pasivos de transporte público y uso de tiempo, sino también datos exógenos disponibles al momento de la realización del trabajo, como es el caso del uso de suelo en Gatineau. Los resultados obtenidos ayudan a entender el comportamiento general de la población en distintas dimensiones (temporal, dinámica de viajes en la ciudad, propósitos y perfiles de actividades, etc.), y revelan profundas diferencias entre los comportamientos de las ciudades de Santiago y Gatineau: los horarios de realización de actividades son distintos, como también lo es el horario diurno laboral en general, funcionando el trabajo en la ciudad canadiense mayormente hasta las 16:00 horas y llegando la gente a su hogar alrededor de las 16:30-17:00, mientras que en la ciudad chilena las actividades laborales comienzan a declinar entorno a las 19:00, y la hora de llegada al hogar es alrededor de las 19:30-20:00. Se han ilustrado en todas las aplicaciones parte de la data e información posible de recabar, principalmente por razones de extensión, sin embargo la riqueza de las bases de datos hace posible un grado elevado de desagregación temporal y espacial, a raíz del cual la mayoría de los cruces posibles entre días, propósitos, ciudades, viajes y etapas son factibles.



Los criterios de identificación presentes en el Módulo de Estimación de Viajes y Actividades (AIM) constituyen un aporte valioso a las formas previas de identificación existentes, correspondientes a las utilizadas en Munizaga y Palma (2012). A saber:

- Criterio de los 30 minutos
- Criterio de las 2 transacciones consecutivas en metro
- Criterio de las 2 transacciones consecutivas en el mismo recorrido de bus
- Criterio de la distancia euclidiana vs. distancia en ruta
- Criterio del bus no abordado

En particular, los patrones de distancia euclidiana y distancia en ruta, que aborda la problemática del encadenamiento de viaje, y el patrón del bus no abordado, que añade información adicional al proceso de estimación, referente a la frecuencia observada e inferencia a partir de esta de la presencia del individuo en el paradero, constituyen un avance en estimación de actividades y viajes.

Los criterios utilizados en el modelo para el Módulo de Asignación de Propósitos, basados en la interpretación de las ventanas de tiempo de las actividades, y del uso de suelo para el caso de la ciudad de Gatineau, constituyen el primer hito metodológico en estimar actividades con propósitos en base a un sistema tecnológico de cobro de transporte público, y han sido elaborados considerando el criterio de los modeladores y también literatura referente a uso de tiempo. Estos se podrán ir refinando y mejorando en la medida que exista una mayor cantidad de formas de validar y observar exógenamente el comportamiento de uso de tiempo de la población, ya sea a través de datos pasivos o encuestas. Igualmente, la inclusión de información adicional referente a uso de suelos para el caso de Santiago sería una contribución valiosa al mejoramiento de la metodología.

Los aportes de la metodología propuesta hacen principalmente referencia a la obtención de información valiosa de viajes en transporte público en base a información pasiva; el aprovechamiento de esta información apunta en gran medida a complementar y eventualmente reemplazar mayormente la principal forma alternativa de medición: las encuestas. Estas poseen numerosos sesgos referentes a respuestas y muestreos, tienen un elevado costo, representan una población mucho más pequeña y describen solamente el momento o corte temporal en el cual se desarrolló la toma de la misma. Los datos pasivos por otro lado, aportan datos masivos y dinámicos en el tiempo a un menor costo y con menores errores y sesgos, no obstante poseen la desventaja de no poder caracterizar demográficamente al usuario. Es en este aspecto que las encuestas pueden ser utilizadas como un complemento.

La obtención de Matrices Origen/Destino a partir de la metodología presentada es directa, en la medida que el formato entregado para las salidas de viajes y actividades posee

localización detallada. Más aún, la caracterización de cada viaje con el propósito que lo motiva permite la elaboración de Matrices Origen/Destino de propósitos, y con el grado de desagregación temporal que se desee. Esto representa una herramienta de planificación muy valiosa para las autoridades de transporte público, ya que contarían con una matriz Origen/Destino representativa de la totalidad de los usuarios del sistema, y dinámica puesto que las posibilidades computacionales al momento de desarrollar este trabajo permiten obtener a lo menos una serie de matrices por propósitos al mes, en lugar de una cada varios años, como es usual cuando se obtienen de encuestas tradicionales.

La validación endógena demuestra ser una herramienta útil de cara a la depuración del software asociado a la metodología y a la detección de errores atribuibles a la data de entrada; la validación exógena es en definitiva la vía para ratificar que los postulados de la metodología representan, con un margen de error aceptable, la realidad del usuario de transporte público. No obstante lo anterior, una validación exógena definitiva se debe realizar en el futuro, cuando un match perfecto entre los registros de las bases de datos utilizadas y las encuestas sea realizable.

Una línea de investigación que no ha sido mayormente tocada en el presente trabajo, pero que sin embargo constituye un campo muy interesante de explorar, lo constituye la dimensión espacial de los viajes y actividades. Un análisis profundo en este tópico podría revelar hallazgos importantes en las áreas de localización de actividades, además de una valiosa descripción de la componente espacial de la demanda de transporte y de servicios, y la consecuente importancia que esto posee para la autoridad planificadora de transporte público y autoridades comunales.

# Referencias Bibliográficas

- Agard, B, Morency, C, Trépanier, M (2009) “Mining Smart Card Data from an Urban Transit Network”, IGI Global, 1-11.
- Bagchi, M, White, PR (2005) “The Potential of Public Transport Smart Card Data”, Transport Policy 12, 464-474.
- Beltrán, P, Cortes, C, Gschwender, A, Ibarra, R, Munizaga, M, Ortega, M, Palma, C, Zúñiga, M (2011) “Obtención de información valiosa a partir de datos de Transantiago”, XV Congreso Chileno de Ingeniería de Transporte.
- Chriqui, C, Robillard, P (1975) “Common Bus Line”, Transportation Science 9, 115-121.
- Chu, KKA, Chapleau, R (2008) “Enriching Archived Smart Card Data for Transit Demand Modeling”, Transportation Research Record 2063/2008, 63-72.
- Chu, KKA, Chapleau, R (2010) “Augmenting Transit Trip Characterization and Travel Behavior Comprehension”, Transportation Research Record 2183/2010, 29-40.
- Devillaine, F, Munizaga, M, Trépanier, M (2012a) “Detection of Public Transport User Activities Through the Analysis of Smartcard Data”, Transportation Research Record 2012 (en imprenta).
- Devillaine, F, Munizaga, M, Palma, C, Zúñiga, M (2012b) “Obtaining a Reliable Origin Destination Matrix From Massive Smartcard Data: Application to Santiago”, aceptado para publicación en el libro de la conferencia ISCTSC 2011.
- Díaz, G, Gómez-Lobo, A, Velasco, A (2004) “Micros en Santiago: de enemigo público a servicio público”, Revista Estudios Públicos CEP Chile n° 96.
- Finkenzeller, K (2003) “RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification”, John Wiley & Sons, 2da edición.

- Hofmann, M, O'Mahony, M (2005) "Transfer Journey Identification and Analyses form Electronic Fare Collection Data", Proceedings of the 8<sup>th</sup> International IEEE Conference on Intelligent Transportation Systems, Vienna, Austria, September 13-16, 2005.
- Jara-Díaz, SR, Munizaga, M, Greeven, P, Guerra, R, Axhausen, K (2008) "Estimating the value of leisure from a time allocation model". *Transportation Research B* 42, 946-957.
- Lianfu, Z, Shuzhi, Z, Yonggang, Z, Ziyin, Z (2007) "Study on the Method of Constructing Bus Stops OD Matrix Based on IC Card Data", *Wireless Communications, Networking and Mobile Computing WiCom 2007*, 3147-3150.
- Ministerio de Vivienda y Urbanismo (2009) "Manual de vialidad urbana: Recomendaciones para el diseño de elementos de infraestructura vial urbana (REDEVU)", disponible en <http://www.minvu.cl>
- 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 smart card data from Santiago, Chile", accepted for publication at *Transportation Research C* 24, 9-18.
- Muñoz, JC, Gschwender, A (2008) "Transantiago: A Tale of Two Cities", *Research in Transport Economics* 22, 45-53.
- Olguín, J, Jara-Díaz, S, Munizaga, M (2009) "Análisis de Patrones de Actividades a partir de la EOD 2001", *Ingeniería de Transporte* V13 N4, 31-38.
- Park, JY, Kim, DJ, (2008) "The potential of using the smart card data to define the use of public transit in Seoul", *Transportation Research Record No. 2063*, Transportation Research Board of the National Academies, Washington, DC, pp. 3-9.
- Pirou, C (2008) "Caractérisation des Lieux d'Activités à Partir de Données de Cartes à Puce de Transport Collectif", Memoria presentada para la obtención del título de magíster en ciencias aplicadas en ingeniería industrial, Ecóle Polytechnique de Montréal, Abril de 2008.

- Sectra (2001) “Encuestas de origen y destino de viajes del gran Santiago 2001”, [http://www.sectra.gob.cl/Datos\\_e\\_Informacion\\_Espacial/Gran\\_Santiago/encuestas\\_movilidad.html](http://www.sectra.gob.cl/Datos_e_Informacion_Espacial/Gran_Santiago/encuestas_movilidad.html)
- Transantiago (2011) “Usos de tarjeta Bip informe generación tablas consulta”, versión web, [http://www.coordinaciontransantiago.cl/corporativo/index.php?option=com\\_content&view=article&id=136&Itemid=29](http://www.coordinaciontransantiago.cl/corporativo/index.php?option=com_content&view=article&id=136&Itemid=29).
- Trépanier, M, Barj, S, Dufour, C, Poilpré, R. (2004) “Examination of the potential use of smart card fare collection system in urban transportation”, Congr s annuel de 2004 de l’Association des transports du Canada, Qu bec.
- Tr panier, 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.
- Tr panier, M, Morency, C (2010) “Assessing Transit Loyalty With Smart Card Data”, Presentado en la 12<sup>th</sup> World Conference on Transport Research, Lisboa, 2010.
- Utsonomiya, M, Attanucci, J, Wilson, N (2007) “Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning”, Transportation Research Record n  1971, 119-126.
- Zhao, J, Rahbee, A (2007) “Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems”, Computer-Aided Civil and Infrastructure Engineering 22, 376-387.

**Anexo A:**

**CD con software Python para la implementación de la metodología**