

VALIDACIÓN DE ESTIMACIONES HECHAS A PARTIR DE TRANSACCIONES BIP!

Marcela Munizaga, Universidad de Chile, mamuniza@ing.uchile.cl

Flavio Devillaine, SUBUS, fmdevilla@gmail.com

Claudio Navarrete, SECTRA, claudionavarrete@gmail.com

Diego Silva, DTPM, diego.i.silva.lopez@gmail.com

RESUMEN

Se presenta la validación de matrices Origen-Destino de viajes en transporte público obtenidas de transacciones bip! y GPS de buses. Estas matrices, constituyen una valiosa fuente de información para gestión y planificación, pero hasta ahora no estaban validadas. En este trabajo se verifica los supuestos y resultados del método usando tres fuentes de información: la misma base de datos usada para realizar estimaciones, una EOD de viajes en Metro, en que se registra el número de tarjeta para un grupo de usuarios, y una muestra de voluntarios. Los resultados son muy positivos, obteniéndose porcentajes de estimación correcta cercanos al 90%.

Palabras clave: tarjeta de pago, matrices OD, transporte público

ABSTRACT

We present the validation of public transport OD matrices obtained from smartcard and gps data. These matrices are very valuable for management and planning, but until now, they were not validated. In this work we verify the assumptions and results of the method using three sources of information: the same database used to make the estimations, a Metro OD survey, where the card number is registered for a group of users, and a sample of volunteers. The results are very positive, with percentages of correct estimation around 90% in all cases.

Keywords: smartcard data, OD matrices, public transport

1. INTRODUCCIÓN

Obtener información de la estructura de viajes a partir de datos de transacciones es una tendencia creciente a nivel mundial. Muchos investigadores han visto la oportunidad de obtener información de alta calidad a bajo costo, y han desarrollado herramientas para procesar los datos generados automáticamente por los dispositivos tecnológicos que operan en los sistemas de transporte. Sin embargo, los mismos autores reconocen que la estimación de paradero de bajada, destino de viaje, ruta y actividades a partir de datos de transacciones está sujeta a muchos supuestos, y debe ser validada. En este trabajo se presenta un proceso de validación para las metodologías de estimación de paradero de subida, paradero de bajada y elección de ruta en el caso de Transantiago (sistema de transporte público de Santiago, Chile), realizado a partir de tres fuentes de información: información de la misma base de datos utilizada para hacer las estimaciones (validación endógena), información de una encuesta origen-destino (EOD) aplicada a 300.000 usuarios de Metro (donde una submuestra contiene el número de tarjeta bip! del usuario que contestó la encuesta) y entrevistas personales realizadas a una pequeña muestra de voluntarios que entregaron información detallada y que permite contrastar sus respuestas con las estimaciones que el modelo hace para ellos. La validación endógena permite detectar errores y mejorar la metodología. La EOD general de Metro contiene información que se puede utilizar para validar la elección de ruta dentro de la red, pero además, para el sub-conjunto de observaciones en que se registró el número identificador de la tarjeta, se puede estudiar los supuestos más fuertes del modelo, como son la estimación de paradero de bajada y estimación de propósito de viaje. La muestra de voluntarios permite todo lo anterior y además explorar otras líneas de investigación exploratoria, como la estimación de uso de tiempo y zona de residencia de los usuarios. A continuación se sintetiza los aspectos más relevantes de la metodología a ser validada, cuyo detalle se encuentra en las publicaciones citadas. Luego se presenta la validación con las distintas fuentes de información. Finalmente se presenta algunas conclusiones.

Descripción de la metodología a validar

Munizaga y Palma (2012) proponen un método para observar transacciones en el sistema de transporte público y estimar la secuencia de viajes, basado en la secuencia de transacciones de una misma tarjeta. Dado el sistema de pago, sólo se observan transacciones de subida. Los principales supuestos que hacen Munizaga y Palma (2012) son:

- Cada etapa de viaje se inicia en el instante y lugar en que ocurre la validación.
- El fin de una etapa de viaje se encuentra en la parada o estación más conveniente para alcanzar la posición de la siguiente transacción. Esta será la más cercana en el caso de Metro, o la que minimiza tiempo generalizado (función que suma el tiempo de viaje y el tiempo de caminata ponderados de acuerdo a las preferencias de los usuarios) para las transacciones en bus y en zona paga. En este último caso, se considera líneas comunes para identificar los posibles servicios. Al interior de la red de Metro, se supone que la elección es determinística (mínimo tiempo de viaje).
- Los viajes se definen como secuencias de etapas con menos de 30 minutos entre el fin de una etapa y el comienzo de la siguiente, y sin transacciones consecutivas en Metro o en la misma línea de buses.

Usando estos supuestos, se estima paradero de bajada para más del 80% de las transacciones. Las matrices OD lucen razonables, siendo mucho más densas que las obtenidas a partir de encuestas. Sin embargo, Munizaga y Palma (2012) reconocen que se debe usar métodos más sofisticados para identificar viajes y etapas.

La correcta identificación de viajes y etapas es crucial para obtener matrices OD confiables. Los orígenes y destinos son lugares donde los usuarios realizan actividades para satisfacer necesidades, mientras que los transbordos son sólo la consecuencia de la interacción entre la red de transporte y esas necesidades de los usuarios. Siguiendo el análisis realizado por Devillaine et al. (2013), exploramos esta diferenciación y proponemos nuevas reglas. Estas nuevas reglas son evaluadas utilizando las muestras de validación exógena.

Utilizando los resultados de Munizaga y Palma (2012), Devillaine et al. (2012) propone un método para estimar ubicación, duración y propósito de actividades, que también se aplica a la base de datos del sistema de transporte público de la localidad de Gatineau (Provincia de Quebec, Canadá), utilizando los resultados de Trépanier et al. (2007). La información detallada de la muestra de voluntarios permite explorar la validez de esos supuestos para el caso de Santiago.

3. Validación endógena

Siguiendo la línea de Devillaine et al. (2013) se realiza una validación endógena, es decir, se analiza la información de la misma base de datos para analizar la validez de los supuestos y detectar comportamientos anómalos, con el objetivo final de proponer mejoras metodológicas.

3.1. Estimación de paradero de bajada

3.1.1 Distancia de caminata

Munizaga y Palma (2012) proponen usar 1.000m como el límite de distancia caminable, es decir, la búsqueda de una posición-hora estimada de bajada se realiza dentro de ese umbral. Si la posición de la siguiente transacción está a más de 1.000m de la ruta del bus o Metro según corresponda, se asume que hay una etapa faltante, lo cual puede deberse a que se utilizó otro modo de viaje (taxi por ejemplo) o la no validación (evasión). Este caso queda codificado como “Distancia no caminable” y es la causa más frecuente de falla en la estimación (sobre 7% del total de transacciones). Exploramos la sensibilidad de ese parámetro, aplicando el mismo procedimiento anterior, pero sin la restricción de distancia máxima, lo cual arroja estimaciones de paradero de bajada distantes a más de 1.000m de la posición de la siguiente transacción. Sorprendentemente, encontramos que la distribución de esos casos no era homogénea en la ciudad, los barrios con más actividad comercial, como en centro de la ciudad, presentaban una mayor concentración de conexiones de más de 1.000m. Analizando los casos en detalle, encontramos casos en que dado que la red de transporte público es más densa, los usuarios pueden llegar al centro por un eje en la mañana, y dejar el centro tomando bus/metro en otro sector en la tarde (evitando los tramos más congestionados de la red). Esto frecuentemente implica distancias mayores a 1.000m entre el paradero de bajada en la mañana, y el de subida en la tarde. En consecuencia, sugerimos utilizar distintos parámetros para áreas comerciales y residenciales, calibrando el parámetro d para cada caso.

3.1.2. Primera transacción del día

La metodología propuesta por Munizaga y Palma (2012) utiliza la primera transacción del día para estimar paradero de bajada de la última transacción de ese mismo día, asumiendo que hay un ciclo que se inicia y termina en el mismo punto, presumiblemente el hogar del usuario. Si este procedimiento falla, el método toma la primera transacción del día siguiente. Por default, los días comienzan a las 0:00 y terminan a las 23:59. Sin embargo, al mirar la distribución temporal de las transacciones, mostrada en la Figura 1, se observa que usualmente a la medianoche aún hay cierto nivel de actividad que corresponde más bien al fin del ciclo del día previo que al inicio del ciclo del día siguiente. En consecuencia, proponemos redefinir el punto donde termina un día y comienza el siguiente como aquel en que el número de transacciones es mínimo. En nuestros datos, eso ocurre alrededor de las 4:00AM.

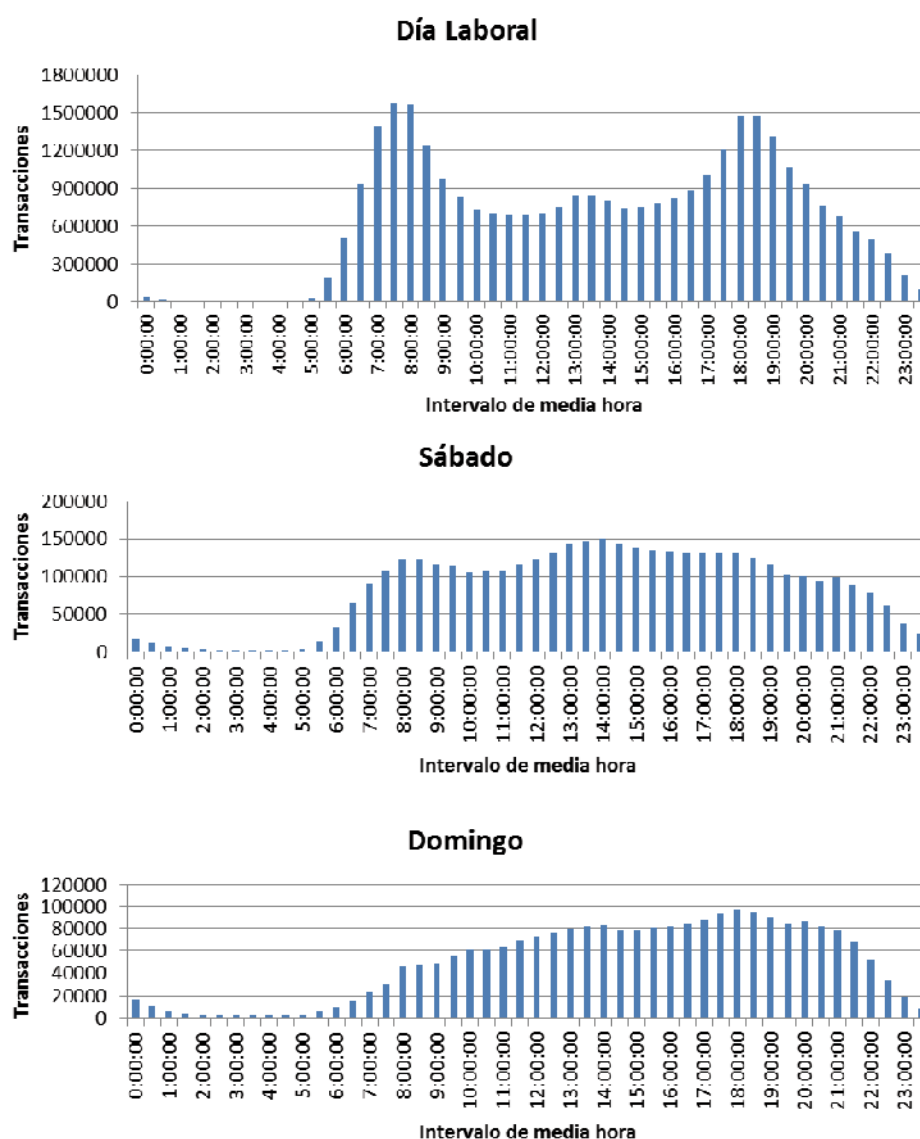


Figura 1: distribución temporal de las transacciones

3.1.3. Transacción única del día

La segunda causa más relevante de falla en el procedimiento de estimación de bajada es la presencia de tarjetas que se observan sólo una vez en un día en particular, lo cual es el caso en el 5% del total de transacciones. Al observar la distribución temporal de las transacciones únicas del día (Figura 2), se puede ver que un porcentaje importante de ellas ocurre en el periodo fuera de punta tarde y punta tarde. Esto sugiere la posibilidad de estimar paradero de bajada usando la primera transacción del día siguiente para cerrar el ciclo. En nuestro caso eso fue posible para 7% de las transacciones únicas del día.

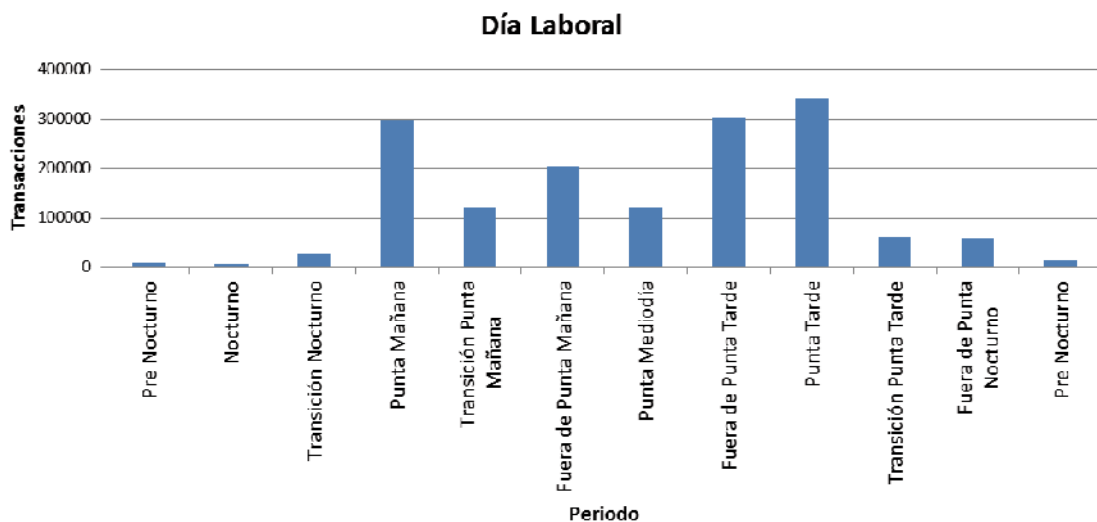


Figura 2: distribución temporal de las transacciones únicas del día

3.2. Identificación de las etapas de viaje

Munizaga y Palma (2012) proponen usar dos reglas simples para identificar un destino donde el usuario accede a realizar una actividad. La primera es la regla de la media hora, criterio que indica que si la diferencia entre el instante de bajada (estimado) de una etapa de viaje, y el instante en que se realiza la siguiente transacción con la misma tarjeta es de 30 minutos o más, entonces se supone que en ese lapso de tiempo se realizó una actividad. Consecuentemente, la etapa previa a esa actividad es etiquetada como fin de viaje, y queda separada de la etapa siguiente, que será el inicio de otro viaje. El segundo criterio propuesto por Munizaga y Palma (2012) se activa si una tarjeta (usuario) realiza dos transacciones consecutivas en un mismo servicio de buses (mismo sentido o sentido contrario) o ambas en Metro, este criterio se aplica en forma independiente del tiempo transcurrido entre la bajada estimada y la siguiente transacción. El fundamento de este criterio es que es muy poco probable que un usuario baje de un bus y se vuelva a subir a un bus del mismo servicio como parte de una estrategia de viaje. Lo más probable es que se trate de un punto en el cual el usuario va a realizar una actividad que le reporta utilidad.

Distancia en ruta/distancia euclidiana

Para verificar la validez de estos supuestos, analizamos la relación entre la distancia recorrida en ruta, y la distancia euclidiana en el par OD. Al analizar los datos encontramos muchos casos razonables en que la distancia en ruta es levemente mayor a la distancia euclidiana, y también algunos casos menos razonables en que la distancia en ruta es mucho mayor que la distancia euclidiana. La Figura 3 ilustra porqué una diferencia muy grande entre ambas es poco razonable. En este caso, muy probablemente se trata de dos viajes y no de uno.

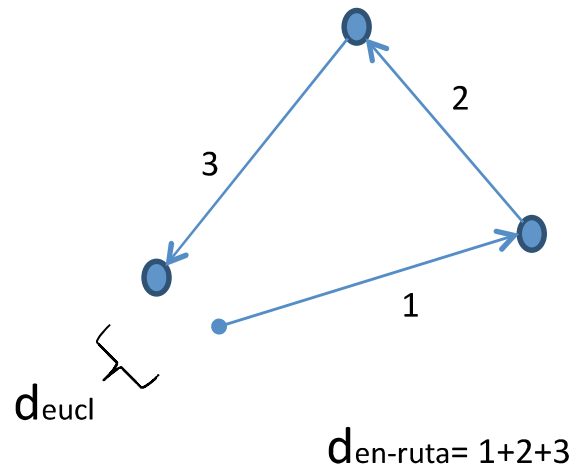


Figura 3: Ejemplo de un viaje de tres etapas que posiblemente está mal estimado

Proponemos usar un criterio para identificar viajes que considere la razón entre distancia en ruta y distancia euclidiana. Este criterio permitirá tomar en cuenta un fenómeno observado, que con la actual metodología no es recogido, que es que los usuarios de transporte público en ocasiones modifican el viaje que realizan habitualmente para acceder a su actividad principal, para encadenar actividades cortas en el trayecto (pagar una cuenta, compras menores, trámites, etc.). Este comportamiento, que habitualmente implica desviarse de la ruta habitual o de la ruta más directa, es muy razonable pues le permite al usuario ahorrar tiempo y dinero, y creemos que puede ser capturado usando un filtro de distancia.

Definiendo:

$$f_d = \frac{d_{en-ruta}}{d_{euclidiana}} \quad (1)$$

Una primera observación es que los casos en que $f_d < 1$ corresponden a errores en la estimación de las distancias, ya que la distancia en ruta no puede ser menor que la distancia euclidiana. Estos casos, aunque existen, son muy escasos, y se deben principalmente a errores en las coordenadas de los paraderos o en la información del GPS del bus. En ese estudio se define como umbral de tolerancia $f_d < 0.98$, para hacerse cargo de errores menores en la aproximación de la posición, la precisión del GPS y la propagación de estos. En cuanto al umbral para identificar viajes asociados a encadenamiento de actividades breves, después de un análisis estadístico de los datos,

proponemos usar $f_d > 2$ como umbral. La Figura 4 muestra que la vasta mayoría de los viajes están bajo ese umbral. Para validar este parámetro, se requiere información adicional.

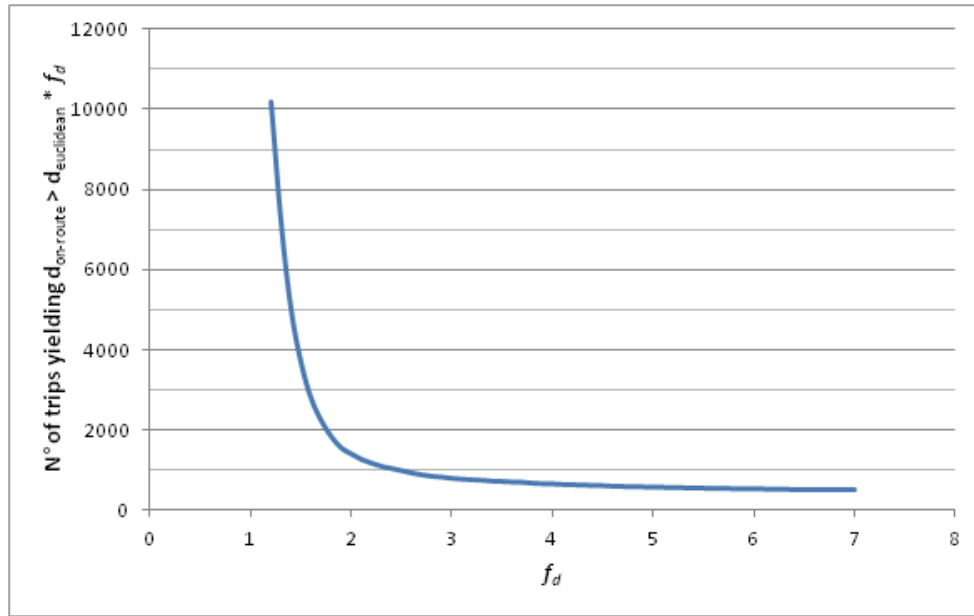


Figura 4: Número de viajes con distancia en ruta/distancia euclidiana bajo f_d vs f_d

Cada viaje que arroja f_d sobre el umbral es separado en diferentes viajes, que deben ser identificados a partir de la información disponible. El procedimiento es muy sencillo cuando el viaje original tiene dos etapas, cada una de ellas es codificada como un viaje de una etapa. El caso de viajes con tres o más etapas es algo más complejo, ya que hay distintas maneras de cortar el viaje original. Para ese caso proponemos el método descrito más abajo:

Considere un viaje con dos o más etapas, tal que la relación entre distancia en ruta y distancia euclidiana sobrepasa el umbral $f_d > 2$. Sea C el conjunto de separaciones posibles en viajes (soluciones). Se define la verosimilitud de la distancia de la solución $i \in C$ como:

$$DL_{sol_i} = - \sum_{viaje\ j \in sol_i} |f_{d_j} - f_{d_j}^{real}| \quad (2)$$

Donde f_{d_j} es la razón f_d del viaje j , y $f_{d_j}^{real}$ es definido como:

$$f_{d_j}^{real} = \frac{1}{\#N_j} \cdot \sum_{viaje\ k \in N_j} f_{d_k} \quad (3)$$

Donde N_j es el conjunto de viajes con igual número de etapas que j , que cumple con

$$0.98 \leq f_d \leq 2 \quad (4)$$

Por definición, la verosimilitud de la distancia de la solución es negativa. En consecuencia, dado un cierto viaje que cumple con $f_d > 2$, la solución cuya verosimilitud de la distancia es máxima, será la solución que minimice las razones de distancia en ruta/distancia euclidiana de entre las soluciones que cumplen con las restricciones (ecuación 4).

Criterio de bus no-abordado

La segunda mejora metodológica propuesta se relaciona con el uso de información de frecuencias para mejorar la identificación de destinos donde se realizan actividades. Este criterio se basa en reconocer el hecho de que si un usuario no aborda un bus que pasa por el paradero durante el periodo analizado, es posible que esto se deba a que está realizando una actividad en el entorno del paradero, y no necesariamente esperando en este. En consecuencia, proponemos usar los datos de GPS de los buses para verificar si buses del mismo servicio que tomó el usuario pasaron por el paradero durante el periodo en que el usuario estaba supuestamente esperando el bus. Se debe introducir alguna flexibilidad, pues puede haber otras razones para que un usuario no aborde un bus, por ejemplo puede que el bus pase lleno, o que no se detenga en el paradero. Proponemos considerar un umbral de tres buses que pasan por el paradero durante el periodo. Si el usuario no se subió al tercer bus del servicio observado en el paradero, asumimos que estaba realizando una actividad en el entorno y no esperando en el paradero.

4. VALIDACIÓN EXÓGENA CON ENCUESTAS DE METRO

Descripción de la muestra

Cada año Metro de Santiago realiza estudios de demanda que incluyen encuestas OD. En el año 2010, decidieron incluir, por primera vez, una pregunta acerca del ID de su tarjeta bip! Esto fue hecho como un experimento exploratorio, por lo cual la pregunta se incluyó sólo para una submuestra. Tomando algunos resguardos sobre la confidencialidad de esta información, los datos nos fueron entregados para validar los supuestos planteados por nuestros métodos. Para realizar este proceso de validación, aplicamos todos nuestros métodos a las bases de datos de transacciones y posicionamiento de la misma semana en que se aplicó la encuesta. Los datos disponibles en la encuesta incluyen todos los detalles del viaje que estaba realizando la persona en el momento de ser encuestada en una estación de Metro. Dada la forma de aplicar la encuesta, todos los viajes observados tienen una etapa en Metro, y la información sobre las etapas previas es revelada (ya ocurrieron) mientras que la información de eventuales etapas posteriores es declarada (están por ocurrir).

La información recolectada incluye instante de tiempo y lugar en que se tomó la encuesta, tipo de usuario (regular, estudiante, adulto mayor), edad, origen del viaje (zona, esquina más cercana), id de la tarjeta bip!, modo usado para acceder al Metro (caminata, bicicleta, auto, servicio de bus), estación de metro de destino (línea, estación), estación usada para transferir entre líneas, destino final del viaje (zona, esquina más cercana), propósito del viaje, ingreso. Después de aplicar algunos filtros de consistencia básica para validar los datos provenientes de encuestas, obtuvimos una muestra de 1.350 etapas realizadas en buses y Metro, que corresponden a 882 viajes registrados en 882 encuestas. Al cruzar esta información con la base de datos de transacciones, encontramos algunos id que no existen en la base de datos, y otros que si bien existen, no registraron ninguna transacción en esa semana. Luego de este filtro, nos quedamos con 684 tarjetas que podían ser usadas en el proceso de validación. Sin embargo, había algunos casos en que si bien la tarjeta existía, y presentaba transacciones en esa semana, no era posible encontrar las transacciones que debieran corresponder a la encuesta, es decir, no había una transacción en la

estación de Metro declarada como estación de ingreso en el periodo en que se realizó la encuesta. Al filtrar también por este efecto, se obtiene una muestra final de 601 tarjetas validables.

En las encuestas se declara sólo una etapa de acceso a Metro y sólo una de egreso (por viaje), pero no es claro en los casos en que hay más de una etapa de acceso o egreso cuál de ellas se declara. En consecuencia, para efectos de análisis, se extraen de la base de datos de transacciones hasta tres etapas previas y posteriores a la etapa de Metro declarada. En la Tabla 1 se muestra el número de etapas encontradas en la base de datos de transacciones, identificando también aquellas en las cuales la metodología de Munizaga y Palma (2012) permitió estimar paradero de bajada.

Tabla 1: Descripción de la muestra de validación

	Pre-metro etapa 3*	Pre-metro etapa 2*	Etapas pre-metro declarada	Etapas Metro	Etapas post-metro declarada	Post-metro etapa 2*	Post-metro etapa 3*	Total
Con estimación de bajada	1	14	88	520	121	40	3	787
Sin estimación de bajada	1	4	8	76	37	12	1	139
Total	2	18	96	596	158	52	4	926
Porcentaje con estimación	50.0	77.8	91.7	87.3	76.6	76.9	75.0	85.0

*: No declarada

El porcentaje de éxito en la estimación de bajada es levemente superior a los valores reportados en Munizaga y Palma (2012) para la base de datos completa. Esto se debe a que la muestra no es representativa de la población, sino que está sesgada a favor de Metro, donde el porcentaje de estimación es mayor. Al comparar por tipo de transacción (metro, bus, zona paga), los valores son similares.

Validación de paradero de subida

La validación de la posición de subida fue hecha a través de una comparación de la posición declarada en la encuesta (esquina más cercana) y la estimación realizada por la metodología de Munizaga y Palma (2012). En el caso de etapas en Metro, la coincidencia es exacta en 100% de los casos, lo cual no es sorprendente. En el caso de transacciones en bus y zona paga, encontramos algunos casos en que la coincidencia es exacta, i.e. el paradero o la zona paga en que el pasajero abordó es el más cercano a la esquina declarada en la encuesta. También encontramos algunos casos en que la intersección declarada en la encuesta era cercana a la parada en que se abordó el bus, pero no la más cercana. Esto usualmente pasaba en casos en que la esquina reportada en la encuesta era una esquina más conocida que la más cercana. En ambos casos se consideró que la estimación es correcta. Se encontró algunos errores en paradas complejas, como la Estación de Intercambio Modal La Cisterna, donde hemos detectado algunos errores en los datos de GPS; probablemente debido a la operación subterránea de buses. El porcentaje global de estimación correcta de paradero de subida es **98,9**.

Validación de paradero de bajada

La validación de la estimación de paradero de bajada es crucial, ya que en ella se basa el resto de la metodología. Sólo es posible realizar la validación en los casos en que el método obtiene una estimación del paradero de bajada, que corresponde a un total de 715 transacciones. Aplicando el mismo criterio definido para la definición de estimación correcta en subidas, encontramos un total de 602 casos de estimación lo que corresponde a un **84,2%**. La principal fuente de error (12,5%) es el uso de modos no integrados, como taxi o auto acompañante, y la no validación en alguna etapa del viaje, que impiden al método reproducir la secuencia de viajes. Hay unos pocos casos (1,8%) donde el paradero de bajada es distinto al declarado porque en la siguiente etapa, que es declarada, el usuario no validó, o bien decidió cambiar de estrategia de viaje después de contestar la encuesta. Hay una observación (0,2%) donde el método falla al parecer debido a que los parámetros que ponderan el tiempo de viaje y el de caminata no recogen bien las preferencias de esos usuarios.

Validación de elección de ruta

La EOD de Metro también contiene información acerca de la ruta elegida al interior de la red de Metro. El modelo de Munizaga y Palma asigna ruta a las transacciones de metro asumiendo elección determinística basada en mínimo costo. La información disponible nos permite explorar si ese supuesto es razonable, o si es necesario incorporar modelos estocásticos de elección de ruta. Para este análisis podemos usar todas las observaciones de la EOD, independiente de si tenemos el id de esos usuarios o no, ya que la ruta elegida por los usuarios debe ser comparada con la ruta que le asignaría el modelo para ese par OD, que es única y conocida. Este análisis se realizó con una muestra de 130.000 observaciones, y a continuación se presenta los resultados preliminares. Mirando las rutas elegidas por los usuarios encontramos tres casos:

1. Pares OD donde todos los usuarios eligen la misma ruta.
2. Pares OD donde hay una ruta claramente dominante, pero algunos usuarios eligen otra(s) ruta(s).
3. Pares OD donde claramente los usuarios eligen más de una ruta.

Los casos 1 y 2 son los más frecuentes, lo cual es positivo porque implica que en la mayoría de los casos el supuesto de elección determinística es adecuado. Además, en la mayoría de los casos en que los usuarios eligen una única ruta, esta es la que el modelo les asigna. Sin embargo, hay algunos casos en que los usuarios eligen una única ruta que no es la asignada por el modelo. Esto claramente se debe a un error de implementación, que puede ser corregido mejorando las estimaciones de tiempos entre estaciones y penalización de trasbordos. Por otra parte, hay un número significativo de casos tipo 3, en que el enfoque actual no es adecuado, y se debe implementar un modelo de elección estocástico.

4. VALIDACIÓN EXÓGENA CON VOLUNTARIOS

Para complementar el análisis de validación, se reclutó una muestra de 53 voluntarios, la mayoría estudiantes, quienes nos entregaron su número de tarjeta y fueron contrastados con las estimaciones hechas por el modelo para una semana en particular. A ellos se les pidió validar las

estimaciones para su caso particular. Durante la semana analizada, los voluntarios hicieron 885 transacciones, correspondientes a 586 viajes. La validación de la identificación viajes y etapas mostró que el método identificaba correctamente los viajes en un **90,0%** de los casos (527) y fallaba en 56. Había también 3 casos en que la validación no se pudo hacer porque el voluntario no recordaba el viaje. Las principales razones de falla del método son descritas en la Tabla 2.

Tabla 2: Causas de falla en el procedimiento de estimación de etapas y viajes

Causa de falla	Frecuencia
Debido a congestión de buses y pasajeros en el paradero, el usuario abordó el cuarto bus que pasó desde su llegada al paradero. El método corta si pasan más de tres.	12
Una etapa intermedia no tenía estimación de bajada, y la etapa siguiente quedó automáticamente codificada como primera etapa.	9
Estimación incorrecta debido al criterio de la relación de distancias.	8
Actividad extremadamente breve, imposible de detectar.	8
Error de propagación debido a que el viaje anterior no fue correctamente estimado.	8
Errores de implementación.	7
Tiempo de espera mayor a 30 min debido a baja frecuencia de buses o a que los buses iban llenos y no se pudo abordar.	4

De los 527 viajes correctamente identificados, el método utilizado para estimar propósito podía realizar estimación en 448, porque los restantes 79 no tenían estimación de bajada. De esos 448, el método estimaba correctamente propósito en 352 (79%). Las principales razones de error en la estimación se reportan en la Tabla 3.

Tabla 3: Causas de error en el método de asignación de propósito

Causa de error	Frecuencia
Actividad "Otro" de duración mayor a dos horas	26
Un viaje al hogar de duración menor a dos horas, que no es el último del día, queda catalogado como un viaje con propósito "Otros" en vez de "Hogar"	16
Actividad trabajo de duración menor a dos horas (típicamente asociado a viajes de trabajo más que viajes al trabajo)	14
Actividad trabajo realizada por un usuario estudiante	12
Actividad estudio de duración menor a dos horas (día laboral)	10
Actividad "Otros" realizada al final del día (queda clasificado como "Hogar")	7
Actividad estudio realizada por usuarios de tarjeta regular	5
Actividad estudio realizada al final del día	3
Actividad estudio realizada de duración menor a dos horas (fin de semana)	1
Viaje al hogar a las 2AM clasificado como Trabajo (el nuevo día comienza justo después de la medianoche)	1
Único viaje (en un día en particular) codificado como "Hogar"	1

Algunos de los errores reportados en las tablas 2 y 3 pueden ser corregidos con mejoras de implementación, modificando los parámetros e implementando procesos de chequeo automáticos. Por ejemplo, la re-definición de la frontera entre un día y el siguiente como las 4:00 AM en vez

de la medianoche propuesta a partir de la validación endógena. Otros errores pueden ser corregidos procesando la información disponible y adaptando la metodología para tratar casos particulares, como por ejemplo incorporar información histórica de perfiles de carga para definir el número de buses no abordados que implica el corte de etapa/viaje. También es posible incorporar información del lugar de residencia para distinguir actividades “Hogar” de “Otros”, “Trabajo” o “Estudio”, dependiendo de la localización en que estas ocurren. Tanto la Zona de residencia como los Perfiles de carga pueden ser estimados a partir de los datos, y corresponden a líneas de investigación actualmente en desarrollo (Ibarra *et al*, 2013; Amaya y Munizaga, 2013). Por otra parte, existen errores que no es posible detectar o corregir sin información exógena, como por ejemplo los que se deben a una etapa de viaje no observada, o el uso de tarjeta de estudiante para realizar actividades de trabajo. Sin embargo, el porcentaje de estimación correcta es muy alto en todas las etapas del proceso, lo cual muestra que la información obtenida es confiable.

6. CONCLUSIONES

Hemos explorado la confiabilidad del método propuesto por Munizaga y Palma (2012) para estimar flujos OD de viajes en transporte público y su nivel de servicio a partir de datos de transacciones y GPS. La validez de los supuestos se analiza en base a tres fuentes de información: los mismos datos usados para desarrollar el modelo, una Encuesta Origen Destino detallada aplicada a usuarios de Metro, incluyendo el id de la tarjeta para un grupo de ellos, y un grupo de voluntarios que accedieron a participar en un experimento de validación.

Todas las etapas de este proceso mostraron ser valiosas. Gracias a este proceso pudimos identificar errores de implementación que son fáciles de corregir; proponer mejoras metodológicas que van a contribuir a mejorar la calidad de los resultados, y además, esto nos da una idea inicial sobre la confiabilidad de los resultados, que es muy positiva. Las más importantes mejoras metodológicas propuestas son la flexibilización de parámetros como la distancia caminable, la redefinición del límite temporal entre un día y el siguiente a las 4AM, la posibilidad de utilizar las transacciones únicas del día realizadas durante la tarde cruzando con información del día siguiente, y un nuevo método para identificar las etapas de un viaje, distinguiendo actividades de transbordos al utilizar el criterio de distancia recorrida versus distancia euclidiana y considerando la frecuencia de buses observada en el paradero.

En cuanto al ejercicio de validación de las estimaciones hechas por el modelo, encontramos que el paradero de subida es estimado correctamente en un 98,9% de los casos, el paradero de bajada es correctamente identificado en un 84,2% de los casos, el corte de etapas de viaje es realizado correctamente en un 90% de los casos, y el propósito es correctamente estimado en un 79%. Estos valores son extraordinariamente positivos, y nos muestran que la información generada es altamente confiable. Esperamos en el futuro cercano tener disponible una muestra representativa de la población de Santiago a través de la EOD 2012, que nos permitirá validar y estimar estos porcentajes para distintos segmentos de la población, como por ejemplo estudiantes y trabajadores, o tramos de ingreso.

En resumen, los resultados que ya son muy buenos, incluso pueden ser mejorados, implementando las mejoras metodológicas propuestas.

AGRADECIMIENTOS

Financiamiento: Fondecyt (1120288), Fondef D10I-1002, ISCI (ICM P-05-004-F, CONICYT FBO16). Agradecemos especialmente la colaboración de Transantiago y de Metro.

REFERENCIAS

Amaya, M. y Munizaga, M. (2013) Estimación de zona de residencia en base a sistemas tecnológicos de transporte público. Enviado al XVI Congreso Chileno de Ingeniería de Transporte.

Devillaine, F., Munizaga, M.A. y Trepanier, M. (2012) Detection activities of public transport users by analyzing smart card data. **Transportation Research Record**, 2276, 48-55.

Devillaine, F., Munizaga, M.A., Palma, C. y Zúñiga, M. (2013) Towards a reliable Origin-Destination matrix from massive amounts of Smartcard and GPS data: application to Santiago En: Zmud, J., Lee-Gosselin, M., Munizaga, M.A. and Carrasco, J.A. (Eds.) **Transport Survey Methods; Best Practice for Decision Making**, Emerald, 695-710.

Ibarra, R., Munizaga, M., Devillaine, F. y Gschwender, F. (2013) Methodology and analysis of load profiles and passenger flows in a complex public transport system. TRISTAN 8, San Pedro de Atacama, Chile Julio.

Munizaga, M.A. y Palma, C. (2012) Estimation of a disaggregate multimodal public transport origin-destination matrix from passive Smart card data from Santiago, Chile. **Transportation Research C**, 24(12), 9-18.

Trépanier, M., Tranchant, N y Chapleau, R (2007) Individual trip destination estimation in a transit smart card automated fare collection system. **Journal of Intelligent Transportation Systems**, 11, 1–14.