



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**Obtención de información para la gestión del Transantiago a través
datos operacionales.**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

DANIEL ANDRES FISCHER MONTT

**PROFESOR GUÍA:
MARCELA MUNIZAGA MUÑOZ**

**MIEMBROS DE LA COMISIÓN:
PATRICIO CONCA KEHL
JOSÉ MOSQUERA CADIZ**

**SANTIAGO DE CHILE
OCTUBRE 2010**

Resumen Ejecutivo

En el sistema de transporte público se ha implementado el uso de una tarjeta de pago llamada bip!, la cual tiene una tasa de penetración cercana al 100% y se han instalado GPS en buses. Los datos generados por estas dos herramientas son almacenados y son fuente de mucha información útil en la gestión y diseño de este servicio público el cual transporta diariamente alrededor de dos millones de personas.

El propósito de este estudio es contribuir a la gestión del Transantiago, utilizando una metodología implementada en la Universidad de Chile, con el fin de estimar la demanda de transporte público. Para esto se utiliza una metodología existente en la literatura la cual es adaptada y mejorada.

El procesamiento de datos consiste en aplicar diversos filtros y procedimientos para identificar el origen y destino de cada etapa de viaje. Se trabajó con una base de datos entregada por Transantiago, correspondiente a las transacciones de la semana del 20 al 26 de marzo del 2009.

Con esta metodología se puede estimar la demanda de transporte a lo largo de un día utilizando información generada por la operación del sistema de transporte.

Para poder utilizar esta estimación en el diseño y gestión del transporte público fue necesario realizar un post procesamiento de los datos obtenidos con la metodología para estimar subidas y bajadas aplicando transformaciones para reflejar demandas reales de transporte con distintos niveles de agregación.

Como resultado se obtuvo una matriz origen-destino que puede ser desagregada incluso a nivel de paradero.

La principal conclusión de este trabajo es que el uso de esta metodología permite utilizar los datos generado por la operación del Transantiago, disminuyendo los costos de realizar mediciones y aumentando los tamaños muestrales.

Contenido

Resumen Ejecutivo	2
Tabla de figuras	4
Índice de Tablas	4
Índice de Ilustraciones	5
Índice de Ecuaciones.....	5
Índice de Gráficos	5
1. Introducción	6
2. Antecedentes	7
2.1 Transantiago	7
2.2 Descripción de la metodología existente.....	13
Fuente: Mora 20102.3 Mejoras a la Metodología (MESB)	18
2.3 Mejoras a la Metodología (MESB)	19
3. Metodología	23
3.1 Transformación de salida del programa	25
3.2 Factores de expansión	30
3.3 Evasión en la Metodología de Estimación de subidas y Bajadas	33
4. Resultados Obtenidos	36
4.1 Indicadores de estimaciones de subidas y bajadas	36
4.2 Factores de Expansión	39
4.3 Estimación de la Demanda.....	45
4.4. Análisis de sensibilidad	51
4.5 Evaluación de la estimación de la demanda.....	52
4.6 Evaluación Beneficios de Metodología	54
5. Conclusiones.....	58
6. Bibliografía	60
7. Apoyos Institucionales	62
Anexos.....	63
A. Zonas de influencia.....	63
B. Últimos desplazamientos.....	67
C. Encabezados Salida	68

Tabla de figuras

Índice de Tablas

Tabla 1: Asignación de Comunas a Área	8
Tabla 2: Tasa de evasión y número de mediciones por horario y operador.....	13
Tabla 3: Principales tablas utilizadas por la metodología.	18
Tabla 4: parámetros utilizados por MESB	22
Tabla 5: Bloques horarios para día laboral del Transantiago.....	31
Tabla 6: Tasa de estimación de etapas por día	36
Tabla 7: Tasa de estimación de viajes por día.....	36
Tabla 8: Tasas de estimación de etapas por forma de pago	38
Tabla 9: Promedio de factores de expansión para viajes.....	39
Tabla 10: Desviaciones estándar de factores de expansión para viajes	40
Tabla 11: Factores de expansión F2	41
Tabla 12: Transacciones sin paradero de origen y cantidad del factor de expansión teórico por esta causa.....	42
Tabla 13: Matriz Origen-Destino por área en miles de viajes	45
Tabla 14: Análisis de sensibilidad para tiempo de espera.....	52
Tabla 15: Demanda de transporte estimada en EOD 2001 (Sectra 2001)	53
Tabla 16: Demanda de transporte estimada por metodología.....	53
Tabla 17: Cambio porcentual con respecto a EOD 2001.....	53
Tabla 18: Error esperado en mediciones	55
Tabla 19: Porcentaje del presupuesto mal asignado por errores muestrales, para distintos niveles de agregación	55
Tabla 20: Costo del error muestral en Porcentaje del costo total de Transantiago	55

Índice de Ilustraciones

<i>Ilustración 1: Zonificación Estras</i>	9
Ilustración 2: Esquema de funcionamiento de Metodología propuesta por Trépanier	14
Ilustración 3: Esquema de funcionamiento del módulo de zonas pagas.....	20
Ilustración 4 : Zona paga a menos de 500 metros de inicio de ruta	21
Ilustración 5: Diagrama de flujo de criterio para de discriminar entre trasbordo y destino.	27
Ilustración 6: Efecto de múltiple pago en MESB	28
Ilustración 7: esquema de múltiple pago ya procesado	28
Ilustración 8: Detección de múltiples Pagos	28
Ilustración 9: Múltiples Pagos en Cercanía	29
Ilustración 11: Distorsión en demanda estimada con MESB	34
Ilustración 12: diagrama de post procesamiento de datos.....	35
Ilustración 13: Transacciones sin paradero asignado	43

Índice de Ecuaciones

Ecuación 1: Intervalo de confianza para poblaciones infinitas.	54
---------------------------------------------------------------------	----

Índice de Gráficos

Gráfico 1: Distribución de tasas en estimación de bajadas, con post procesamiento	37
Gráfico 2: Última Etapa vs Resto cada 100 metros.....	44
Gráfico 3: Histograma de Hora de inicio de viajes	46
Gráfico 4: Histograma de Hora de fin de viajes.....	46
Gráfico 5: Histograma de hora de inicio de viajes.....	48
Gráfico 6: Orígenes y Destinos en el sector centro de Santiago para el día lunes.....	49
Gráfico 7: Orígenes y Destinos en el sector occidente de Santiago para el día lunes	50
Gráfico 8: Porcentaje Estimado, con respecto a la distancia Máxima caminable cada 100 metros	51

1. Introducción

En Santiago de Chile, al igual que en otras ciudades del mundo como Washington, Tokyo, Chicago, San Francisco, New York, Changchun (China), entre otros se ha implementado con distintos niveles de penetración tarjetas “Smartcard” para efectuar pagos en el sistema de transporte público, convirtiendo esta modalidad en la tendencia mundial.

En Transantiago se utiliza una tarjeta (Smartcard) llamada bip! y es el único medio de pago en buses urbanos, mientras que en metro, se puede pagar con tarjeta bip! o comprando un boleto sencillo. Al usar la tarjeta bip! para el pago, se permite hacer transbordos sin costo adicional, beneficio que no existe para el boleto unitario del Metro. Dado esto, el porcentaje de personas que utiliza boleto unitario en metro es sólo el 1% (Munizaga et al 2010)

Por otra parte, los buses son monitoreados a través de un sistema GPS, que registra la posición geográfica de los buses cada 30 segundos.

Las coordenadas entregadas por los GPS y las transacciones realizadas con la tarjeta bip! son guardadas como datos operacionales generando semanalmente 35 millones de transacciones bip con una coordenada GPS asociada (Mora 2010).

Esto constituye una fuente de información pasiva que puede ser de gran valor y bajo costo, si se logra implementar procedimientos adecuados que permitan extraer la información de los datos a esto se añade que la información necesaria para la modelación del transporte público es tradicionalmente muy costosa y difícil de obtener. En la Universidad de Chile se ha trabajado en el desarrollo de metodologías que permiten utilizar esta información para la gestión y desarrollo del transporte público.

Una de estas metodologías fue desarrollada para estimar orígenes y destinos de viajes pagados con tarjeta bip!, esto con el fin de encontrar una estimación de la demanda de transporte.

El propósito de esta memoria es estimar la demanda total de transporte utilizando la información extraída con la metodología, contribuyendo a optimización y efectividad del sistema de transporte.

El objetivo principal de este estudio es generar información útil para la gestión y optimización del sistema de transporte público, en particular obteniendo una estimación de la demanda en forma de matriz origen destino de viajes. Para lograr este objetivo se realizarán mejoras a la metodología existente, tales como aplicación de factores de expansión y de viajes incorrectamente estimados.

Los resultados obtenidos serán validados y contrastados con las mediciones existentes a la fecha y se estimarán los beneficios obtenidos al utilizar esta metodología.

Se realizará una evaluación de la herramienta implementada, identificando sesgos, errores en los resultados y proponiendo mejoras a esta. Aquellas mejoras posibles de implementar serán

realizadas, poniendo énfasis en el post procesamientos de los resultados obtenidos con la metodología implementada en la Universidad de Chile.

Aquellos sesgos o errores identificados, serán tratados en lo posible para minimizar su efecto sobre la estimación de la demanda de transporte.

La estructura del informe es la siguiente:

En el capítulo 2 se presentan los antecedentes de esta investigación tocando, principalmente tres temas, el primero es el servicio de transporte público existente en Santiago, el cuales llamado Transantiago y el segundo la descripción de la metodología ya existente que permite estimar las subidas y bajadas de los usuarios del transporte público y para terminar, con una descripción de las modificaciones y mejoras que se realizaron a esta con el fin de estimar la demanda.

El tercer capítulo contiene una descripción de la metodología implementada para estimar la demanda en base a las estimaciones de subidas y bajadas. En el capítulo 4 se presenta los resultados obtenidos con la metodología propuesta, los cuales incluyen la estimación de la demanda, análisis de sensibilidad sobre los parámetros y una evaluación económica de los beneficios que esta metodología podría traer.

En el capítulo 5 se presenta las conclusiones de la investigación, continuado por la bibliografía utilizada, apoyos institucionales y para terminar los anexos.

2. Antecedentes

2.1 Transantiago

Transantiago es un sistema de transporte público urbano integrado que opera en el área metropolitana de la ciudad de Santiago, capital de Chile. Comenzó el 10 de febrero del 2007, implementando el uso de un sistema tarifario integrado con medio de pago exclusivo mediante la tarjeta bip!. Transantiago reformó las rutas de buses de la capital, diseñando un sistema basado en uso de alimentadores y troncales integrado con el Metro de Santiago, diferenciándose así del sistema anterior que obedecía mas a un sistema puerta a puerta.

2.1.1 Economías de escala y Externalidades del transporte público

El transporte público es un caso de economías de escala, un sistema con más usuarios, requerirá de más buses y por ende, se tendrán mejores frecuencias, además el costo marginal social de proveer un nuevo viaje es menor que el del transporte privado llamado Efecto Mohring (Morandé et al 2007).

Dada las externalidades del transporte público, las políticas a llevar a cabo deben.

- Incentivar el sistema público para disminuir las externalidades negativas del transporte privado.
- Incentivar el uso del transporte público para sacarle provecho a las economías de escala (externalidades positivas).

Un sistema de transporte con externalidades positivas justifica la Subvención estatal del transporte público.

2.1.2 Antes y después del Transantiago

Antes de la implementación de Transantiago, circulaban por la ciudad las denominadas “micros amarillas” que operaban concesiones de recorridos diseñados por el gobierno, la tarifa era única y se pagaba por etapa de viaje.

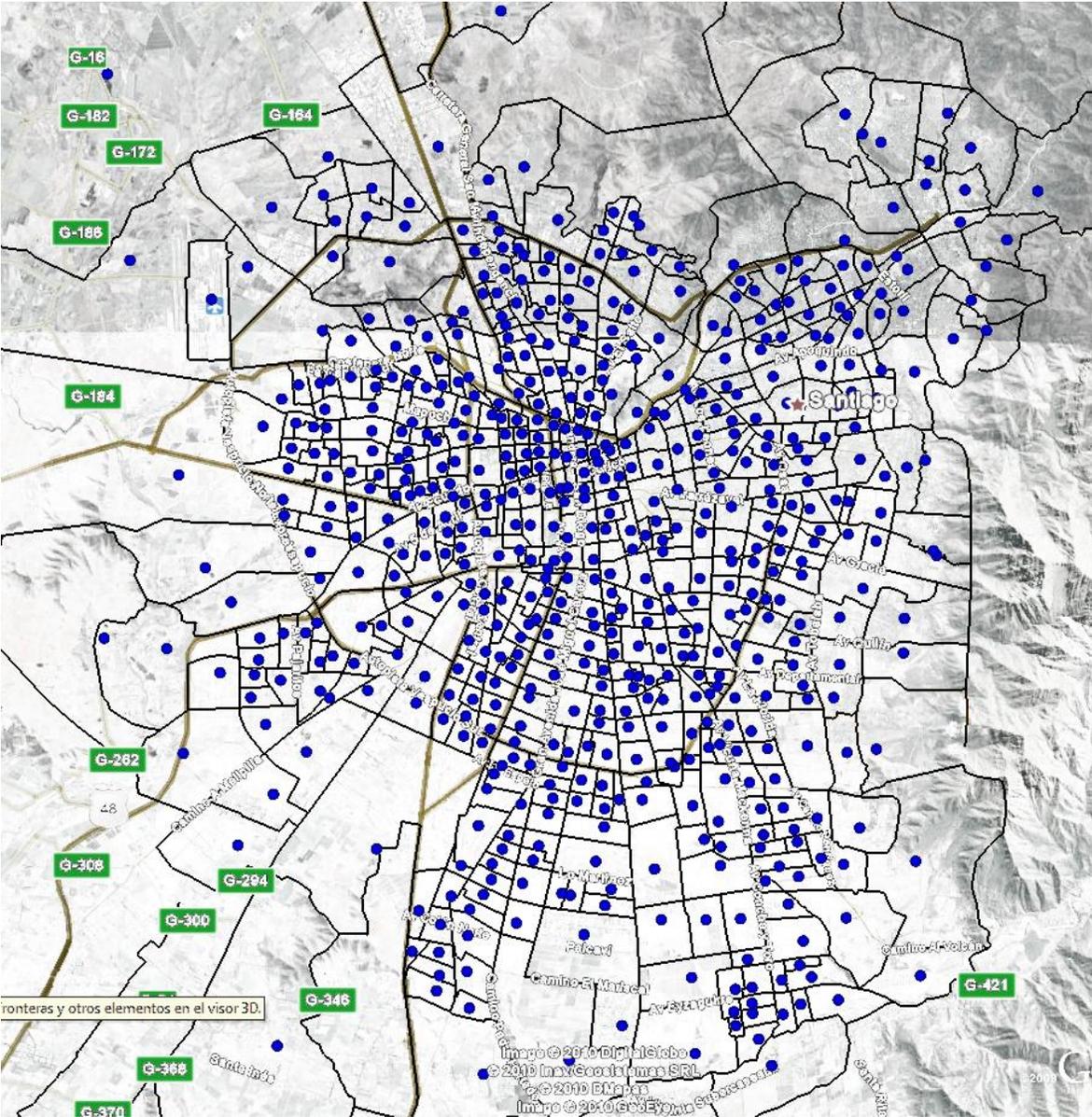
Los niveles de contaminación (acústica y atmosférica), congestión en las vías principales, deficiencia en la calidad de los buses y las peligrosas competencias de los microbuseros por los pasajeros provocaron que el 2002 el transporte público tuviera una muy mala evaluación (11,2 en escala de 1 a 100). Con el fin de reducir la contaminación y mejorar la calidad de vida de los santiaguinos se propuso diseñar un sistema de transporte integrado que luego sería llamado Transantiago. Con el fin de optimizar los recorridos, el sistema se dividió en dos tipos de servicios, llamados alimentadores y troncales, que funcionarían en conjunto con Metro. Los alimentadores circularían dentro de zonas geográficas que se describen en la Tabla 1, con el objetivo de proveer de pasajeros a los troncales y Metro que realizarían los servicios largos, se pagaría por viaje, pudiendo realizar todas las combinaciones necesarias para alcanzar el destino de su viaje.

Tabla 1: Asignación de Comunas a Área

Sector	Comunas
A	Santiago
B	Conchalí, Huechuraba, Independencia, Quilicura, Recoleta, Renca
C	Las Condes, Lo Barnechea, Providencia, Vitacura
D	La Reina, Macul, Ñuñoa, Peñalolén
E	La Florida, La Granja
F	Puente Alto
G	El Bosque, La Cisterna, La Pintana, San Bernardo, San Ramón
H	Lo Espejo, Pedro Aguirre Cerda, San Joaquín, San Miguel
I	Cerrillos, Estación Central, Maipú
J	Cerro Navia, Lo Prado, Pudahuel, Quinta Normal

Para diseñar el Transantiago, se utilizó como entrada al modelo de optimización una matriz origen destino desagregada a nivel de zonas Estras descritas geográficamente en la Ilustración 1, correspondientes a agrupaciones de suelos de usos comunes que dividen Santiago en 618 zonas.

Ilustración 1: Zonificación Estras



2.1.3 Tarjeta Bip!

La tarjeta bip! es el más importante medio de pago, con formato sin contacto, que se utiliza en el sistema de transportes de Santiago Transantiago. Su emisión y mantención se encuentran a cargo del Administrador Financiero de Transantiago. El usuario debe pre cargar un monto en la tarjeta, la cual al ser validada descuenta el costo del pasaje.

Al validarse la tarjeta en los puntos de pago, el validador muestra una luz verde para las transacciones correctas y una luz roja para las incorrectas, una validación incorrecta, se puede deber a falta de saldo, mal funcionamiento de la tarjeta o un error ocasional requiriéndose una nueva validación.

Las validaciones pueden realizarse en Buses, Zonas Pagas o Metro. Para el caso del bus se valida al momento del abordaje, mientras que para Zona paga y Metro, se válida para entrar al área donde se realiza el abordaje.

2.1.4 AFT Administradora Financiera del Transantiago

El Administrador Financiero de Transantiago (AFT) es el ente encargado de recaudar, administrar y distribuir los ingresos entre los operadores del Transantiago además de encargarse de la comercialización y distribución de la tarjeta Bip!.

Los dueños del AFT son:

- BancoEstado (21% de participación)
- Banco de Chile (20% de participación)
- BCI (20% de participación)
- Banco Santander Chile (20% de participación)
- Promotora CMR Falabella (9,5% de participación)

También se encuentra la empresa Sonda (9,5% de participación), que facilita la gestión del sistema de validación y pagos mediante la Tarjeta bip!, además de Administrar el Centro de Información y Gestión (CIG).

2.1.5 Situación Actual del Transantiago

El Transantiago cuesta 53,3 mil millones de pesos chilenos mensuales en pago a operadores, de los cuales una parte es subvencionada por el estado. Mensualmente se realizan 150 millones de validaciones, con pago promedio a operadores de \$404 por validación (Coordinación de Transantiago 2010) . Diariamente se transportan 1,9 millones de pasajeros para un día laboral, 1,1 y 0,9 millones el día sábados y domingos, realizándose 3,9 millones de viajes diarios en un día laboral, 2,3 y 1,4 millones en días sábados y domingo respectivamente, lo que en promedio

significa 1,55 validaciones de bip! por viaje. Por ende en promedio un viaje cuesta \$628 para el sistema de transporte público y el estado debe subvencionar el restante (elaboración propia¹).

Con el objetivo de disminuir la subvención, se ha incrementado el precio del pasaje a \$500 para buses y \$580 para metro en horario punta y no se descarta una nueva alza en los precios. (Diario la Nación (2) 2010).

Criteria Research realizó un estudio en el que asignó nota 4,3 a Transantiago (Diario la Nación 2010), una mejora sustentable en comparación 2,7 recibido un año después de la puesta en marcha (Diario la Nación 2008).

2.1.6 Evasión en Transantiago

Los sistemas de transporte público suelen estar sujetos a evasión y es considerado normal en cifras bajas. Sin embargo la evasión global en Transantiago hoy en día llega al 17%, pero si se descuentan los viajes en metro, donde prácticamente no hay evasión, la cifra alcanza órdenes del 25% del total de las etapas para el transporte en buses. (Fundación Chile 2010)

Los perfiles de los evasores son variados y tienen comportamientos distintos al momento de evadir, según un estudio realizado por Collect GFK (2010) estos pueden ser clasificados en cuatro tipos:

- Evasor Involuntario
 - No logra cargar su tarjeta en su barrio por falta de oferta / cuotas. Pide al chofer que lo lleve en su primer viaje, y luego carga en el punto más cercano que encuentra en su recorrido, generalmente en el metro.
 - Se debe principalmente a la falta de puntos de recarga o baja cuota entregada a almacenes con recarga Bip!.
- Evasor Prepotente
 - Escuda su evasión como protesta al sistema. Se escuda en que el sistema no funciona, que el chofer no gana con el pasaje, etc.
 - Este usuario es uno de los más criticados socialmente.
 - Son considerados una minoría dentro de los evasores
- Evasor Oportunista
 - Es evasor cuando se le presenta la posibilidad. Generalmente es un evasor que actúa en grupos y que intenta pasar desapercibido, no es confrontacional y se aprovecha de situaciones de colapso del sistema para evadir el pago de su pasaje

¹ Datos obtenidos utilizando datos entregados por Transantiago y el informe financiero de este.

- Este comportamiento aparece principalmente en los horarios punta y paraderos con mucha demanda, como por ejemplo: Paraderos cercanos a colegios
- Evasor por Ahorro
 - Conoce el sistema y utiliza los medios disponibles para ahorrar. Este evasor conoce el sistema y ha buscado las formas de utilizar el sistema como medio de ahorro, frecuentemente se ve en grupos familiares. Aquí encontramos la *estrategia de “los cuatro minutos”*, donde el usuario marca su pasaje y pide al conductor marcar el pasaje de su(s) acompañante(s) luego de 4 minutos para que el pasaje tenga un costo 0. De esta forma, cumple con el sistema, pero logra ahorrar pasajes en su grupo familiar.
 - Este es el evasor que más daño causa financieramente al Transantiago, porque a los operadores se les paga por bip!, recibiendo solo un pago el sistema, pero facturando 2 al operador.

La evasión es un grave problema para el financiamiento del Transantiago y tiene consecuencias en la gestión de este a través de información de transacciones, ya que no todos los viajes son detectados, pudiendo sub estimar la demanda de transporte.

Fundación Chile, realiza mediciones periódicas de la tasa de evasión (porcentaje del total de etapas de viaje efectuadas sobre buses, en que no existió una validación o ésta no fue exitosa) en el sistema de transporte. La más actualizada a la fecha es la número 65, realizada entre el 6 al 21 de abril del 2010. Se encuestaron 775.090 usuarios sobre 28.538 buses. La metodología utilizada consiste en que el encuestador se sube a un bus por 20 minutos y cuenta la gente que validó correctamente su tarjeta bip! y el total de personas que suben al bus (Fundación Chile 2010). Esta muestra es bastante poco representativa, por lo que para obtener mejores datos se utilizaron todas las muestras del año 2009, en Tabla 2 se puede ver los resultados obtenidos de estas mediciones además del número de buses medidos en cada operador y horario, se puede apreciar que no existen mediciones en horas punta, principalmente en hora punta mañana, además la muestra es bastante pequeña.

Tabla 2: Tasa de evasión y número de mediciones por horario y operador

Hora/Evasión y Mediciones	1	2	3	4	5	B	C	D	E	F	G	H	I	J	Total
1:00:00		0.00%				0.00%									0.00%
		1				1									2
5:30:00		20.51%	15.94%	28.57%		50.00%			23.53%	33.33%	0.00%				23.48%
		1	2	1		1			1	1	1				8
6:30:00	6.25%	12.50%			0.00%	0.00%	28.57%	25.00%				18.18%			12.09%
	2	1			1	1	1	1				1			8
8:30:00	15.78%	12.51%	14.73%	13.14%	14.49%	12.60%	5.21%	10.99%	14.14%	20.38%	19.95%	12.01%	12.66%	8.94%	13.46%
	135	161	137	144	126	146	136	141	135	158	115	143	101	101	1879
9:30:00	13.45%	13.27%	16.90%	12.92%	13.75%	11.79%	3.10%	7.61%	9.17%	22.32%	20.46%	10.87%	10.58%	9.54%	12.24%
	76	72	62	51	57	69	72	72	69	56	46	51	40	53	846
12:30:00	13.08%	12.40%	14.21%	12.29%	11.75%	10.63%	4.28%	9.00%	12.91%	17.11%	19.69%	13.68%	11.10%	11.53%	12.13%
	315	286	245	276	199	314	292	273	234	221	192	175	156	160	3338
14:00:00	15.59%	13.66%	17.07%	15.83%	12.44%	14.84%	4.27%	13.49%	19.63%	21.73%	27.56%	16.10%	13.11%	14.83%	15.61%
	107	116	79	86	71	169	99	107	97	89	90	84	75	80	1349
17:30:00	15.40%	16.50%	18.44%	16.19%	14.42%	17.77%	6.75%	12.92%	18.32%	21.71%	27.69%	15.83%	15.68%	18.11%	16.84%
	250	282	217	225	165	292	230	257	244	237	222	267	247	284	3419
20:30:00	13.54%	14.93%	19.51%	12.36%	10.08%	13.05%	7.71%	11.10%	15.36%	18.91%	22.75%	14.76%	12.50%	12.18%	14.13%
	221	226	166	197	165	234	205	186	165	191	195	162	191	191	2695
21:30:00	9.24%	6.95%	10.62%	8.13%	13.16%	3.59%	4.72%	7.98%	16.26%	13.49%	13.68%	12.50%	10.82%	15.76%	9.20%
	11	17	6	13	13	19	31	14	6	7	15	2	14	13	181
23:00:00				0.00%											0.00%
				2											2
Total	14.23%	14.00%	16.66%	13.58%	12.60%	13.53%	5.55%	10.96%	15.34%	19.87%	23.15%	14.33%	13.14%	13.73%	14.19%
	1117	1163	914	995	797	1246	1066	1051	951	960	876	885	824	882	13727

Fuente: Fundación Chile (2010)

2.2 Descripción de la metodología existente

Con el fin de entregar información útil para la gestión y mejora del sistema de transporte público, en la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile se ha estado trabajando en distintas metodologías utilizando la información obtenida con la operación del sistema, una de ellas tiene como objetivo encontrar la matriz origen destino del sistema. Esta información se obtiene estimando la el origen y el destino de cada pasajero.

En Transantiago se monitorean los buses a través de un sistema GPS y se registran las coordenadas de los buses continuamente además las transacciones realizadas con la tarjeta bip! son almacenadas con su lugar de pago y hora asociada. Estos datos son utilizados como entrada para la metodología descrita a continuación.

Para comprender la metodología existente es necesario mostrar las definiciones realizadas en Mora (2010), las cuales se utilizaron en esta memoria.

- **Origen:** Posición donde comienza el viaje.
- **Destino:** Posición donde termina el viaje.
- **Viaje:** Desplazamiento entre un origen y un destino para realizar una cierta actividad en el destino, en un cierto período. Puede estar compuesto por etapas.
- **Etapas:** Segmento de viaje que se identifica al existir diferentes modos de transporte ó servicios utilizados. La presencia de etapas en el viaje constituyen transbordos.
- **Punto de subida:** Posición donde comienza la etapa.
- **Punto de bajada:** Posición donde termina la etapa.

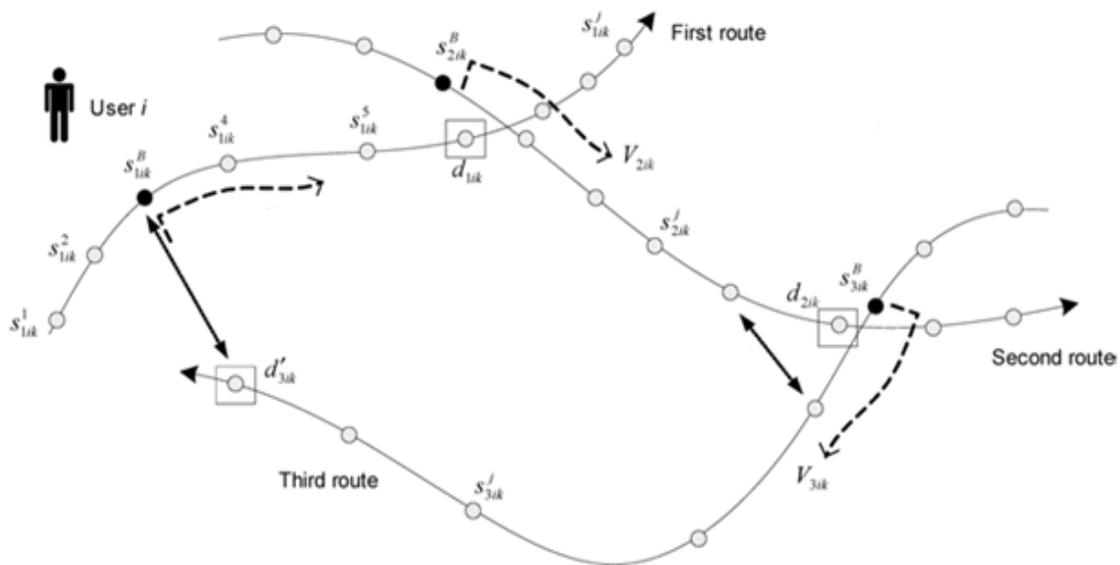
- **Hora origen:** Hora en que se realizó la transacción en el origen del viaje.
- **Hora destino:** Hora estimada en que se llegó al destino del viaje.
- **Hora punto de subida:** Hora en que se realizó la transacción que da inicio a la etapa.
- **Hora punto de bajada:** Hora estimada en que se terminó la etapa.
- **Tiempo de viaje:** Duración del viaje, diferencia de tiempo entre hora origen y hora estimada de destino.
- **Tiempo de transbordo:** Diferencia de tiempo entre etapas de un viaje. Compuesto por tiempo de espera y tiempo de caminata.

2.2.1 Metodología de estimación de viajes

Al tener un sistema de monitoreo de GPS para buses y pago electrónico con smartcards, resulta fácil estimar los puntos de subida a los buses, como el paradero más cercano al pulso GPS más próximo al momento de la validación, la dificultad se centra en estimar las bajadas. La metodología está basada en una propuesta por Trépanier et al (2007) que a grandes rasgos propone estimar la bajada con el punto de la ruta del bus más cercano a la siguiente transacción.

Esta metodología es descrita esquemáticamente en Ilustración 2. Los puntos negros representan las estimaciones de subida, y los puntos con un cuadrado son los puntos donde se estima que el pasajero se baja del medio de transporte para realizar un trasbordo o alguna actividad, mientras que los círculos grises son las paradas posibles. Un supuesto de esta metodología es que al final del día el usuario vuelve al lugar de origen de la primera transacción.

Ilustración 2: Esquema de funcionamiento de Metodología propuesta por Trépanier



Fuente: Trepanier et al (2007)

Con esta metodología se estiman puntos de bajada y subida indistintamente si son viajes o etapas de este.

Munizaga et al (2010) proponen refinar la metodología propuesta por Trépanier et al (2007) buscando no el paradero más cercano a la siguiente transacción si no aquel que representa el menor tiempo generalizado de viaje (Mora 2010), considerando tanto el tiempo de viaje como la caminata entre el punto estimado de bajada y la posición de la siguiente subida. Para ser consistente con la teoría de comportamiento de viajeros, se considera una penalización del tiempo de caminata equivalente a 2,5 veces el tiempo de viaje en bus.

2.2.2 Distintos medios de pago en la metodología

Para las distintas formas de pago y transporte, existen variaciones en la metodología propuesta por Munizaga et al (2010). Por ejemplo, cuando un usuario valida en una zona paga, se desconoce qué bus toma, pero se conoce la coordenada de subida y un subconjunto de buses asignados a esa zona paga, por otro lado, cuando toma el metro, tampoco se conoce qué tren tomo o que combinaciones hizo. Cuando un pasajero toma un bus, este puede tener o no asignado el servicio identificado que está entregando. Para cada una de las situaciones, se utilizan variaciones de la metodología propuesta, pero en todas se minimiza el tiempo generalizado de viaje.

2.2.2.1 Buses

Transantiago tiene metodologías implementadas para determinar qué servicio está siguiendo cada bus y así medir la tasa de cumplimiento. Por lo tanto es posible conocer el servicio, rutas y paraderos de estos buses. Sin embargo, no son completamente efectivas y en algunos los casos, no se logra saber qué servicio está realizando el bus.

Cuando el pasajero sube al bus y realiza una validación exitosa con su tarjeta bip!, es posible identificar ese bus a través del id del identificador y seguir su posición usando información de los pulsos GPS. De este modo se puede seguir al pasajero a lo largo de la etapa. Luego se estima el punto de baja minimizando el tiempo generalizado de viaje hasta la siguiente transacción.

El subconjunto de punto evaluados como candidatos a bajada, son los puntos de la trayectoria que están dentro de un radio de 1000 metros de la siguiente transacción, este valor es un parámetro que representa la distancia caminable para los usuarios.

2.2.2.2 Metro

La metodología aplicada en el caso de transacciones en estaciones de Metro funciona de la misma forma para predecir las estaciones de bajada, pero en este caso a diferencia del caso anterior, se desconoce el vehículo al que el pasajero se subió y su trayectoria. Para hacer frente a este problema, se estima ruta como aquella con menor tiempo de viaje.

En este cálculo, no se consideran los tiempos generalizados de caminata dentro de la estación o trasbordo en las estaciones, solo se considera el tiempo en el tren.

2.2.2.3 Zona Paga

En el caso de Zonas Paga, el problema es algo más complejo ya que muchas ya que muchas de ellas ocupan varios servicios, por lo que no es posible determinar a priori qué bus (ni a qué servicio) se subió el usuario. Para esto se supone que los usuarios van a elegir la línea o el conjunto de líneas que minimiza el tiempo generalizado de viaje según la definición utilizada en Chriqui et al (1975).

Es un supuesto que el pasajero conoce el set de servicios que puede tomar. Este set se define por la intersección entre los que tienen servicio en la zona paga y los que pasan a menos de 1000 metros de la siguiente transacción. El pasajero no sabe cuánto demorará a priori el viaje, pero en esperanza minimiza su tiempo generalizado de viaje.

El tiempo generalizado esperado de viaje es la suma del tiempo esperado de espera, el tiempo esperado de viaje y el tiempo esperado de caminata. Estos tiempos son determinados de la siguiente forma.

La esperanza del tiempo de espera se calcula en base a la frecuencia, suponiendo llegadas de buses con distribución uniforme

- Esperanza de tiempo de espera: $\left\{ \frac{k}{f_1}, \frac{k}{f_2}, \frac{k}{f_3}, \dots, \frac{k}{f_n} \right\}$

Donde k : constante de espera = $\frac{1}{2}$

f_i : frecuencia servicio i (teórica), estimada a partir de la velocidad promedio observada de la ruta, en horario establecido para zona paga.

Los tiempos de viaje esperado son obtenidos con mediciones utilizando los pulsos GPS (Cortés et al 2010).

$$\min_x TVE(r, b) = \min_x \left\{ \frac{k + \sum_i tv_i \cdot f_i \cdot x_i}{\sum_i f_i \cdot x_i} \right\}$$

$$s. a. \quad x_i = \{0,1\}$$

El Tiempo de caminata es calculado considerando la distancia desde el paradero a la siguiente transacción y una velocidad de caminata obtenida de la bibliografía (Mora 2010). Este caso se diseñó pero no se implementó.

2.2.3 Datos utilizados por MESB

La información original disponible se encuentra organizada por día, horario, operador y tipo de servicio (alimentador o troncal), en las siguientes 5 carpetas.

- **BIPS:** Registro de transacciones BIP realizadas en un período de tiempo dado, incluyendo campos de información como el número de identificación de la tarjeta (ID), fecha, hora en que se realiza la transacción (tiempo), sitio de la transacción (patente bus, validador zona paga, estación de metro), monto de la transacción, número de secuencia de transacción, tipo de transacción (uso, carga o consulta. Solo fueron entregadas las de uso), tipo de contrato (escolar básico, media, superior, normal).
- **Posicionamiento:** Información GPS por placa patente única del bus (PPU), incluye los campos latitud, longitud, fecha y hora, PPU, concesionario, ignición (motor de bus prendido (sí/no)). Cabe mencionar que los registros GPS son emitidos usualmente cada 30 segundos, sin embargo también existen emisiones en otros intervalos de tiempo, principalmente cada 10 y 300 segundos para situaciones de emergencia o motor apagado respetivamente.
- **Actual programa de operación:** Programa de operación para el sistema de transporte de la ciudad, vigente para el período en estudio, en el cual se encuentran unidad de negocio, frecuencias programadas, velocidad media por servicio-sentido en cada período de asignación de servicio.
- **Asignación de servicio a buses:** Estimación realizada por Transantiago del servicio-sentido que realizó cada bus durante el período de estudio. Incluye la distancia recorrida, el tiempo en servicio, y velocidad media resultante para cada bus en cada servicio-sentido.
- **Paradas y zonas pagas:** Información sobre la ubicación y tipo de paradas, para estaciones de transbordo y paradas sencillas. Las zonas pagas presentan: código de parada, horario de operación, servicios que la utilizan, estimación del porcentaje de transacciones correspondientes a cada servicio, asignación de validadores a zonas pagas.

Para integrar toda la información en grandes bases de datos, es necesario procesar la información original y crear nuevas tablas que combinan distintos campos de información provenientes de las carpetas anteriormente citadas².

Tabla 3: Principales tablas utilizadas por la metodología.

Transaccion Bip	Estaciones de Metro	Rutas Metro	Paradero
ID Bip	Nombre estación	Inicio	id_Paradero
Tiempo	Línea	Fin	Servicio
Sitio(metro, bus, zp)	Posición	Pasos	Posición
pago		id_estacion	
operador		Estación	
posición		Posición	
servicio		Tiempos Operación	

Posicionamiento	Zonas Pagas	Posición_ZP
Patente	Validador_zp	cod_zp
tiempo GPS	cod_zp	Posición
Posicionamiento	cod_Parada	
Operador	Unidad de negocio	
	Horario	
	Servicio	

Fuente: Mora 2010

² Proceso desarrollado por el analista de sistemas Mauricio Zuñiga en el marco del desarrollo del Proyecto Bicentenario de Ciencias y Tecnología (PBCT). El detalle del procedimiento se encuentra reportado en el “Informe proceso de modificaciones base de datos Transantiago” documento interno del PBCT.

2.3 Mejoras a la Metodología (MESB)

Con el fin de obtener una mejor estimación de la demanda de transporte, se realizaron mejoras a la metodología existente. Estas mejoras se realizaron de dos formas, la primera como un post procesamiento de las estimaciones de subidas y bajadas, mientras que aquellas que no se pudieron implementar de este modo, se implementaron dentro la metodología existente. Los cambios fueron desde agregar nueva información a la salida del software hasta implementar partes faltantes de la metodología.

Los cambios realizados dentro de la metodología existente se explican en este capítulo.

2.3.1 Filtrado de pulsos GPS

Una metodología para filtrar pulsos GPS es fundamental y esta ya existía, pero se modificó, agregando nuevos filtros y ajustando parámetros y se diseñó así los siguientes filtros para aceptar un pulso GPS como correcto:

- El pulso debe estar dentro de Santiago
- La transacción debe estar entre 2 pulsos con menos de 31 segundos de distancia
- La transacción debe estar al menos a 20 segundos del pulso anterior o posterior.
- 2 pulsos no pueden estar a más de 500 metros.
- Un pulso no puede estar a más de 100 metros de la ruta del bus (solo en caso de bus con servicio).

2.3.2 Implementación del módulo de Zonas Pagas

Las transacciones en zonas pagas representan solo el 5,9% del total de las transacciones validas del Transantiago. Están instaladas en los lugares de mayor demanda, siendo fundamental considerarlas en la estimación de la demanda de transporte.

Utilizando la metodología propuesta en la revisión bibliográfica, esta se implementó de modo que se pudiera reconocer en detalle que bus tomó cada pasajero. Así, del conjunto de servicios que minimizan el tiempo generalizado esperado de viaje, se seleccionan todos los que han pasado frente a la zona paga desde 20³ segundos después del bip! hasta 30 minutos después. Un bus ha pasado frente a la zona paga si tiene un pulso GPS a 100 metros de la zona paga, Si el bus se detuvo en la zona paga, este con certeza tendrá al menos un pulso GPS y en la práctica se ven acumulaciones de pulsos en estos lugares.

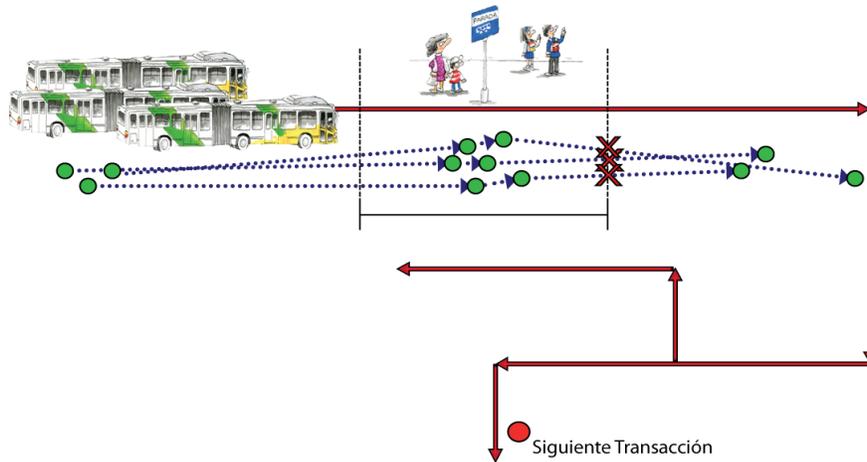
En Ilustración 3 se grafica la metodología utilizada en las Zonas Pagas. Se considera como instante de inicio del viaje cuando el bus se aleja a mas de 100 metros de la zona paga (marcado con cruz

³ Esos 20 segundos son para asegurar que el pasajero realmente pudo tomar el bus y este no halla partido.

rojas). Se halla o no subido antes el pasajero al bus, debido a que los buses esperan en las zonas pagas para evitar bunchs⁴ y coordinar las frecuencias.

Cabe destacar que en la decisión del pasajero no se consideran los perfiles de carga, se supone que este tomará el primer bus del conjunto que minimiza su esperanza de tiempo generalizado de viaje.

Ilustración 3: Esquema de funcionamiento del módulo de zonas pagas



2.3.2.1 Dificultades, Zona paga en Estación de intercambio modal.

Las zonas de intercambio modal presentan como problemática particular que están bajo techo, imposibilitando el funcionamiento adecuado de los GPS, donde se pierde el rastreo del bus. Para corregir este problema se buscan pulsos GPS de buses asignados a la zona de intercambio modal a 500 metros de diámetro en vez de 100, y se monitorea la entrada y salida de buses dentro de esa área.

Las zonas de intercambio modal fueron marcadas manualmente.

2.3.2.2 Dificultades, Zona Paga en inicio de ruta.

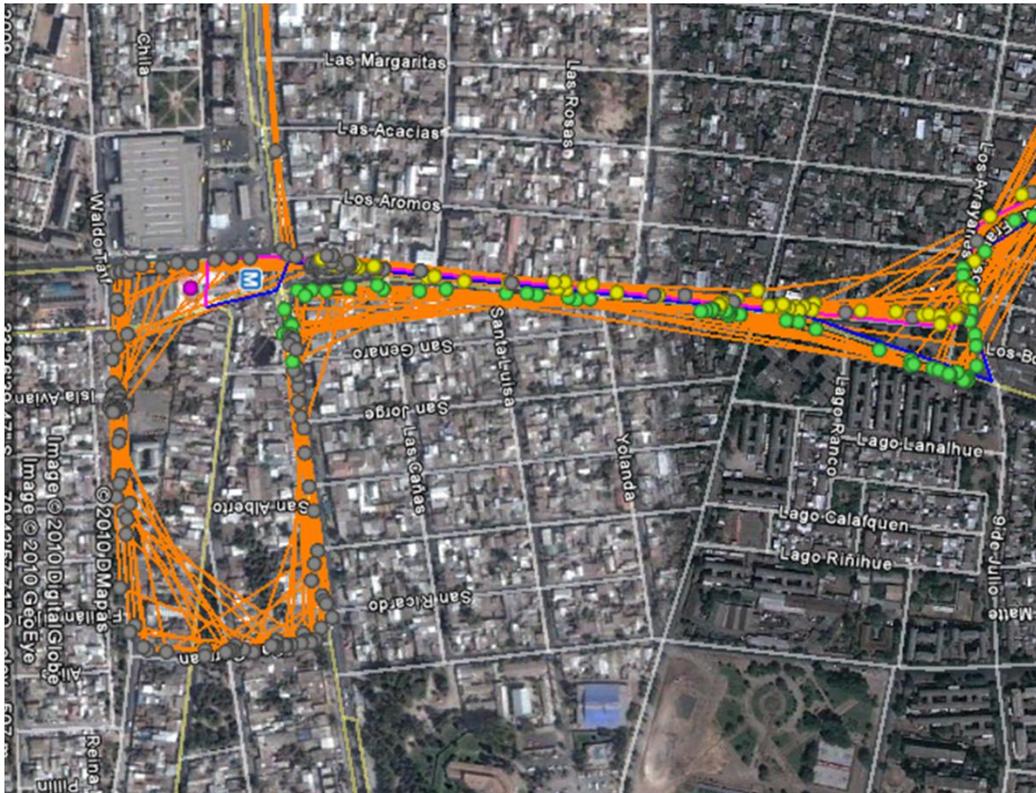
Debido las imperfecciones en la metodología de Transantiago para asignar servicio a buses y pulsos GPS, en algunos inicios de la ruta los pulsos GPS asociados al bus no tienen servicio asignado, esta situación se puede apreciar en

Ilustración 4, donde en la zona paga ningún bus con servicio asignado pasa por ella. Al utilizar la metodología como fue descrita, las zonas pagas cercanas a los inicios de ruta y filtrar los buses con servicio asignado a esa zona paga, dará como resultado un conjunto vacío. Para tratar esta

⁴ Se llama bunching o apelonamiento de buses al fenómeno que se produce cuando dos o más buses del mismo servicio viajan juntos.

situación las zonas pagas que están a 500 metros del inicio de la ruta, se seleccionan además los buses sin servicio que pasaron por la zona paga y se ven secuencialmente los pulsos GPS en adelante, hasta encontrar un servicio asignado. Si coincide con los de la zona paga, ese bus es considerado dentro del conjunto de elección del usuario.

Ilustración 4 : Zona paga a menos de 500 metros de inicio de ruta



- Línea naranja: una Pulsos GPS de un bus
- Punto Amarillo: Recorrido de bus en sentido contrario
- Punto Gris: Pulso GPS sin servicio asignado
- Punto Verde: Pulso GPS con servicio asignado
- Punto Morado: Zona Paga

2.3.3 Parámetros utilizados en la MESB

Para MESB, existe una serie de parámetros, estos parámetros fueron discutidos y calibrados con el fin de obtener mejores resultados, en Tabla 4 se presentan los parámetros utilizados en la estimación utilizada en la muestra utilizada

Tabla 4: parámetros utilizados por MESB

Parámetro	Unidades	Descripción
1000	Metros	Distancia considerada caminable
2.5		Costo relativo entre caminar y utilizar el bus (Tiempo Generalizado)
1.17	m/s	Velocidad de caminata (Tiempo Generalizado)
20	metros	Tamaño de zona paga
1000	Metros	Radio máximo para encontrar rutas en zonas pagas
40	Metros	Radio inicial búsqueda paraderos sin servicio
400	Metros	Radio máximo búsqueda paraderos sin servicio
40	Metros	Distancia umbral máxima para considerar un paradero dentro de la ruta
100	Metros	Distancia umbral máxima para considerar una transacción este fuera de ruta
50	Metros	Metros que se consideran validados para asignar una transacción al siguiente paradero en la ruta y no el ya pasado.
20	Segundos	Tiempo máximo entre transacción y pulsación mas cercana (Filtrado)
31	Segundos	Tiempo máximo entre dos pulsos GPS (Filtrado)
500	Metros	Distancia máxima entre dos pulsos GPS (Filtrado)
100	Metros	Distancia a ruta de la primera posición GPS (Filtrado)
30	Minutos	Tiempo máximo de espera por bus entre etapas de viaje

2.3.4 Cambios en información en Salida

Las estimaciones de subida y bajada son hechas a nivel de paradero, pero cada paradero corresponde a distintas zonas de Santiago. Dependiendo del nivel de abstracción, puede ser desde zona Etraus a unidad de negocio. A cada paradero se le asignó una zona a través de un método geométrico, sin importar si el paradero o estación estaba justo entre dos o más zonas, se asignó solo una. Así cada subida y bajada, además tiene la zonificación a la que pertenece, realizando agrupaciones de demanda por zona.

La entrada para el método geométrico fueron los polígonos disponibles en Transantiago.

3. Metodología

La estimación de la demanda de transporte se construye sobre la estimación de subidas y bajadas de los pasajeros en el sistema de transporte público, se realizan rutinas para mejorar la estimación y medir la calidad de esta, intentando corregir sesgos y errores. Para comprender mejor la metodología se definieron nuevos conceptos además de los definidos en Mora 2010, estos se disponen a continuación.

Definiciones

- **Tasa de evasión** se define como el porcentaje del total de etapas de viaje efectuadas sobre buses, en que no existió una validación o ésta no fue exitosa.
- Una **Línea** es un servicio de transporte del Transantiago, este se caracteriza por tener una ruta y paraderos definidos, una línea puede tener varios vehículos. En particular en el caso de Transantiago puede ser Una línea de Metro o Bus.
- **Matriz Origen-Destino:** matriz con demanda agregada de transporte, esta al menos debe tener dos dimensiones, origen y destino para los viajes. Puede ser desagregada (por ejemplo temporalmente) dependiendo del modelo a utilizar
- **Pasajero:** cliente que está recibiendo un servicio de transporte. Este es identificado por su Bip! través de sus transbordos.
- **Zona:** un polígono geográfico que encierra un área terrestre, por ejemplo una zona, puede ser "Sector Oriente" o menos agregada "Vitacura".
- **Factor de Expansión:** Representatividad de una muestra dentro de un muestreo. Se utiliza para muestreos no homogéneos.
- **MESB:** Metodología de estimación de subidas y bajadas, utilizado para referirse a la metodología implementada en la división de transporte
- **Caso Representativo:** etapas o viajes con una estimación válida y utilizadas para expandir aquellos casos en que no se pudo realizar la estimación

Motivación de Factores de Expansión

Al realizar un muestreo con distintos tipos de individuos para estimar una variable aleatoria, en caso de conocer los totales reales de cada uno de los tipos de individuos, puede balancearse la muestra para obtener un mejor estimador.

Por ejemplo al realizar un estudio de mercado para un nuevo producto, se quiere estimar el tamaño del público objetivo, se encuestan 100 hombres y 50 mujeres, mientras que la relación

real de hombres y mujeres es 50% y 50% respectivamente, claramente esta medición no cumple los supuestos de IID, para corregir este efecto, se aplican factores de expansión.

Si la población es N , el factor de expansión para un hombre será $N/(100*2)$ y para una mujer $N/(50*2)$, asignándole una representatividad a cada uno de los encuestados. En particular es importante realizar esta expansión si se notan tasas de aceptación distintas entre los 2 grupos .

Esta metodología es aplicada en las mediciones para la gestión del transporte público para expandir las mediciones hechas en terreno. Existen distintas formas de expandir la muestra, la última tendencia indica que la mejor opción es expandir por características demográficas del encuestado. (Ortúzar et al 2008)

En la estimación de subidas y bajadas no se pose información demográfica de los usuarios y no se pudo aplicar la metodología utilizada en la EOD 2001, debido a que se poseen orígenes de viajes, no hogares. Debido a la diferencia en la naturaleza de la información, pero para hacer un símil con esta se utilizó el origen del viaje y la hora en que se realizó la transacción, la metodología aplicada está detallada más adelante en el ítem 3.2.

Volumen de datos de entrada

Para estimar la matriz origen destino previamente se utilizó la salida del programa que estima el paradero de bajas y subidas de las etapas, para luego agrupar aquellas etapas que corresponden a un viaje. Para estimar las subidas y bajadas, se utiliza la metodología de estimación de subidas y bajas (MESB) de la división de transporte de la Universidad de Chile.

Debido al volumen de datos que acumulan en una semana, procesarlos en un motor de bases de datos normal como son los basados en lenguajes SQL y modelos entidad-relación sería muy ineficiente. Esto se debe a que la tabla resultante de estimar subidas y bajadas para una semana tiene un tamaño de 15 gb y que los registros deben relacionarse con el siguiente o anterior, como por ejemplo al encontrar todas las etapas de un viaje. Estas operaciones son bastantes complejas a nivel de uso CPU y de lenguaje en un motor SQL.

Para estimar que plataforma de datos utilizar se realizó un cálculo con varios supuestos. Si cada transacción toma $\log(n)$ para encontrarse, para encontrar las transacciones de cada usuario en un caso muy eficiente tomaría $(\text{Días}) \times (\text{Número de usuarios}) \times \log(\text{Transacciones}) = K \times 240.000.000$, mientras que realizar una lectura secuencial tomaría 35.000.000, en la práctica la metodología implementada toma 30 horas en un Pentium Dual Core SU4100, de haber optado por un motor SQL, este proceso tomaría del orden de 300 horas.

Se optó por crear una plataforma de datos que fuera eficiente en el uso de memoria y que permitiera trabajar eficientemente con el subconjunto de etapas correspondientes a solo un usuario, un problema bastante más pequeño que trabajar con toda la base de datos. Para encontrar eficientemente las transacciones los datos se cargan en formato CSV, y están ordenados

por ID de tarjeta bip y timestamp (fecha y hora) de transacción, pudiendo así procesar secuencialmente la base de datos evitando búsquedas en la tabla completa.

3.1 Transformación de salida del programa

Con el fin de mejorar la estimación de la demanda de transporte, se realizan transformaciones a los resultados obtenidos con la MESB, estas transformaciones se realizan en un post proceso de los datos y serán integrados posteriormente a esta.

Las distintas transformaciones previas a la estimación de la demanda del sistema de transporte, son descritas a continuación.

3.1.1 Zonas de influencia

Con el fin de obtener datos agregados, a cada paradero de subida y bajada se asigna una zona, la cual puede ser una comuna o zona ESTRAUS entre otras. Esta es asignada geoméricamente con las coordenadas GPS del paradero utilizando el polígono que describe esta zona. Al asignar una zona geográfica a un paradero, se sesga la muestra debido a que un pasajero con punto de bajada entre dos o más zonas geográficas, podría tener como destino cualquiera, este problema se consideró particularmente importante para las estaciones de metro, debido al volumen de pasajeros que utilizan este servicio.

La importancia de este efecto es que los datos a nivel agregado serán utilizados más tarde en los factores de expansión.

Para eliminar este efecto, se definieron áreas de influencia para las estaciones de metro, en que una zona está dentro del área de influencia de una estación de metro si está a menos de 1000 metros y no hay otra estación de metro más cercana. También se consideró la existencia de obstáculos geográficos como ríos y grandes terrenos cerrados.

Las subidas y bajadas, se reparten homogéneamente en las zonas de influencia de la estación. Se decidió utilizar este criterio por dos razones: Es el que utiliza SECTRA para las estaciones de metro entre comunas y no se poseía más información sobre los usos de los suelos ni mediciones en las estaciones.

En anexo A, se entrega un listado de las zonas de influencia de cada estación de metro. Estas tablas fueron hechas manualmente con la proyección de los polígonos sobre Google Earth.

3.1.2 Criterio de Destino o Transbordo

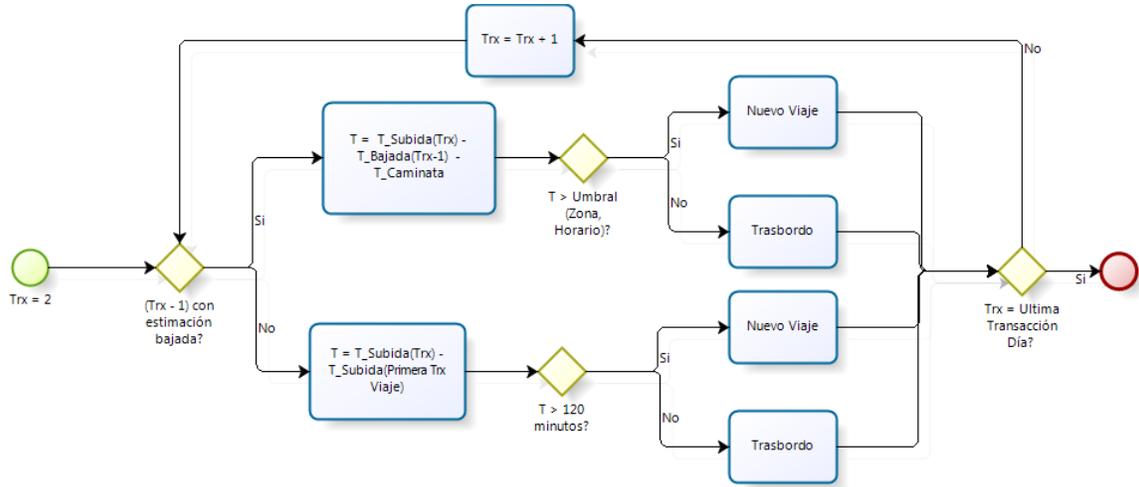
Con la metodología existente se estiman paraderos de bajada y subida indistintamente si son viajes o etapas, mientras que para generar la matriz origen destino, se requieren los orígenes y destinos de los viajes.

En este trabajo se decidió aplicar un criterio simple para distinguir transbordo de un viaje con etapas de un punto de destino donde el usuario va a realizar una actividad.

Si el tiempo entre la hora estimada de bajada y la siguiente transacción es mayor a un cierto umbral, se considera que es un destino donde el usuario realizó una actividad y a continuación hay un nuevo viaje. Por el contrario, si el tiempo es menor se asume que se realiza un transbordo. Este umbral es la función de tres parámetros: La hora de la siguiente transacción, la comuna de la siguiente transacción y de la distancia caminada entre la bajada y subida estimada. Este se descompone en dos factores, el primero es un parámetro obtenido de una tabla por comuna y refleja el tiempo de espera estimado en el paradero, el cual se estableció en 30 minutos para todos los casos en espera de ser calibrado en futuras investigaciones. El segundo factor estimación del tiempo de caminata entre la bajada y la subida, este tiempo se estima como el desplazamiento en línea recta desde la bajada a la subida estimada, la velocidad de caminata es de $\frac{1.5 m}{seg} * \frac{\pi}{4}$, donde 1,5 es la velocidad promedio lineal de una persona (Mora 2010) y $\frac{\pi}{4}$ que corresponde a un factor de corrección dado que en la ciudad, no siempre se puede caminar en línea recta correspondiente a el valor medio de la función $\text{Cos}(x) + \text{Sin}(x)$ entre 0 y $\frac{\pi}{4}$.

La metodología existente no siempre logra estimar los paraderos de subida o bajada, siendo imposible estimar tiempo entre la subida y la bajada, para estos casos se utiliza un segundo criterio, en que si la primera transacción del viaje fue hace más de dos horas de la hora de subida, se considera un nuevo viaje, en caso contrario, una nueva etapa del mismo viaje.

Ilustración 5: Diagrama de flujo de criterio para de discriminar entre trasbordo y destino.



Debido a que no era el fin de esta investigación no se implementaron mejores criterios, tales como tomar en cuenta la secuencia de servicios tomados, como repetición de servicios o múltiple uso del metro. Para mejorar este criterio se pueden tomar dos líneas de trabajo. La primera consiste en calibrar los tiempos de espera máximo en función de las frecuencias y usos de suelo. La segunda consiste en una nueva metodología que considera el número de buses que pasan por el paradero de subida. Si han pasado muchos buses que servían al usuario entre el momento estimado de llegada al paradero y la siguiente transacción, significa que el pasajero realizó alguna actividad y no un trasbordo. El tiempo estimado de caminata se puede refinar estimando la trayectoria de caminata. Estas dos últimas metodologías propuestas podrían mejorar el criterio, pero requerirían de mucho tiempo de CPU. Este tema está siendo abordado por una tesis de Magister.

3.1.3 Múltiple Pagos

Existen usuarios que realizan múltiples pagos con una tarjeta bip!, en un mismo servicio, dentro de una pequeña ventana de tiempo, para pagarle a un tercero o por error. Sea cual sea el caso, para evitar subestimar la demanda debe considerarse en la estimación de la matriz como más de un pasajero.

La metodología existente detecta estos casos como pasajeros se bajan y se suben rápidamente en el mismo bus, algunos repiten este comportamiento a lo largo del día. En Ilustración 6 se representa gráficamente esta situación.

Ilustración 6: Efecto de múltiple pago en MESB



Para aquellos casos en que se detecta un múltiple pago, se consideran viajes grupales y se considera como etapa representativa la ultima transacción del pago múltiple, transformando la situación representada en Ilustración 6 a Ilustración 7, Se consideraron dos casos distintos de múltiple pago, el primero representa aquellos viajes grupales como criterio de detección de múltiple pago, se utilizó la coincidencia en el id de lugar de pago, por ejemplo, patente del bus, estación de metro o id de zona paga, la metodología para detectarlos es detallada en Ilustración 8. Este primer filtro recorre las transacciones que realiza un usuario, al detectar que la siguiente transacción es en el mismo lugar que la evaluada, se anula la primera transacción y a la siguiente se pondera de modo que represente a ambas.

Ilustración 7: esquema de múltiple pago ya procesado

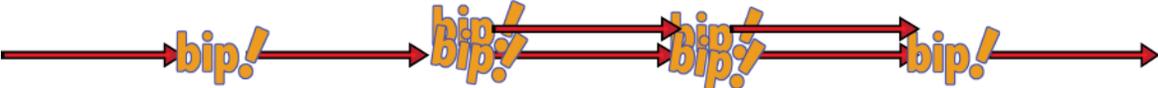
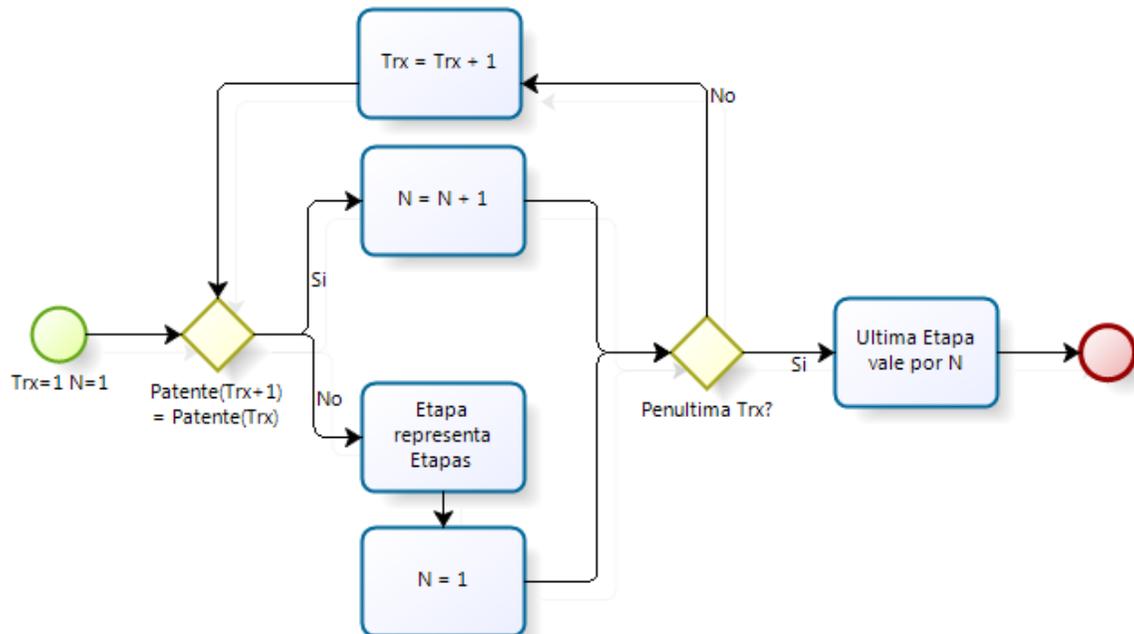
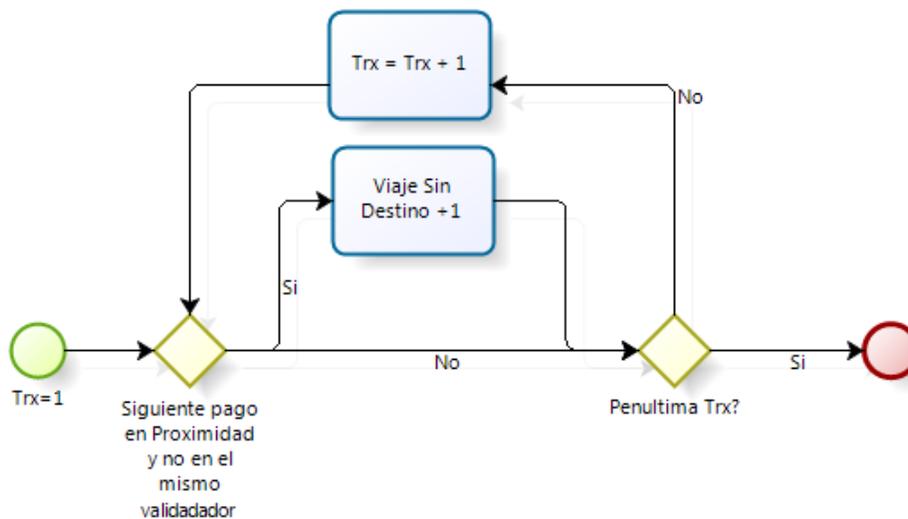


Ilustración 8: Detección de múltiples Pagos



En Ilustración 9 se muestra un segundo filtro y su objetivo es detectar y dan tratamiento a casos en que un usuario paga por otro, y luego toma su propio bus, en este caso, no se validó la tarjeta en el mismo bus, pero se validó más de una vez en una proximidad temporal y geográfica, en estos casos no se puede detectar el destino del viaje y se trata como viaje sin destino estimado.

Ilustración 9: Múltiples Pagos en Cercanía



3.2 Factores de expansión

La metodología para estimar subidas y bajadas tiene una alta tasa de éxito, de alrededor del 80% de las etapas, con una muestra correspondiente al 80% de la demanda registrada en la plataforma informática de Transantiago, intentar estimar la demanda de transporte sin realizar un tratamiento para dar cuenta de esta situación, provocaría una subestimación del sistema, pudiendo traer consecuencias en la atención al cliente y la calidad de vida de los usuarios de este.

La solución trivial es aplicar un factor de expansión a toda la muestra, pero la tasa de éxito en las estimaciones no es constante y aplicar este tratamiento, sobre estimaría la demanda en lugares con mejor tasa y la subestimaría en los casos contrarios, obteniendo una mala estimación de la demanda. Por lo tanto, se optó por aplicar factores que disminuyeran estos efectos, asignando a cada etapa y viaje correctamente estimados una representatividad en función de datos disponibles aun cuando no se pudo estimar el destino de una etapa o viaje.

3.2.1 Elección de datos para factores de expansión

Un requisito fundamental para la elección de los factores de expansión es que se basen en información disponible en la muestra y se conozcan los totales correspondientes a cada subconjunto, pueden subdividirse a niveles atómicos corriéndose el riesgo de dejar grupos sin un caso representativo y subestimando la estimación o seleccionar grupos demasiado grandes u homogéneos y mantener los sesgos en la muestra.

En la encuesta origen Destino del 2001, se encuestó al 16% de la población de Santiago en encuestas a hogares. Para complementar los datos obtenidos, se realizó una Encuesta Origen Destino de Intercepción (EODI) de buses del transporte público por cada bus encuestado, se encuestó uno de cada 2 pasajeros y uno de cada 3 buses, y se aplicaron factores de expansión para homogeneizar la muestra. Obteniendo factores de alrededor de 6 para cada encuestado. (Consortio Fernández et al 2001)

En el caso de la EODI⁵, se utilizó el servicio y el perfil de carga de los buses para encontrar los factores de expansión, esta información no está disponible en la metodología de estimación de subidas y bajadas, en particular los perfiles de carga y servicios para zonas pagas o buses sin servicio asignado, por lo que se debió utilizar datos distintos.

Se consideró importante incluir en los factores de expansión una dimensión horaria y una dimensión diaria, principalmente por la importancia que tiene estimar la demanda correctamente a lo largo del día. Se utilizaron los bloques horarios definidos por Transantiago para un día laboral, que son más extensos para los horarios de menos demanda. En Tabla 5 se muestran estos bloques horarios.

⁵ Encuesta Origen Destino de Intercepción, esta es realizada sobre buses

Tabla 5: Bloques horarios para día laboral del Transantiago

Bloque	Desde	Hasta
1	0:00:00	1:00:00
2	1:00:01	5:30:00
3	5:30:01	6:30:00
4	6:30:01	8:30:00
5	8:30:01	9:30:00
6	9:30:01	12:30:00
7	12:30:01	14:00:00
8	14:00:01	17:30:00
9	17:30:01	20:30:00
10	20:30:01	21:30:00
11	21:30:01	23:00:00
12	23:00:01	23:59:59

Para considerar distintos datos a utilizar como agrupaciones para estimar factores de expansión, estos deben cumplir requisitos, entre los cuales está ser conocido en la mayoría de los casos (99%), para saber a qué grupo pertenecerá muestra y además debe presentar varianza entre grupos en la tasa de estimación, dado que de lo contrario no tendría caso incluirlo. Dentro de los datos que cumplen los primeros requisitos se encuentran:

- Forma de pago
- Número de etapas
- Zona estraus de origen
- Paradero de origen

En cuanto a la agregación espacial, se decidió utilizar la zona estraus de origen, esto debido a que si se utilizara el paradero de origen, se tendría $10000 \cdot 12$ factores de expansión diarios provocando una alta probabilidad de no tener caso representativo para los horarios de baja demanda, por eso se prefirió utilizar la zonificación estraus de subida, que tendría $600 \cdot 12 = 7200$ factores diariamente. Se consideró importante incluir este dato en los factores de expansión, para tener certeza de que la matriz está balanceada al menos en los puntos de origen. Al realizar pruebas se encontraron grupos sin caso representativo, por lo que no se agregaron más datos o sub-agrupaciones.

En cuanto a los otros candidatos, como forma de pago o número de etapas no pudieron ser incluidos debido a que aparecían grupos sin casos representativos, en especial en los viajes de varias etapas.

Para obtener el total de los viajes y dar tratamiento a los grupos sin caso representativo, se creó un segundo factor de expansión de dimensiones bloque horario inicio viaje y día de la semana,

estos datos están disponibles en el total de las transacciones. De todos mundos este factor debe ser pequeño

Así se definió F1 = F1(zona origen, bloque horario de inicio de viaje, día de semana de inicio de viaje) y otro de solo F2 = F2(bloque horario de inicio de viaje, día de semana de inicio de viaje). El factor de expansión de cada viaje o etapa correctamente estimado se define como la multiplicación de los dos anteriores.

Factor por Origen y Hora y Día (F1)

Sea:

- i = Zona estraus origen del viaje
- j = Zona estraus del destino del viaje
- t = Bloque Hora Inicio Viaje
- d = Día del Inicio del viaje

$$F1(i, t, d) = \frac{\sum_j \text{Viajes}_{i,j,t,d}}{\sum_{j \neq \text{null}} \text{Viajes}_{i,j,t,d}}$$

Ecuación 1

Factor por Hora y Día (F2)

$$F2(t, d) = \frac{\sum_{i,j} \text{Viajes}_{i,j,t,d}}{\sum_{i \neq \text{null}, j \neq \text{null}} \text{Viajes}_{i,j,t,d} * FExp(i, t, d)}$$

Ecuación 2

Las transacciones, a las que no se les pudo asignar subida o no se encontró caso representativo son tratadas con este factor.

Existen casos sin paradero de origen y por ende sin zona estraus de origen pero si poseen coordenada de origen y pueden representar casos en que los paraderos no están ingresados en el sistema, o una parada en un lugar que no corresponde, estos casos son tratados con F2. Podría mejorarse marginalmente la estimación si a estos casos se les asignara geoméricamente la zona estraus a la que pertenecen, quitando así peso a F2 en beneficio de F1.

3.2.2 Transacción Única en el día

Las tarjetas que solo realizan una transacción en el día no permiten la estimación de paradero de bajada. Se decidió no darles un tratamiento especial e incluirlas en el factor de expansión como cualquier otro viaje al que no se le pudo estimar el destino, estos representan el 8,2% de la

muestra de viajes y pueden deberse a un cambio en la tarjeta bip! el uso de otro medio de transporte. Todos ellos efectivamente realizaron solo una transacción en el día, lo que no sabemos es si esto se debe a que:

- Realizaron un único viaje en el día y fue de una sola etapa.
- Realizaron un único viaje en el día y fue de una sola etapa en transporte público (El resto en otros medios de transporte).
- Realizaron más de un viaje en transporte público, pero con distintas tarjetas bip!.

Se pueden nombrar muchas causas pero el tratamiento para todas fue aplicar solo un factor de expansión como al resto de los viajes.

3.2.3 Información agregada sobre la estimación

Para determinar la calidad de la estimación, se generaron indicadores, los cuales son determinados de la siguiente forma:

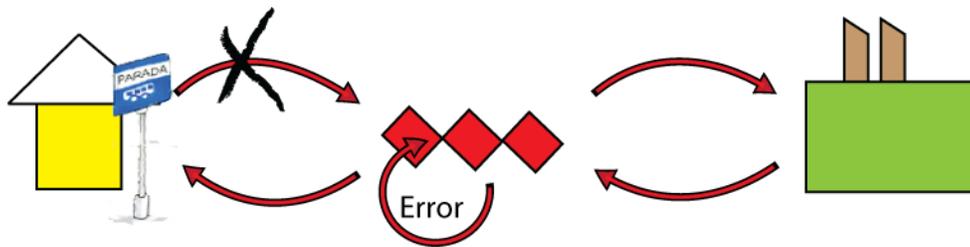
- Número de Viajes (NV)
- Numero de Etapas (NE)
- Número de viajes Estimados (NVE)
- Número de etapas Estimadas (NEE)
- Transbordos = $NE - NV$ (TR)
- Transbordos Estimados = $NEE - NVE$ (TRE)
- Tasa Estimación viajes = NE/NV
- Tasa Estimación Traslados = TRE/TR
- Tasa Estimación Etapas = NEE/NE

3.3 Evasión en la Metodología de Estimación de subidas y Bajadas

La evasión no es un problema menor, como se dijo en la sección de antecedentes, se estima en un 14% según las mediciones del año 2009. Hacen falta datos para poder estimar la demanda de los evasores. Además las mediciones de evasión son bastante pobres por lo que se decidió no considerar la evasión. Principalmente por que las mediciones son por etapa y no viaje.

Los dos principales efectos de la evasión sobre estimación de la demanda a través de datos operacionales son la sub estimación y la distorsión de la demanda, este efecto es el más grave y difícil de tratar. En Ilustración 10 se muestra un ejemplo de una persona que evade en la primera etapa del día. Al regresar tendrá una etapa con un desplazamiento nulo, dado que es su coordenada de origen.

Ilustración 10: Distorsión en demanda estimada con MESB



En los antecedentes se describieron los distintos tipos de evasores. En algunos casos, se puede realizar correcciones.

Aquellos que evaden en la primera etapa del día por no tener saldo, podrían identificarse con la información de las validaciones no exitosas sin embargo esta información no se registra, pero se podría comenzar a almacenar. Si la tarjeta es validada nuevamente lejos de donde apareció sin saldo, es posible que esto represente una evasión. Este caso se puede identificar sin realizar nuevas mediciones distintas a las obtenidas con la operación del sistema.

El evasor duro, que protesta porque el sistema es malo y no valida su bip! en ninguna de sus etapas de viaje, puede tratarse con factores de expansión que ponderen por la tasa de evasión, son relativamente fácil de tratar, porque solo causan una sub estimación de la demanda, no una distorsión de esta.

Aquellos que no pagan debido a que no pueden por que el bus está muy lleno, son un caso particularmente significativo en algunas horas y lugares. Subestimar la demanda en estos puntos es particularmente malo, dado que son paraderos con problemas. Las subidas en estos tendrían que medirse exógenamente para realizar las correcciones a la demanda.

Los dos casos anteriores requieren de mediciones en terreno y no están dentro del alcance de esta investigación.

Cualquiera de las soluciones propuestas, eliminan el síntoma pero no la causa, la verdadera solución es disminuir la evasión, las estrategias pueden ser variadas:

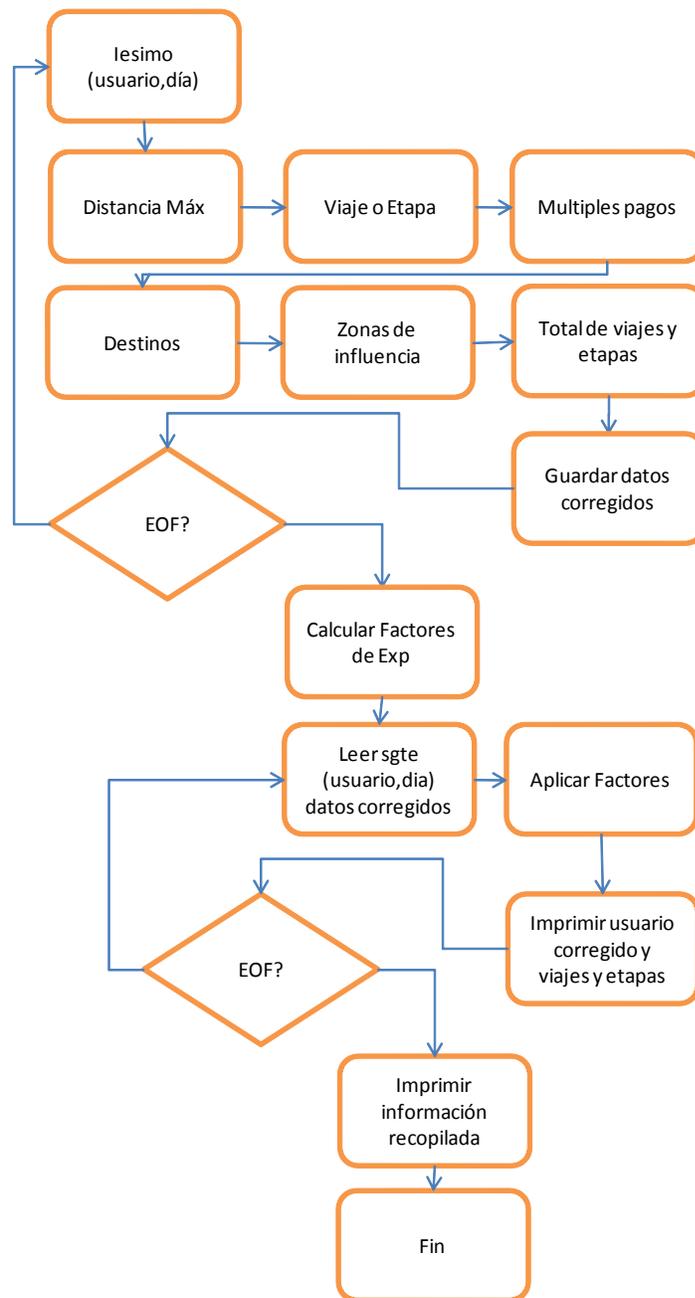
- Campañas publicitarias
- Explicar que con las validaciones se estiman los buses necesarios
- Aplicando multas y proveyendo de mas inspectores
- Aumentando los puntos de recarga

Estos son solo ejemplos, y no es un objetivo de esta memoria analizarlos.

3.3.1 Flujo de post proceso

A grandes rasgos, el flujo del procesamiento de las estimaciones subidas se presenta en Ilustración 11, se procesa de uno en uno, cada usuario y día de este, aplicándosele filtros y correcciones.

Ilustración 11: diagrama de post procesamiento de datos



4. Resultados Obtenidos

Se aplicó la metodología a una muestra correspondiente a 35 millones de transacciones y se obtuvieron los siguientes resultados.

4.1 Indicadores de estimaciones de subidas y bajadas

4.1.1 Tasa de Estimación

En la mayoría de los casos se puede estimar la las subida y bajada de una etapa de viaje. Las tasas de éxito en las estimaciones de la MESB con post procesamiento de datos se resumen en Tabla 7 y Tabla 6 y se pueden apreciar tasas de estimación constantes a lo largo de la semana, excepto para el día domingo, se puede resaltar que la tasa de estimación de etapas es mayor que la de viajes, más adelante se puede ver que se debe a que se estiman casi la totalidad de los trasbordos, mientras que los destinos de viaje son más difíciles de estimar.

Se cree que la disminución en la tasa de estimación en los fines de semana, se debe al uso de transporte mixto de transporte público y otros como transporte privado. Esto no es un gran problema, debido a que el fin de semana el sistema de transporte no está saturado haciendo más importante las frecuencias que los perfiles de carga al momento de realizar modelameintos.

Tabla 6: Tasa de estimación de etapas por día

Día	Estimado	Total	Porcentaje Estimado
lun	4857874	6021068	80.68%
mar	4872006	5989129	81.35%
mié	4873764	5988084	81.39%
jue	4790673	5916961	80.97%
vie	4901284	6158213	79.59%
sáb	2676018	3518450	76.06%
dom	1600226	2179183	73.43%

Tabla 7: Tasa de estimación de viajes por día.

Día	Total	Estimado	Porcentaje Estimado
lun	2752811	3756231	73.29%
mar	2780588	3736716	74.41%
mié	2772044	3730907	74.30%
jue	2716022	3687118	73.66%
vie	2778858	3871203	71.78%
sáb	1468288	2216929	66.23%
dom	869360	1392983	62.41%

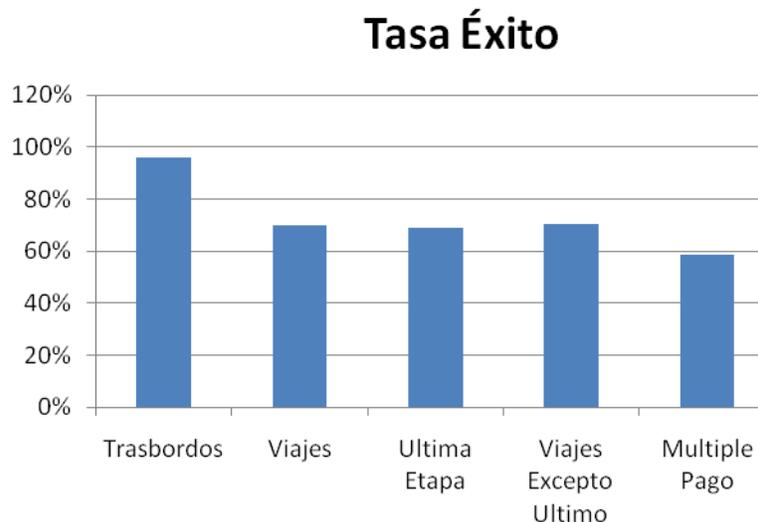
- Un viaje se considera correcto, si este tiene estimada la subida de la primera etapa, la bajada estimada de la última etapa y hora estimada de bajada de la última etapa.
- Una etapa se considera Correcta, si tiene estimada la subida, bajada y la hora de bajada.

4.1.2 Tasas de estimación en distintas instancias

Las estimaciones de el punto de bajada tienen distintas tasas de éxito, dependiendo de la situación en la que se dé, en Gráfico 1 se muestra que trasbordos se encuentra la mayor tasa de éxito en la estimación de bajadas, siendo cercano al 95%, mucho mayor a la tasa de estimación de viajes, la cual es de alrededor del 70%. Además es importante tomar en cuenta que la tasa de estimación de la última etapa del día es idéntica al de destino de viajes, validando el supuesto de regreso a casa.

Resultaría interesante evaluar el efecto de utilizar la primera transacción del día siguiente en vez de la primera del día para evaluar los cambios en la estimación de la demanda, como también resultaría interesante evaluar cuando estos dos puntos son idénticos.

Gráfico 1: Distribución de tasas en estimación de bajadas, con post procesamiento



- Tráspodo: estimación de bajada entre dos etapas de un viaje, se incluyen los múltiples pagos.
- Viajes: Estimación correcta de viajes, se incluyen los múltiples pagos
- Última etapa: Estimación correcta de última etapa del día.
- Viajes Excepto Último: es la resta entre los viajes y los última Etapa, se incluyen múltiples pagos.

4.1.3 Estimaciones por tipo de pago

La metodología para estimar las subidas y bajadas de etapas, está compuesto por cuatro módulos, de los cuales dos son para la estimación de bajadas de buses con validación dentro de ellos, otro para metro y el cuarto para validaciones en zona paga.

Las tasas de éxito en estimación de etapas se presentan en Tabla 8, en particular destaca la baja tasa de estimación para Zona paga, en particular para el día domingo lo cual no es muy relevante dado que los días domingo casi no operan las Zonas Pagas.

La tasa de estimación de la Zona Paga es la menor de todas y se debe a la complejidad que tiene la implementación de esta metodología, en particular en la detección del bus que tomó el pasajero.

Tabla 8: Tasas de estimación de etapas por forma de pago

Rótulos de fila	Valores		
	Suma de Estimado	Suma de Total	Estimado
BUS	17466544	22381487	78.04%
lun	2878290	3653436	78.78%
mar	2894521	3641816	79.48%
mié	2901479	3644024	79.62%
jue	2849601	3599256	79.17%
vie	2932438	3753065	78.13%
sáb	1855469	2492648	74.44%
dom	1154746	1597242	72.30%
METRO	9709335	11268559	86.16%
lun	1697075	1937897	87.57%
mar	1698098	1928871	88.04%
mié	1695069	1929403	87.85%
jue	1668888	1906359	87.54%
vie	1704755	1997315	85.35%
sáb	804395	998217	80.58%
dom	441055	570497	77.31%
ZP	1395966	2121042	65.82%
lun	282509	429735	65.74%
mar	279387	418442	66.77%
mié	277216	414657	66.85%
jue	272184	411346	66.17%
vie	264091	407833	64.75%
sáb	16154	27585	58.56%
dom	4425	11444	38.67%

4.2 Factores de Expansión

Los factores de expansión promedio presentados en Tabla 9, para cada bloque horario y día de la semana, se puede apreciar que son bastante más pequeñas que los utilizados en la EDO 2001, los cuales eran del orden de 6, reflejando la mayor representatividad en los datos. Para los días sábado y domingo el factor de expansión es algo mayor, al igual que para los horarios desde 23:00:01 a 23:59:59 y 1:00:01 a 5:30:00.

Tabla 9: Promedio de factores de expansión para viajes

Hora/Día	lun	mar	mier	jue	vie	sab	dom	Total
1:00:00	1.63	1.55	1.63	1.64	1.60	2.21	2.42	1.83
5:30:00	1.82	1.68	1.69	1.70	1.76	2.11	2.49	1.90
6:30:00	1.58	1.45	1.44	1.45	1.50	1.62	1.90	1.56
8:30:00	1.49	1.40	1.44	1.44	1.44	1.60	1.72	1.50
9:30:00	1.40	1.37	1.37	1.42	1.42	1.58	1.72	1.47
12:30:00	1.37	1.36	1.35	1.36	1.40	1.56	1.65	1.43
14:00:00	1.39	1.37	1.37	1.38	1.41	1.55	1.65	1.44
17:30:00	1.44	1.43	1.43	1.45	1.46	1.59	1.66	1.49
20:30:00	1.48	1.45	1.45	1.46	1.54	1.70	1.78	1.55
21:30:00	1.54	1.47	1.47	1.48	1.67	1.81	1.81	1.61
23:00:00	1.49	1.46	1.46	1.49	1.71	1.93	1.86	1.63
23:59:59	1.75	1.59	1.63	2.01	2.01	2.16	1.91	1.87
Total	1.52	1.46	1.47	1.52	1.57	1.78	1.87	1.60

Las Desviaciones estándar de los distintos factores de expansión para cada zona geográfica presentados en Tabla 10. Las desviaciones estándar con respecto al tamaño de los factores de expansión son grandes, reflejando considerables diferencias en las tasas de estimación dependiendo de la coordenada de origen. De esto se puede concluir que aplicar factores de expansión por origen geográfico ayudó a balancear la muestra.

Tabla 10: Desviaciones estándar de factores de expansión para viajes

Desvest	lun	mar	mier	jue	vie	sab	dom	Total
1:00:00	1.05	0.88	0.91	0.83	0.92	1.45	1.33	1.14
5:30:00	1.05	1.01	0.95	0.86	1.13	1.20	1.56	1.16
6:30:00	0.35	0.26	0.26	0.30	0.34	0.47	1.24	0.58
8:30:00	0.91	0.20	0.69	0.70	0.20	0.45	0.37	0.57
9:30:00	0.59	0.21	0.20	0.78	0.37	0.46	0.47	0.49
12:30:00	0.20	0.28	0.23	0.22	0.41	0.50	0.44	0.36
14:00:00	0.36	0.22	0.26	0.22	0.19	0.27	0.34	0.29
17:30:00	0.19	0.23	0.29	0.32	0.21	0.44	0.62	0.37
20:30:00	0.18	0.26	0.33	0.22	0.28	0.36	0.39	0.32
21:30:00	0.24	0.26	0.20	0.32	0.56	0.59	0.50	0.44
23:00:00	0.28	0.26	0.21	0.48	0.69	0.77	0.66	0.56
23:59:59	0.89	0.54	0.54	0.98	0.91	0.74	0.99	0.84
Total	0.62	0.46	0.50	0.60	0.62	0.75	0.87	0.66

En Tabla 11 se presentan los valores para F2 para viajes resultó pequeño en comparación con F1. Este es un poco mayor para los horarios de baja demanda y se debe a la falta de casos representativos para los viajes sin destino, se puede concluir que haber utilizado bloques horarios más grandes para los horarios de baja demanda fue mejor que haber utilizado una equipartición. Este factor de expansión balancea la muestra para aquellos casos en que no se pudo estimar el origen o no existía caso representativo para los viajes sin destino por lo que es bueno que resultaran considerablemente menores que F1.

Tabla 11: Factores de expansión F2

Valor	lun	mar	mier	jue	vie	sab	dom	Promedio
1:00:00	1.027	1.020	1.023	1.008	1.016	1.009	1.011	1.016
5:30:00	1.017	1.023	1.026	1.017	1.008	1.006	1.016	1.016
6:30:00	1.008	1.006	1.007	1.007	1.007	1.011	1.012	1.008
8:30:00	1.007	1.006	1.006	1.005	1.007	1.009	1.008	1.007
9:30:00	1.006	1.005	1.006	1.005	1.006	1.006	1.007	1.006
12:30:00	1.005	1.005	1.005	1.004	1.005	1.006	1.006	1.005
14:00:00	1.005	1.005	1.004	1.003	1.005	1.005	1.006	1.005
17:30:00	1.004	1.005	1.005	1.004	1.004	1.005	1.006	1.005
20:30:00	1.005	1.005	1.005	1.005	1.004	1.005	1.006	1.005
21:30:00	1.004	1.004	1.004	1.003	1.004	1.005	1.005	1.004
23:00:00	1.005	1.005	1.004	1.003	1.005	1.006	1.005	1.005
23:59:59	1.007	1.015	1.009	1.008	1.012	1.011	1.010	1.010
Promedio	1.008	1.009	1.009	1.006	1.007	1.007	1.008	1.008

4.2.1 Sin coordenada de origen

Los errores en la predicción de los paraderos de bajada no solo se deben a que los usuarios no se comportan según los supuestos de la metodología. También existen errores en los datos recibidos, en particular en las coordenadas GPS, las cuales tienen un rango de error de 10 metros y debieran ser pulsos cada 30 segundos y existen paradero no ingresados al sistema, lo que causa tener coordenada de subida, pero no zona o paradero.

Para estimar el paradero de subida, se busca el paradero con parada asignada al bus más cercano dentro de un radio. Debido a que hay paraderos no ingresados en las bases de datos, se forman conglomerados de transacciones sin paradero de subida asignado

En la Ilustración 12 las transacciones sin paradero son marcados con puntos rojos y azules. Se puede ver que están principalmente en la periferia de Santiago, donde se sabe de la existencia de paraderos no ingresados en las bases de datos. Por otro lado, también podrían ser pulsos GPS erróneos. Debido a que se desconoce la causa, estos viajes o etapas, fueron tratados con el factor de expansión F2. Estos casos representan el 0,5% de la muestra, siendo poco importantes para la estimación de datos agregados.

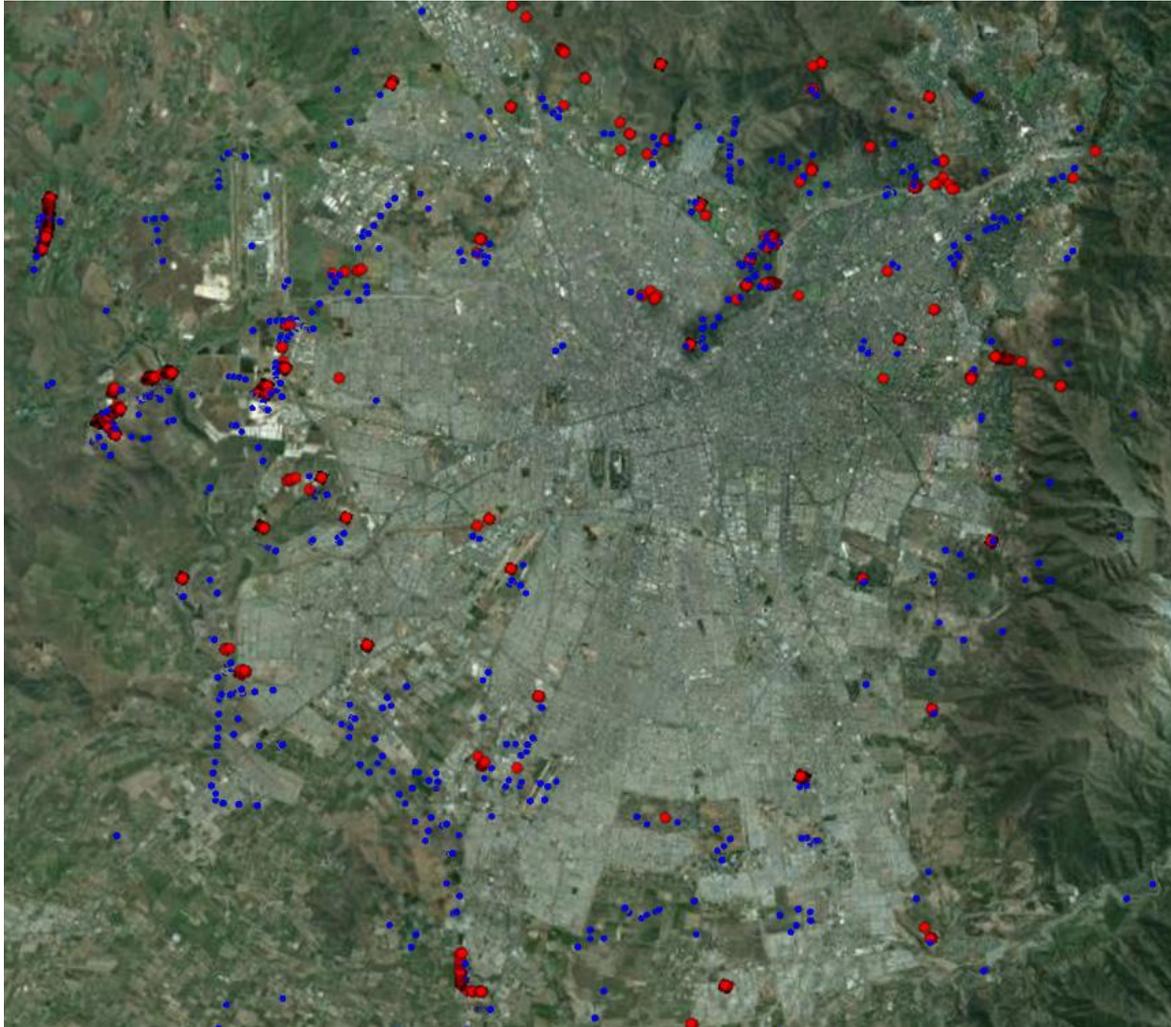
En la Tabla 12 se muestran los porcentajes de subidas no estimadas y el factor de expansión necesario para tratar estos casos, estos en general representan la mitad de F2, se puede concluir que en este factor la mitad de las causas es no tener caso representativo y la otra no tener coordenada de origen.

En total existen 101.752 casos en que no se pudo estimar paradero de origen y representan menos del 0,5% del total de las transacciones. De estos casos en que no se pudo estimar el paradero de subida 43.643 no poseían coordenadas de origen asociada, el resto no tenía un paradero habilitado a la línea cercano a la transacción, lo que se pudo deber a que el paradero no estaba ingresado en el sistema, que el bus paró en un lugar que no correspondía o que el pulso GPS asignado a la transacción era erróneo y no fue detectado por el proceso de filtrado.

Tabla 12: Transacciones sin paradero de origen y cantidad del factor de expansión teórico por esta causa.

<i>Día</i>	<i>Sin Origen</i>	<i>Total Viajes</i>	<i>Porcentaje</i>	<i>Factor Exp</i>
<i>lun</i>	17781	3738450	99.52%	1.00478
<i>mar</i>	17074	3719642	99.54%	1.00461
<i>mié</i>	16984	3713923	99.54%	1.00459
<i>jue</i>	13991	3673127	99.62%	1.00382
<i>vie</i>	18076	3853127	99.53%	1.00471
<i>sáb</i>	11059	2205870	99.50%	1.00504
<i>dom</i>	6787	1386196	99.51%	1.00492

Ilustración 12: Transacciones sin paradero asignado

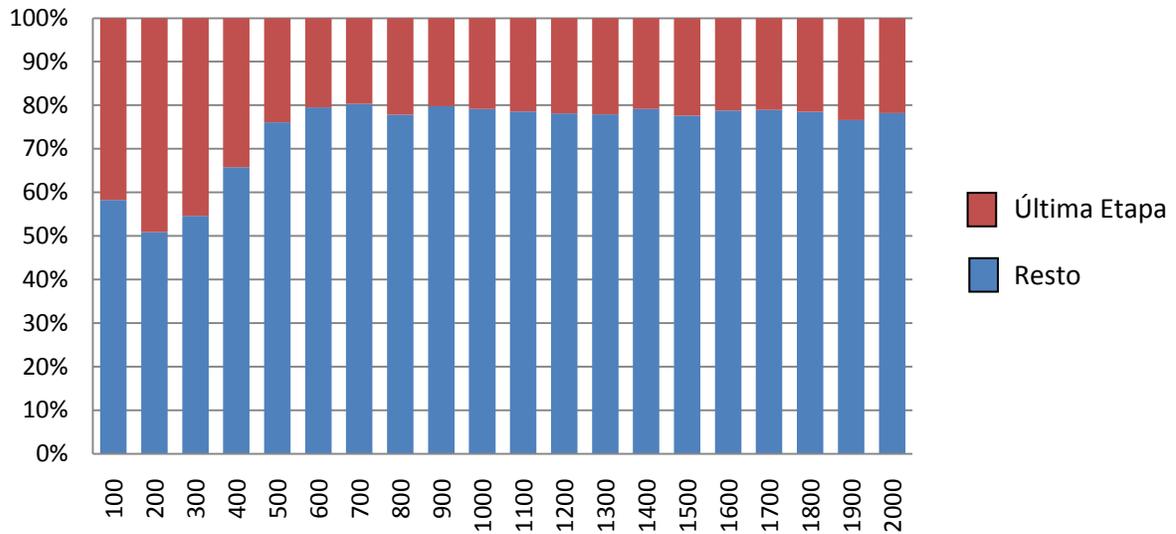


4.2.3 Evasión en primera etapa

Con el ejemplo de Ilustración 10 que explicaba el efecto de una evasión en la primera etapa del día. Para explorar esa situación se realizó un análisis, en que para cada etapa se calculó el desplazamiento, se agrupó cada 100 metros y se marcó con un flag si era la última etapa del día. En Gráfico 2 se comparan las porciones de desplazamiento de última etapa con el resto de los desplazamientos, para desplazamientos de más de 600 metros, la proporción es 20%-80% para últimas etapas y resto, mientras que en desplazamientos menores a 300 metros, hay un considerable aumento de últimas etapas, en relación con el resto, llegando a 50%-50%, mostrando que efectivamente puede existir una distinción en la estimación de la demanda. Estos números pueden representar tanto evasión, como simplemente el uso de otro medio de transporte, distinto al Transantiago, en la primera etapa. Este resultado se puede interpretar más como una

medida de insatisfacción, dado que agrupa la evasión y el uso de medios de transporte alternativos.

Gráfico 2: Última Etapa vs Resto cada 100 metros



De este análisis se concluye que efectivamente la evasión en la primera etapa puede distorsionar la estimación de la demanda. Para tratar esta situación se requieren datos que actualmente no existen, en el capítulo 3.3 se explica con más detalle esta situación.

También se analizó para qué servicios se realizaban viajes de menos de 300 metros en la última etapa, con el fin de encontrar posibles lugares donde los pasajeros están evadiendo o utilizan otro medio de transporte en la mañana y pueden utilizarse para medir satisfacción de los pasajeros, esta tabla se encuentra en el disco adjunto.

4.3 Estimación de la Demanda

La estimación de la demanda se realizó por paradero, pero con el fin de mostrar resultados obtenidos se resumió en bloques horarios y áreas geográficas.

El formato de la salida del software se puede ver en los anexos.

4.3.1 Estimación por Origen y Destino

En Tabla 13 se muestra la estimación del matriz origen destino agregado por área de Transantiago para el día lunes, esta se presenta en miles de viajes. La zona A correspondiente al centro de la ciudad, abarca la mayor cantidad de orígenes y destinos, presentando la mayor demanda de transporte.

Tabla 13: Matriz Origen-Destino por área en miles de viajes

	Destino	Centro	Norte	Occidente	Oriente	Sur	Sur Oriente	Total
Origen		A	B	I J	C D	E G H	D E F	
Centro	A	198	84	101 79	135 39	8 55 52	31 42 36	860
Norte	B	93	155	17 15	42 9	1 11 10	6 6 5	371
Occidente	I	109	18	156 24	56 12	1 9 15	7 6 5	417
	J	78	14	27 72	41 7	0 4 6	4 3 2	259
Oriente	C	148	38	48 37	183 32	5 27 22	33 31 31	637
	D	42	8	10 7	33 34	2 8 6	23 12 11	197
Sur	E	11	2	1 1	7 2	5 5 4	2 4 2	44
	G	62	12	9 4	34 9	6 74 25	7 10 9	262
	H	55	10	13 5	23 6	3 23 45	6 11 9	207
Sur Oriente	D	32	6	7 3	34 23	1 6 6	32 13 10	174
	E	45	6	6 3	33 12	4 9 10	15 45 21	209
	F	39	6	4 2	35 12	2 8 9	11 20 73	221
Total		913	357	398 252	656 197	40 239 210	176 205 215	3859

4.3.3 Estimación de Hora de Origen y Destino

En los Gráfico 3 y Gráfico 4 se presentan los histograma con las horas estimadas de inicio y fin de viaje, con el fin de contrastar el desfase que tienen principalmente en los horarios punta, donde un gran desfase en los picks, representa un tiempo de viaje grande. El pick del origen en la mañana es a las 7:30, mientras que el de llegada es a las 8:30, el de salida de la tarde es a las 18:30, mientras que el de llegada es a las 19:00, se puede apreciar que las horas de llegada en la tarde están mucho más distribuidas que las de salida.

Gráfico 3: Histograma de Hora de inicio de viajes

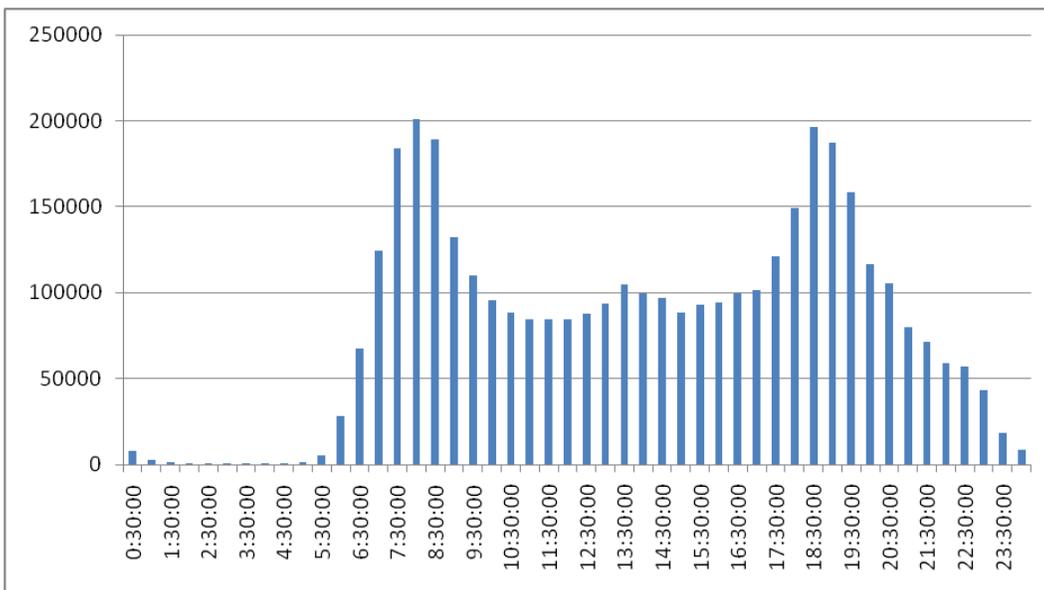
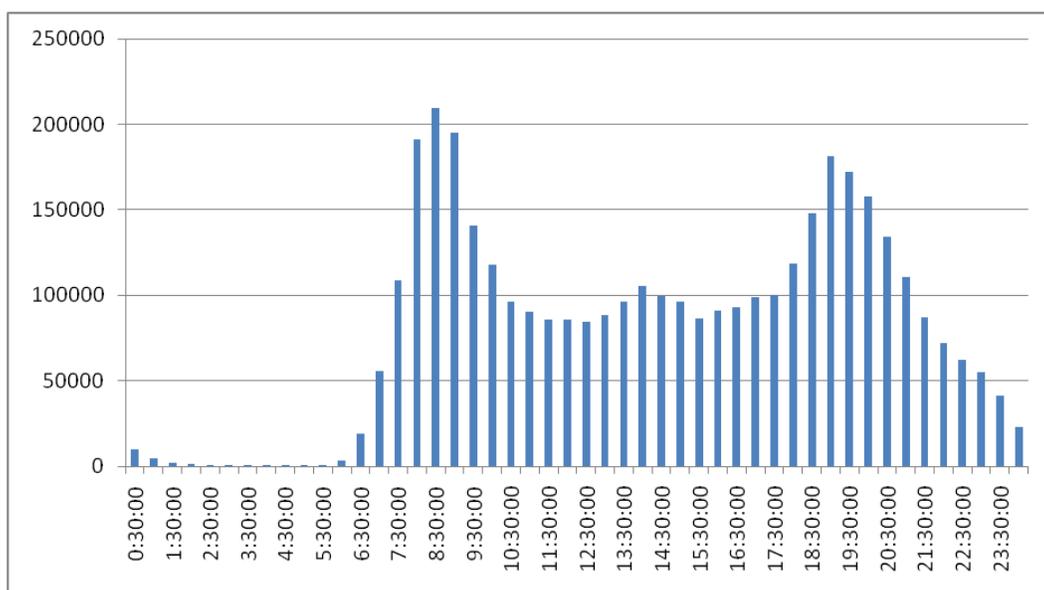


Gráfico 4: Histograma de Hora de fin de viajes



4.3.4 Estimación para distintos días de la semana

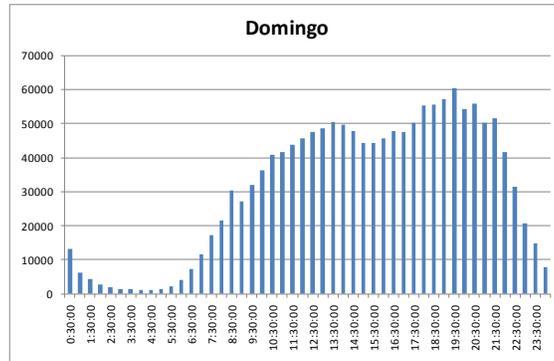
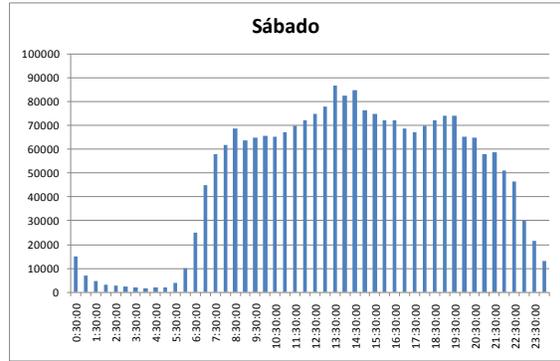
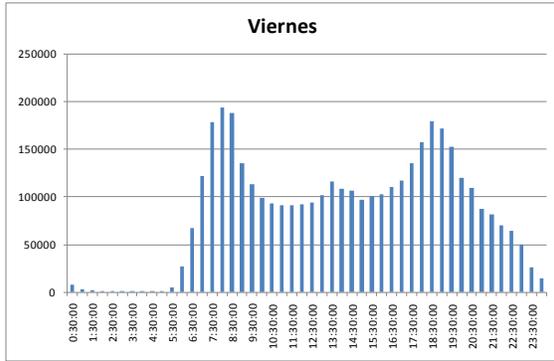
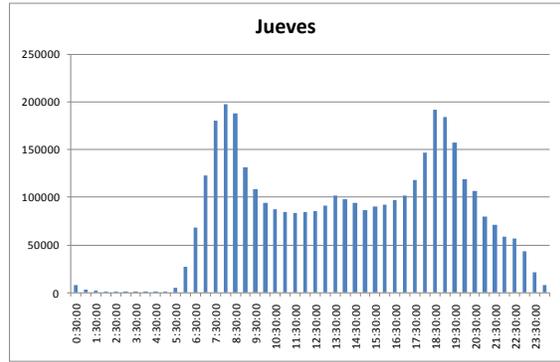
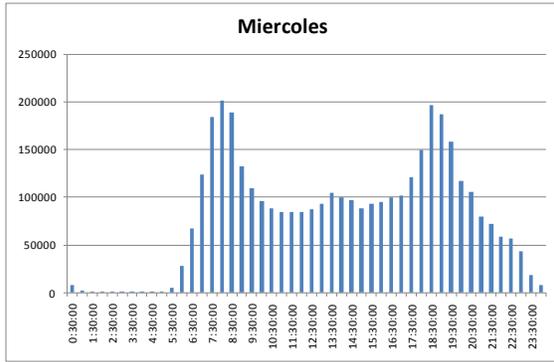
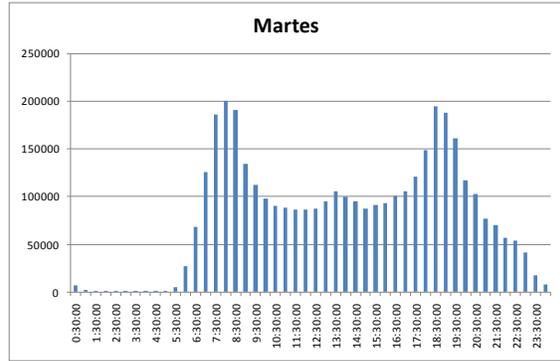
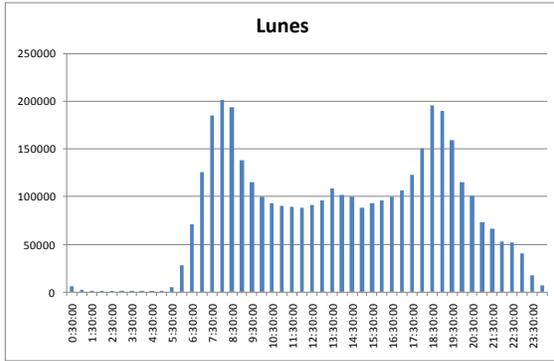
El comportamiento de la demanda de transporte no es idéntico para todos los días de la semana, en particular, los días lunes, martes, miércoles y jueves son muy similares, mientras que para los días viernes, sábado y domingo hay comportamientos distintos a lo largo del día, en Gráfico 5 se puede ver el comportamiento de la demanda para los distintos días de la semana.

En particular se destaca que el pick de la tarde es más distribuido que el resto de los días de la semana.

Para el día sábado y domingo, la demanda es completamente distinta a los días de semana, el sábado tiene un máximo a las 13:30 aunque la demanda es relativamente plana, mientras que el día domingo tiene los máximos a las 13:30 y 19:30.

También se puede apreciar que la demanda en horario punta el fin de semana no alcanza si quiera a la demanda en el horario de medio día en la semana.

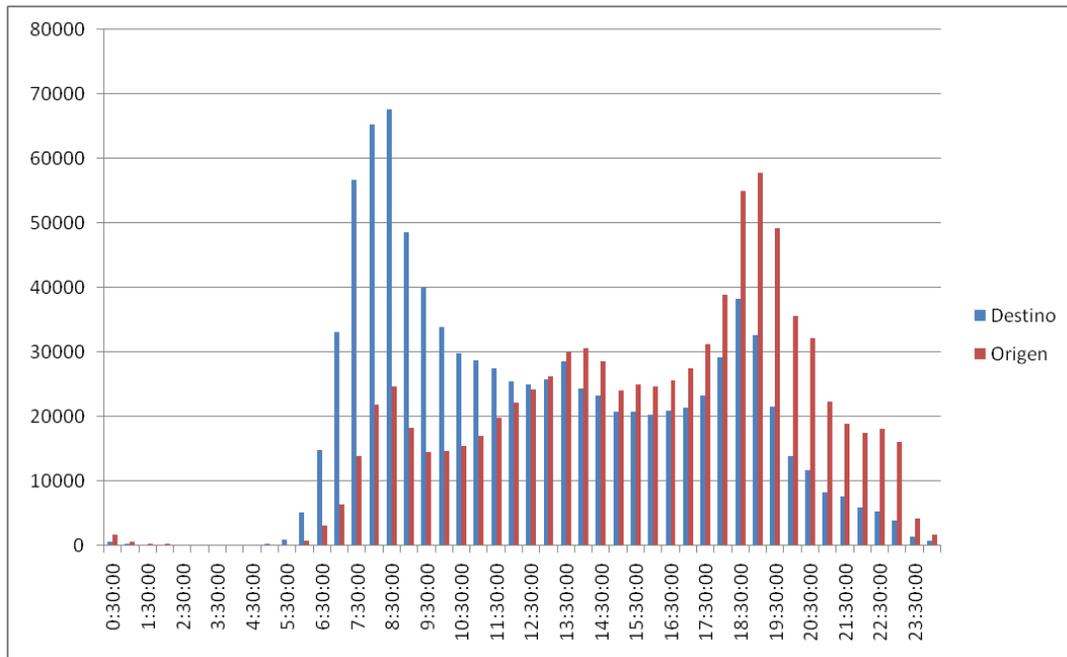
Gráfico 5: Histograma de hora de inicio de viajes



4.3.5 Comportamiento de demanda a lo largo del día en distintas zonas geográficas

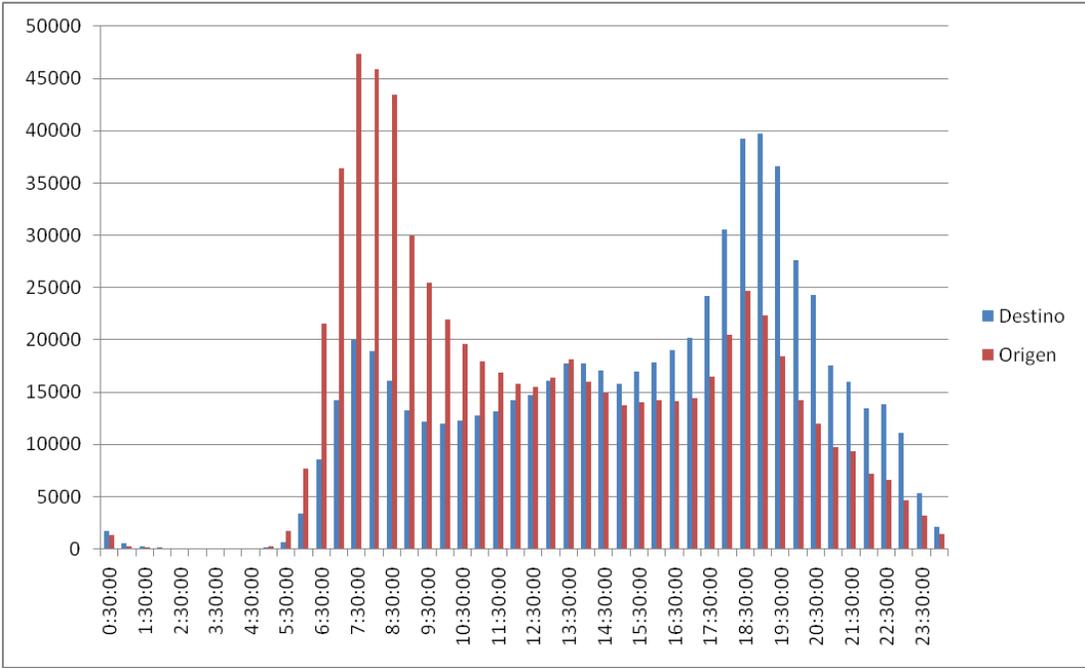
En las distintas zonas geográficas de Santiago, el comportamiento de la demanda es distinto a lo largo del día, por ejemplo en Gráfico 6, se muestra la demanda para los orígenes y destinos de viajes en el centro de Santiago, la mayoría de los destinos se concentran en la mañana, mientras que los orígenes están concentrados en la tarde, de esto se puede concluir que en general la gente tiende a ir al centro en la mañana y al final de la jornada de trabajo estos dejan el lugar.

Gráfico 6: Orígenes y Destinos en el sector centro de Santiago para el día lunes



En el sector occidente de Santiago se aprecia exactamente el comportamiento contrario presentado en Gráfico 6, estos resultados se presentan en Gráfico 7 y se debe a que el sector occidente corresponde a un sector más residencial que comercial.

Gráfico 7: Orígenes y Destinos en el sector occidente de Santiago para el día lunes



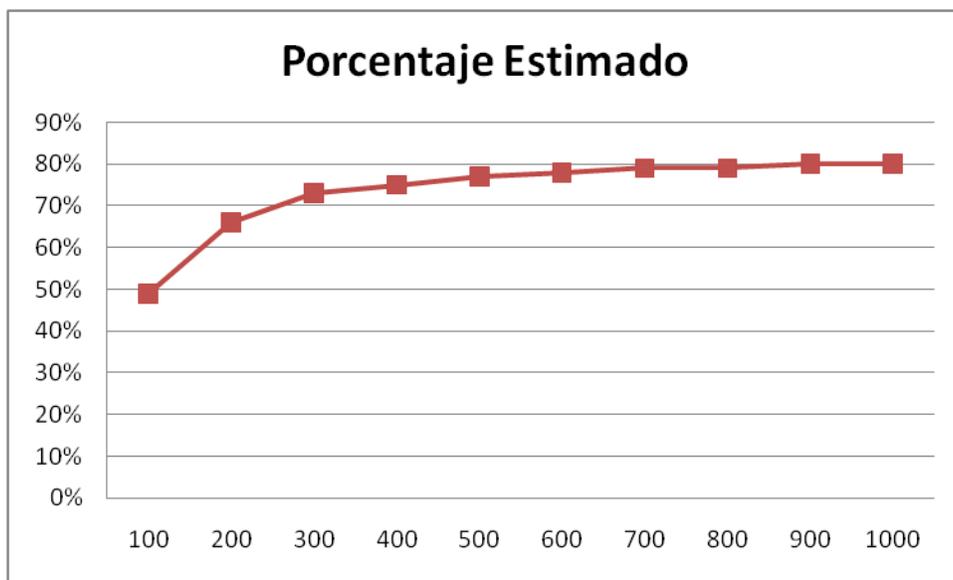
Se eligieron estos dos casos debido a la diferencia que tienen, para así mostrar el valor que tiene poder generar histogramas de demanda por sector. No todos los sectores tienen comportamientos tan distintos entre orígenes y destinos

4.4. Análisis de sensibilidad

4.4.1. Distancia caminable

La distancia caminable, es un parámetro que representa la distancia máxima de búsqueda de paradero de bajada con respecto al de subida posterior⁶, Una distancia máxima grande permite estimar más bajadas, pero con menor certidumbre, por otro lado, una distancia pequeña estima el paradero con mucha certidumbre, pero estima menos paraderos En Gráfico 8 se ve el efecto de disminuir la distancia caminable sobre la tasa de estimación de bajadas.

Gráfico 8: Porcentaje Estimado, con respecto a la distancia Máxima caminable cada 100 metros



NOTA: Los análisis de sensibilidad se realizaron sobre una muestra de 50.000 transacciones.

El porcentaje de estimación es bastante insensible a disminuciones a este parámetro, el cual fue fijado en 1000 metros, se consideraron reducciones de este, pero no aumentos debido a que la distancia medida es una distancia euclidiana y en la realidad el pasajero pudo realizar caminatas más largas.

⁶. Página 17

4.4.2. Tiempo entre viaje-etapa

Este parámetro es la estimación de tiempo de espera máximo que el pasajero realiza en el paradero, se estableció en 30 minutos y es explicado en el capítulo 2.3 Mejoras a la Metodología (MESB). Al disminuir este tiempo, aumentará la cantidad de viajes, debido a que si un pasajero espera mucho tiempo el bus, podría confundirse un trasbordo con una actividad, en caso de aumentar este parámetro, podría caerse justamente en el caso contrario.

Al variar el parámetro de tiempo de espera de buses, como se ve en Tabla 14, el cociente entre viajes y etapas cambia.

Tabla 14: Análisis de sensibilidad para tiempo de espera

Minutos	Viaje/Bips!	Cambio %
15	66.0%	
20	63.9%	3.3%
25	62.4%	2.3%
30	61.3%	1.8%
35	60.4%	1.5%
40	59.7%	1.2%
45	59.0%	1.1%
50	58.4%	1.0%
55	57.9%	1.0%

La sensibilidad en torno a 30 minutos es bastante baja, en particular comienza a aumentar para los 15 minutos, por lo que dar 30 minutos como parámetro evita confundir trasbordos con orígenes.

4.5 Evaluación de la estimación de la demanda

Es difícil evaluar los efectos del post procesamiento de MESB y la metodología utilizada será a nivel de comuna contar el porcentaje de viajes que cambian su origen o destino a nivel agregado y se comparará con el valor total del Transantiago.

Gracias a la búsqueda de etapas por múltiples pagos, se incrementa el número de etapas en 687.114, representando el 1.91% de las etapas semanales, de los cuales 404611 tienen destino estimado con un costo mensual asociado de 700 millones de pesos.

Al comparar utilizar factores de expansión en vez dividir la matriz por la tasa de estimación, se obtiene un cambio del 2,77% de los viajes a nivel de comuna, con un costo mensual asociado de 1600 millones de pesos mensuales.

El cambio en la matriz a nivel de comuna debido a aplicar la distribución de bajadas de metro es de un 10% de la matriz, el más alto y representa 5.000 millones de pesos mensuales.

4.5.1 Comparación con EOD 2001

La información disponible para la EOD 2001 estimaba un mayor número de viajes que la encontrada por la metodología. Para hacer comparables las cifras, ambas tablas se convirtieron a porcentajes del total de viajes del día. En las Tabla 15 y Tabla 16 se presentan los datos para los mismos niveles de agregación espacial para un día laboral

Tabla 15: Demanda de transporte estimada en EOD 2001 (Sectra 2001)

EOD 2001	Norte	Occidente	Oriente	Centro	Sur	Sur-Oriente	Total
Norte	6.50%	0.91%	1.17%	2.29%	0.74%	0.50%	12.11%
Occidente	1.02%	10.60%	2.25%	3.95%	1.09%	0.74%	19.66%
Oriente	1.15%	2.14%	5.74%	2.85%	1.42%	2.84%	16.15%
Centro	2.38%	4.11%	2.70%	2.52%	2.84%	2.31%	16.85%
Sur	0.70%	1.21%	1.54%	2.77%	9.69%	1.76%	17.66%
Sur-Oriente	0.51%	0.70%	2.83%	2.26%	1.75%	9.53%	17.57%
Total	12.25%	19.67%	16.23%	16.63%	17.53%	17.68%	100.00%

Tabla 16: Demanda de transporte estimada por metodología

Estimado	Norte	Occidente	Oriente	Centro	Sur	Sur Oriente	Total
Norte	4.02%	0.85%	1.34%	2.41%	0.61%	0.47%	9.70%
Occidente	0.84%	6.99%	2.96%	4.68%	0.94%	0.70%	17.11%
Oriente	1.24%	2.65%	7.52%	4.98%	1.89%	3.74%	22.03%
Centro	2.16%	4.50%	4.51%	5.10%	3.01%	2.85%	22.12%
Sur	0.63%	0.87%	2.15%	3.30%	4.81%	1.53%	13.29%
Sur Oriente	0.45%	0.67%	3.90%	3.01%	1.48%	6.25%	15.75%
Total	9.35%	16.51%	22.37%	23.48%	12.74%	15.55%	100.00%

Para comparar estas cifras, se calculó el cambio porcentual con respecto a la EOD 2001. Como se ve en Tabla 17, los cambios son grandes. Si se supone que la metodología obtiene con los valores correctos y la EOD con errores estadísticos de muestreo, el error promedio sería de 0.89% mientras que el error esperado de la EOD 2001 para 6 zonas es del 2%, reflejando concordancia entre los datos a nivel agregado.

Tabla 17: Cambio porcentual con respecto a EOD 2001

Cambio	Norte	Occidente	Oriente	Centro	Sur	Sur Oriente	Total
Norte	2.48%	0.06%	0.17%	0.12%	0.13%	0.03%	0.50%
Occidente	0.18%	3.61%	0.70%	0.73%	0.16%	0.04%	0.90%
Oriente	0.10%	0.50%	1.78%	2.13%	0.47%	0.90%	0.98%
Centro	0.22%	0.38%	1.81%	2.58%	0.17%	0.54%	0.95%
Sur	0.08%	0.34%	0.62%	0.53%	4.88%	0.23%	1.11%
Sur Oriente	0.06%	0.03%	1.06%	0.76%	0.27%	3.28%	0.91%
Total	0.52%	0.82%	1.02%	1.14%	1.01%	0.84%	0.89%

4.6 Evaluación Beneficios de Metodología

4.6.1 Validez estadística de los datos

Al intentar estimar la matriz, en el fondo se está estimando el parámetro de una distribución de bernoulli (Ruiz 2001), donde cada persona encuestada es un caso, y al encuestarla, esta genera un 1 en la combinación de origen-destino correspondiente y un 0 en el resto, las variables resultan correlacionadas, pero son tantas, que esta es despreciable.

Al realizar esta encuesta a una serie de personas, cada parámetro de la matriz seguirá una distribución binomial (Ruiz 2001), con parámetro p idéntico a la distribución bernoulli.

Cada p_i , es la probabilidad de que una persona que utiliza el transporte público cambie de Zona, dado que p es una probabilidad $\sum_i p_i = 1$.

Por ejemplo suponiendo que la demanda se distribuye homogéneamente, para las 10 aéreas del Transantiago. $p_i = p_j$, $p_i = 1/10$. Para un muestreo como el de la EOD 2001 de 48000 encuestas. El error esperado para las variables estimadas ($z_{\alpha/2}=0.712$)⁷, utilizando el intervalo de confianza (Ecuación 1) el error esperado para las variables es de 3.21%.

Ecuación 1: Intervalo de confianza para poblaciones infinitas.

$$\left(p_n - z_{\alpha/2} \sqrt{\frac{p_n(1-p_n)}{n} \frac{N-n}{N-1}}, p_n + z_{\alpha/2} \sqrt{\frac{p_n(1-p_n)}{n} \frac{N-n}{N-1}} \right)$$

- Donde:
 - P_n : probabilidad real.
 - Z : valor alfa de intervalo de confianza.
 - n : población encuestada.
 - N : Población total.

Con la metodología propuesta, se logra estimar el paradero de bajada alrededor del 70% de los viajes pagados, la evasión llega a alrededor del 25% para los buses, teniendo así origen y destino para el 53% de los viajes realmente efectuados. Debido a que no se sabe qué porcentaje de la evasión corresponde a evasión en solo una etapa de un viaje, se considera el total de viajes como

⁷ $(E(X)|X>0)$

3.8 millones. Los errores esperados de las mediciones se muestran para distintos niveles de desagregación en Tabla 18.

Tabla 18: Error esperado en mediciones

Error %	Plan modernizacion	Dia Representativo	Mes Representativo
N	480000	3800000	15200000
10 Areas	0.9%	0.34%	0.17%
36 Comunas	3.4%	1.20%	0.61%
618 Zonas	58.3%	21%	10.00%
10000 Paraderos	943.2%	344%	172%

Si suponemos que el sistema funciona, pero están mal estimadas las demandas y los servicios son de zona a zona, existirá un porcentaje del presupuesto mal asignado, por exceso de frecuencia o falta de esta, para los distintos niveles de agregación, el porcentaje se presenta en Tabla 19.

Tabla 19: Porcentaje del presupuesto mal asignado por errores muestrales, para distintos niveles de agregación

Costo por ErrorEsperado/Mes	Plan modernizacion	Dia Representativo	Mes Representativo
N	480000	3800000	15200000
10 Areas	1%	0%	0%
36 Comunas	3%	1%	1%
618 Zonas	37%	17%	9%
10000 Paraderos	90%	77%	63%

Pero el Transantiago se diseñó a grandes rasgos con troncales, utilizando como nivel de agregación una vista más macro, como podrían ser las comunas y los alimentadores, diseñados a nivel de zona estraus, al ponderar el número de transacciones realizadas en troncales y Metro con el error a nivel de comuna y las transacciones en alimentadores por el error a nivel estraus, se obtiene Tabla 20 con las estimaciones del costo operacional del error estadístico.

Tabla 20: Costo del error muestral en Porcentaje del costo total de Transantiago

Costo por ErrorEsperado/Dia	Plan modernizacion	Dia Representativo	Semana Representativa
N	480000	3800000	15200000
Porcentaje	11.2%	5.0%	2.6%
Dinero	\$ 6,510,050,459	\$ 2,907,702,340	\$ 1,516,469,805

4.6.2 Sesgos de la estimación

Una fuerte desventaja de la metodología propuesta, son los sesgos provocados por las personas que no siguen el comportamiento de los pasajeros supuesto, dejando fuera de la estimación un 27% de los viajes, esto se debe a: usuarios que utilizan transporte mixto entre público y privado, una persona puede utilizar en un día más de una tarjeta bip!, paraderos no ingresados, usuarios que pagan por otro y un sinfín de casos filtrados previamente.

Estos casos en que el método no logra estimar un paradero de bajada, se realizaron ponderaciones y factores de expansión (la metodología se redefinió y se realizaron cambios en corto plazo), asociando al comportamiento de usuarios similares.

Es difícil validar si estos factores de expansión introducen un error estadístico o un sesgo, pero dado que se puede estimar actualmente el 27% de los casos, estos serán del orden de 1,5 introduciendo sesgos pequeños.

Esta metodología puede validarse a través de mediciones en puntos estratégicos, por ejemplo, se pueden medir todas las subidas y bajadas de una lista de paraderos, para luego comparar con los resultados obtenidos con la metodología, existen mediciones de este tipo, pero no la muestra es contradictoria con la metodología, esto se puede deber a que la medición es de una fecha lejana a la información utilizada en la metodología o que realmente la metodología no es buena. A pesar de que la estimación es por paraderos. Para realizar esta comparación, se deben medir agrupaciones de paraderos cercanos y compararlos con los resultados obtenidos, esto se debe a que podrían existir errores en el paradero asignado a una subida o bajada, pero este debería ser pequeño y se debería asignar a un paradero cercano.

4.6.3 Nivel de Agregación

Gracias a la metodología que utiliza la información transaccional en vez de una EOD, se pueden obtener muestras mucho más grandes, pudiendo la demanda a niveles mucho menos agregados, pudiéndose llegar a un nivel de estimación por paradero de subida y bajada.

4.6.4 Costos

Actualmente en Transantiago se pidió un presupuesto para realizar una medición en 2 horarios punta para obtener una Matriz Origen-Destino como la obtenida con la metodología propuesta, el costo asciende a los 400 millones de pesos, contra 60 millones que ha costado el desarrollo de la herramienta para generarla con información transaccional.

Por otro lado, la herramienta hecha, permite actualizar la matriz periódicamente, para encontrar cambios en la demanda, solo con un costo marginal, dado que los 60 millones son un costo hundido, mientras que los 400 que cuesta una EOD pequeña es marginal (se debe pagar cada vez que se desea actualizar los datos).

En caso de querer implementarse una plataforma que permita generar los datos para la metodología, se requerirá evaluar el costo de:

- Utilizar SmartCard o equivalente para pagos.
- Seguimiento GPS en buses.
- Contratar el servicio de almacenamiento de información transaccional.

4.6.5 Nueva Información

Además de obtener la matriz origen destino, la nueva metodología permite obtener nueva información, como por ejemplo:

- Desagregar temporalmente la matriz.
- Estimar los tiempos de espera.
- Estimar las distancias caminadas.
- Estimar la varianza de los parámetros obtenidos.
- Encontrar tendencias en la demanda de transporte.
- Estimar la elasticidad precio de la demanda de transporte público por origen-destino.

Abriendo un horizonte de nuevas investigaciones y mejoras en la gestión en base a esta nueva información obtenida.

4.6.6 Limitaciones de la metodología

A diferencia de las estimaciones realizadas con mediciones en terreno, con esta metodología no se obtiene información demográfica sobre los usuarios.

La muestra solo incluye a los usuarios del Transantiago, excluyendo alternativas tales como taxi, bus interurbano y taxi colectivo.

No se plantea una metodología para dar un tratamiento al efecto de la evasión sobre la estimación de la demanda.

No se puede utilizar para el diseño de rutas no alcanzadas por el sistema de transporte público.

5. Conclusiones

En este estudio se evaluó y mejoró la metodología de estimación de subidas y bajadas (MESB) logrando estimar el 70% de los orígenes y destino de los viajes registrados en las bases de datos del Transantiago. Considerando la evasión estos representan al menos un 53% del total de la demanda de transporte de Transantiago. Esto permitió disminuir el error muestral de la matriz origen destino con 618 zonas de 53% a 21%.

Las mejoras hechas a la MESB fueron: Considerar el tiempo estimado de caminata en el criterio para diferencias entre transbordo y destino, detección de múltiples usuario asociados a una tarjeta bip!, distribución de demanda en estaciones de metro a zonas de influencia, implementar el módulo de Zonas Pagas al MESB, reparar la equivalencia entre paraderos y zonas, aplicar factores de expansión.

Al considerar el tiempo de caminata en el criterio de trasbordo o destino, se cambió la tasa de etapas por viaje de 1,63 a 1,55 identificando así mejor los viajes.

La detección de múltiples pagos permitió detectar un 1,91% de etapas que fueron realizadas por personas distintas a la asociada a la tarjeta bip!. Anteriormente no se daba ningún tratamiento a este comportamiento. Esto permite evitar la subestimación de la demanda del transporte público.

Anteriormente una estación de metro estaba asociada a una zonificación geográfica, asignando todas las subidas o bajadas a esta, esto en particular no correspondía al comportamiento real de los usuarios en estaciones entre dos zonificaciones. Gracias a la distribución de la demanda de metro, se mejoró esta situación.

En la implementación anterior del MESB, las etapas con inicio en zona paga no eran estimadas y correspondían al 5,9% del total de las transacciones. La existencia del módulo de Zona Paga, permite estimar subida, bajada y bus asociado al 65% de estos.

Con la aplicación de factores de expansión se logró balancear con respecto a los orígenes las distintas tasas de estimación de etapas y viajes y estimando el total de la demanda. Anteriormente solo se reflejaba el 70% de los viajes y aquellas zonas con menor tasa de estimación de destinos eran subestimadas. Con la expansión se logra reflejar el 100% de la demanda capturada por los datos operacionales de Transantiago, esto excluye la evasión.

Al reparar la equivalencia entre paradero y zonificaciones se eliminaron inconsistencias en estas, además de agregar nuevas zonificaciones con distintos niveles de agregación, las cuales fueron utilizadas para encontrar los factores de expansión.

Los resultados obtenidos con esta metodología son congruentes con los obtenidos en la EOD 2001 (Sectra 2001). Debido a que el error muestral esperado para la EOD 2001 es de un 2% al considerar

6 zonas y al compararlo con los resultados obtenidos este tiene menos de un 1% de diferencia. Se estima que debido a errores muestrales el 11,2% del presupuesto del Transantiago es mal asignado en comparación al 5% si se utilizaran las demandas estimadas con esta metodología, significando mensualmente tres mil quinientos millones de pesos.

Realizar una medición de demanda de transporte para solo 2 horas punta cuesta 400 millones de pesos, mientras que todo el desarrollo de la metodología ha costado solo 60 millones de pesos. Esto sin considerar los costos para obtener los datos.

Las limitaciones de este estudio son: la falta de una metodología para tratar la evasión en el sistema, escasas de información demográfica de la demanda, la medición está sesgada a la demanda del Transantiago y no considera el uso de otros medios de transporte como taxi, automóvil, bicicleta, etc.

Se propone como futura línea de investigación: cambiar los bloques horarios utilizados en los factores de expansión a los correspondientes a cada día de la semana, aplicar herramientas más avanzadas para encontrar los destinos no estimados, como por ejemplo, revisar el comportamiento de otros días. Mejorar el criterio para diferencial entre trasbordo y destino, mejorar la estimación de tiempo de viaje en Metro, mejorar la tasa de estimación de zona paga. Las líneas de investigación más importantes son: encontrar una metodología para realizar mediciones y dar un tratamiento adecuado a la evasión del pago y validar la estimación de la demanda con una muestra realizada el mismo día de los datos de entrada al modelo.

6. Bibliografía

Barry, J.J., Newhouser, R., Rahbee, A. y Sayeda, S. (2002), Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record* 1817, 183-187.

Buneman, K. (1984), Automatic and passenger-based transit performance measures, *Transportation Research Record* 992, 23-28.

Chapleau, R. y Chu, K.K. (2007). Modeling transit travel patterns from location-stamped smart card data using a disaggregate approach. Presented at the 11th World Conference on Transportation Research, June 24-28 2007, Berkeley, CA.

Chapleau, R., Trépanier, M. y Chu, K.K. (2008). The ultimate survey for transit planning: Complete information with smart card data and GIS. Presented at the 8th International Conference on International Steering Committee for Travel Survey Conferences, Lac d'Annecy, France.

Chriqui, C. and Robillard, P. (1975) Common bus line. *Transportation Science* 9, 115-121.

Collect Gfk – Luminae (2010), Levantamiento Etnográfico Pasajeros Transantiago, Santiago, Chile

Consorcio Fernández & de Cea Ingenieros Ltda. - CIS Asociados Consultores en Transporte S. A. (2001), Análisis Modernización Transporte Público, V Etapa, Santiago, Chile Capitulo 1-4.

Coordinación de Transantiago – Gobierno de Chile (2010), Informe Financiero de Transantiago. Periodo Junio 2007 – Marzo 2010, Santiago, Chile

Coordinación de Transantiago (2010) – Gobierno de Chile, Informe Financiero de Transantiago. Periodo Junio 2007 – Marzo 2010 (2010), Santiago, Chile

Cortés C., Gibson J., Gschwender A., Munizaga M. y Zúñiga M. (2010), bus Commercial Speed Diagnosis base on GPS monitored data, *Transportation Research C*. En Prensa

Cruz, Carlos (2002), Transporte urbano para un Nuevo Santiago, Ministerio de Obras Públicas, Transporte y Telecomunicaciones, Santiago de Chile.

Devillaine F. (2010). Informe de Practica de Vacaciones CI59A Universidad de Chile, Departamento Ingeniería Civil, Universidad de Chile.

Diario la Nación (2008) – La Nación, edición 8 de febrero del 2008 – Santiago, Chile Febrero 2008.

Diario la Nación (2010) – La Nación, edición 19 de junio del 2010 – Santiago, Chile Junio 2010.

Diario la Nación 2 (2010) – La Nación, edición 17 de julio del 2010 – Santiago, Chile Julio 2010.

Díaz, G. Gómez-Lobo, A. Velasco, A (2004) Micros en Santiago: De Enemigo Público a Servicio Público. Centro de Estudios Públicos Chile.

Fundación Chile (2010), Pago y Evasión en Transantiago Encuesta N°65 versión 1, Santiago, Chile

Furth, P.G., Hemily, B.J., Muller, T.H.J. y Strathman, J.G. (2006) Uses of archived AVL-APC data to improve transit performance and management: Transportation Research Board, TCRP Report No. 113.

Lianfu, Z., Shuzhi, Z., Yonggang, Z. y 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.

Mora P. (2010). Informe de Generación de datos de patrones de viaje a partir de transacciones bip Periodo: 2009 Universidad de Chile, Departamento Ingeniería Civil, Universidad de Chile.

Morandé, F. Doña, J. (2007) Transantiago: el remedio que está matando al paciente. Departamento de Economía de la Universidad de Chile.

Muller, T.H.J. y Furth, P.G. (2001) Trip time analyzes: Key to transit service quality. Transportation Research Record 1760, 10-19.

Munizaga, M. Palma, C. Mora, P. (2010). Public transport OD matrix estimation from smart card payment system data. Universidad de Chile, Departamento Ingeniería Civil, 2010, Universidad de Chile.

Ortúzar D., Willumsen J., Luis G. (2008), Modelos de Transporte, Publican Universidad Cantabria, España.

Ruiz, M., Morcillo, M. García, J. Castillo, C. (2000) "Curso de Probabilidad y Estadística", Ed. Universidad de Málaga / Manuales

SECTRA (2001) Encuesta Origen Destino de Viajes. Santiago, Chile.: Pontifica Universidad Católica de Chile. Informe Final Para SECTRA 2001.

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.

Unidad de servicios de Transporte (2010) , USOS DE TARETAS BIP Semana 20 al 26 de Marzo 2009 Informe Generación Tablas de Consulta, Santiago, Chile.

Unidad de servicios de Transporte (2010), Usos de tarjetas BIP Semana 20 al 26 de marzo del 2009, informe de generación de consulta, Santiago, Chile

Zhao, J., Rahbee, A. y Wilson, N. (2007) Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Computer-Aided Civil and Infrastructure Engineering 22, 376-387.

7. Apoyos Institucionales

- División de transporte, Ingeniería Civil, Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.
- Transantiago.
- Proyecto Fondecyt 1090204.
- Instituto Milenio Sistemas Complejos de Ingeniería.

Anexos

A. Zonas de influencia

A1. ESTRAUS

Estación	ESTRAUS					
Alcántara	184	192				
Baquedano	159	275	238			
Bellas Artes	278	238	237			
Bellavista de la Florida	542	403	543			
Cal y Canto	276	236	7	8		
Camino Agrícola	390	284	283			
Carlos Valdivinos	280	384	390	283		
Cementerios	497	498	47			
Cerro Blanco	49	47	8			
Ciudad del Niño	318	319	320	317		
Cristóbal Colón	197	199	167	610		
Cumming	243	244	240	241		
Departamental	317	318	319	320		
Dorsal	40	41	596			
Ecuador	71	73	69	70		
Einstein	44	45	40	41		
El Golf	184	192				
El LLano	322	312				
El Parrón	326	327	328	329	330	
Elisa Correa	562	458	555	560		
Escuela Militar	185	192	184	577		
Estación Central	500	72	563	247	250	
Francisco Bilbao	199	223	169	168		
Franklin	272					
Grecia	213	394	398	387	219	
Hospital Sótero del Río	524	562	546	418		
Irrazával	209	259	262	267		
La Cisterna	329	331	333	332	335	334
La Granja	297	300	414	299	413	
La Moneda	277	245	254	253		
Las Mercedes	419	422				
Las Rejas	112	69	73	68	113	
Las Torres	541	557	570	389		
Lo Ovalle	325	326	327			
Lo Vial	315	316	317	318		

Los Héroes	249	254	253	252		
Los Leones	161	158				
Los Orientales	213	394	397	395		
Los Presidentes	387	400	512	386	457	
Los Quillayes	595	551	458	553	554	
Macul	393	541	570	406	405	404
Manuel Montt	157	467	165	160	155	
Mirador	523	403				
Neptuno	108	109	110	111	112	
Ñuble	215	268	216	221	274	
Pajaritos	112	111	66	67		
Parque Bustamante	163	164	259	258		
Parque O'higgins	260	515	265	261	266	
Patronato	50	51	8			
Pedrero	392	286	403	292	520	
Pedro de Valdivia	467	158	162	156	166	
Plaza de Armas	276	246	277	236		
Plaza de Puente Alto	425	575	426	428	430	
Plaza Egaña	208	228	214	229	230	
Príncipe de Gales	223	511	226	227	222	
Protectora de la Infancia	615	418	419			
Quilín	539	401	577	540	538	389
Quinta Normal	242	563	247	239		
República	428	251				
Rodrigo de Araya	384	280	220	279		
Rojas Magallanes	546	547	545	548		
Rondizonni	264	515	265	272	271	358
Salvador	160	155				
San Alberto Hurtado	71	74	70			
San Joaquín	519	286	285	391		
San Miguel	313	314				
San Pablo	108	599	105	106	99	
San Ramón	307	491	308	310	309	
Santa Ana	244	245				
Santa Isabel	163	259	258	164		
Santa Julia	410	545				
Santa Lucía	237	256	258			
Santa Rosa	311	491	568	298		
Simón Bolívar	208	465	227	228		
Tobalaba	162	158	184	192		
Toesca	253	252	254	257	255	251
Trinidad	549	415				

Unión Latino Americana	251	250	247	248
Universidad Católica	275	238	256	
Universidad de Chile	255	246	254	253
Universidad de Santiago	72	77	563	
Vespucio Norte	35	31	36	
Vicente Valdés	544	571	572	
Vicuña Mackenna	542	543	405	
Zapadores	37	36	596	

A2. Comunas

Estacion	Comunas			
Alcántara	LAS CONDES			
Baquedano	PROVIDENCIA	SANTIAGO		
Bellas Artes	SANTIAGO			
Bellavista de la Florida	LA FLORIDA			
Cal y Canto	SANTIAGO			
Camino Agrícola	SAN JOAQUÍN	MACUL		
Carlos Valdovinos	SAN JOAQUÍN	MACUL		
Cementerios	RECOLETA			
Cerro Blanco	RECOLETA			
Ciudad del Niño	SAN MIGUEL			
Cristóbal Colón	LAS CONDES	PROVIDENCIA		
Cumming	SANTIAGO			
Departamental	SAN MIGUEL			
Dorsal	RECOLETA			
Ecuador	ESTACIÓN CENTRAL			
Einstein	RECOLETA			
El Golf	LAS CONDES			
El LLano	SAN MIGUEL			
El Parrón	LA CISTERNA			
Elisa Correa	LA FLORIDA	PUENTE ALTO		
Escuela Militar	LAS CONDES			
Estación Central	SANTIAGO	ESTACIÓN CENTRAL		
Francisco Bilbao	PROVIDENCIA	LAS CONDES	LA REINA	
Franklin	SANTIAGO			
Grecia	ÑUÑO A	PEÑALOLÉN	MACUL	
Hospital Sótero del Río	PUENTE ALTO			
Irrázaval	ÑUÑO A	SANTIAGO		
La Cisterna	LA CISTERNA			
La Granja	LA GRANJA			
La Moneda	SANTIAGO			

Las Mercedes	PUENTE ALTO		
Las Rejas	ESTACIÓN CENTRAL	LO PRADO	
Las Torres	PEÑALOLÉN	MACUL	
Lo Ovalle	LA CISTERNA		
Lo Vial	SAN MIGUEL		
Los Héroes	SANTIAGO		
Los Leones	PROVIDENCIA		
Los Orientales	ÑUÑO A	PEÑALOLÉN	
Los Presidentes	PEÑALOLÉN	MACUL	
Los Quillayes	LA FLORIDA		
Macul	PEÑALOLÉN	MACUL	LA FLORIDA
Manuel Montt	PROVIDENCIA		
Mirador	LA FLORIDA		
Neptuno	LO PRADO		
Ñuble	ÑUÑO A	SANTIAGO	
Pajaritos	LO PRADO	ESTACIÓN CENTRAL	
Parque Bustamante	PROVIDENCIA	SANTIAGO	
Parque O`higgins	SANTIAGO		
Patronato	RECOLETA		
Pedrero	MACUL	LA FLORIDA	SAN JOAQUÍN
Pedro de Valdivia	PROVIDENCIA		
Plaza de Armas	SANTIAGO		
Plaza de Puente Alto	PUENTE ALTO		
Plaza Egaña	ÑUÑO A	LA REINA	
Príncipe de Gales	ÑUÑO A	LA REINA	
Protectora de la Infancia	PUENTE ALTO		
Quilín	PEÑALOLÉN	MACUL	
Quinta Normal	SANTIAGO		
República	SANTIAGO		
Rodrigo de Araya	SAN JOAQUÍN	MACUL	
Rojas Magallanes	LA FLORIDA		
Rondizonni	SANTIAGO		
Salvador	PROVIDENCIA		
San Alberto Hurtado	ESTACIÓN CENTRAL		
San Joaquín	SAN JOAQUÍN	MACUL	
San Miguel	SAN MIGUEL		
San Pablo	LO PRADO		
San Ramón	SAN RAMÓN		
Santa Ana	SANTIAGO		
Santa Isabel	PROVIDENCIA	SANTIAGO	
Santa Julia	LA FLORIDA		
Santa Lucía	SANTIAGO		

Santa Rosa	LA GRANJA	SAN RAMÓN
Simón Bolívar	ÑUÑO A	LA REINA
Tobalaba	LAS CONDES	PROVIDENCIA
Toesca	SANTIAGO	
Trinidad	LA FLORIDA	
Unión Latino Americana	SANTIAGO	
Universidad Católica	SANTIAGO	
Universidad de Chile	SANTIAGO	
Universidad de Santiago	ESTACIÓN CENTRAL	
Vespucio Norte	RECOLETA	
Vicente Valdés	LA FLORIDA	
Vicuña Mackenna	LA FLORIDA	
Zapadores	RECOLETA	
-	-	

B. Últimos desplazamientos

B1. Por Operador

Es Ultimo?	Corto	Normal
A1		
0	2.88%	97.12%
1	12.14%	87.86%
A2		
0	3.78%	96.22%
1	13.35%	86.65%
A3		
0	2.30%	97.70%
1	14.49%	85.51%
A4		
0	1.92%	98.08%
1	14.39%	85.61%
A5		
0	2.73%	97.27%
1	16.44%	83.56%
A6		
0	1.81%	98.19%
1	9.81%	90.19%
A7		
0	1.77%	98.23%
1	10.28%	89.72%
A8		
0	2.95%	97.05%

1	13.01%	86.99%
A9		
0	2.46%	97.54%
1	11.03%	88.97%
Metro		
0	0.40%	99.60%
1	6.34%	93.66%
T1		
0	2.35%	97.65%
1	11.22%	88.78%
T2		
0	2.10%	97.90%
1	9.95%	90.05%
T3		
0	2.03%	97.97%
1	9.32%	90.68%
T4		
0	2.08%	97.92%
1	7.54%	92.46%
T5		
0	2.09%	97.91%
1	8.81%	91.19%

C. Encabezados Salida

Los encabezados para los archivos de salida demasiado grandes para abrirse en Excel se presentan en formato SQL, para ser cargados en un servidor y analizados.

C1. Salida principal del Post Proceso

La salida del post procesamiento son una serie de archivos con información sobre la demanda y la calidad de la estimación.

Se optó por una salida más completa que la matriz origen destino, pero que fuera fácil de convertir en una matriz origen destino con el nivel de agregación deseado. En los anexos⁸ se incluyen los encabezados, pero la salida a este nivel es confidencial, por ende solo se presenta una matriz origen destino con distintos niveles de agregación, hora de inicio, y fin de viaje, agrupada cada 30 minutos.

⁸ C. Encabezados Salida Página 90

Nota: Todos los numéricos agrupados, por ejemplo bloque horario, distancia caminada, en el valor de la pestaña, se ve el máximo valor que puede tener un elemento de ese conjunto, para conocer cuál es el menor, basta con ver el grupo anterior.

- OD_Expandida.csv contiene el origen, destino y factor de expansión de cada viaje, además de otros datos interesantes
- OD.csv contiene el origen y el destino, sin factores de expansión ni viajes no estimados filtrados para cada viaje, además de otros datos interesantes.
- OD_Etapa_Expandida.csv contiene para cada etapa, su factor de expansión para la matriz origen destino por etapa, además de otros datos complementarios.
- OD_Etapa.csv, origen y destino para cada etapa, pero sin factores de expansión, las etapas que tienen múltiple pago, son marcadas con 0 y las que valen por más de una, tienen un número mayor.
- Final.csv, Es un extracto de columnas consideradas importantes del archivo de input para el post procesamiento de datos, además de corregir algunas celdas y agregar los factores de expansión en 2 nuevas columnas.

C2. Salida con Indicadores sobre la estimación

- Resumen.csv, Incluye el resumen de 3 módulos
 - CaminataMax, en caso de existir una caminata, determina cuantas estaban sobre el umbral y bajo el umbral, para cada zona estraus.
 - MultiplePago
 - Asignado: se encontró patrón de viaje grupal y se asignó bajada
 - SinDestino: No se pudo asignar destino al múltiple pago y se agregó al factor de expansión
 - Sin Trasladarse: se detectó un múltiple pago por no trasladarse y se agrega al factor de expansión como viaje sin destino
 - Destinos: Muestra indicadores sobre los porcentajes globales de estimación.
- factorExpansion.csv: valores de los factores de expansión para los viajes, un 0 significa sin al menos un caso representativo (F1).
- FactorExpansionEtapa.csv: valores de los factores de expansión para las etapas, un 0 significa sin al menos un caso representativo (F1).
- FactorDiarioViaje.csv: factor de expansión para viajes, sin considerar el origen (F2).
- FactorDiarioEtapa.csv: factor de expansión para etapas, sin considerar el origen (F2).
- EtapaTotal.csv: Total de etapas.
- ViajeTotal.csv : Total de viajes.
- TotalViajesConDestino.csv: Total de viajes correctamente estimadas.
- TotalEtapasConDestino.csv: Total de etapas correctamente estimadas.
- ViajeExpandidosEnPrimeraExpansion.csv: Total viajes obtenidos al aplicar (F1)
- EtapaExpandidosEnPrimeraExpansion.csv: Total de Etapas obtenidas al aplicar (F1)

C3. Salida con datos obtenidos con la metodología

- Caminata.csv: Numero de caminatas estimadas cada 100 metros.
- DesplazamientoEtapa.csv: largo de desplazamientos de etapas de viajes estimados, para últimas etapas del día y para el resto de las etapas.
- entreViajes.csv: Tiempo entre 2 etapas, sea trasbordo o actividad, cada 1 minuto, hay tiempos negativos por: errores en estimación.
- entreViajesHora.csv: idéntico a entreViajes.csv, pero cada 1 hora
- tiempoEspera.csv: Tiempo estimado de espera, idéntico a entreViajes.csv, pero descontado el tiempo estimado de caminata
- UltimoDespazamiento.csv: Agrupación de desplazamiento de viajes para menos de 300 metros y más de 300, y última etapa y resto de etapas.

OD_Expandida.csv

Bip BIGINT, id tarjeta Bip!

xO double precisión, coordenada x origen

yO double precisión ,coordenada y origen

xD double precisión, coordenada x destino

yD double precision, coordenada y destino

ParaderoOrigen varchar(50), paradero origen

ParaderoDestino varchar(50), paradero destino

ZonaOrigen varchar(50), zona estraus origen

ZonaDestino varchar(50), zona estraus destino

ComunaOrigen varchar(50), comuna origen

ComunaDestino varchar(50), comuna destino

SecParaderos varchar(1000), secuencia de paraderos de subida y bajada

SecServicio varchar(1000), secuencia de servicios

SecTipoPago varchar(1000), secuencia de tipo de pago (ZP,METRO,BUS)

SecOperador varchar(1000), secuencia de operadores

SecTipoOperador varchar(1000), secuencia tipo de operador (Metro, Alimentador, Troncal)

FechaInicio varchar(50), fecha y hora de transacción

FechaFin varchar(50), Fecha y hora estimada de bajada

Dia varchar(10), Día de la semana de la transacción

DiaBajada varchar(10), Día de la semana en que se estimó la bajada

bloqueHora varchar(20), bloque horario cada 30 minutos de hora de subida

bloqueHoraFin varchar(20), bloque horario cada 30 minutos de hora estimada de bajada

Cantidad double precision, factor de expansión del viaje

OD.csv

Bip BIGINT, id tarjeta Bip!

xO double precisión, coordenada x origen

yO double precisión ,coordenada y origen

xD double precisión, coordenada x destino
yD double precision, coordenada y destino
ParaderoOrigen varchar(50), paradero origen
ParaderoDestino varchar(50), paradero destino
ZonaOrigen varchar(50), zona estaus origen
ZonaDestino varchar(50), zona estaus destino
ComunaOrigen varchar(50), comuna origen
ComunaDestino varchar(50), comuna destino
SecParaderos varchar(1000), secuencia de paraderos de subida y bajada
SecServicio varchar(1000), secuencia de servicios
SecTipoPago varchar(1000), secuencia de tipo de pago (ZP,METRO,BUS)
SecOperador varchar(1000), secuencia de operadores
SecTipoOperador varchar(1000), secuencia tipo de operador (Metro, Alimentador, Troncal)
FechaInicio varchar(50), fecha y hora de transacción
FechaFin varchar(50), Fecha y hora estimada de bajada
Dia varchar(10), Día de la semana de la transacción
bloqueHora varchar(20), bloque horario cada 30 minutos de hora de subida
Cantidad double precision, factor de expansión, en este caso solo por cuantas personas vale el viaje

OD_Etapa_Expandida.csv

Bip BIGINT, id tarjeta Bip!
xO double precisión, coordenada x origen
yO double precisión ,coordenada y origen
xD double precisión, coordenada x destino
yD double precision, coordenada y destino
ParaderoOrigen varchar(50), paradero origen
ParaderoDestino varchar(50), paradero destino
ZonaOrigen varchar(50), zona estaus origen
ZonaDestino varchar(50), zona estaus destino
ComunaOrigen varchar(50), comuna origen
ComunaDestino varchar(50), comuna destino
tipoOperador varchar(20), tipo de operador (Metro, Troncal, Alimentador)
Servicio varchar(20), servicio de transporte de la etapa
UnidadNegocio varchar(10), unidad de negocio
tipoLugarPago varchar(10), Tipo de lugar de pago (ZP, Bus,Metro)
FechaInicio varchar(50), Fecha y hora en que se hizo la transacción
FechaFin varchar(50), fecha y hora estimada de bajada
Dia varchar(10), díade la semana en que se realizó la transacción
DiaBajada varchar(10), Día de la semana en que se estimó la bajada
bloqueHora varchar(20), bloque horario de hora de transacción
bloqueHoraFin varchar(20), bloque horario de hora de bajada
Cantidad double precisión, factor de expansión de la etapa

OD_Etapa.csv

Bip BIGINT, id tarjeta Bip!
xO double precisión, coordenada x origen
yO double precisión, coordenada y origen
xD double precisión, coordenada x destino
yD double precisión, coordenada y destino
ParaderoOrigen varchar(50), paradero origen
ParaderoDestino varchar(50), paradero destino
ZonaOrigen varchar(50), zona estaus origen
ZonaDestino varchar(50), zona estaus destino
ComunaOrigen varchar(50), comuna origen
ComunaDestino varchar(50), comuna destino
tipoOperador varchar(20), tipo de operador (Metro, Troncal, Alimentador)
Servicio varchar(20), servicio de transporte de la etapa
UnidadNegocio varchar(10), unidad de negocio
tipoLugarPago varchar(10), Tipo de lugar de pago (ZP, Bus, Metro)
FechaInicio varchar(50), Fecha y hora en que se hizo la transacción
FechaFin varchar(50), fecha y hora estimada de bajada
Dia varchar(10), día de la semana en que se realizó la transacción
bloqueHora varchar(20), bloque horario de hora de transacción
Cantidad double precisión, factor de expansión de la etapa

Final.csv

tiempo character varying(100), Fecha y hora de la transacción
id bigint, ID tarjeta bip!
pago smallint, Monto pago asociado a transacción
x double precision, coordenada x de subida
y double precision, coordenada y de subida
servicio character varying(20),
tipo_transporte character varying(20),
tipo_operador character varying(500),
unidad_negocio character varying(100),
serv_un_zp character varying(500),
nviaje integer,
netapa integer,
x_bajada double precision,
y_bajada double precision,
t_bajada character varying(100),
par_subida character varying(30),
par_bajada character varying(30),
comuna_subida character varying(20),
comuna_bajada character varying(20),
estaus_618_subida character varying(10),
estaus_618_bajada character varying(10),
ExpViaje double precision,
ExpEtapa double precisión,

