



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

**MODELO DE INFERENCIA ZONAL BASADO
EN INFORMACIÓN OBSERVADA EN
PARADEROS DE TRANSPORTE PÚBLICO**

SEBASTIÁN JORGE TAMBLAY MOËNNE

Tesis para optar al grado de

Magíster en Ciencias de la Ingeniería

Profesores Supervisores:

PATRICIA VIVIANA GALILEA ARANDA

JUAN CARLOS MUÑOZ ABOGABIR

Santiago de Chile, Abril, 2015

© 2015, Sebastián Tamblay M.



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

**MODELO DE INFERENCIA ZONAL
BASADO EN INFORMACIÓN
OBSERVADA EN PARADEROS DE
TRANSPORTE PÚBLICO**

SEBASTIÁN JORGE TAMBLAY MOËNNE

Tesis presentada a la Comisión integrada por los profesores:

PATRICIA VIVIANA GALILEA ARANDA

JUAN CARLOS MUÑOZ ABOGABIR

JUAN DE DIOS ORTÚZAR SALAS

MARCELA ADRIANA MUNIZAGA MUÑOZ

LEONARDO VANZI

Para completar las exigencias del grado de

Magíster en Ciencias de la Ingeniería

Santiago de Chile, Abril, 2015

A mis abuelitas, padres, hermanos y
sobrinos.

AGRADECIMIENTOS

En primer lugar, quiero agradecer a mi familia, pareja, compañeros y amigos (tanto a los del Departamento de Transporte y Logística como a los de siempre) por el apoyo y compañía que me han brindado en este proceso.

Además, deseo agradecer a aquellos profesores que me aconsejaron y ayudaron a desarrollar esta tesis de la mejor manera posible. En especial, a mis profesores supervisores, que me han guiado no solo en esta investigación, sino también en otras esferas de mi desarrollo profesional. Asimismo, les agradezco a los demás miembros de la Comisión, que contribuyeron revisando mi trabajo y entregando sus ideas y comentarios para mejorarlo.

También, quiero darle las gracias al equipo del proyecto que acompañó esta tesis. Me enorgullece haber trabajado con ustedes y espero que el fruto contribuya a mejorar el transporte público de nuestra ciudad.

Por último, deseo destacar el aporte del Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF) y la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), que financiaron este postgrado a través del proyecto D10I1049 “Una herramienta táctico-estratégica de gestión y planificación de sistemas de transporte público urbano” y la beca CONICYT-PCHA/MagísterNacional/2014 – 22141442, respectivamente. Asimismo, agradezco el apoyo del Centro de Desarrollo Urbano Sustentable (CEDEUS), en diversas fases de este proceso.

ÍNDICE GENERAL

	Pág.
DEDICATORIA.....	i
AGRADECIMIENTOS	ii
ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS.....	vii
RESUMEN.....	ix
ABSTRACT.....	x
1 INTRODUCCIÓN	1
1.1 Objetivos	2
1.2 Alcances	2
1.3 Estructura	3
2 MARCO TEÓRICO.....	4
2.1 Estimación de matrices O-D zonales	4
2.1.1 Modelos gravitacionales.....	5
2.1.2 Estimación utilizando conteos de flujo	6
2.2 Estimación de matrices de viajes entre paraderos de transporte público	8
2.3 Extensiones y aplicaciones de matrices estimadas de viajes entre paraderos para Santiago.....	12
2.3.1 Estimación de propósitos de viajes	12
2.3.2 Estimación de la zona de residencia.....	13
2.4 Modelos de elección discreta	15

2.4.1	Teoría de la utilidad aleatoria.....	15
2.4.2	Modelo Logit Multinomial.....	16
2.4.3	Modelo Logit Jerárquico	17
2.5	Estimación de modelos de elección con variables de atracción.....	19
3	FORMULACIÓN DEL MODELO Y MÉTODOS DE ESTIMACIÓN	23
3.1	Modelo de inferencia zonal.....	23
3.2	Estimación por máxima verosimilitud	25
4	DATOS Y ENTRADAS DEL MODELO PARA SANTIAGO DE CHILE	27
4.1	Información origen-destino de viajes entre paraderos	27
4.2	Costo de elección de paraderos	28
4.3	Información de uso de suelo	30
4.4	Zonificación	32
4.5	Encuesta	35
5	PROCESAMIENTO Y CONSTRUCCIÓN DE LA BASE DE DATOS.....	40
5.1	Asignación de manzanas a zonas	40
5.2	Asignación de zonas de origen y destino a encuestas	40
5.3	Creación de centroides y arcos de acceso	42
6	GENERACIÓN Y DESCRIPCIÓN DEL BANCO DE ESTIMACIÓN	49
6.1	Procesamiento de la información de las encuestas	49
6.2	Descripción estadística del banco de estimación	50
6.2.1	Variables geográficas y de la red de transporte público	51
6.2.2	Variables de uso de suelo.....	56

6.2.3	Variables socioeconómicas y del viaje	58
7	MODELACIÓN Y RESULTADOS	63
7.1	Modelos de inferencia zonal iniciales	63
7.1.1	Calibración de arcos de acceso	65
7.2	Modelo de inferencia zonal definitivo	67
7.3	Formulaciones alternativas.....	71
7.3.1	Diferenciación de tarjetas de estudiante.....	71
7.3.2	Constantes modales para metro.....	73
7.3.3	Formulación Logit Jerárquica	75
7.3.4	Estimación diferenciando datos de punta mañana	75
7.3.5	Estimación de modelos considerando variables socioeconómicas	77
7.4	Reconstrucción de matriz origen-destino zonal	77
8	CONCLUSIONES Y RECOMENDACIONES.....	83
	BIBLIOGRAFIA	88
	ANEXOS	92
	ANEXO A: Estimación por minimización de errores cuadráticos	93
	ANEXO B: Encuesta de preferencias reveladas	96
a.	Cuestionario definitivo.....	96
b.	Lugares de realización de la encuesta	98
	ANEXO C: Zonificación.....	102
	ANEXO D: Ejemplos de aplicación del modelo definitivo	105

ÍNDICE DE TABLAS

	Pág.
Tabla 2-1: Criterios de asignación de propósitos para día laboral	13
Tabla 4-1: Destinos no agrícolas de bienes raíces	31
Tabla 6-1: Distribución de propósitos de viaje por horario	62
Tabla 7-1: Modelo inicial (MNL1)	65
Tabla 7-2: Resultados iteración de atracción para arcos de acceso	66
Tabla 7-3: Modelo definitivo (MNL5).....	68
Tabla 7-4: Modelo diferenciando estudiantes (MNL5e).....	72
Tabla 7-5: Modelo con costos diferenciados entre modos (MNL5c)	74
Tabla 7-6: Modelo diferenciando punta mañana (MNL5am)	76

ÍNDICE DE FIGURAS

	Pág.
Figura 2-1: Ejemplo de estructura jerárquica de HL.....	18
Figura 4-1: Mapa de Santiago destacando las zonas encuestadas.....	37
Figura 4-2: Mapa de Santiago con orígenes de encuestas reportados.....	38
Figura 4-3: Mapa de Santiago con destinos de encuestas reportados	39
Figura 5-1: Criterio de asignación de zonas a encuestas.....	41
Figura 5-2: Criterio de asignación de paraderos a zonas	43
Figura 5-3: Porcentaje de observaciones correctamente representadas por la red y total de arcos de acceso, para distintos valores de radios de influencia de metro	44
Figura 5-4: Porcentaje de observaciones correctamente representadas por la red y total de arcos de acceso, para distintos valores de radios de influencia de bus	45
Figura 5-5: Criterio de cálculo de distancias de acceso	47
Figura 6-1: Mapa de Santiago con orígenes de encuestas válidas	51
Figura 6-2: Mapa de Santiago con destinos de encuestas válidas.....	52
Figura 6-3: Número de observaciones por cantidad de zonas asociadas	54
Figura 6-4: Número de observaciones por distancia de arcos de acceso en metros	55
Figura 6-5: Número de observaciones por distancia de caminata real en metros.....	56
Figura 6-6: Porcentaje de superficie construida por tipo de uso de suelo.....	57
Figura 6-7: Avalúo fiscal promedio (\$2014) de <i>m</i> ² construido por uso de suelo	58
Figura 6-8: Distribución por rango etario de la muestra y población	60
Figura 6-9: Distribución por ingresos personales de la muestra	60

Figura 6-10: Distribución horaria de realización de encuestas de la muestra.....	61
Figura 7-1: Función de costo inversa f_{mnl} del modelo MNL5.....	70
Figura 7-2: Totales de viajes generados por zona, según el modelo MNL5.....	80
Figura 7-3: Totales de viajes atraídos por zona, según el modelo MNL5	81
Figura 7-4: Número de viajes generados y atraídos por macro-zona, punta mañana	82
Figura A-1: Cuestionario final encuesta en paraderos, hoja 1	96
Figura A-2: Cuestionario final encuesta en paraderos, hoja 2	97
Figura A-3: Mapa de la zona encuestada en Maipú.....	98
Figura A-4: Mapa de la zona encuestada en Peñalolén.....	98
Figura A-5: Mapa de la zona encuestada en La Florida.....	99
Figura A-6: Mapa de la zona encuestada en Ñuñoa.....	99
Figura A-7: Mapa de la zona encuestada en Metro La Cisterna.....	100
Figura A-8: Mapa de la zona encuestada en Metro Escuela Militar	100
Figura A-9: Mapa de la zona encuestada en Metro Vespucio Norte	101
Figura A-10: Mapa de la zona encuestada en Metros Elisa Correa, Los Quillayes y San José de la Estrella	101
Figura C-1: Ejemplo partición de zonas sector Tobalaba – Escuela Militar.....	103
Figura C-2: Zonificación definitiva, distinguiendo en colores las seis macro-zonas	104
Figura D-1: Ejemplo asignación de viajes a zonas, destino estación Tobalaba.....	105
Figura D-2: Ejemplo asignación de viajes a zonas, origen estación Tobalaba	106
Figura D-3: Ejemplo asignación de viajes a zonas, destino estación San Joaquín	107
Figura D-4: Ejemplo asignación de viajes a zonas, origen estación San Joaquín.....	108

RESUMEN

La recolección de información origen-destino para una ciudad es una importante pero habitualmente costosa tarea. De esta manera, existe la necesidad de desarrollar metodologías más eficaces y económicas para descubrir los patrones de viaje de los ciudadanos. En esta línea, esta tesis presenta una metodología genérica que permite inferir las zonas de origen y destino para un viaje observado entre dos paraderos de transporte público (*i.e.*, paraderos de buses o estaciones de metro) utilizando información de la red y uso de suelo. El modelo de inferencia zonal propuesto sigue un enfoque desagregado tipo Logit que incluye variables de tamaño. El modelo permite estimar una matriz origen-destino zonal para la ciudad si se cuenta con información pasiva de viajes entre paraderos (recolectada por un sistema de pago con tarjeta inteligente).

La metodología es aplicada al período punta mañana de Santiago con el objetivo de alimentar la herramienta computacional de planificación del proyecto D10I1049 para el Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF). Para estimar el modelo, fue necesario procesar y unificar distintos datos provenientes de diversas fuentes; incluyendo una encuesta realizada en paraderos, información de uso de suelo y una matriz de viajes entre paraderos de transporte público. Además, se construyó un sistema de 1.171 zonas para la ciudad, incluyendo la definición de sus arcos de acceso con una metodología original para identificarlos y asignarles sus distancias correspondientes.

Respecto a los resultados de la aplicación, se logró estimar modelos que capturan los efectos esperados del uso de suelo y de las variables de costo, con parámetros significativos y de signos correctos. Además, se presentan formulaciones alternativas que exploran posibilidades de mejora del modelo. Por último, cabe mencionar las numerosas implicancias de política de esta investigación. La matriz resultante permitirá evaluar, a través de la herramienta computacional, cambios mayores al sistema de transporte que modifiquen los paraderos escogidos por los usuarios, como por ejemplo extensiones de rutas, relocalización de paraderos, nuevos recorridos o nuevas líneas de metro.

ABSTRACT

The collection of origin-destination data for a city is an important but often costly task. This way, there is a need to develop more efficient and inexpensive methods of collecting information about citizens' travel patterns. In this line, this thesis presents a generic methodology that allows to infer the origin and destination zones for an observed trip between two public transport detentions (*i.e.*, bus stops or metro stations) using land use and network information. The proposed zonal inference model follows a disaggregated Logit approach including size variables. The model enables the estimation of an origin-destination matrix for a city, if trip information passively collected by a smart-card payment system is available.

The methodology is applied to the Santiago de Chile's morning peak period, with the purpose of serving as input for the planning computational tool being developed by Project D10I1049 of the *Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF)*. To estimate the model, information had to be gathered from different sources and processed into a unified framework; data included a survey conducted at public transport stops, land use information, and a stop-to-stop trip matrix. Additionally, a zonal system with 1.171 zones was constructed for the city, including the definition of its access arcs along with an original methodology to identify and assign them their associated distances.

Regarding results, the estimated models manage to capture the expected effects of land use and cost variables on trip generation and attraction, with significant parameters and correct signs. Furthermore, alternative formulations are presented to explore potential improvements for the model. Lastly, we highlight the numerous policy implications of our research. The resulting matrix will allow to evaluate, through the computational tool, major changes to the transport system that modify the stop where users start or end their trips; such as route extensions, relocation of bus stops, or designing new bus or metro lines.

1 INTRODUCCIÓN

Uno de los factores clave para una correcta planificación de sistemas de transporte urbano es disponer de una matriz origen-destino (O-D) confiable de la ciudad de interés (Ortúzar y Willumsen, 2011). Sin embargo, en urbes cambiantes y con zonificaciones complejas, es difícil y costoso mantener esta información al día. Adicionalmente, incluso gastando grandes cantidades de dinero en encuestas origen-destino en hogares y ruta, no es posible garantizar que las matrices resultantes sean precisas si no se incluye información adicional (Cools et al., 2010). Lo anterior evidencia la necesidad de contar con metodologías de recolección de datos sobre los patrones de viaje de los ciudadanos, más eficientes y efectivas que las encuestas O-D a nivel de ciudad.

Por otro lado, en la última década se ha observado una tendencia creciente a complementar los sistemas de transporte público con tecnologías de información, tales como sistemas de recolección automática de pago tarifario y de localización automática de vehículos, los cuales entregan información valiosa sobre los patrones de viaje de los ciudadanos.

En sistemas en que los pasajeros son forzados a validar su pasaje al subir y al bajar en cada etapa de viaje (ya sea en bus o en metro), obtener una matriz O-D de paradero a paradero es directo. En cambio, en sistemas donde los viajeros deben validar el pasaje solo al subir, no se deja registro del paradero de bajada. Sin embargo, desarrollos recientes han permitido estimar una alta proporción de ellos; en particular, Munizaga y Palma (2012) desarrollaron una metodología para Santiago y Gordon et al. (2013) para la ciudad de Londres. Estas matrices, si bien valiosas, no pueden ser utilizadas directamente para estimar los impactos de cambios en el sistema de transporte que modifiquen los paraderos en que los usuarios comienzan y/o terminan sus viajes; por ejemplo, cambios en la estructura tarifaria, relocalización de paradas, diseño de nuevos servicios o nuevas líneas de metro. Para estudiar estos cambios, es necesario disponer de una matriz origen-destino zona a zona, que permita a los usuarios modelados escoger sus paraderos iniciales y finales de viaje.

En particular, la red de transporte público de Santiago de Chile es extensa y está en permanente cambio. En este contexto, el sistema se beneficiaría de contar con mejor información sobre su demanda y viajes para apoyar las decisiones de gestión y planificación. Desde el 2007, con la implementación de Transantiago, el sistema de transporte de la capital cambió radicalmente, reestructurándose el servicio de buses por completo e incluyendo nuevos operadores y tecnologías, entre otros cambios relevantes (Muñoz y Gschwender, 2008). Además, los cambios en la red santiaguina han continuado hasta la actualidad y se proyectan a futuro, con numerosas modificaciones a los servicios de buses, extensiones de la red de Metro y nuevos desarrollos urbanos.

1.1 Objetivos

En vista de lo anterior, el objetivo general de esta tesis es desarrollar una metodología que permita construir una estimación de la matriz O-D zonal de transporte público para una ciudad, utilizando como base una matriz de viajes de paradero a paradero. El modelo propuesto permite inferir la zona de origen y destino para cada viaje observado entre dos paraderos o estaciones de metro, usando un modelo Logit con variables de uso de suelo e información de transporte público.

Cabe destacar que el modelo de inferencia calibrado para Santiago, junto con su respectiva matriz O-D zonal resultante de esta tesis, serán usados para el proyecto D10I1049 del Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF) “Una herramienta táctico-estratégica de gestión y planificación de sistemas de transporte público urbano”. Esta iniciativa busca generar una herramienta computacional que permita predecir el comportamiento de los usuarios de transporte público frente a cambios en la red, permitiendo mejorar el diseño y operación del sistema de transporte de Santiago.

1.2 Alcances

El modelo formulado es genérico y podría ser aplicado (tras su respectiva calibración) a cualquier ciudad y período de interés, siempre y cuando se cuente con la información

necesaria. Esta tesis en particular, contiene la calibración del modelo para el período de punta mañana de la ciudad de Santiago.

Adicionalmente, este documento informa sobre otras tareas predecesoras a la calibración que fueron necesarias para posibilitar la estimación y aplicación del modelo. Entre ellas, se destaca la realización de una encuesta de preferencias reveladas, la definición de la zonificación, y la creación de centroides y arcos de acceso para la red de modelación.

1.3 Estructura

Esta tesis está estructurada en ocho capítulos. El segundo capítulo presenta el marco teórico y la revisión bibliográfica, exhibiendo metodologías y aplicaciones de estimación de matrices O-D zonales y entre paraderos a partir de transacciones con tarjetas inteligentes, junto con una exploración de modelos de elección discreta y su aplicación a modelos de distribución de viajes. A continuación, el tercer capítulo incluye la formulación del modelo de inferencia zonal, junto con su procedimiento de estimación y reconstrucción de la matriz O-D zonal. El capítulo 4 introduce las distintas fuentes de datos, describiendo las entradas que tendrá el modelo. El procesamiento de estos datos para la construcción de la red y la base de datos se detalla en el capítulo 5. Luego, el capítulo 6 presenta y analiza estadísticamente el banco de estimación definitivo. La siguiente sección muestra los resultados de la estimación del modelo de inferencia, incluyendo el proceso de calibración de arcos de acceso y formulaciones alternativas del modelo, además de la aplicación en la herramienta y consiguiente reconstrucción de la matriz O-D de Santiago. Finalmente, el capítulo 8 contiene las conclusiones y lineamientos para investigación futura.

2 MARCO TEÓRICO

Como ya se adelantó, este capítulo contiene una revisión bibliográfica de metodologías de estimación de matrices origen-destino (O-D) relacionadas al problema enfrentado por esta tesis. En otras palabras, dada una división de la ciudad en zonas, se busca encontrar la cantidad de viajes entre cada zona i y cada zona j , definida como T_{ij} . Adicionalmente, se explora en profundidad el modelo de construcción de matrices de viajes entre paraderos, cuyos resultados son utilizados como entrada del modelo de inferencia zonal, junto con algunas de sus aplicaciones. Finalmente, se presenta una revisión de modelos de elección discreta, centrada en modelos Logit y su aplicación a reconstrucción de matrices O-D.

2.1 Estimación de matrices O-D zonales

Entre los numerosos problemas a los que se enfrentan los planificadores de transporte, uno de los más desafiantes (y potencialmente costosos de solucionar) corresponde a la determinación de la demanda por viajes entre zonas de una ciudad. Dada la complejidad de este objetivo y su envergadura, suele ocurrir que la información O-D contenga errores o se encuentre incompleta. Por esta razón, se recomienda aprovechar y combinar las fuentes de información disponibles que sean razonablemente confiables (Ben-Akiva, 1987). En esta línea, la problemática de estimación de matrices O-D ha sido abordada uniendo distintas fuentes de información y siguiendo diversos enfoques, presentados en las próximas secciones.

La forma general en que se ha abordado el tema es siguiendo el modelo clásico de transporte de cuatro etapas (generación, distribución, partición modal y asignación), presentado, entre muchos otros, por Ortúzar y Willumsen (2011). Luego de recopilada la información base, la etapa de generación consiste en estimar la cantidad de viajes generados y atraídos por cada una de las zonas del área de estudio; lo que puede lograrse con regresiones lineales zonales, cuyas variables dependientes corresponden a la cantidad de viajes generados o atraídos y las independientes a variables socioeconómicas o de uso de suelo de las zonas de análisis, por ejemplo. El próximo paso, la etapa de distribución,

permite repartir estos viajes a sus orígenes y destinos correspondientes, resultando una matriz O-D de zona a zona, equivalente al objetivo de esta tesis. En las etapas restantes los viajeros son asignados a los distintos modos de transporte y rutas posibles.

Es importante destacar que las distintas etapas del modelo presentan interacciones entre ellas. Por ejemplo, los resultados de la etapa de asignación podrían afectar la partición modal de la región de estudio (una alta congestión en un modo podría incentivar a que algunos usuarios se cambien a otro). De esta manera, si bien el modelo clásico de transporte constituye un punto de partida válido, es necesario estar conscientes de que su aplicación secuencial corresponde a una simplificación de la realidad (Ortúzar y Willumsen, 2011).

2.1.1 Modelos gravitacionales

Una de las primeras soluciones al problema de estimación de matrices origen-destino corresponde a los modelos gravitacionales de transporte, que llevan este nombre por su similitud con la teoría de física clásica homónima. Tradicionalmente han sido utilizados para resolver la etapa de distribución de viajes, asumiendo conocidas las cantidades de viajes atraídos y/o generados por cada zona, resultando en modelos simplemente o doblemente acotados, respectivamente (Cochrane, 1975).

Una de sus formulaciones más sencillas, perteneciente al caso doblemente acotado, es explicada por Erlander y Stewart (1990) y presentada, en forma simplificada, a continuación.

En primer lugar, suponemos que la cantidad de viajes entre las zonas i y j responden a la siguiente fórmula gravitacional:

$$T_{ij} = A_i B_j f_{ij} \quad (2.1)$$

donde A_i y B_j corresponden a medidas de los potenciales de generación y atracción de las zonas i y j , así como también a factores de balance que aseguran el cumplimiento de las

restricciones; T_{ij} representa la cantidad (desconocida) de viajes de la zona i a la zona j ; f_{ij} es una medida inversa del costo de viajar entre estas zonas o función de separación.

Si asumimos conocidos los totales de viajes originados y atraídos por cada zona, esta fórmula gravitacional está sujeta a las restricciones:

$$\sum_j T_{ij} = O_i \quad \forall i \quad (2.2)$$

$$\sum_i T_{ij} = D_j \quad \forall j \quad (2.3)$$

Donde O_i y D_j corresponden a los totales de viajes con origen en i o destino en j , respectivamente.

La necesidad de cumplir estas restricciones genera una interdependencia entre los factores A_i y B_j , que puede solucionarse con procedimientos iterativos en que: dado un conjunto de valores para la medida inversa del costo f_{ij} , comenzar con $B_j = 1$, encontrar A_i y usar estos valores para re-estimar B_j , repitiendo hasta alcanzar la convergencia. Finalmente, se utilizan los valores definitivos de A_i , B_j y f_{ij} para calcular los viajes T_{ij} entre cada par de zonas (Ortúzar y Willumsen, 2011).

2.1.2 Estimación utilizando conteos de flujo

Otra arista de esta área de estudio corresponde a la actualización de la información origen-destino previamente obtenida. En efecto, se hace necesario actualizar los datos sobre los patrones de viajes de la región de estudio, para responder al crecimiento y evolución que esta puede presentar con el paso de los años.

Esto se ha hecho, principalmente para el caso de transporte privado, uniendo matrices origen-destino obtenidas *a priori* (a partir de encuestas origen-destino a gran escala y

modelos gravitacionales, por ejemplo) con conteos de flujo vehicular (ver, por ejemplo, Bell, 1983; Bierlaire y Toint, 1995; Van Zuylen y Willumsen, 1980).

Un ejemplo de formulación siguiendo esta lógica clásica se presenta en Bierlaire y Toint (1995), donde se busca encontrar los elementos T_{ij} de la matriz O-D objetivo; minimizando la “distancia” entre la nueva matriz y una matriz *a priori*, y entre los conteos de flujos observados y los predichos tras la asignación de la nueva matriz. Matemáticamente, podemos expresar el problema de la siguiente manera:

$$\underset{\{T_{ij}\}}{\text{minimizar}} \sum_{i \in O, j \in D} w_{ij}^t (T_{ij} - t_{ij})^2 + \gamma \sum_{a \in A} w_a^v (V_a - v_a)^2 \quad (2.4)$$

donde O y D corresponden a los potenciales nodos generadores y atractores de viajes, respectivamente; A es el conjunto de arcos para los que se tiene conteo de flujos; t_{ij} representa los viajes entre el origen i y el destino j en la matriz *a priori*; V_a simbolizan los viajes en el arco a resultantes de asignar la matriz T objetivo (mediante una asignación a rutas de costo mínimo, por ejemplo), mientras que v_a indican los conteos observados en los arcos; se incluyen también parámetros de peso w_{ij}^t y w_a^v que indican la confianza relativa que se tiene en cada elemento de la matriz *a priori* y del conteo de flujo; finalmente, γ corresponde al peso global relativo asignado a los conteos de tráfico en comparación con la matriz *a priori*.

Por el lado del transporte público, foco de esta tesis, han surgido recientemente modelos orientados a la reconstrucción de matrices O-D. No obstante, el volumen de trabajos reportados es menor que para su contraparte privada, lo que puede ser parcialmente atribuido a la mayor dificultad de los modelos de elección de ruta asociados (ver, por ejemplo, Lam et al., 2003; Fisk, 1989; y Florian y Chen, 1995).

En particular, Lam et al. (2003) siguen una formulación similar a la expresada en la Ecuación (2.4), pero adaptada para el caso del transporte público. La principal diferencia entre ambas formulaciones es el modelo de asignación subyacente (es decir, la manera en

que se calculan los términos V_a de la Ecuación (2.4)). En el trabajo propuesto en Lam et al. (2003), la asignación se realiza siguiendo un modelo de equilibrio de usuarios estocásticos (SUE por su sigla en inglés) tipo Logit (para más información sobre el SUE revisar Maher et al., 2001; o Yang et al., 2001). Por último, otro aspecto distinto es la manera de definir los viajes y conteos de flujos, que en el caso de transporte privado pueden ser simplemente conteos vehiculares recolectados con espiras magnéticas, mientras que para el transporte público se requiere saber la cantidad de pasajeros utilizando los buses o trenes en cada segmento de línea de la red.

2.2 Estimación de matrices de viajes entre paraderos de transporte público

Por otro lado, para el caso de la ciudad de Santiago, Munizaga y Palma (2012) desarrollaron un modelo capaz de estimar matrices de viajes entre paraderos y estaciones de metro. Como se adelantó previamente, para lograrlo se debe primero identificar los paraderos y tiempos de subida y bajada, asociados a cada transacción de tarjeta inteligente. Lo anterior se realiza considerando información de las tarjetas Bip!, la geo-codificación de la red de transporte público y los GPS de los buses. Esta información es utilizada como entrada para la metodología de estimación, la cual considera además una serie de supuestos que serán explicados a continuación.

El principio general de la metodología corresponde a identificar el lugar y servicio al que cada usuario subió, y su correspondiente punto de bajada se asume como el más cercano en la ruta al próximo punto de subida (para la última validación del día, en cambio, se asume que el punto de bajada se encuentra cerca de la primera validación de ese mismo día). Por supuesto, “más cercano” no se refiere simplemente a la distancia física de caminata, sino que debe considerar el tiempo generalizado de viaje desde un paradero a otro; es decir, considera tanto tiempo de viaje en vehículo como caminando (este último es penalizado para representar la preferencia de los usuarios por el tiempo de viaje en vehículo por sobre el de caminata). Un ejemplo que ilustra la necesidad de considerar un tiempo generalizado, es un viaje por un corredor en el que transita un servicio de ida y vuelta. Si escogiéramos como punto final el paradero más cercano al próximo paradero de

subida, la bajada se asignaría exactamente al paradero donde se realiza la siguiente validación. Esto tiene el problema de que asumiría que el usuario pasa frente a su destino, pero prefiere esperar que el bus termine el recorrido, dé la vuelta y vuelva a llegar al destino, en vez de bajarse y cruzar la calle.

Para implementar lo anterior, se requiere definir una distancia de caminata máxima desde el paradero de bajada al de próxima subida, la que para Munizaga y Palma (2012) fue parametrizada en 1.000 metros. En los casos en que no se encuentra un paradero de bajada a menos de mil metros de la próxima subida, se considera que se perdió una etapa o viaje, puesto que probablemente el usuario no caminó más de eso sino que utilizó otro modo, tarjeta Bip! o evadió el pago de su pasaje.

Como es de esperar, la metodología propuesta por el artículo tiene algunas diferencias para identificar paraderos de subida y bajada para etapas de viajes en buses, metro y zonas pagas.

Para viajes en metro, las estaciones de subida y bajada se pueden obtener directamente aplicando la metodología anterior, pero solo para viajes que pagaron con tarjeta inteligente. Es decir, no es posible encontrar la estación de bajada para un usuario que pagó con boleto de papel (que corresponden a un pequeño porcentaje en el horario de punta mañana). Además, no es posible identificar la ruta que siguieron para llegar de una estación a otra, puesto que para realizar transbordos en Metro no es necesario validar nuevamente la tarjeta. Para solucionar este problema, las autoras asumieron un comportamiento de ruta de costo mínimo de acuerdo al algoritmo de Dijkstra (1959).

Por otro lado, el bus al que subió un pasajero que validó en una zona paga no puede ser obtenido directamente, sino que tiene que ser estimado. En este caso, se identifican primero todas las rutas que pasan por la zona paga y tienen algún paradero a distancia caminable del lugar donde se hizo la próxima validación. Luego, si existen dos o más rutas, se hace el supuesto de que el usuario se sube al primer bus que pasa de su conjunto de líneas comunes, definido según Chriqui y Robillard (1975).

Una vez estimados los tiempos y posiciones de subida y bajada de cada transacción, hay que realizar un procesamiento adicional para obtener una matriz de viajes entre paraderos y estaciones, puesto que es necesario diferenciar entre viajes (movimientos de un origen a un destino con el objetivo de realizar una actividad en el último) y etapas de viaje (movimientos en un servicio en particular, ya sea de bus o Metro).

Para ello, Munizaga y Palma (2012) realizan el simple supuesto de que si una persona (tarjeta Bip!) se mantiene por más de 30 minutos en un lugar, es un destino en el que se realizó alguna actividad (de lo contrario, es una estación de transbordo). Esta suposición no permite detectar actividades muy cortas y falla si hubo tiempos de espera muy altos en un paradero. Por otra parte, para los casos en que la posición de bajada no se puede estimar, se asume que la próxima transacción es el comienzo de un nuevo viaje, siempre que esté a más de dos horas de distancia de la anterior. Además, se reconoce que es posible incluir criterios adicionales para mejorar la estimación, como por ejemplo, si dos transacciones seguidas son realizadas en Metro o en la misma línea de buses, se puede asumir que hubo un destino entre ellas, puesto que es poco probable que una persona baje y vuelva a subir en cualquiera de estos casos, perdiendo tiempo y volviendo a pagar la tarifa sin necesidad de hacerlo.

Respecto a las limitaciones de este procesamiento, cabe destacar que la metodología no es capaz de asignar paraderos de subida y bajada a todos los viajes, sino que lo logra para cerca del 80% de las transacciones. Para solucionar este problema, se hace el supuesto de que los viajes no reconstruidos tienen la misma distribución que los observados en su zona de origen (la que es asignada simplemente a la zona que físicamente contiene al paradero de inicio), por lo que son expandidos en esta proporción. Además, no es posible asegurar que todos los viajes reconstruidos hayan sido asignados correctamente a sus paraderos iniciales y finales (o que sean realmente etapas de viajes y no viajes propiamente tal). Por último, es importante notar que el modelo se puede aplicar solamente a los viajes que validaron con su tarjeta; es decir, los usuarios que evaden la tarifa no quedan registrados en la matriz de salida. Esto puede generar distorsiones importantes en el caso de Santiago,

que presenta niveles de evasión promedio cercanos al 20% (pero considerablemente mayores en algunos sectores de la capital, ver DTPM, 2014).

Por otra parte y siguiendo la misma línea, Devillaine (2012) presenta algunas mejoras a la metodología de reconocimiento de viajes de Munizaga y Palma (2012); implementando el criterio de dos transacciones seguidas en Metro o una misma línea de buses previamente expuesto. Además, identifica y elimina parte de los viajes infactibles que pueden haberse generado, considerando la imposibilidad geométrica de un viaje en que la distancia euclidiana entre sus paraderos inicial y final sea mayor que la distancia recorrida en ruta. Por último, asume que un usuario que está esperando en el paradero no deja pasar más de tres buses que le sirvan; de lo contrario, se encontraba realizando una actividad y por lo tanto no corresponde a un transbordo sino a un viaje. Este último criterio se justifica en la experiencia de los modeladores, quienes consideran infrecuente que un pasajero tenga que dejar pasar más de tres buses por razones de capacidad.

En comparación con la metodología anterior, Devillaine (2012) logra estimar un 14% más de viajes para la base de datos de comparación de Santiago. Esto es especialmente relevante puesto que la mayor cantidad de viajes estimados proviene de identificar como viajes, trayectos que previamente estaban identificados como etapas de viaje. De esta manera, el modelo propuesto por Devillaine (2012) logra representar de mejor manera la partición de etapas de viajes, indicando por ejemplo una mayor proporción de viajes de una etapa. Además, mejora la proporción de viajes que poseen estimaciones de tiempos de viaje, entre otras diferencias.

Recientemente, Munizaga et al. (2014) complementaron ambos trabajos, proponiendo nuevas mejoras metodológicas que permiten disminuir errores y aumentar la cantidad de viajes estimados. En primer lugar, se sugiere revisar las distancias de caminata máxima, puesto que estas pueden variar según el uso de suelo de la zona (sin embargo, este análisis requeriría una base de datos exógena, no disponible actualmente). Segundo, se reconoce que existe una cantidad no despreciable de viajes que ocurre pasada la media noche, posiblemente identificados de manera incorrecta por la metodología previa; por ello, se

propone redefinir el período de transición entre días para las 4:00 am (donde el número de transacciones es mínimo). Adicionalmente, el artículo propone correcciones para las tarjetas que se observan solo una vez en un día, estimando su punto de bajada con información de la primera transacción del próximo día. Finalmente, se presentan mejoras al proceso de identificación de etapas de viaje, descomponiendo aquellos viajes en que la distancia entre origen y destino en ruta es mucho mayor que la euclidiana (si un usuario usa una ruta extremadamente indirecta, es probable que haya realizado una actividad intermedia en alguno de sus puntos de transbordo).

2.3 Extensiones y aplicaciones de matrices estimadas de viajes entre paraderos para Santiago

Los resultados del modelo de estimación de matrices de viajes entre paraderos y estaciones de Metro, presentado en el subcapítulo anterior, representan una valiosa fuente de información sobre el comportamiento de los usuarios del sistema de transporte. En particular, en las siguientes secciones se exhiben dos aplicaciones que buscan extraer conclusiones sobre las actividades e individuos detrás de las transacciones; un modelo de estimación de propósito de viaje y un modelo de estimación de zona de residencia.

2.3.1 Estimación de propósitos de viajes

Una vez identificadas las etapas y los viajes, junto con sus consiguientes puntos y tiempos de subida y bajada, Devillaine (2012) propone un módulo de asignación de propósito de viajes. Este algoritmo, podría ayudar a completar pasivamente información de actividades, relevante para encuestas y estudios origen-destino en general. La Tabla 2-1 presenta los criterios definidos para viajes en días laborales, en los que se enfoca el modelo propuesto en esta tesis.

Lo anterior se basa en literatura previa relacionada y en el criterio de los modeladores, quienes reconocen la posibilidad de refinar estos criterios en la medida en que se disponga de una mejor forma de validar los resultados, así como también de mayor información referente a uso de suelo, por ejemplo.

Tabla 2-1: Criterios de asignación de propósitos para día laboral

Propósito	Criterios de asignación
Trabajo	<ul style="list-style-type: none"> • Tipo de tarjeta es adulto • La actividad dura más de 5 horas • El viaje anterior a la actividad no fue el último del día
Estudio	<ul style="list-style-type: none"> • Tipo de tarjeta es estudiante • La actividad dura más de 5 horas • El viaje anterior a la actividad no fue el último del día
Hogar	<ul style="list-style-type: none"> • El viaje anterior a la actividad fue el último del día
Otros	<ul style="list-style-type: none"> • La actividad dura entre 1 minuto y 5 horas • El viaje anterior a la actividad no fue el último del día

Fuente: Devillaine (2012)

2.3.2 Estimación de la zona de residencia

Tomando como base las matrices de viajes entre paraderos construidas siguiendo la metodología previamente presentada, Amaya (2013) propone dos tipos de análisis de patrones de viajes y conductas de usuarios, comparando entre: (1) cortes temporales y (2) distintas zonas de la ciudad para un mismo corte temporal; siendo este último de especial interés como referencia para esta tesis. Para realizar este análisis, propone una

metodología de estimación de zona de residencia, expuesta de manera sintetizada en Amaya y Munizaga (2013).

La metodología se basa en el supuesto de que el principal modo de transporte de un usuario frecuente es el transporte público, y que la primera transacción de cada día es cercana a su hogar. De esta manera, si las primeras transacciones de cada día se encuentran dentro de un cierto radio caminable (parametrizado en 500 metros), se considera que el individuo vive dentro de ese radio y se le asigna la zona correspondiente. En el caso de estudio de Amaya y Munizaga (2013), se propuso utilizar la zonificación sugerida por Jara-Díaz et al. (2013), que divide la ciudad de Santiago en seis macro-zonas: Centro, Norte, Sur, Poniente, Oriente y Suroriente.

Usando estos supuestos, se logra estimar el lugar de residencia para cerca del 25% de las tarjetas Bip! de Santiago. Sin embargo, esto podría extenderse si se consideran períodos de tiempos mayores. Adicionalmente, una validación exploratoria muestra un gran porcentaje de aciertos en la zona de residencia, aunque sin validez estadística (Amaya y Munizaga, 2013).

Uno de los casos que esta metodología no podría capturar es cuando un usuario utiliza regularmente un modo no integrado en Transantiago (como auto o taxi colectivo) para realizar la primera etapa de su viaje, y luego se conecta con la red Bip! y llega a destino. En un caso como este (común en algunas estaciones de la capital, ver Metro, 2011), la zona de residencia quedaría asignada a las cercanías de donde realiza su primera validación con la tarjeta inteligente y no al hogar real.

Por otro lado, Amaya y Munizaga (2013) sugieren que variando el radio caminable la metodología es aplicable para otras zonificaciones, como nivel comunal o de zona ESTRAUS (ver, por ejemplo, Fernández y De Cea, 2005). No obstante, identifican complicaciones de aplicarlo a una zonificación como esta última, puesto que un número importante de zonas tienen formas no cóncavas y existe una gran variabilidad en su

tamaño. Para enfrentar esta problemática, las autoras plantean estudiar alguna metodología de estimación de residencia con áreas de influencia de límites dinámicos.

Sin embargo, es importante considerar que la mayoría de los bordes zonales de ESTRAUS están definidos en calles importantes de la ciudad, por lo que un gran número de estaciones de Metro y paraderos quedan ubicados entre dos o más zonas. De esta manera, incluso considerando áreas o radios dinámicos extremadamente pequeños, en muchos casos no sería posible garantizar una zona de residencia de manera unívoca para una zonificación tan detallada como la de ESTRAUS.

2.4 Modelos de elección discreta

Como fue mencionado y se profundiza en el subcapítulo 2.5, los modelos de elección discreta tipo Logit han sido aplicados al problema de estimación de matrices origen-destino. En vista de esto, se hace necesario presentar una breve revisión de ellos en este apartado, incluyendo sus supuestos, deducción y algunas de sus formas más relevantes. Para más detalles sobre la formulación Logit, referirse a Daly (1979), McFadden (1979) u Ortúzar (2000).

2.4.1 Teoría de la utilidad aleatoria

Esta familia de modelos asume como punto de partida la teoría de la utilidad aleatoria, la que supone que la utilidad (o beneficio asociado al consumo de un bien) se deriva de los atributos del bien, y no del bien en sí mismo. En esta misma línea, suponiendo un individuo q con un conjunto de alternativas A_q de elección, podemos suponer que su utilidad U_{iq} asociada a la opción i es:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \quad (2.5)$$

donde V_{iq} es la componente sistemática o representativa de la utilidad, función de los atributos medidos y que se asume conocida; y ε_{iq} es una componente aleatoria que permite representar atributos no especificados, errores de medición y gustos específicos de los

individuos, entre otros detalles que escapan a la modelación. El subíndice q en los distintos términos representa que los distintos atributos y valoraciones pueden variar de individuo en individuo (Domencich y McFadden, 1975).

Siguiendo la teoría económica clásica, cada individuo escoge (sin ninguna aleatoriedad) la alternativa que le reporta una utilidad más alta. Sin embargo, como el modelador es incapaz de observar perfectamente este proceso de decisión, se introducen los términos ε_{iq} para explicar comportamientos aparentemente irracionales.

De esta manera, un individuo escoge la alternativa i (de su conjunto de alternativas A_q) si y solo si:

$$U_{iq} \geq U_{jq}, \forall j \in A_q \quad (2.6)$$

O, equivalentemente:

$$V_{iq} - V_{jq} \geq \varepsilon_{jq} - \varepsilon_{iq}, \forall j \in A_q \quad (2.7)$$

Puesto que no conocemos los términos aleatorios, no es posible saber con seguridad si estas desigualdades se cumplen o no; por lo tanto, la elección de cada individuo es expresada en términos probabilísticos.

Ahora, si suponemos alguna distribución conocida para los ε_{iq} , podemos calcular las probabilidades de elección de cada alternativa.

2.4.2 Modelo Logit Multinomial

En el caso más sencillo, se asume una distribución de los términos aleatorios *i.i.d.* Gumbell, obteniéndose la siguiente probabilidad de elección de la opción i (asumiendo individuos iguales):

$$Prob(i) = \frac{e^{\beta V_i}}{\sum_{j \in A} e^{\beta V_j}} \quad (2.8)$$

Esta formulación corresponde al modelo Logit Multinomial (MNL, por su abreviación en inglés) y es una de las más utilizadas en la práctica. Cabe destacar que los parámetros de la función de utilidad representativa no pueden ser estimados de manera independiente del parámetro β , por lo que el modelo suele ser reescrito obviando este último:

$$Prob(i) = \frac{e^{V_i'}}{\sum_{j \in A} e^{V_j'}} \quad (2.9)$$

Dentro de las propiedades del MNL, la más relevante es la de independencia de alternativas irrelevantes, que se puede expresar en el siguiente cociente:

$$\frac{Prob(i)}{Prob(j)} = e^{\beta(V_i - V_j)} \quad (2.10)$$

Luego, la proporción de elección entre dos alternativas depende únicamente de sus utilidades representativas, y puede ser calculada sin tomar en cuenta el resto de las alternativas. Esta propiedad, si bien puede parecer ventajosa, puede conducir a predicciones sesgadas si hay correlación entre alternativas (considerar el famoso ejemplo del bus azul y bus rojo; Mayberry, 1973).

2.4.3 Modelo Logit Jerárquico

Por otro lado, el modelo Logit Jerárquico (HL, por el inglés *Hierarchical Logit*) permite tratar estructuras de alternativas en que algunas de ellas presentan correlación entre sí (Williams, 1977). Para ello, esta formulación agrupa las alternativas más semejantes entre ellas en jerarquías o nidos, como muestra la Figura 2-1.

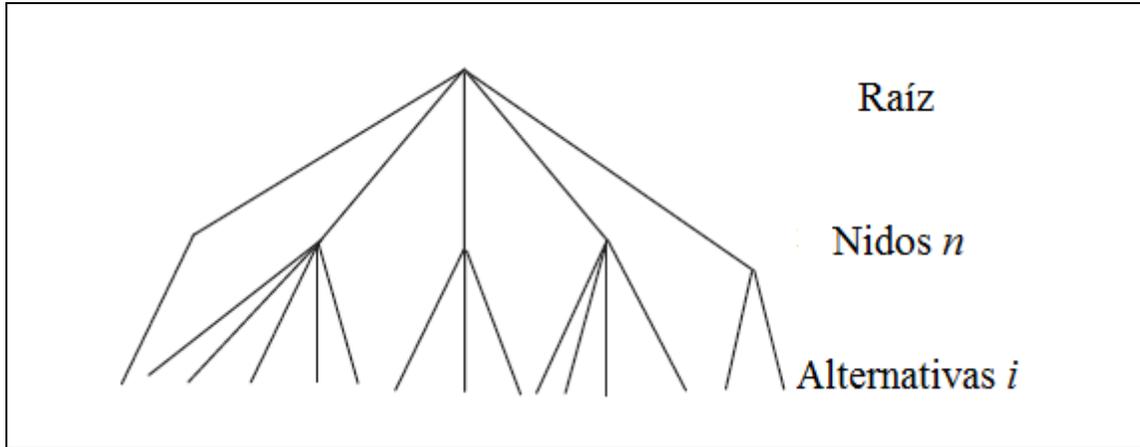


Figura 2-1: Ejemplo de estructura jerárquica de HL

Fuente: Ortúzar y Willumsen (2011)

Consideremos el caso sencillo de una elección en dos niveles como el representado por la imagen, asumiendo que las alternativas i pertenecientes a un nido n se encuentran correlacionadas entre sí. En este caso, las utilidades asociadas a elegir las alternativas i_1 e i_2 del nido n_1 están dadas por:

$$U_{n_1, i_1} = V_{n_1} + V_{i_1, n_1} + \varepsilon_{n_1} + \varepsilon_{i_1, n_1} \quad (2.11)$$

$$U_{n_1, i_2} = V_{n_1} + V_{i_2, n_1} + \varepsilon_{n_1} + \varepsilon_{i_2, n_1} \quad (2.12)$$

donde se puede observar que existe una fuente de correlación entre ambas, proveniente del residuo ε_{n_1} (en el caso de que este residuo sea igual a 0, el modelo HL colapsa al MNL). No obstante, siguiendo esta formulación es igualmente posible encontrar expresiones cerradas para la probabilidad de escoger una alternativa i_1 perteneciente a un nido n_1 , como:

$$Prob(n_1, i_1) = Prob(n_1) * Prob(i_1/n_1) \quad (2.13)$$

Desarrollando:

$$Prob(n_1, i_1) = \frac{e^{\beta(V_{n_1} + V_{n_1}^*)}}{\sum_n e^{\beta(V_n + V_n^*)}} * \frac{e^{\lambda V_{n_1, i_1}}}{\sum_{i \in n_1} e^{\lambda V_{n_1, i}}} \quad (2.14)$$

donde (siendo λ el equivalente al parámetro β pero de la jerarquía inferior):

$$V_n^* = \left(\frac{1}{\lambda}\right) \ln \sum_{i \in n} e^{\lambda V_{n, i}} \quad (2.15)$$

Es importante destacar, que para que el modelo sea internamente consistente, se debe cumplir que:

$$\beta \leq \lambda \quad (2.16)$$

Por último, es importante notar que este modelo no permite tratar casos en que exista correlación entre alternativas pertenecientes a nidos diferentes.

2.5 Estimación de modelos de elección con variables de atracción

Una de las razones por la que los modelos Logit han adquirido tanta atención es su versatilidad. En efecto, su formulación puede ser aplicada a una variedad de contextos y procesos de elección, que no necesariamente traten de usuarios escogiendo el bien que maximice su nivel de utilidad.

En esta línea, Daly (1982) demuestra que un modelo gravitacional agregado puede ser representado en forma Logit sin ninguna pérdida de generalidad y sin imponer restricciones sobre las formas funcionales de sus parámetros.

El procedimiento utilizado por el autor considera como ejemplo un modelo gravitacional simplemente acotado, definido por:

$$T_{ij} = A_i B_j f_{ij} \quad (2.17)$$

Sujeto a la restricción de orígenes:

$$\sum_j T_{ij} = O_i \quad \forall i \quad (2.18)$$

Este modelo es de tipo agregado, pues predice el número total de viajes entre cada par de zonas y no la decisión de cada individuo. Sin embargo, la probabilidad de que un viajero individual que se encuentra en la zona i elija como destino a la zona j se puede expresar de manera desagregada como:

$$Prob(j/i) = \frac{T_{ij}}{O_i} = \frac{B_j f_{ij}}{\sum_j B_j f_{ij}} \quad (2.19)$$

Luego, si definimos:

$$V_k = \ln(B_k f_{ik}) \quad (2.20)$$

Obtenemos un modelo Logit para la decisión de destino de los viajeros de la zona i :

$$Prob(j/i) = \frac{e^{V_j}}{\sum_k e^{V_k}} \quad (2.21)$$

Una forma de explicar y comprender este modelo, proviene de considerar que la elección real del destino de cada viajero se encuentra oculta en la agregación zonal. Así, por ejemplo, un usuario viaja a una zona k debido a que desea visitar un centro comercial que se encuentra en esa zona y no por ir a esa zona como fin último. De esta manera, cada zona representa una agregación de un gran número de alternativas elementales (*e.g.*, oficinas de trabajo u hogares), lo que a su vez fuerza la inclusión de variables de tamaño en los parámetros B_k (tales como número de edificios de oficinas o superficie construida educacional, por ejemplo), que indiquen no la “calidad” sino la cantidad de alternativas elementales existentes en la zona k .

En este sentido, usualmente se requiere que la probabilidad de elegir una zona como destino sea directamente proporcional a sus variables de tamaño, lo que asegura que el modelo sea independiente de la agregación de las alternativas elementales (es decir, independiente de la zonificación utilizada). En otras palabras, si por ejemplo dos zonas se fusionan, es de esperar que la cantidad de viajes atraídos a la nueva zona sea igual a la suma de los viajes atraídos a las zonas originales (Daly, 1982).

Para que se cumpla lo anterior, la forma funcional del parámetro de atracción debe ser:

$$\ln(B_k) = \theta_1 x_{1k} + \dots + \theta_r x_{rk} + \theta_{r+n+1} \ln(\theta_{r+1} S_{1k} + \dots + \theta_{r+n} S_{nk}) \quad (2.22)$$

donde las variables x indican la calidad de la elección k , mientras que las S son variables de tamaño que describen la cantidad de alternativas elementales en la zona. Cabe destacar que por razones de identificabilidad, uno de los ponderadores asociado a una variable de tamaño no puede ser estimado y se fija en 1.

Por su parte, el parámetro θ_{r+n+1} se introduce por consistencia, pero puede ser reinterpretado considerando los modelos HL presentados en la sección anterior. Este proceso de elección es equivalente a un modelo HL en que el primer nivel es la elección de zona, y en el segundo se escoge la alternativa elemental dentro de la zona. La teoría indica que este coeficiente debe estar entre 0 y 1; un valor menor a 1 indica que los límites zonales tienen alguna importancia para los viajeros por sí mismos, mientras que si vale 1 el modelo se vuelve independiente de la zonificación utilizada para su estimación (Daly, 1982). En relación con esto último, es importante destacar que esta proporcionalidad puede mantenerse o no dependiendo del parámetro f_{ij} , por lo que debe ser calibrado adecuadamente al modificar la zonificación.

Por último, un punto importante es que cuando distintos viajeros son atraídos a distintas alternativas elementales, el modelo presentado es solo una aproximación del modelo correcto. Por esto, si existen datos suficientes para diferenciar a distintos tipos de usuarios

y lo que buscan en sus viajes, estos deberían incluirse para representar fielmente el proceso de elección (Daly, 1982).

3 FORMULACIÓN DEL MODELO Y MÉTODOS DE ESTIMACIÓN

Como ya fue mencionado, el objetivo de esta tesis es desarrollar una metodología para estimar la matriz origen-destino (O-D) zonal para un sistema de transporte público, sin la necesidad de aplicar grandes encuestas O-D sino que usando información recolectada pasivamente. Esto se logra con el modelo de inferencia zonal formulado a continuación, que asigna probabilidades de zonas de origen y destino a viajes entre paraderos de la red, provenientes de una matriz exógena. Una vez deducido, se presenta el procedimiento de estimación mediante máxima verosimilitud (usando información de encuestas) que fue utilizado en esta tesis. Adicionalmente, el Anexo A presenta una metodología de estimación alternativa correspondiente a minimización de errores cuadráticos (usando información pasiva de viajes entre paraderos de transporte público).

3.1 Modelo de inferencia zonal

El modelo de inferencia zonal propuesto permite obtener la probabilidad de que un viaje observado entre dos paraderos k y l (como sus puntos de subida y bajada, respectivamente), haya sido originado en la zona i y se dirija a la zona j . Esta probabilidad se define como $Prob\left(\frac{ij}{kl}\right)$.

Para obtener estas probabilidades para cada paradero, estación de metro, y zona, seguimos a Daly (1982) y formulamos este particular modelo gravitacional:

$$T_{ij}^{kl} = A_i B_j f_{ij}^{kl} \quad \forall (i, j) \wedge (k, l) \quad (3.1)$$

Sujeto a las restricciones:

$$\sum_m \sum_n T_{mn}^{kl} = T_{kl} \quad \forall (k, l) \quad (3.2)$$

donde A_i y B_j corresponden a medidas de los potenciales de generación y atracción de las zonas i y j , ligadas a su información de uso de suelo; T_{ij}^{kl} representa la cantidad (desconocida) de viajes de la zona i a la zona j que usan los paraderos k y l como sus puntos de subida y bajada, respectivamente; f_{ij}^{kl} es una medida inversa del “costo” de escoger los paraderos k y l para realizar un viaje entre i y j ; por último, T_{kl} corresponde a la cantidad de viajes observados entre los paraderos k y l , que se asume conocida para toda la red.

De esta manera, para un viaje que comienza en el paradero k y termina en l , la probabilidad de que su origen sea la zona i y su destino la zona j puede ser expresada como:

$$Prob(ij/kl) = \frac{T_{ij}^{kl}}{T_{kl}} \quad (3.3)$$

Adicionalmente, si reemplazamos las Ecuaciones (3.1) y (3.2) en (3.3), obtenemos:

$$Prob(ij/kl) = \frac{A_i B_j f_{ij}^{kl}}{\sum_m \sum_n A_m B_n f_{mn}^{kl}} \quad (3.4)$$

Si ahora definimos:

$$V_{mn/kl} = \ln(A_m B_n f_{mn}^{kl}) \quad (3.5)$$

Lo anterior puede ser representado por un modelo desagregado Logit, de la siguiente manera:

$$Prob(ij/kl) = \frac{e^{V_{ij/kl}}}{\sum_m \sum_n e^{V_{mn/kl}}} \quad (3.6)$$

Es importante destacar que este modelo (que incluye variables de tamaño) se obtiene sin ninguna pérdida de generalidad y sin imponer restricciones en la especificación de la función de costos a utilizar (Daly, 1982).

Finalmente, al definir T_{ij} como la cantidad de viajes de la zona i con destino en la zona j , es posible reconstruir cada celda de la matriz O-D zonal objetivo, a partir del comportamiento de viajes observado entre los paraderos y estaciones de la red, como muestra la Ecuación (3.7):

$$T_{ij} = \sum_k \sum_l Prob(ij/kl) * T_{kl} \quad \forall(i,j) \quad (3.7)$$

En consecuencia, la tarea remanente consiste en encontrar la forma apropiada para los parámetros A_m , B_n y f_{mn}^{kl} de las funciones de “utilidad” expresadas en la Ecuación (3.5), tomando las consideraciones adecuadas al incluir las variables de tamaño, siguiendo la metodología de Daly (1982) expuesta en el subcapítulo 2.5.

3.2 Estimación por máxima verosimilitud

Si se cuenta con información de encuestas, es posible estimar distintas especificaciones del modelo utilizando el método de máxima verosimilitud (Ortúzar y Willumsen, 2011). Consiguientemente, es necesario definir en primer lugar cada observación q obtenida de los encuestados. En esta línea, cada uno de los viajes registrados debe tener zonas y estaciones asignadas para sus orígenes y destinos. Además, se debe encontrar los subconjuntos m y n como se definen por la Ecuación (3.6), correspondientes a las zonas conectadas con los paraderos iniciales k y finales l . Una vez realizado, la ecuación de verosimilitud se plantea de la siguiente manera:

$$L(\theta) = \prod_q Prob(i_q j_q / k_q l_q) = \prod_q \frac{e^{V_{i_q j_q / k_q l_q}}}{\sum_{m_q} \sum_{n_q} e^{V_{m_q n_q / k_q l_q}}} \quad (3.8)$$

Donde los subíndices q indican que los conjuntos de zonas y estaciones (tanto de origen como destino), difieren para cada observación q . Luego, la Ecuación (3.8) es maximizada en los parámetros θ de la función de utilidad V_{ij} (*i.e.*, los parámetros relacionados a A_i , B_j y f_{mn}^{kl}), con lo que es posible deducir y comparar distintas especificaciones del modelo de inferencia zonal.

4 DATOS Y ENTRADAS DEL MODELO PARA SANTIAGO DE CHILE

En este capítulo se presentan las distintas fuentes de información que fueron utilizadas para construir la base de datos para la calibración y aplicación del modelo para la ciudad de Santiago. Las distintas entradas y definiciones previas, necesarias para la construcción del modelo, son: información de viajes entre paraderos; información sobre el costo de elegir cada paradero o estación; información de uso de suelo; zonificación a utilizar; y finalmente, la encuesta de preferencias reveladas sobre viajes en Transantiago, realizada a fines del 2013.

4.1 Información origen-destino de viajes entre paraderos

El modelo de inferencia zonal presentado opera asignando los viajes observados entre un par de paraderos a sus zonas conectadas, mediante un modelo de tipo Logit. De esta forma, se tomará la información de las matrices de viajes entre paraderos de transporte público estimadas según la metodología presentada en el subcapítulo 2.2, las que serán usadas en los parámetros T_{kl} , definidos en el capítulo anterior.

Es importante recalcar que la forma en que estas matrices se construyeron, se traduce en algunas limitaciones para el modelo desarrollado en esta tesis. Dos problemas particularmente relevantes surgen de los viajes evadidos y de aquellos que tienen alguna etapa en un modo no integrado a Transantiago

En primer lugar, puesto que no se tiene información en la matriz T_{kl} sobre los viajes evadidos, no es posible aplicar el modelo de inferencia zonal sobre ellos (pues corresponden a viajes no observados). Por lo tanto, la matriz resultante tras la aplicación del modelo de inferencia zonal subestimaría los viajes totales en transporte público de la ciudad, representando en cambio únicamente a los viajes que pagan la tarifa.

Semejantemente, el viaje de un usuario que cada día utiliza un modo no integrado a Transantiago (como taxi colectivo o automóvil privado) para acercarse a su paradero de inicio, quedará asignado por el modelo de esta tesis a una de las zonas aledañas a este

último, en vez de a su origen real. Este efecto se traduce en que la matriz origen-destino (O-D) resultante de la aplicación del modelo de inferencia zonal subestimaré los viajes originados en las zonas donde comienzan viajes multimodales en taxi colectivo o automóvil y sobreestimaré los viajes originados en las zonas donde se realizan los transbordos desde estos modos a los integrados en Transantiago.

Por último, usuarios que no evadan su viaje completo sino solo alguna de sus etapas (como es el caso de la denominada evasión blanda) pueden producir distorsiones similares a las asociadas a viajes multimodales.

4.2 Costo de elección de paraderos

Para formular el modelo de inferencia zonal, es necesario conocer cómo los usuarios toman sus decisiones de viaje e incluir esta información correctamente. En esta línea, podemos observar que el modelo gravitacional base contiene variables f_{ij}^{kl} , ligadas a los costos percibidos por los usuarios de escoger las estaciones k y l para realizar un viaje entre las zonas i y j .

Dentro de los alcances de esta tesis, se explorará la alternativa de calcular f_{ij}^{kl} utilizando las distancias de acceso y egreso como una aproximación del costo de escoger los paraderos k y l .

Un procedimiento alternativo, propuesto para alguna investigación futura, consiste en relacionar las variables f_{ij}^{kl} con las probabilidades de escoger los paraderos k y l para viajar entre i y j , definidas como $Prob(kl/ij)$, puesto que cuando una alternativa de ruta es percibida como costosa, la probabilidad de elegir sus paraderos asociados debe ser baja para un individuo racional (y para rutas percibidas como de bajo costo, las probabilidades de escogerla deben ser altas).

Para ello, se vuelve necesario incluir un modelo de elección de paradero y ruta al proceso de calibración. El modelo propuesto se encuentra actualmente en desarrollo y corresponde a una formulación de tipo Logit, basada en Raveau et al. (2011), que incluye no solo

variables tradicionales como tiempos de viaje y tarifas, sino que también considera variables novedosas relacionadas a la topología de la ruta y al confort del viaje, entre otras.

Los primeros resultados de este modelo se presentan en Abud (2015) y corresponden a una formulación jerárquica, donde los usuarios escogen primero el paradero a utilizar para entrar en la red de transporte público, y luego la ruta que seguirán para llegar a su destino. En este sentido, el paradero de bajada queda implícito en la probabilidad de elección de cada ruta y puede ser rescatado directamente.

Esta construcción jerárquica permite aplicar el modelo a dos niveles; desde una matriz origen-destino zonal o una de paradero a paradero. En efecto, si se considera conocida una matriz T_{ij} (que indica los viajes entre las zonas i y j), la aplicación de un modelo de elección de paradero de inicio y de ruta permite asignar estas demandas y traducirlas a flujos en la red. Por otro lado, si tomamos como punto de partida una matriz T_{kl} (con los viajes entre los paraderos o estaciones k y l), es posible asignar estas demandas a la red mediante el nido inferior del modelo Logit, es decir, utilizando el modelo de elección de ruta desde cada uno de los paraderos que contienen viajes.

Si consideramos el primero de estos casos, el modelo de elección se traduce en que los usuarios deciden los paraderos y ruta que usarán para viajar entre dos zonas en particular. En consecuencia, podemos rescatar la probabilidad de elección $Prob(kl/ij)$ implícita, para ser incluida en el cálculo de f_{ij}^{kl} , usando la Ecuación (4.1):

$$Prob(kl/ij) = \sum_{\forall r: (k=i(r) \& l=f(r)) \in R_{ij}} Prob(r/ij) \quad (4.1)$$

donde $Prob(r/ij)$ representa la probabilidad de que un pasajero que viaja de la zona i a la zona j use la ruta r (del conjunto de rutas R_{ij}); mientras que $i(r)$ y $l(r)$ denotan los paraderos iniciales y finales de la ruta r , respectivamente.

Actualmente, el modelo de elección de ruta se encuentra en una etapa de calibración, en que se están recalculando y validando algunos de sus parámetros usando la herramienta del proyecto FONDEF, por lo que no está listo para ser incluido en las predicciones de inferencia zonal.

Es importante destacar que existe una dificultad adicional para calibrar los parámetros de la función de costos utilizando el modelo de elección de paradero y ruta: la presencia de variables endógenas en este último. Como ya fue señalado, el modelo incluye variables relacionadas con la comodidad de cada ruta, que a su vez están ligadas a la cantidad de personas que la utilizan (de existir congestión de transporte público, una ruta se vuelve menos atractiva por existir una menor probabilidad de viajar sentado, por ejemplo).

Lo anterior se traduce en que las probabilidades $Prob(kl/ij)$ dependan de T_{ij} ; en otras palabras, uno de los parámetros de entrada depende de los resultados finales del modelo de inferencia zonal, generando un problema de endogeneidad entre ambos. Para solucionar esta dificultad, se propone calcular las probabilidades $Prob(kl/ij)$ necesarias para calibrar (*i.e.*, las asociadas a paraderos y zonas observadas en la encuesta) en un paso previo a la estimación del modelo de esta tesis, basándose en los flujos resultantes de la asignación de la matriz de viajes entre paraderos obtenida siguiendo a Munizaga y Palma (2012), Devillaine (2012) y Munizaga et al. (2014), y no en una asignación de la matriz zonal T_{ij} .

4.3 Información de uso de suelo

Los potenciales de generación y atracción de cada zona (A_i y B_j , respectivamente) se encuentran ligados a la información de uso de suelo de cada una de ellas; esto permite representar, en el modelo, que una zona con más hogares genere más viajes o una zona con más trabajos atraiga más viajes en el horario punta mañana, por ejemplo.

La información disponible para este estudio corresponde a registros del Servicio de Impuestos Internos (SII, 2014), sobre cada manzana de la ciudad de Santiago; esta incluye

el número de unidades, la superficie total construida y el total de avalúo fiscal en pesos a la fecha (\$), para cada destino no agrícola según los códigos definidos en la Tabla 4-1.

Tabla 4-1: Destinos no agrícolas de bienes raíces

Código	Destino
C	Comercio
D	Deporte y recreación
E	Educación y cultura
G	Hotel, motel
H	Habitacional
I	Industria
K	Bienes comunes
L	Almacenaje y bodega
M	Minería
O	Oficina
P	Administración pública y defensa
Q	Culto
S	Salud
T	Transporte y telecomunicaciones
V	Otros destinos no considerados
W	Sitio eriazo
Z	Estacionamiento

Fuente: Servicio de Impuestos Internos (2014)

Esta información es recolectada y actualizada periódicamente por el Servicio de Impuestos Internos con propósitos regulatorios y taxativos, por lo que podría estar disponible para los planificadores de transporte prácticamente sin costo alguno.

Es relevante notar que algunas variables tradicionales en los modelos de transporte no están directamente disponibles en esta información, tales como población, ingreso o disponibilidad de vehículos particulares. Sin embargo, es posible argumentar que estas variables se encuentran ligadas a información que sí se dispone; por ejemplo, la población está relacionada al número de unidades habitacionales, mientras que el ingreso y la disponibilidad de automóviles se asocian al avalúo fiscal de las viviendas.

Si bien estas variables tradicionales podrían obtenerse de información censal recolectada por el Instituto Nacional de Estadísticas (INE), el Censo más actualizado disponible corresponde al del año 2002 (INE, 2002), puesto que tras la problemática asociada al Censo 2012, sus resultados no han sido completamente validados y su acceso *web* fue deshabilitado por el INE. Además, los censos son realizados cada 10 años en Chile, lo que implicaría que el modelo no se podría aplicar con datos actualizados por varios años (entre dos censos consecutivos). En consecuencia, se prefirió utilizar solo la información del SII, por un criterio de consistencia y aplicabilidad futura.

4.4 Zonificación

Como se especificó anteriormente, el modelo de inferencia zonal propuesto asigna viajes observados a las zonas cercanas, siguiendo las probabilidades resultantes de los cálculos Logit; en esta línea, un elemento clave a definir es la zonificación de ciudad a utilizar. Debido a que el modelo desarrollado por esta tesis tiene como finalidad ser aplicado en la herramienta computacional resultante de este FONDEF, la decisión de la zonificación seleccionada fue realizada en conjunto con el equipo de proyecto (así como muchas otras decisiones de modelación).

En primer lugar, como A_i y B_j están calculados en base a información de uso de suelo para cada zona, es fundamental que la zonificación escogida sea consistente con la

información disponible de este tipo, es decir, con la base de datos del SII. Sin embargo, esta consideración prácticamente no genera restricciones sobre la zonificación a escoger, puesto que se cuenta con información de uso de suelo a nivel de manzana; por lo que solo se exige que cada zona corresponda a la agregación de algún número de manzanas completas.

Adicionalmente, es esencial que la zonificación sea lo suficientemente detallada como para que sus zonas sean caminables y no excesivamente grandes, permitiendo la aplicación de un modelo de elección de rutas desde su centroide sin generar errores de agregación mayores.

Considerando estos factores, se decidió tomar como punto de partida la zonificación propuesta por la Encuesta Origen-Destino (EOD) 2006, que se basa en encuestas realizadas en el hogar e información adicional recolectada principalmente desde 2001 a 2006 (SECTRA, 2008). Esta zonificación divide a la región en 779 zonas y cada una de ellas consiste en una agregación de manzanas, lo que permite obtener fácilmente las variables de uso de suelo a nivel zonal.

Sin embargo, numerosos lugares de la ciudad han crecido y cambiado bastante desde la realización de la EOD 2006, por lo que los bordes zonales tuvieron que ser actualizados. Además, se tomó la decisión de dividir zonas, puesto que muchas de ellas resultaban demasiado grandes para aplicar el modelo de elección de rutas desde ellas correctamente.

La tarea de actualizar y densificar la zonificación fue abordada por esta tesis, incluyendo la gestión del proceso, la definición de criterios estandarizados de revisión de zonas, la identificación de las zonas más problemáticas (que requerían mayor atención) y el diseño de parte de los cortes zonales.

Respecto a las limitaciones de este proceso de zonificación, y debido a consideraciones computacionales de la herramienta resultante del Proyecto FONDEF, se impuso un número máximo de 1.200 zonas finales. Esta restricción, si bien limita el nivel de detalle que se puede alcanzar con las distintas zonas, se justifica para evitar tiempos

excesivamente altos de ejecución de procesos y algoritmos que debe realizar la herramienta.

En cuanto a los criterios de revisión, se definieron las características de una zona ideal, las que trataron de ser aplicadas a la zonificación final, pero siempre teniendo en cuenta el número máximo de zonas. Estos criterios se detallan de manera resumida a continuación, donde cada zona idealmente debe cumplir:

- Sus bordes deben corresponder preferentemente a calles.
- Ser lo más homogénea posible (respecto a su uso de suelo).
- Mantener una forma compacta y simétrica.
- Excluir terrenos vacíos.
- De ser más atractiva (es decir, tener más viviendas o atractores de viajes), tener menor tamaño (mayor nivel de detalle).
- Sus viajeros asociados deben tener comportamientos de transporte similares.

Este último criterio es particularmente importante para los efectos del proyecto, por lo que amerita ser explicado con mayor profundidad. Recordemos, en primer lugar, que el propósito de este FONDEF es construir una herramienta que prediga el comportamiento de los usuarios de transporte público de Santiago. Para modelar las decisiones de los ciudadanos, la ciudad es dividida en zonas que poseen a su vez centroides, los que se supone generan y atraen todos los viajes de la zona; de esta manera, todos los viajes que salen o entran a una zona lo hacen a través de los paraderos o estaciones conectados (por arcos de acceso) a su centroide. Así, un pasajero modelado que quiere viajar entre dos zonas, observa las posibles rutas que usan los paraderos asociados a estas zonas y escoge entre ellas según el modelo de elección de ruta y paradero descrito en el subcapítulo 4.2.

En vista de lo anterior, es necesario que cada una de las zonas resultantes posea alternativas de paraderos y rutas que permitan representar el comportamiento real de las personas que en ella habitan. En este sentido, por ejemplo, se dividieron zonas donde los usuarios podrían elegir entre tres ejes paralelos para realizar su viaje, puesto que el

comportamiento normal de cada persona es que considere como máximo solo los dos más cercanos, lo que fue validado con la encuesta (ver subcapítulo 4.5). Similarmente, zonas muy largas fueron divididas, con el objetivo de cargar sus paraderos asociados correctamente y poder representar los viajes en servicios alimentadores para llegar de un extremo a otro de ellas.

Por último, cabe destacar que para poder llevar a cabo estas divisiones y actualizaciones, el trabajo fue apoyado por la estudiante de Geografía de la Pontificia Universidad Católica Melissa Miranda; contratada por el Proyecto FONDEF para esta tarea dados sus conocimientos de *softwares* de manejo y edición de sistemas de información geográfica.

De esta manera, la zonificación propuesta llegó a tener 1.176 zonas, tras la creación de nuevas zonas y la remoción de zonas no urbanas (las zonas no servidas por Transantiago fueron eliminadas para efectos de este proyecto, puesto que el modelo propuesto busca asignar viajes observados en paraderos a zonas cercanas) y vacías o de áreas verdes (las que se supuso atraen una cantidad despreciable de viajes en el horario punta mañana, lo que fue confirmado con la encuesta). Para más información sobre la zonificación u observar gráficamente sus resultados, consultar el Anexo C.

4.5 Encuesta

Las observaciones a utilizar en la estimación por máxima verosimilitud provienen de la encuesta de preferencias reveladas realizada en paraderos y estaciones de metro durante noviembre del año 2013, que fue elaborada principalmente con el objetivo de calibrar el modelo de elección de paradero y ruta presentado en Abud (2015). En vista de que se decidió aprovechar esta recopilación de datos para calibrar también el modelo de inferencia zonal, como parte del trabajo de esta tesis se apoyó la realización de la encuesta, incluyendo desde la fase de mediciones piloto, toma de datos definitivos y el posterior procesamiento computacional de la información.

El cuestionario definitivo se diseñó tomando en cuenta los resultados de la fase cualitativa y las mediciones piloto, y se presenta en el Anexo B-a (para más detalles sobre las fases

previas consultar Abud, 2015). Como es posible observar, las preguntas se centran en reconstruir con el mayor nivel de detalle posible el viaje que realizó el encuestado, incluyendo su origen y destino, así como también los servicios y paraderos que utilizó. Además, se incluyen preguntas de contexto y propósito del viaje, de tipo socioeconómico y sobre las alternativas que el usuario conoce y considera.

Debido a que se buscaba representar la elección de viajes durante la mañana, las encuestas se realizaron entre 7:00 y 12:00 am. Este criterio, si bien incluye más horas que el período punta mañana propiamente tal, se definió por razones prácticas relacionadas con el número de encuestas necesario para obtener una adecuada significancia de los parámetros a estimar.

Los lugares fueron seleccionados procurando cubrir sectores diversos de la capital, donde se generara una gran cantidad de viajes y en los que hubiera varias alternativas de rutas. Además, se seleccionaron tres grandes centros de transbordo de la ciudad (las estaciones de Metro de Vespucio Norte, La Cisterna y Escuela Militar), para interceptar en ellas usuarios con orígenes más diversos. La Figura 4-1 muestra un mapa con las áreas cubiertas por los distintos días de encuestas (para una visualización con mayor nivel de detalle, consultar el Anexo B-b).

Como resultado, se obtuvo 2.207 encuestas previas a su digitación y procesamiento; sin embargo, no todas ellas pudieron ser utilizadas para la calibración del modelo de elección de paradero y ruta de la herramienta FONDEF (por su parte, el modelo de inferencia zonal agregó restricciones adicionales para las observaciones a incluir), por encontrarse incompletas o poco claras en algunas preguntas clave (para más información referente a las encuestas utilizadas por cada estudio, consultar Abud, 2015 y/o el capítulo 6 de este documento).

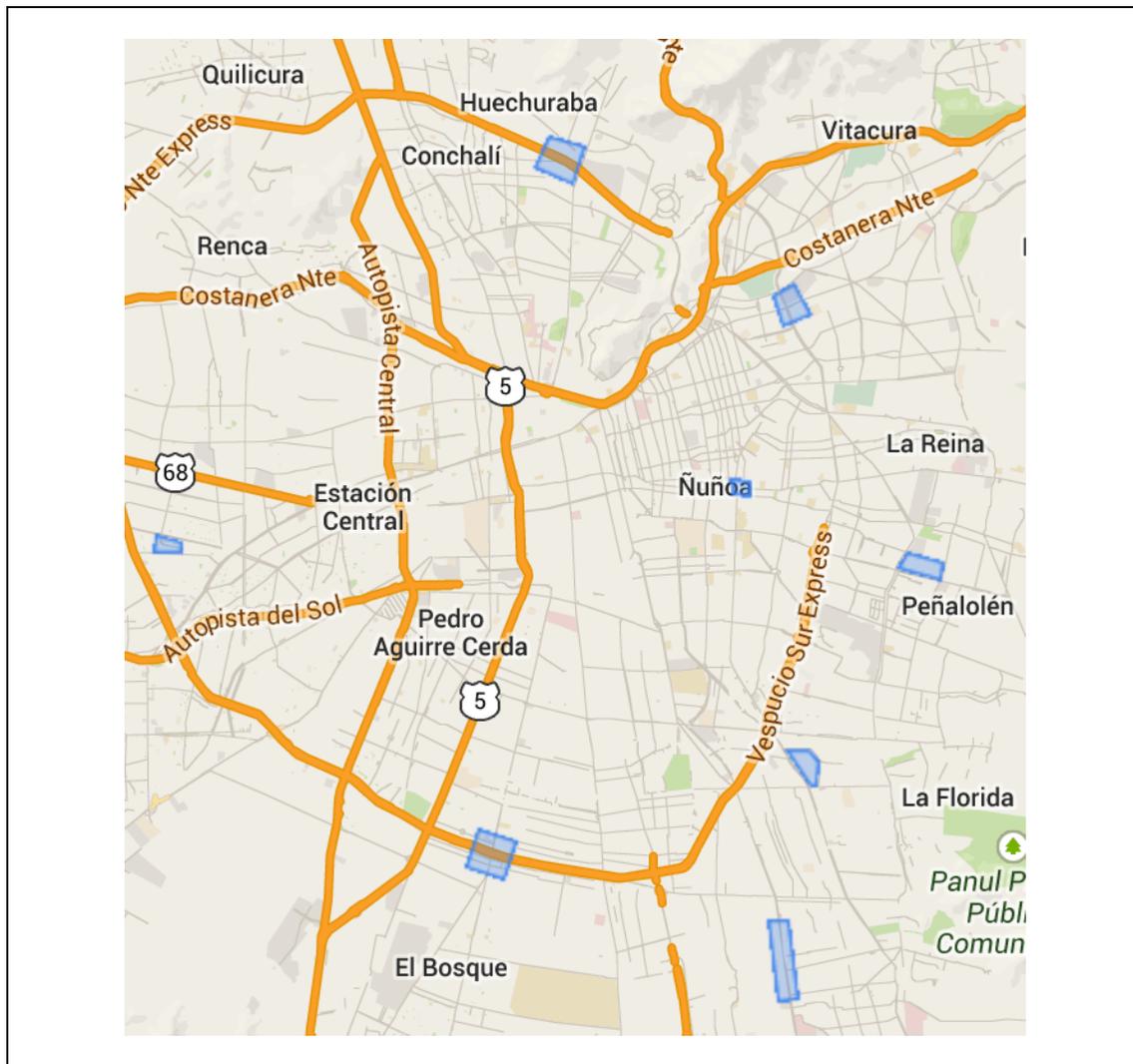


Figura 4-1: Mapa de Santiago destacando las zonas encuestadas

Fuente: Abud (2015)

A continuación, las direcciones y rutas declaradas en las encuestas válidas fueron georeferenciadas. Para ello, se identificaron las estaciones de metro y paraderos de bus usados, y las coordenadas asociadas a los orígenes y destinos reales de los viajes, reportados típicamente en términos de intersecciones y direcciones (Figuras 4-2 y 4-3).

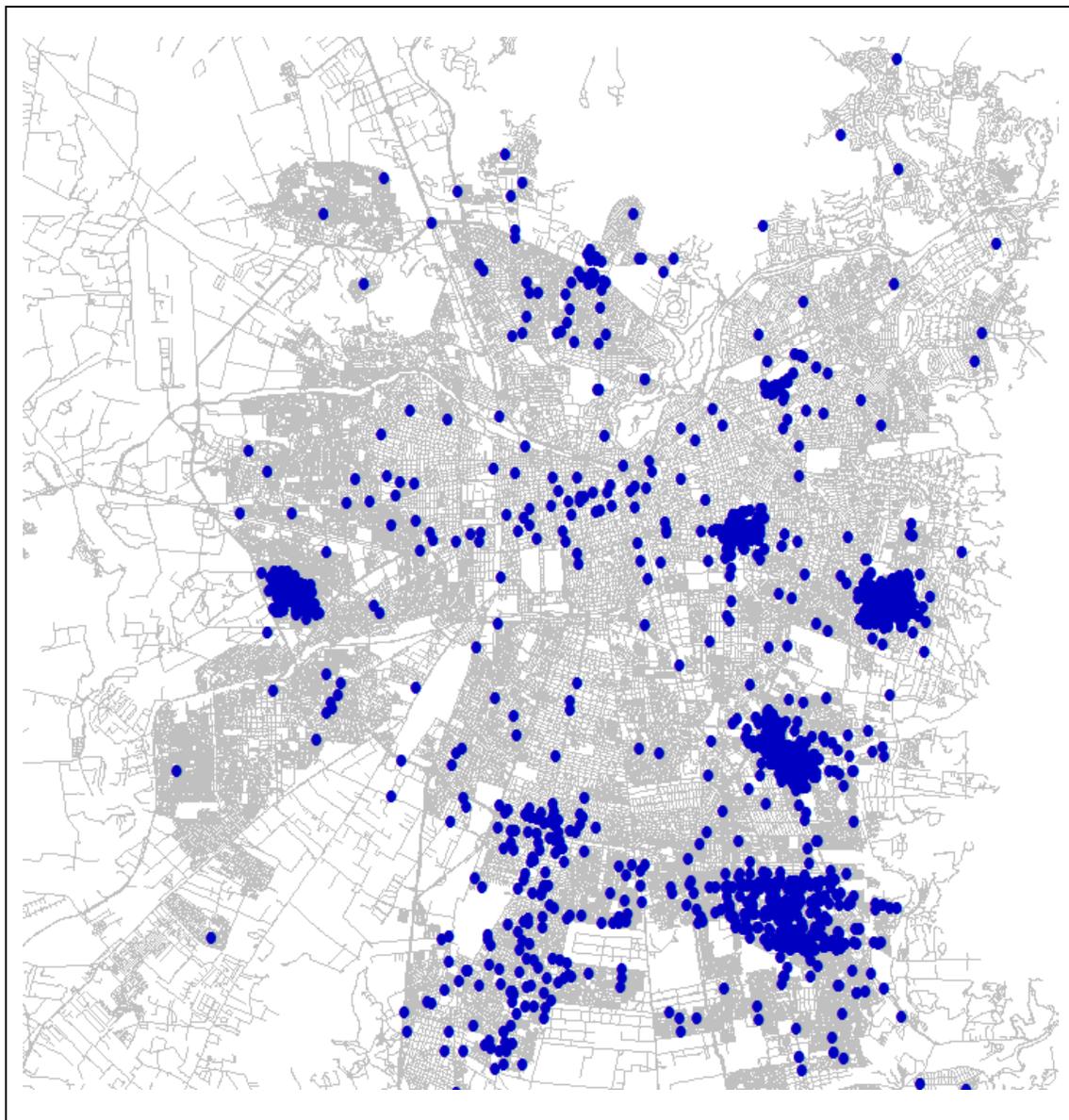


Figura 4-2: Mapa de Santiago con orígenes de encuestas reportados

Fuente: Elaboración propia en base a *TRANSCAD*

En concordancia con los objetivos de la encuesta, de las Figuras 4-2 y 4-3 se puede observar una amplia cobertura de los distintos polos de atracción y generación de viajes en Santiago. Sin embargo, como es lógico dado que la encuesta se trató de realizar cerca del origen del entrevistado, existe una menor variedad espacial en orígenes que en destinos de viajes.

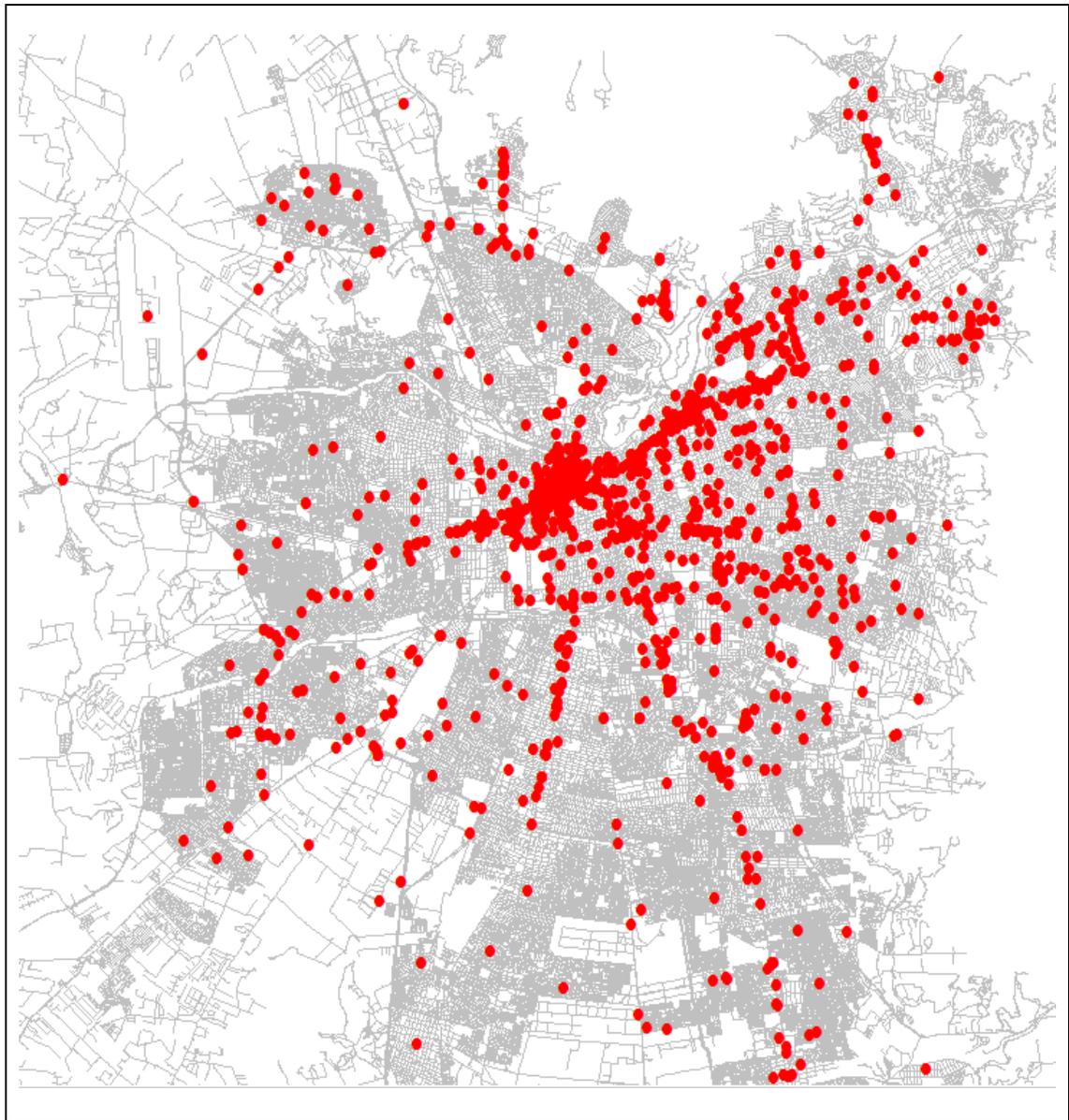


Figura 4-3: Mapa de Santiago con destinos de encuestas reportados

Fuente: Elaboración propia en base a *TRANSCAD*

5 PROCESAMIENTO Y CONSTRUCCIÓN DE LA BASE DE DATOS

Como es posible concluir del capítulo anterior, el modelo de esta tesis requiere una gran cantidad de datos de entrada provenientes de diversas fuentes. Por esta razón, fue necesario un procesamiento adicional para consolidarlos en un banco de datos común, acorde a las necesidades del proyecto. Adicionalmente, se depuró la información de uso de suelo pues algunas de sus manzanas presentaban inconsistencias menores. Las tareas previamente descritas se detallan a continuación.

5.1 Asignación de manzanas a zonas

Como se explicó anteriormente, las zonas propuestas corresponden a agregaciones de las más de 50.000 manzanas definidas según el Servicio de Impuestos Internos (SII). Por este motivo, se construyó un diccionario que relaciona cada código de manzana con la zona que lo contiene.

Esta relación, en la mayoría de los casos, es unívoca y directa; no obstante, existía un porcentaje (menor al 1%) de manzanas que estaban repetidas en la base de datos geográfica y pertenecían a más de una zona a la vez. Cada una de estas manzanas fue analizada y transformada según fuese necesario en dos o más pseudo-manzanas que permitiesen repartir el uso de suelo entre las distintas zonas a las que pertenecía. Así, cada manzana (y pseudo-manzana) del diccionario final tiene solo una zona asociada.

Por otro lado, hubo casos en que una zona parecía inicialmente vacía pues no tenía manzanas asociadas (y por lo tanto no atraería ni generaría viajes); no obstante, al observar las imágenes satelitales se constataba que esto era un error pues la zona tenía una gran parte construida. Esto se solucionó generando manzanas ficticias con una información de uso de suelo estimada observando zonas cercanas de características similares.

5.2 Asignación de zonas de origen y destino a encuestas

Dado que el modelo de inferencia zonal funciona en términos agregados, fue necesario realizar un procesamiento adicional a las encuestas. En este, cada coordenada de origen y

destino fue asignada a su zona correspondiente, simplemente encontrando la única zona que la contenía.

Sin embargo, es importante notar que parte de los encuestados fueron reticentes a entregar direcciones específicas, reportando intersecciones de calles como referencias de orígenes y/o destinos. Considerando además que la mayoría de los bordes zonales están definidos por calles, no fue posible asignar todas las coordenadas unívocamente a zonas.

La Figura 5-1 muestra un ejemplo de la problemática anterior, donde los puntos numerados corresponden a coordenadas de origen reportadas y las líneas gruesas delimitan zonas.



Figura 5-1: Criterio de asignación de zonas a encuestas

Fuente: Elaboración propia en base a *TRANSCAD*

En este caso, podemos ver un usuario que describió su origen como la intersección de calles asociada a la coordenada 1525; como esta se encuentra en la intersección de tres zonas, sin información adicional, no es posible asignarle una zona de origen única. Por otro lado, las cuatro coordenadas restantes (562, 641, 756 y 854) no presentan este problema puesto que están claramente dentro de una zona.

Recordando la metodología de estimación propuesta en el subcapítulo 3.2, cada observación debe tener zonas y paraderos de origen y destino únicos, por lo que se descartaron las observaciones incompletas. Este problema afectó distintos propósitos de viaje de manera diferente; por ejemplo, los estudiantes y quienes se dirigían a centros de salud tendieron a reportar direcciones más específicas y por lo tanto, más fáciles de asignar a zonas unívocas.

Finalmente, usando la notación definida en el capítulo 3, cada encuesta fue traducida a un viaje observado entre las zonas i y j , usando las estaciones de acceso k y l , como sus puntos de inicio y término del viaje, respectivamente.

5.3 Creación de centroides y arcos de acceso

Una etapa fundamental en la realización del proyecto FONDEF fue la definición de centroides para las zonas, con sus respectivos arcos de acceso a los paraderos y estaciones cercanos. Esta tarea fue abordada en esta tesis y se detalla a continuación.

Los centroides fueron dibujados en el centro geométrico de cada zona para simplificar la visualización de los mapas. No obstante, el centroide es una construcción teórica y su posición no es necesariamente relevante. Por el contrario, lo realmente importante es la definición de sus arcos de acceso (y egreso) y el cálculo de sus distancias asociadas.

En esta línea, la primera tarea consistió en definir qué paraderos de buses y estaciones de metro estarían conectadas por arcos de acceso con cada una de las zonas de la ciudad. Para decidirlo se observó en primera instancia, que muchos pasajeros encuestados usaban estaciones o paraderos que se encontraban fuera de sus zonas de origen o destino (cerca

de un 75% de los encuestados que usaron metro accedieron desde una zona exterior, mientras que el porcentaje análogo para bus corresponde al 45%); esto es bastante factible considerando que las zonas son relativamente pequeñas (muchas de ellas miden cerca de un kilómetro de largo). De esta forma, se determinó que los arcos de acceso debían construirse no solo a paraderos que se encontraran dentro de una zona, sino también a aquellos que se encontraran dentro de una extensión imaginaria de sus bordes, definida como radio de influencia (r). La Figura 5-2 muestra que un centroide (representado por el triángulo) se encontrará conectado con todos los paraderos que se encuentren dentro de su zona (línea continua) y también con todos los paraderos que se encuentren fuera de la zona pero a menos de r metros de los límites zonales.

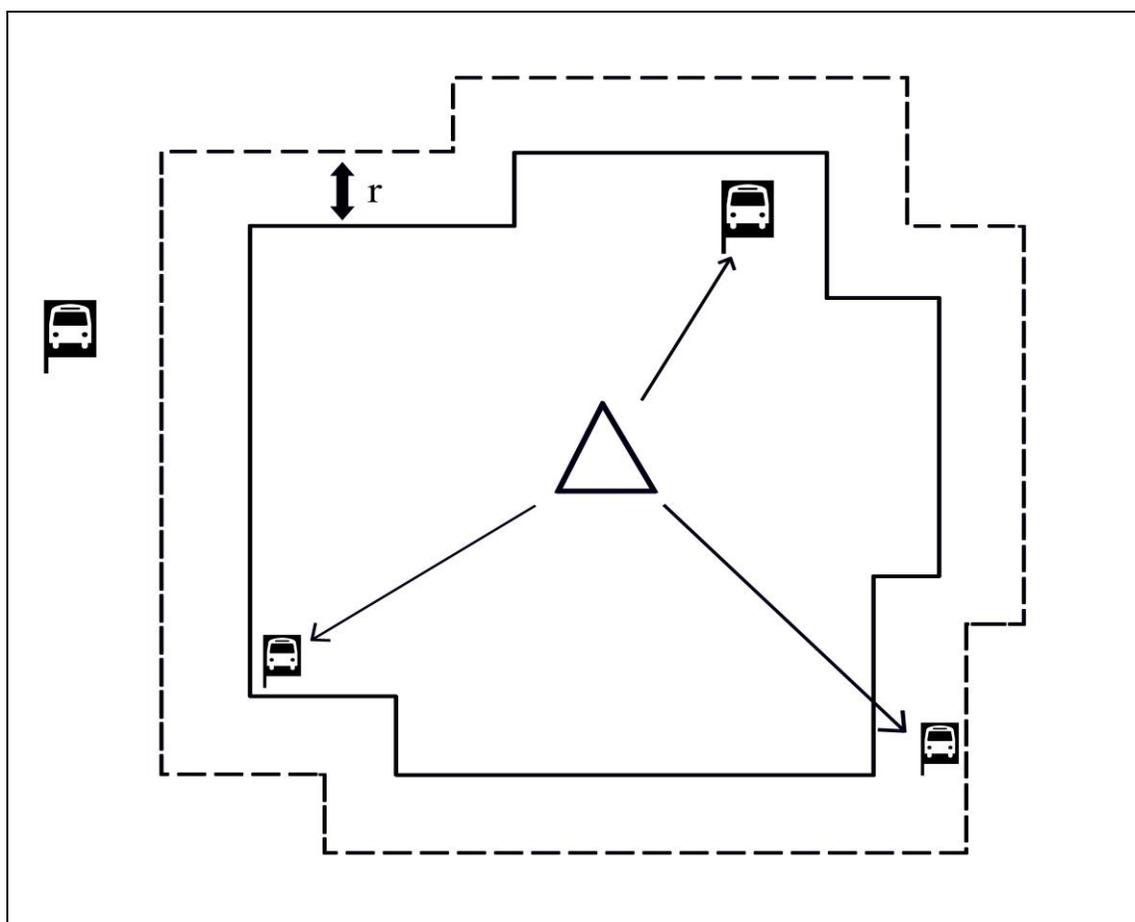


Figura 5-2: Criterio de asignación de paraderos a zonas

Fuente: Elaboración propia

Luego de descartar *outliers*, se determinó cuánto estaba dispuesto a caminar un usuario fuera de su zona para acceder a un paradero de bus o estación de metro. Estos dos modos fueron explícitamente separados puesto que los santiaguinos tienen mayor disposición a caminar cuando se dirigen a una estación de metro, lo que fue confirmado con la encuesta.

En esta línea, tomando los datos de la encuesta en paraderos, se calcularon distancias euclidianas entre cada paradero (o estación) elegido por el usuario a los bordes de la zona de origen o destino correspondiente (a quienes escogieron paraderos o estaciones dentro de la zona de origen o destino se les asignó una distancia nula para este propósito) y se evaluaron distintos valores para los radios de influencia de cada modo. Las Figuras 5-3 y 5-4 muestran el porcentaje de observaciones correctamente representadas por la red (*i.e.*, donde el encuestado usa un paradero conectado con la zona de donde proviene) y el total de arcos generados para distintos valores de radios de influencias, para metro y bus.

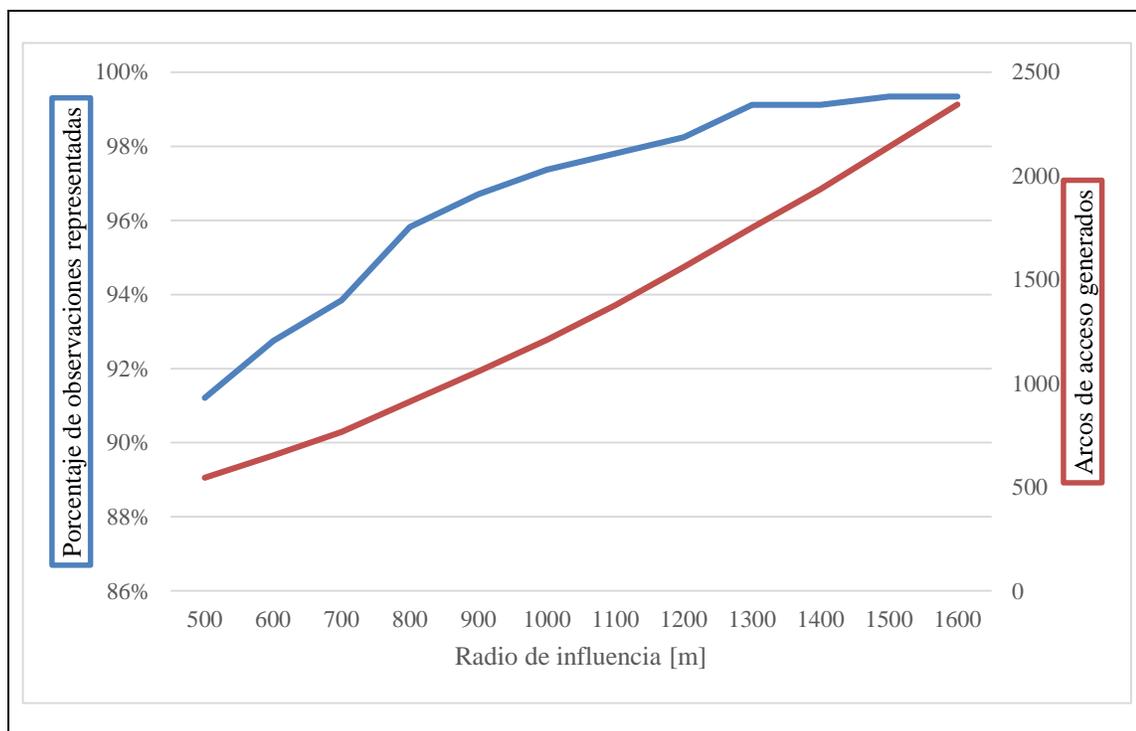


Figura 5-3: Porcentaje de observaciones correctamente representadas por la red y total de arcos de acceso, para distintos valores de radios de influencia de metro

Fuente: Elaboración propia

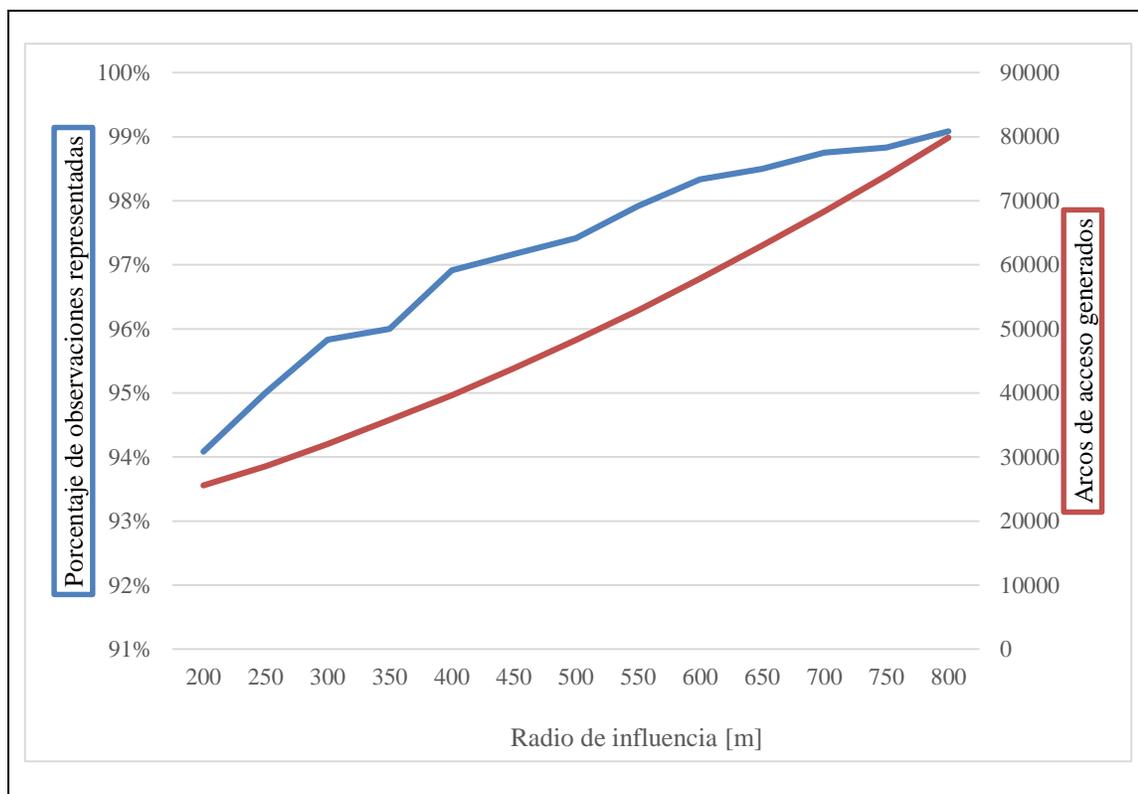


Figura 5-4: Porcentaje de observaciones correctamente representadas por la red y total de arcos de acceso, para distintos valores de radios de influencia de bus

Fuente: Elaboración propia

Tras el análisis de los resultados anteriores se determinó el radio de influencia como el 95% superior de la muestra, resultando una medida de 250 metros para paraderos de bus y 750 metros para estaciones de metro. Esta definición acotada al 95%, permite que la red construida represente las decisiones observadas de la gran mayoría de la muestra y no genere una cantidad excesiva de arcos de acceso que sobrecargue la herramienta computacional. Cabe destacar que este análisis se realizó sobre una muestra de 1.655 caminatas, correspondientes a todas ellas en que se identificó un paradero y zona única de origen o destino.

Finalmente, se añadió a la red los arcos de acceso entre centroides y paraderos que se encontraran a menor distancia que la especificada. Adicionalmente, estas conexiones fueron verificadas manualmente, para asegurarse que todas ellas representaban alternativas racionales de viaje (añadiendo y eliminando arcos según fuese necesario). Recordando el modelo de inferencia zonal, los conjuntos m y n según lo expresado por la Ecuación (3.6) se definen como todas las zonas conectadas por arcos de acceso (o egreso) a los paraderos k y l , respectivamente.

Una vez definido qué zonas estarían conectadas con qué paraderos (tanto para la herramienta computacional como para la calibración del modelo de inferencia zonal), se procedió a asignar un valor de distancia y tiempo de acceso a cada uno de estos arcos.

En este estudio, las distancias de acceso a los paraderos y estaciones se calcularon considerando variables de uso de suelo correspondientes a cada manzana de la zona. Esto se realizó con el propósito de obtener valores precisos para las distancias de acceso, puesto que las zonas no son necesariamente homogéneas y, por ejemplo, la población puede concentrarse en áreas particulares de ellas.

De esta manera, a cada manzana M se le asignó un punto en su centro geométrico con dos diferentes pesos ligados a sus características de uso de suelo; uno para su potencial como generador de viajes (w_M^g , ligado a su uso residencial, *i.e.*, número de unidades habitacionales), y otro para su potencial atractor (w_M^a , ligado al total de metros cuadrados construidos de tipo comercial, industrial, residencial, salud, educacional y de oficinas). La selección de qué variables incluir en cada peso se tomó considerando no solo el criterio de los modeladores, sino también resultados de modelos de inferencia zonal preliminares que no fueron incluidos en este documento debido a su simpleza y carácter exploratorio.

Luego, se midió la distancia entre los centros geométricos de cada manzana y cada paradero de transporte público (ya sea un paradero de bus o una estación de metro) conectado con la zona, como muestra la Figura 5-5.

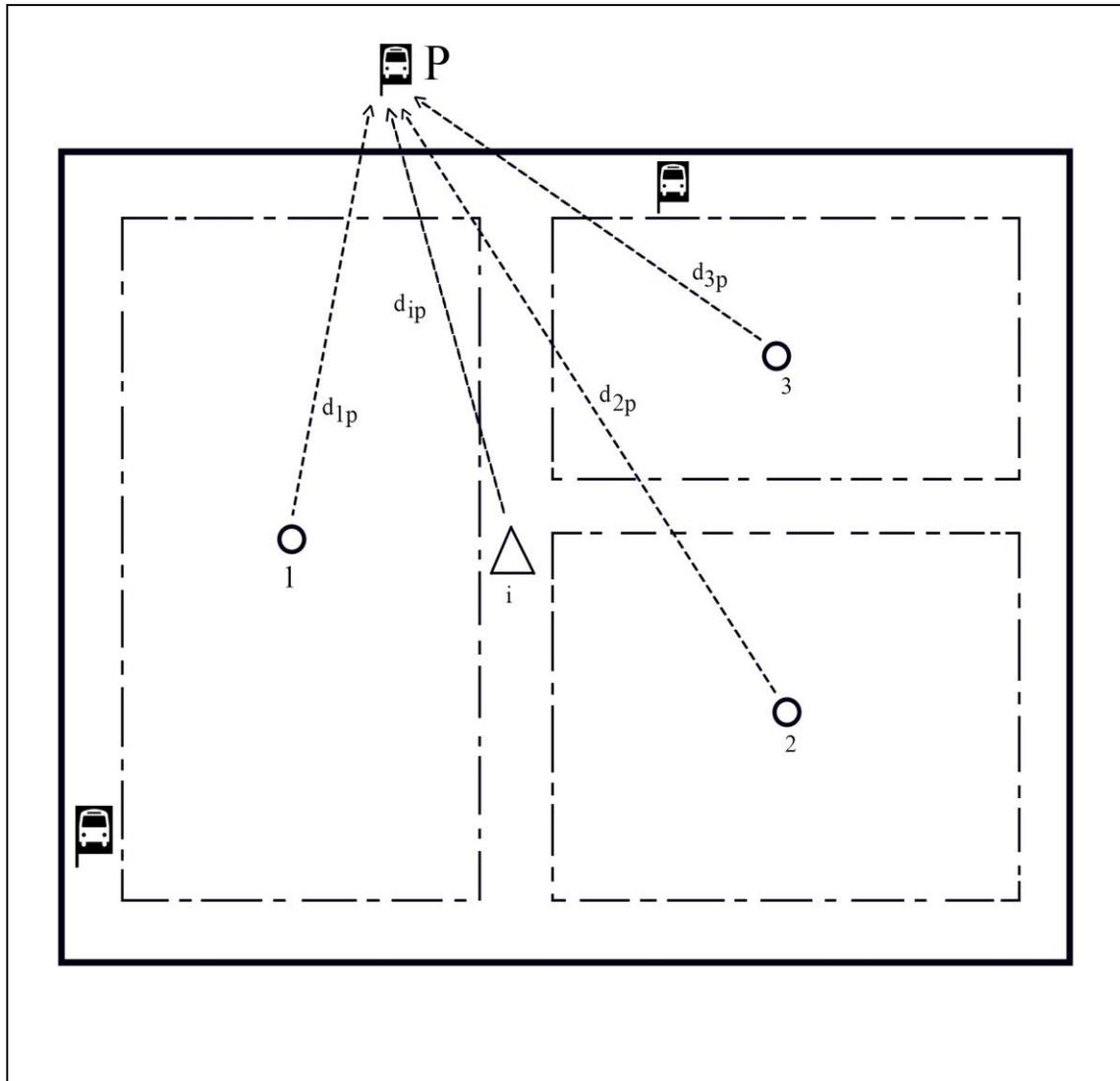


Figura 5-5: Criterio de cálculo de distancias de acceso

Fuente: Elaboración propia

En ella, se representa una zona i que posee tres manzanas, ilustrando el cálculo de la distancia de acceso hacia el paradero p .

$$d_{ip} = \frac{\sum_{M=1,2,3} d_{Mp} * w_M^g}{\sum_{M=1,2,3} w_M^g} \quad (5.1)$$

La Ecuación (5.1) muestra el promedio ponderado explicado, donde w_M^g representa el peso asociado al potencial generador de la manzana M . Análogamente, para el arco de atracción de viajes se usa el peso asociado al potencial atractor w_M^a . Así, se obtuvo dos distancias de conexión para cada paradero con su zona respectiva: una correspondiente al arco generador y otra al arco atractor de viajes de la zona.

Una vez calculadas estas distancias, y asumiendo una velocidad de caminata de 5 km/h (Bohannon, 1997), fue posible obtener los tiempos de acceso desde una zona a cada paradero de bus y estación de metro conectada con ella. Es importante notar que, utilizando la notación de las secciones previas, esta metodología permite que el tiempo (y distancia) de acceso desde la zona i al paradero k sea distinto al valor de egreso que conecta el mismo paradero k con la zona i ; esto posibilita una mejor representación de zonas poco homogéneas, donde se concentre la población en un sector y los atractores de viajes en otro.

6 GENERACIÓN Y DESCRIPCIÓN DEL BANCO DE ESTIMACIÓN

Una vez recopilada y procesada toda la información necesaria, esta fue agrupada y codificada para construir el banco de estimación definitivo. Este capítulo enumera los criterios con que se seleccionaron las entradas válidas de las encuestas, y posteriormente presenta y describe estadísticamente las variables de modelación.

6.1 Procesamiento de la información de las encuestas

Como ya fue adelantado, parte importante de las 2.207 encuestas tomadas en terreno no calificaron como datos válidos para la estimación de los modelos de elección de Abud (2015). En efecto, para un 15% de ellas no fue posible encontrar y geo-referenciar el origen y/o destino, debido a: errores en el proceso de digitación, direcciones reportadas poco precisas o por inconsistencias entre las rutas y direcciones reportadas (por ejemplo, usuarios que afirmaban tomar un servicio en cierto lugar por donde este no transita). Además, cerca de un 10% de los encuestados usaba un modo no integrado con Transantiago (taxi colectivo, bus interurbano o automóvil privado, principalmente) en alguna de las etapas de su viaje, por lo que su ruta no podía ser incluida en la herramienta FONDEF. Por último, aproximadamente un 13% de las encuestas tenía problemas con la ruta reportada (*e.g.*, una etapa faltaba o contenía contradicciones). De esta manera, para el modelo de elección de paradero y ruta se contó con 1.360 encuestas utilizables, lo que se consideró un tamaño razonable para ese proyecto (Abud, 2015).

Además de este primer filtro, el tratamiento agregado del modelo de inferencia zonal añadió restricciones adicionales a los datos, para que estos pudieran clasificarse como utilizables para los propósitos de esta investigación. Los subcapítulos 5.2 y 5.3 ilustran los principales problemas: (a) el problema de asignación de zonas de origen y destino a encuestas y (b) el problema de los arcos de acceso.

Respecto al primero de ellos, finalmente se encontró solo 825 orígenes y 749 destinos asignados a una zona unívoca, correspondientes a 628 pares origen-destino completos (es decir, encuestas para las que sabemos con certeza cuál fue su zona de origen y de destino).

Como es posible observar, esta consideración implicó grandes restricciones al tamaño muestral del estudio, significando una pérdida de cerca del 54% de las encuestas consideradas como válidas para la calibración de los modelos presentados en Abud (2015).

Por otro lado, como se explicó en el subcapítulo 5.3, se generaron los arcos de acceso asociados a la zonificación propuesta considerando ciertas limitaciones computacionales. De esta forma, existe un porcentaje de viajes que no podemos representar con la herramienta, pues utilizan paraderos que no están conectados con sus centroides respectivos (es decir, que caminan distancias de acceso (o egreso) mayores al radio de influencia r). Esta limitación se traduce en una pérdida de 59 registros adicionales (correspondientes a cerca de un 9% de las 628 encuestas restantes), por lo que el banco de estimación definitivo cuenta con 569 observaciones utilizables.

Finalmente, recordando la notación del capítulo 5, se codificó cada observación q para la estimación por máxima verosimilitud procurando que tuviera toda la información necesaria. Así, obtuvimos 569 observaciones con: un paradero (o estación) inicial k y final l ; las zonas de origen i y destino j (correspondientes a las variables dependientes); y los conjuntos de alternativas de origen m y destino n .

Además, cada una de estas zonas tiene asociada su información de uso de suelo correspondiente (a incluir en la estimación de A_m y B_n , recordar el subcapítulo 4.3), así como también un costo asociado a su arco de acceso desde el paradero en cuestión (relacionado con f_{ij}^{kl} , presentado en el subcapítulo 4.2).

6.2 Descripción estadística del banco de estimación

A continuación, se entrega información estadística sobre las variables disponibles en el banco de estimación, describiendo no solo las respuestas obtenidas de las encuestas sino también la red y zonificación usadas para la modelación.

6.2.1 Variables geográficas y de la red de transporte público

Una vez que se descartaron las observaciones no válidas, se obtuvieron nuevos mapas que muestran la repartición espacial de los viajes en la ciudad. Las Figuras 6-1 y 6-2 muestran los orígenes y destinos reportados en las encuestas del banco de estimación.

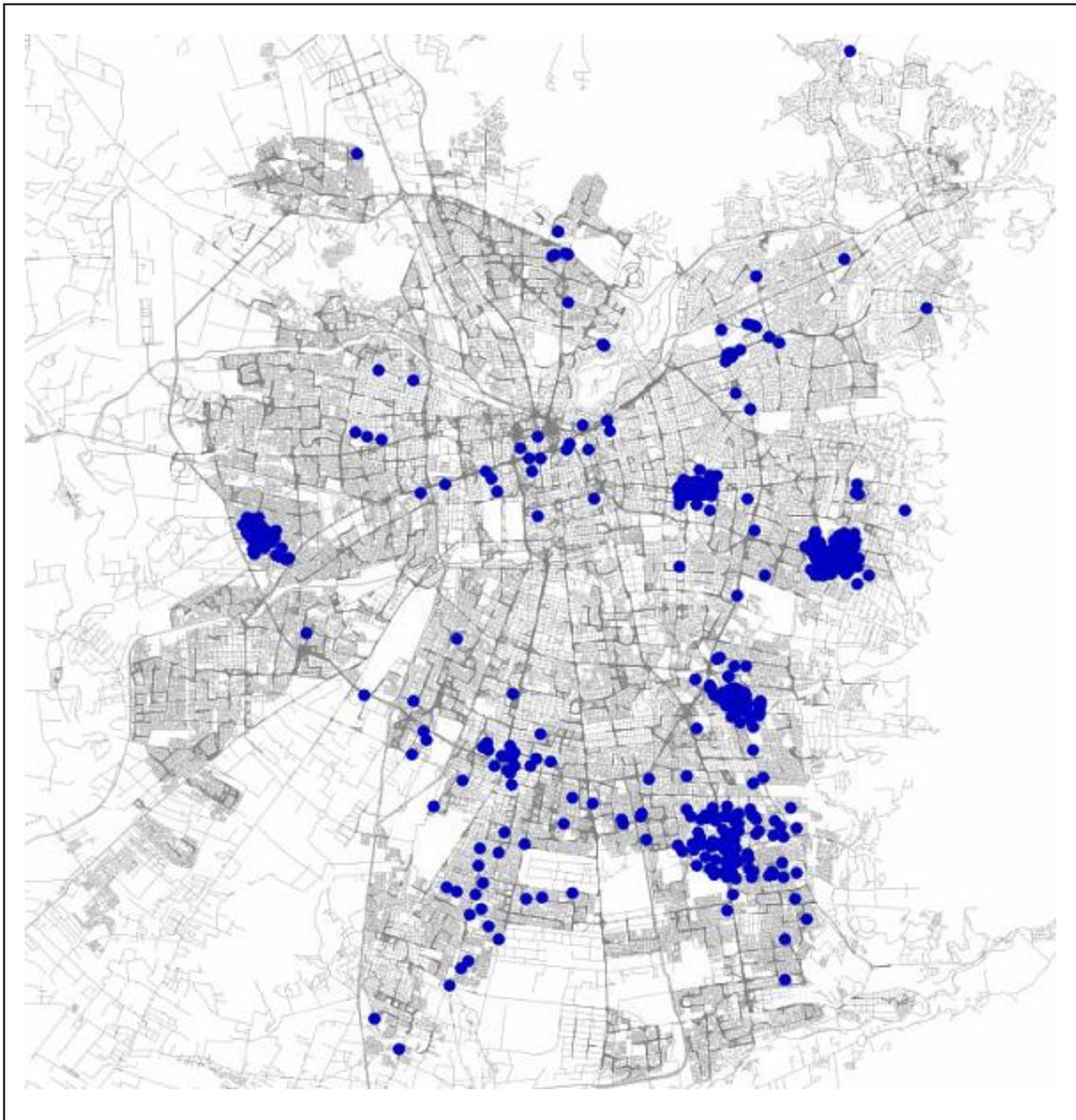


Figura 6-1: Mapa de Santiago con orígenes de encuestas válidas

Fuente: Elaboración propia en base a *TRANSCAD*

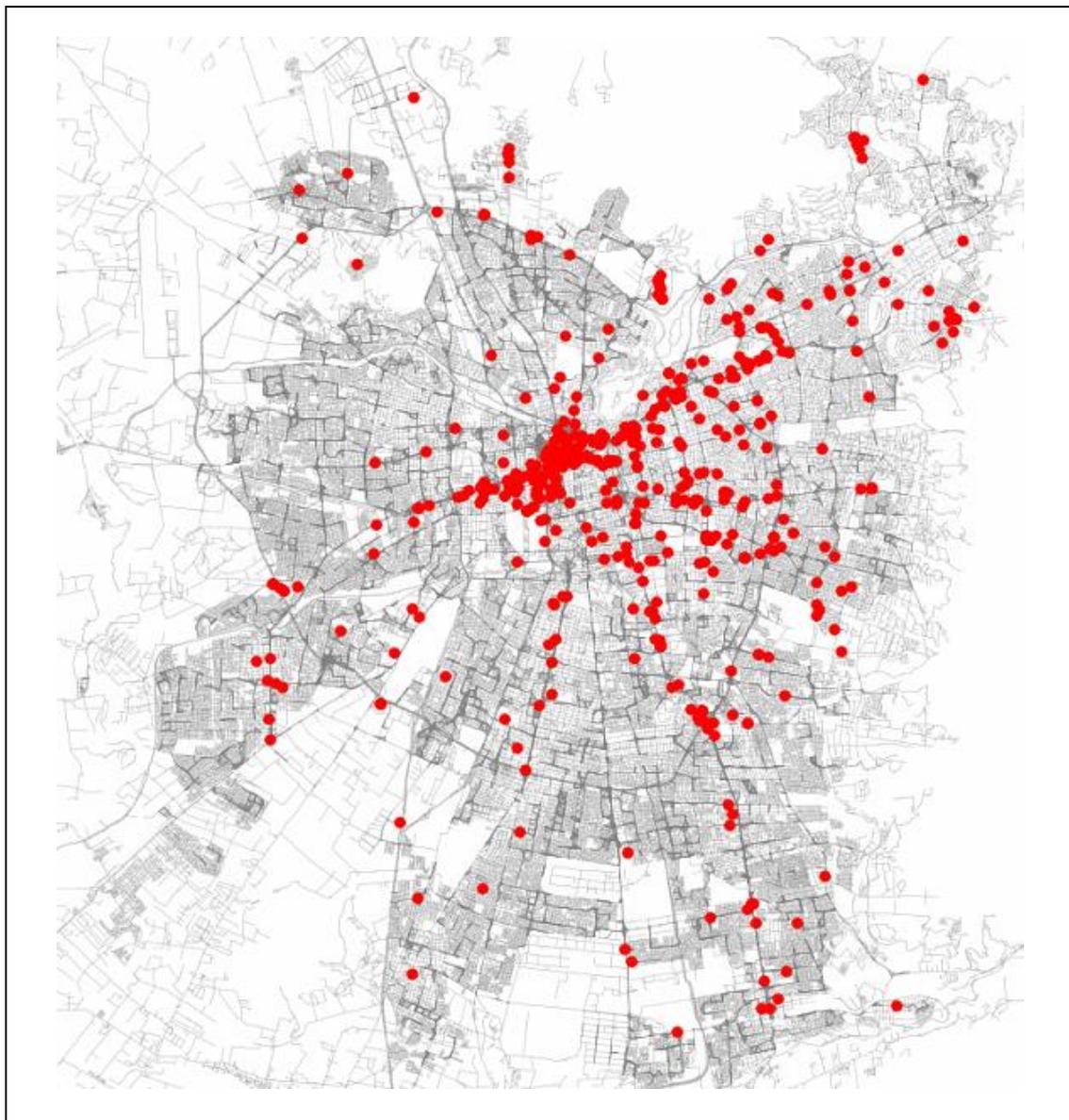


Figura 6-2: Mapa de Santiago con destinos de encuestas válidas

Fuente: Elaboración propia en base a *TRANSCAD*

De ellas, es posible observar que existe una distribución similar a la presentada en el subcapítulo 4.5. En otras palabras, las observaciones consideradas como válidas para este estudio, tienen una repartición espacial semejante al total de las encuestas; los orígenes se encuentran repartidos principalmente cerca de los lugares de realización de la encuesta,

mientras que los destinos se concentran cerca de las líneas de metro, además del sector centro y oriente de la capital.

Dado que estas coordenadas de orígenes y destinos fueron asociadas a sus respectivas zonas, es posible efectuar distintos tipos de análisis relacionados con la red de transporte público y con los arcos de acceso.

En primer lugar, si analizamos los paraderos de inicio y término del viaje, la muestra posee un 54,5% de viajes que empiezan y terminan en bus, 14,1% de viajes que empiezan y terminan en metro y 31,4% de viajes con inicio y fin en distintos modos (de los cuales un 28,5% empieza en una estación de metro y termina en un paradero, mientras que el 71,5% restante parte en un paradero de bus y termina en una estación).

Es importante notar que la matriz de viajes entre paraderos de abril 2013 (correspondiente a la usada en la herramienta FONDEF y que fue calculada con la metodología de Munizaga y Palma, 2012; Devillaine, 2012; y Munizaga et al., 2014) tiene una repartición bastante diferente; en efecto, un 30,5% de sus viajes empieza y termina en metro, un 41% en bus y un 28,5% en modos diferentes (con un 18,9% empezando en metro y terminando en bus, mientras que el 81,1% restante parte en un paradero y baja en una estación).

En este sentido, el banco de estimación tiene el sesgo de que incluir una mayor proporción de viajes que empiezan y terminan en paraderos de buses que la real, así como también una menor proporción de viajes en metro (explicable porque la encuesta se tomó en pocas estaciones de Metro). Por su parte, los viajes con inicio y final en distintos modos tienen una representación similar en ambas muestras, pero con diferencias parciales entre sus orígenes y destinos.

Respecto a la interacción entre la red ficticia (*i.e.*, arcos de acceso y centroides) y las observaciones provenientes de las encuestas, en la Figura 6-3 se presenta el número de zonas asociadas a los paraderos de inicio y fin del viaje.

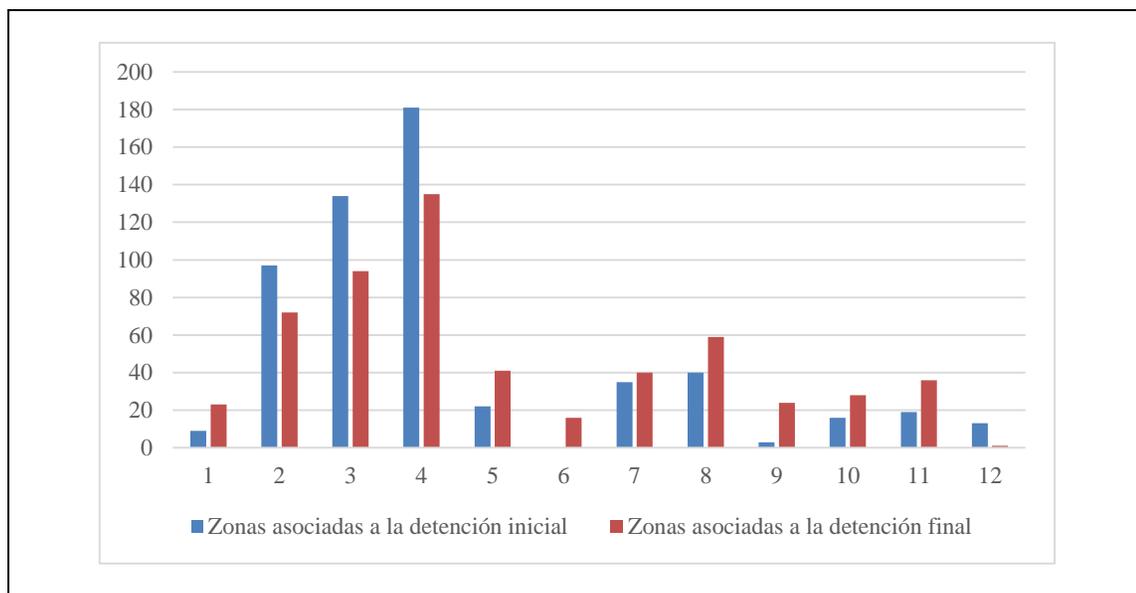


Figura 6-3: Número de observaciones por cantidad de zonas asociadas

Fuente: Elaboración propia

Es posible observar que la mayoría de los paraderos usados se encuentran conectados por arcos de acceso a 2, 3 y 4 zonas (el promedio de zonas asociadas al origen es de 4,49 y al destino de 5,18), pero existen casos en que pueden estar conectadas hasta con 12 zonas (lo que puede ocurrir en estaciones de metro aledañas a zonas pequeñas). Además, existen algunos casos en que desde el paradero inicial o final se tiene una sola alternativa de zona de origen o destino; en estos, los viajes serían asignados sin necesitar el modelo de inferencia zonal.

Analizando también los valores asociados a estos arcos de acceso y egreso, se generó la Figura 6-4 que muestra el número de observaciones por cada categoría de distancia de los arcos, separando entre origen y destino.

La figura muestra que la mayoría de los arcos usados, tanto al origen como al destino, tienen un valor de entre 200 y 700 metros de distancia (el promedio para orígenes es 491,1 metros y para destinos es 484,7 metros). Además, no existen en la muestra arcos menores a 100 metros, lo que se explica por la manera en que estos fueron calculados (al realizar un promedio ponderado entre las distintas manzanas de la zona, el valor resultante agrupa

tanto manzanas cercanas como lejanas, por lo que es difícil que el valor final sea muy bajo). Similarmente, los valores extremos más altos (más de 1.200 metros) son muy pocos y ocurren en zonas muy grandes.

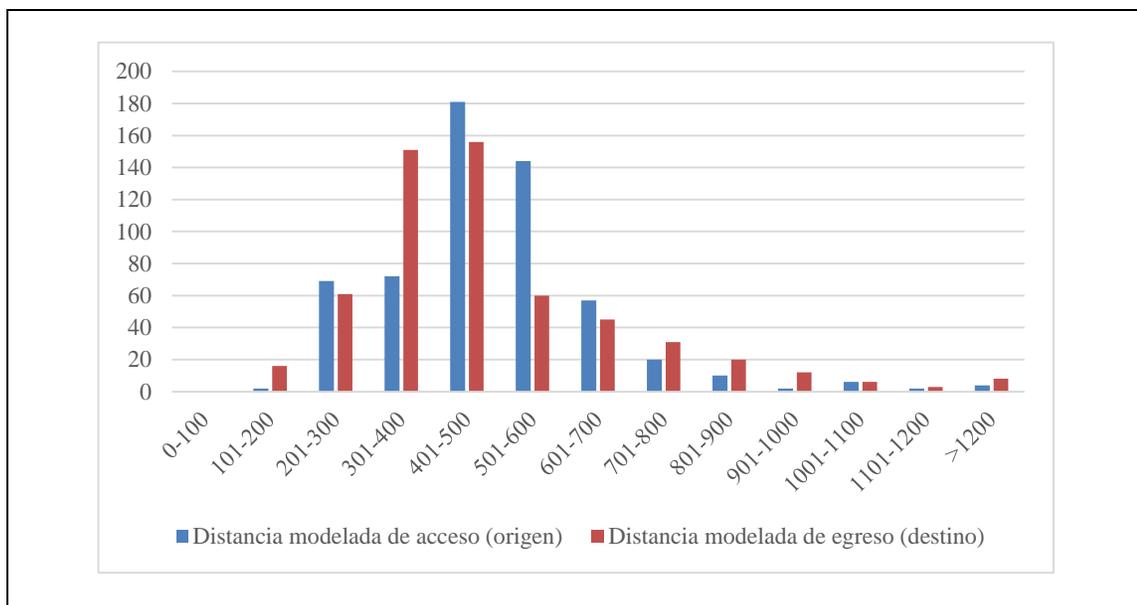


Figura 6-4: Número de observaciones por distancia de arcos de acceso en metros

Fuente: Elaboración propia

De igual manera, se analizaron las distancias de acceso y egreso reales, es decir, cuánto caminó cada uno de los encuestados. La Figura 6-5 muestra, para origen y destino, el número de observaciones por cada categoría de distancia de caminata real.

Esta visualización muestra que la mayoría de los encuestados caminaron menos de 500 metros, tanto al origen como al destino (el promedio para orígenes es 406,5 metros y para destinos es 343,0 metros). Cabe destacar que este análisis abarca solo 543 observaciones de caminatas de origen y 491 de destino, puesto que las restantes correspondían a direcciones (y por lo tanto, caminatas) imprecisas.

Por otro lado, al comparar las distancias modeladas y reales de acceso y egreso (es decir, Figuras 6-4 y 6-5), se constata que las últimas son significativamente menores. Lo anterior se explica puesto que los arcos de acceso y egreso deben representar la distancia promedio

de acceder al paradero desde la zona, mientras que las distancias reales corresponden solo a aquellos usuarios que prefirieron ese paradero por sobre todas sus demás alternativas. Por último, se observa que las caminatas de acceso exceden a las de egreso, explicable por la mayor densidad de la red en las zonas de destino.

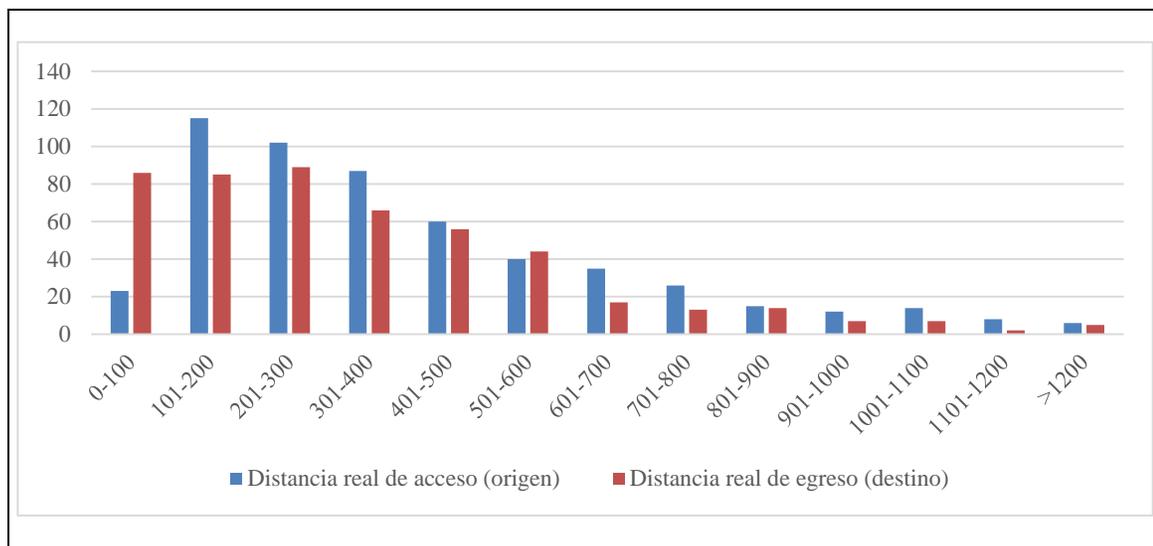


Figura 6-5: Número de observaciones por distancia de caminata real en metros

Fuente: Elaboración propia

6.2.2 Variables de uso de suelo

El uso de suelo de las potenciales zonas de origen y destino es uno de los factores fundamentales para asignar viajes según el modelo de inferencia zonal. Por esto, se hace necesario revisar y presentar las variables más importantes que se asociarán a los potenciales de atracción y generación de las zonas de la ciudad.

La Figura 6-6 muestra las distintas proporciones de cada tipo de uso de suelo presentes en las zonas de origen y destino reportadas por la encuesta, comparándolas con el promedio de zonas de la ciudad.

Es posible observar que las zonas de origen reportadas por los encuestados son principalmente residenciales, superando la participación promedio de este uso de suelo en la ciudad. Similarmente, las zonas de destino presentan porcentajes más altos que el

promedio de uso comercial, educacional, oficinas y salud, pero una menor proporción habitacional que las de origen. Ambos resultados eran esperables dado que el período de análisis corresponde a la mañana, donde la mayoría de los usuarios del sistema viaja desde su hogar a su lugar de trabajo o estudio. Sin embargo, el uso industrial de las zonas de destino es menor al promedio de la ciudad, lo que podría adelantar un bajo poder de atracción de esta variable (o que, simplemente, los encuestados tienen una baja participación de trabajadores industriales).

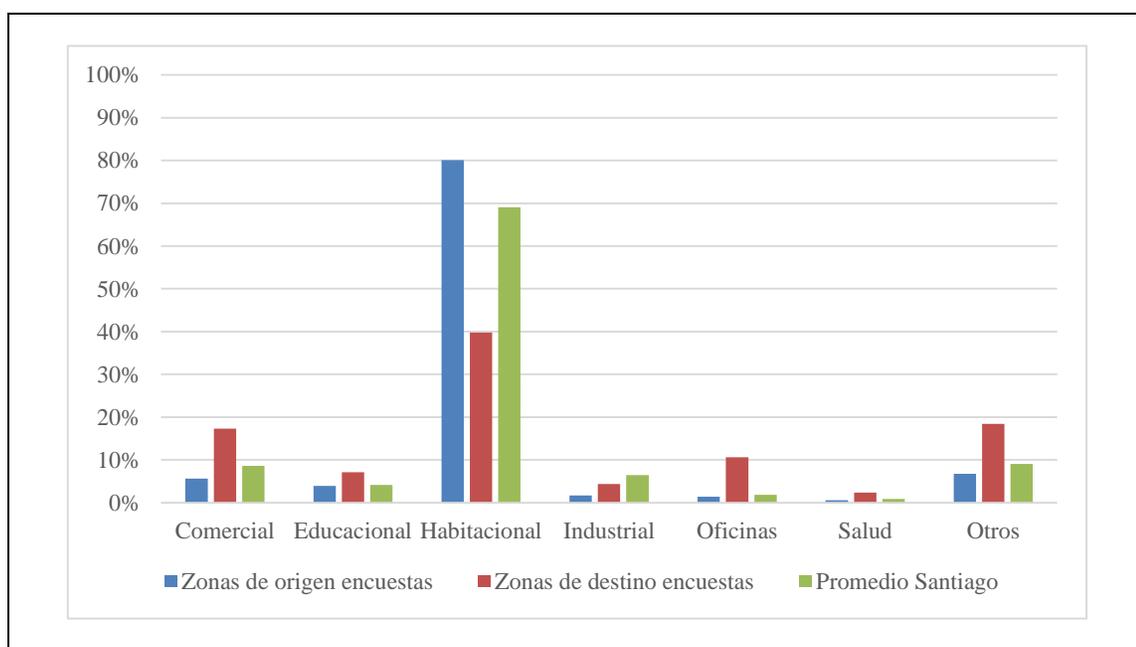


Figura 6-6: Porcentaje de superficie construida por tipo de uso de suelo

Fuente: Elaboración propia en base a SII (2014)

Por otro lado, se analizó también el avalúo fiscal promedio del metro cuadrado construido en las distintas zonas, con el objetivo de caracterizar la muestra y dar luces sobre posibles efectos relacionados al precio de los terrenos (Figura 6-7).

Este gráfico permite notar algunas características interesantes de los datos de la encuesta. Primero, se observa que el precio promedio del terreno habitacional en Santiago es muy similar al de los orígenes reportados en la encuesta, lo que indica que la muestra parece representativa al menos en términos de origen promedio. Por otro lado, se observa una

tendencia general a que los usos de suelo asociados a los destinos de la encuesta tengan una tasación mayor que el promedio de la ciudad; esto podría explicarse por el patrón general de atracción de viajes de Santiago, donde la mayoría de ellos se dirige hacia los sectores centro y oriente de la capital, donde existen terrenos de mayor valor.

Este último efecto es particularmente importante en el destino comercial, educacional y habitacional, donde en los dos primeros casos existen grandes diferencias en las densidades de uso entre establecimientos de la ciudad; respecto al comercio, por ejemplo, resulta ilustrativo comparar un gran centro comercial ubicado en sectores céntricos con un negocio de barrio de las afueras de la ciudad. En cuanto al uso habitacional, esto podría explicarse porque las viviendas de mayor valor atraen una mayor cantidad de viajes de trabajo (de asesoras del hogar u otros servicios, por ejemplo).

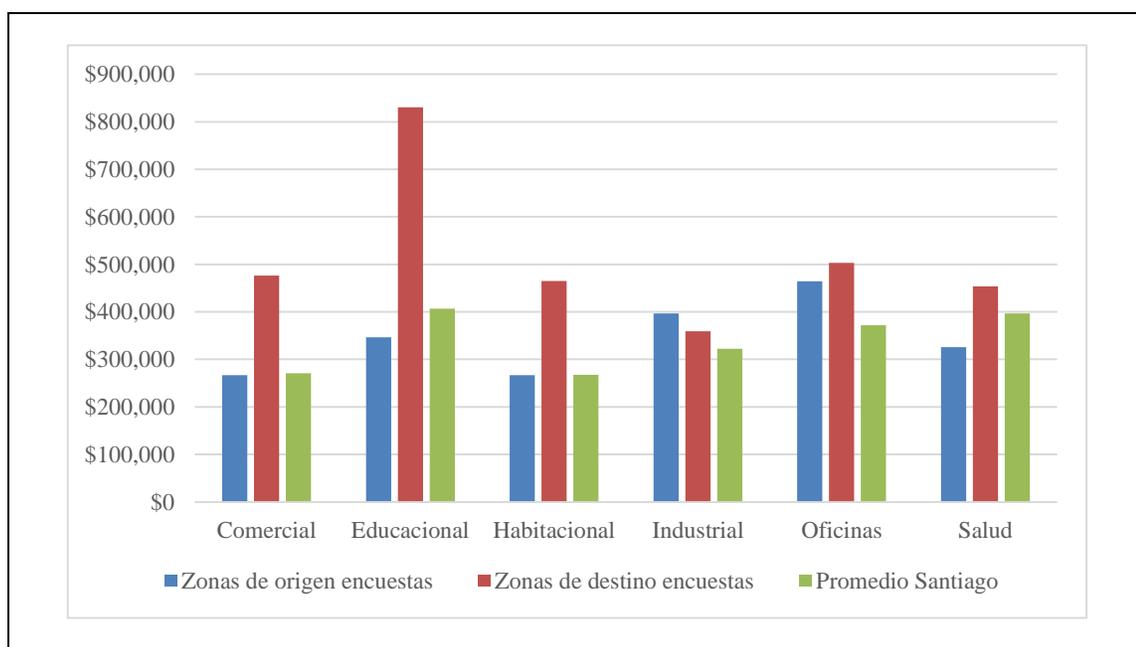


Figura 6-7: Avalúo fiscal promedio (\$2014) de m^2 construido por uso de suelo

Fuente: Elaboración propia en base a SII (2014)

6.2.3 Variables socioeconómicas y del viaje

Dado que las tarjetas electrónicas de Santiago no poseen información socioeconómica sobre sus dueños, los modelos estimados en esta tesis no pueden incluir este tipo de

variables directamente. No obstante, un análisis general de estos datos resulta importante para dar luces sobre la representatividad de la muestra que estamos utilizando para desarrollar los modelos. Es importante destacar que un análisis más detallado sobre estos factores fue realizado por Abud (2015), presentando los resultados de otras preguntas de la encuesta.

En el caso de estudio de esta tesis, el banco de estimación tiene un 51,3% de mujeres y 48,7% de hombres, muy similar a lo reportado por el Instituto Nacional de Estadísticas (INE, 2014) para el año 2013 (en que se realizó la encuesta): 51,1% mujeres y 48,9% hombres.

Respecto al tipo de tarjeta, un 30,6% de los encuestados declaró el pago de tarifa de estudiante, mientras que un 64,3% pagó tarifa normal (el 5,1% restante no respondió o pagó tarifa preferencial de adulto mayor).

La distribución por edades de la muestra, por su parte, se presenta en la Figura 6-8 junto al total de la población para el año 2013. Se pueden observar ciertas diferencias importantes, ya que en la muestra existe una menor proporción de personas menores de 20 y mayores de 60, contrastando con un alto número de viajeros entre 20 y 29 años. Estas diferencias no debiesen preocupar mayormente, puesto que el universo de usuarios de transporte público de la mañana (que están representados por el banco de estimación) difícilmente calza con el total de la población; por ejemplo, los niños pequeños y adultos mayores realizan muy pocos viajes durante el horario de punta mañana. No obstante lo anterior, y en vista de que no se cuenta con un catastro socioeconómico detallado de los viajeros de Transantiago, se decidió presentar datos del INE como una primera aproximación.

Por otro lado, la Figura 6-9 muestra la distribución de los encuestados que declararon ingresos (cerca de un 38% no lo hizo, ya sea porque no responde o no trabaja). De ellos, la moda se encuentra entre \$200.001 y \$300.000, mientras que la mediana está en el intervalo de \$300.001 a \$400.001. Existe también cerca de un 10% de los encuestados,

principalmente interceptados en las comunas de Ñuñoa, Peñalolén, La Florida y Las Condes, que declaró un ingreso personal superior al \$1.000.000.

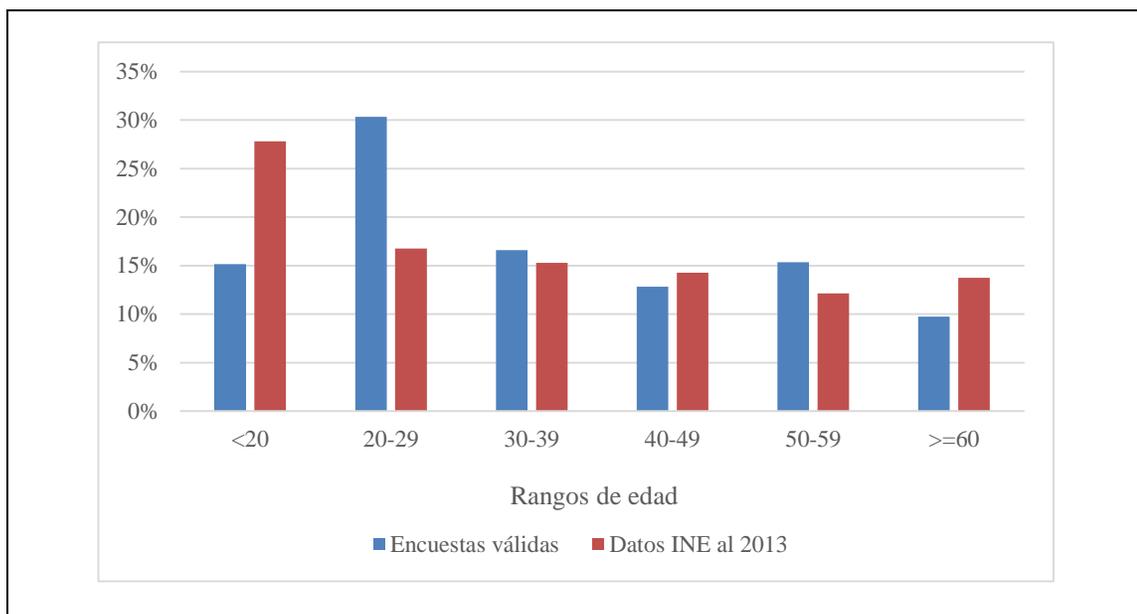


Figura 6-8: Distribución por rango etario de la muestra y población

Fuente: Elaboración propia usando INE (2014)

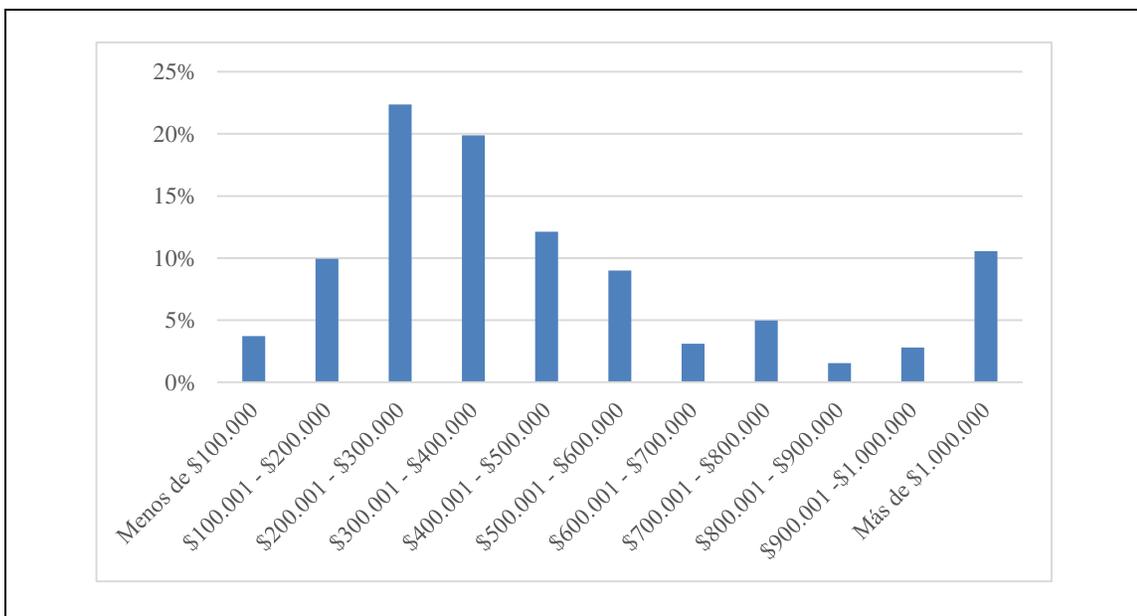


Figura 6-9: Distribución por ingresos personales de la muestra

Fuente: Elaboración propia usando INE (2014)

Otro punto relevante que vale la pena analizar es la hora en que fue tomada la encuesta. Cabe recordar que el propósito de la misma era entregar datos para estimar modelos en el horario de punta mañana; sin embargo, la encuesta fue tomada en un horizonte de tiempo más amplio para obtener un mayor número de encuestas dado el presupuesto disponible. Así, la Figura 6-10 muestra la distribución horaria en que fueron realizadas las distintas encuestas. Es importante notar que esta hora corresponde al momento en que se interceptó al viajero y no al inicio de su viaje (por ejemplo, un viaje puede haber comenzado a las 8:30 am en San Bernardo pero encuestado a las 9:30 am en Escuela Militar).

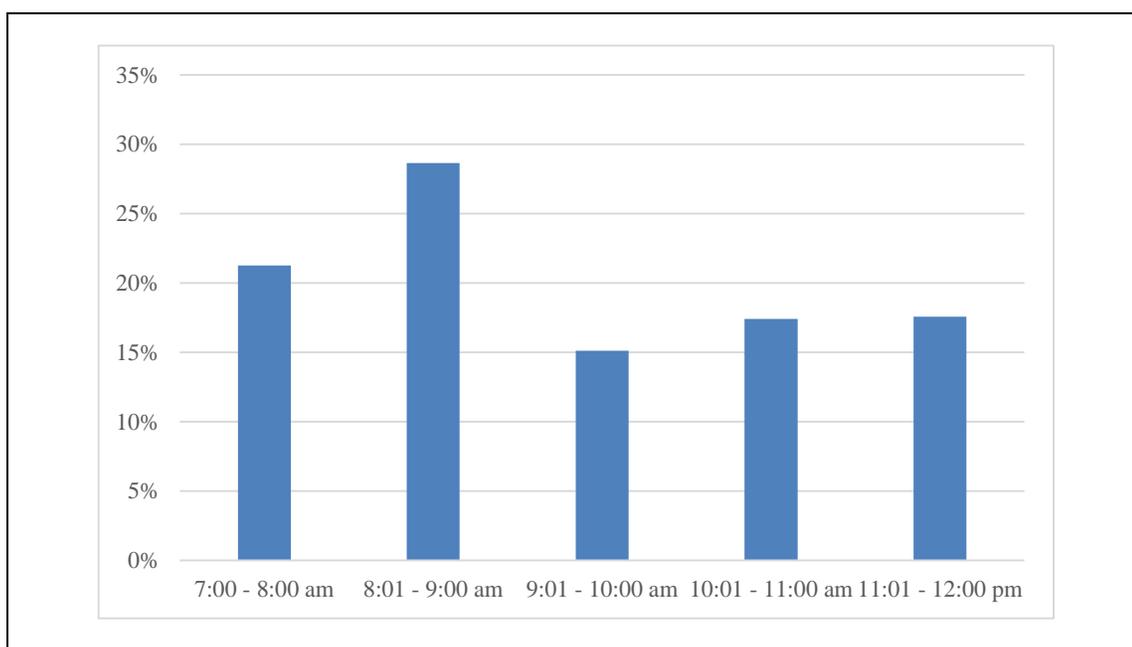


Figura 6-10: Distribución horaria de realización de encuestas de la muestra

Fuente: Elaboración propia usando INE (2014)

Es posible observar que la moda es entre las 8:01 y 9:00 am, y que cerca de un 65% de la muestra fue interceptada antes de las 10:00 am. Considerando que la herramienta FONDEF pretende modelar la punta mañana (perteneciente a los dos primeros períodos), prácticamente un 50% de la muestra cumpliría con estar en este horario.

El hecho anterior preocupa, puesto que presumiblemente los patrones y propósitos de viajes son distintos en el período punta, comparado con el resto de la mañana. Para analizar

este efecto, la Tabla 6-1 presenta la distribución de propósitos de viaje de la muestra completa, comparándola con los datos registrados antes de las 9:00 am.

Tabla 6-1: Distribución de propósitos de viaje por horario

Propósito	Muestra Completa	Antes de las 9:00am
Trabajo	56,7 %	72,4 %
Estudio	25,3 %	23,0 %
Trámites	6,9 %	1,8 %
Salud	3,4 %	0,7 %
Otros	7,7 %	2,1 %

Fuente: Elaboración propia

De la Tabla anterior se puede observar que los viajes registrados antes de las 9:00 am tienen una distribución de propósitos distinta al total de la muestra. En particular, interesa notar que los viajes realizados durante el período punta tienen una mayor proporción de viajes de trabajo y presentan menor cantidad de casos con propósitos secundarios (trámites, salud y otros). Respecto a los viajes de estudio, se puede apreciar que el porcentaje es bastante similar para ambos períodos, aunque ligeramente menor en el período punta.

7 MODELACIÓN Y RESULTADOS

En este capítulo se presentan los resultados de diversos modelos de inferencia zonal para viajes observados en paraderos de transporte público; estos fueron estimados usando *BIOGEME* (Bierlaire, 2003) según la metodología presentada en el capítulo 3. El subcapítulo 7.1 contiene la primera familia de modelos estimados, que se utilizaron además para calibrar los ponderadores asociados a los arcos de acceso. A continuación, se presenta y analiza el modelo de inferencia definitivo (subcapítulo 7.2), estimado con la red corregida. Por su parte, el subcapítulo 7.3 presenta algunas formulaciones alternativas, explorando potenciales mejoras al modelo. Finalmente, el subcapítulo 7.4 presenta la metodología de reconstrucción de la matriz origen-destino zonal, y exhibe sus resultados.

7.1 Modelos de inferencia zonal iniciales

Se estimó y comparó distintas especificaciones del modelo, siguiendo el enfoque de maximización de la verosimilitud y una formulación Logit Multinomial (MNL, por su abreviación en inglés), usando una función de utilidad de la forma general que presenta la Ecuación (7.1).

$$V_{mn/kl} = \ln(A_m) + \ln(B_n) + \ln(f_{mn}^{kl}) \quad (7.1)$$

Más específicamente, las funciones de utilidad fueron expresadas de la siguiente manera:

$$V_{mn/kl} = \theta_O * \ln(\theta_{u_{h_O}} * U_{h_{Om}}) + \theta_D * \ln(\sum_P \theta_{m2_{PD}} * M2_{PDn}) + \ln\left(\frac{e^{(\beta_O * d_{m-k})}}{(d_{m-k})^{\gamma_O}}\right) + \ln\left(\frac{e^{(\beta_D * d_{l-n})}}{(d_{l-n})^{\gamma_D}}\right) \quad (7.2)$$

donde $U_{h_{Om}}$ es el número de unidades habitacionales en la zona de origen m ; $M2_{PDn}$ corresponde al total de metros cuadrados de la clasificación de uso de suelo P en la zona de destino n ; d_{m-k} y d_{l-n} son las distancias de acceso (en kilómetros) desde las zonas de origen y destino hacia los paraderos escogidos (iniciales y finales,

respectivamente); finalmente, θ , β e γ son los parámetros a estimar. Los dos primeros sumandos están relacionados con el poder de generación y atracción de las zonas m y n , respectivamente; mientras que los últimos dos corresponden a la medida inversa del costo f_{mn}^{kl} , donde se usó una función *gamma* y se permitió explícitamente distintos impactos del acceso y egreso en la utilidad.

Es importante recordar que por razones de identificabilidad, no es posible estimar un parámetro para cada variable de tamaño en el origen y el destino, por lo que dos de ellas se fijan a un valor unitario. Además, se fijaron θ_O y θ_D en uno para asegurar que el modelo fuera independiente de la zonificación usada para la estimación (según lo expuesto en el subcapítulo 2.5).

Tras probar distintas especificaciones y estudiar los efectos de cada uso de suelo en la generación y atracción, el mejor modelo estimado se muestra en la Tabla 7-1.

Se puede observar que todos los parámetros tienen los signos esperados, y la mayoría de ellos resulta estadísticamente significativo; sin embargo, el parámetro asociado a los metros cuadrados de uso de suelo industrial en el destino no resultó significativamente distinto de cero (al 90% de confianza). En esta línea, aunque nuestras observaciones no fueron suficientes para identificar efectos significativos de este atractor de baja densidad, se decidió mantener el parámetro en el modelo puesto que tiene el signo correcto y es comúnmente relevante en modelos de distribución de viajes.

Por otro lado, el modelo muestra que los usos de suelo más atractivos son los edificios de salud y educacionales, seguidos por el comercio y las oficinas, mientras que las viviendas e industrias atraen un número relativamente bajo de viajes por metro cuadrado.

Respecto a la función de costo inverso, el parámetro β_O obtuvo un test t de 0,34; en vista de que no resultó significativo y no es un parámetro fundamental (la curva de costo mantiene sin β_O la forma esperada), fue eliminado del modelo.

Tabla 7-1: Modelo inicial (MNL1)

Asociado a	Parámetro	Valor	Test t
Generación	θ_O	1	Fija
	$\theta_{u_{H_O}}$	1	Fija
Atracción	θ_D	1	Fija
	$\theta_{m^2_{C_D}}$ - Comercial m^2	0,194	2,76
	$\theta_{m^2_{E_D}}$ - Educacional m^2	0,568	2,81
	$\theta_{m^2_{H_D}}$ - Habitacional m^2	0,0489	2,55
	$\theta_{m^2_{I_D}}$ - Industrial m^2	0,0235	1,01*
	$\theta_{m^2_{O_D}}$ - Oficinas m^2	0,129	1,91
	$\theta_{m^2_{S_D}}$ - Salud m^2	1	Fija
Costo	γ_O	2,77	14,57
	β_D	2,04	2,27
	γ_D	2,78	5,41
	Log-verosimilitud nula (LVN)	-1.614,9	
Ajuste	Log-verosimilitud final (LVF)	-1.343,1	
	Test de razón de verosimilitud (TRV)	543,6	

Fuente: Elaboración propia

7.1.1 Calibración de arcos de acceso

El modelo presentado contiene ponderadores asociados a los distintos tipos de uso de suelo que pueden atraer viajes. En este sentido, ilustra las diferencias en poder de atracción entre uno y otro, permitiendo encontrar relaciones de proporcionalidad entre ellos.

Se decidió aprovechar esta información para revisar el procedimiento con el que se calcularon los arcos de acceso (en el subcapítulo 5.3), en el que se asumió que todos estos distintos usos de suelo eran igualmente atractivos para los viajeros.

De esta manera, se utilizaron los parámetros de atracción del modelo MNL1 para ponderar los distintos metros cuadrados de cada manzana, en el promedio de distancias para el cálculo de los arcos atractores. A continuación, con estos nuevos arcos de acceso se reestimó el modelo, obteniéndose a su vez nuevos ponderadores a utilizar en el cálculo de arcos de acceso, dando pie a un proceso iterativo de calibración de los mismos. La Tabla 7-2 exhibe los resultados de este proceso iterativo, donde se mantuvieron las mismas variables en cada modelo, actualizándose solamente los ponderadores que correspondan.

Tabla 7-2: Resultados iteración de atracción para arcos de acceso

Asociado a	Parámetro	MNL1 Valor	MNL2 Valor	MNL3 Valor	MNL4 Valor	MNL4 Test t
Generación	θ_O			1		Fija
	$\theta_{u_{HO}}$			1		Fija
Atracción	θ_D			1		Fija
	$\theta_{m2_{CD}}$	0,194	0,241	0,235	0,236	2,56
	$\theta_{m2_{ED}}$	0,568	0,69	0,671	0,673	2,60
	$\theta_{m2_{HD}}$	0,0489	0,0637	0,0619	0,0622	2,40
	$\theta_{m2_{ID}}$	0,0235	0,0284	0,0279	0,028	1,00*
	$\theta_{m2_{OD}}$	0,129	0,18	0,172	0,173	1,86
	$\theta_{m2_{SD}}$			1		Fija
Costo	γ_O			2,77		14,57
	β_D	2,04	1,68	1,64	1,65	1,89
	γ_D	2,78	2,49	2,48	2,48	5,19
Ajuste	TRV	543,6	552,8	553,1	553,0	

Fuente: Elaboración propia

Como es posible observar, se alcanzó una rápida convergencia y las diferencias en los ponderadores de atracción entre el modelo MNL3 y MNL4 son despreciables (menores al 0,5%). Por razones visuales se omitieron los test t de los modelos intermedios en la tabla,

pero se mantienen relativamente estables y similares entre una iteración y otra. Por su lado, las variables asociadas al origen no cambian de una iteración a otra, puesto que la forma de la función de utilidad garantiza que sean separables.

Sumado a la convergencia, el hecho que el test de razón de verosimilitud (y otros tests omitidos en la tabla) mejore rápidamente en las primeras iteraciones, refuerza esta metodología y nos da luces que la relación propuesta entre modelo de inferencia y arcos de acceso existe y es aprovechable.

Siguiendo esta línea, en modelos posteriores se detectó que la variable correspondiente a los metros cuadrados habitacionales en la zona de origen, explicaba mejor los datos que las unidades de este mismo tipo (lo que podría explicarse si existiera una correlación positiva entre el tamaño de la vivienda y cuántas personas viven en ella, por ejemplo).

Finalmente, se realizó una última revisión de la zonificación y red ficticia, en vista de los resultados de asignación preliminares, corrigiéndose detalles menores. En particular, se eliminaron algunos arcos de acceso que constituían alternativas poco realistas, se agruparon zonas poco atractivas y se corrigieron unas pocas manzanas que continuaban estando mal asignadas. Con esto, la red definitiva llegó a tener 1.171 zonas y 33.504 arcos de acceso bidireccionales. Lo anterior significa que, en promedio, cada zona está conectada con poco más de 28 puntos de acceso al transporte público, mientras que el paradero de bus y estación de metro promedio se encuentran asociados a 3 y 8 zonas, respectivamente.

7.2 Modelo de inferencia zonal definitivo

Usando la red definitiva y los arcos de acceso creados según la metodología previamente expuesta, es decir, usando como ponderadores los metros cuadrados habitacionales para el origen y los resultados del modelo MNL4 para el destino, se reestimó el modelo siguiendo la formulación anterior y se actualizaron los arcos de acceso. Los resultados correspondientes a este modelo se exhiben en la Tabla 7-3.

Tabla 7-3: Modelo definitivo (MNL5)

Asociado a	Parámetro	Valor	Test t
Generación	θ_O	1	Fija
	θ_{m2H_O}	1	Fija
Atracción	θ_D	1	Fija
	θ_{m2C_D} - Comercial m^2	0,228	2,53
	θ_{m2E_D} - Educacional m^2	0,689	2,59
	θ_{m2H_D} - Habitacional m^2	0,0623	2,39
	θ_{m2I_D} - Industrial m^2	0,0315	1,04*
	θ_{m2O_D} - Oficinas m^2	0,198	1,93
	θ_{m2S_D} - Salud m^2	1	Fija
Costo	γ_O	2,83	14,75
	β_D	1,71	1,96
	γ_D	2,48	5,24
	Log-verosimilitud nula (LVN)	-1.612,9	
Ajuste	Log-verosimilitud final (LVF)	-1.327,3	
	Test de razón de verosimilitud (TRV)	571,1	

Fuente: Elaboración propia

Se puede observar que este modelo tiene los mejores resultados de verosimilitud presentados hasta ahora, pero presenta algunas disminuciones menores en algunos de sus test t respecto al modelo inicial MNL1. Esto no debiese preocupar, pues se explica dado que parte del poder explicativo del modelo fue traspasado a una mejor construcción de los arcos de acceso y egreso, los que fueron calculados usando información endógena acerca de qué tan atractiva era cada manzana cercana para los viajes que empezaban o terminaban en cada paradero o estación.

Por otro lado, se probó recalcular los arcos de acceso usando estos nuevos ponderadores para iterar nuevamente si fuese necesario, siguiendo la metodología del subcapítulo anterior. Sin embargo, desde la primera corrida las diferencias entre los ponderadores de atracción resultaron despreciables (menores al 0,5%), por lo que se concluyó que el modelo había convergido. Esta rápida convergencia no debiera extrañar, puesto que los últimos cambios realizados a la red y zonificación fueron detalles menores.

Comparando con el modelo inicial MNL1, este nuevo modelo presenta mejores resultados en cuanto a verosimilitud, lo que se puede apreciar al comparar los tests de razón de verosimilitud. En cuanto a sus parámetros, estos mantienen no solo los signos esperados sino también el mismo orden de poder de atracción entre los distintos usos de suelo de destino. Nuevamente, nuestras observaciones no permitieron rescatar un parámetro estadísticamente distinto de cero para el destino industrial (no obstante, se decidió mantenerlo en el modelo por las mismas razones esgrimidas anteriormente).

Analizando los parámetros presentados, podría llamar la atención el gran poder atractivo que ejercen los centros de salud y educacionales (comparados con el uso comercial o de oficinas, por ejemplo). Este efecto era, hasta cierto punto, esperable dada la gran densidad de uso que tienen este tipo de instalaciones. Sin embargo, una mirada más detenida nos hace reconocer que existen efectos adicionales que acentúan estos valores: primero, como ya fue mencionado, se perdió una menor proporción de viajes con propósito estudio o salud, por lo que estos podrían estar sobrerrepresentados en la muestra; asimismo, existe una mayor proporción de viajes relacionados con salud durante el período fuera de punta; por último, no todas las alternativas asociadas a comercio y oficinas representan centros de atracción de viajes, ya que un gran espectro de distintos tipos de edificios caen en estas categorías (es más, algunos de ellos no se encuentran abiertos durante el período de análisis).

El último punto es especialmente relevante, puesto que ofrece una oportunidad de mejorar el modelo, alcanzable con una información zonal más detallada. En efecto, sería esperable obtener mejores modelos, por ejemplo, si se agregara información geo-referenciada que

permitiera distinguir distintos tipos de usos comerciales u oficinas, o que indicara el número de puestos de trabajo registrados en cada zona.

Respecto a la función de costo inverso, nuevamente el parámetro β_0 salió del modelo por no resultar significativo, por lo que la función f_{mn}^{kl} puede expresarse matemáticamente como muestra la Ecuación (7.3).

$$f_{mn}^{kl} = \left(\frac{e^{(\beta_D * d_{l-n})}}{(d_{m-k})^{\gamma_O} * (d_{l-n})^{\gamma_D}} \right) \quad (7.3)$$

Por su parte, la Figura 7-1 muestra un gráfico de esta función, donde las áreas oscuras representan costos totales más altos.

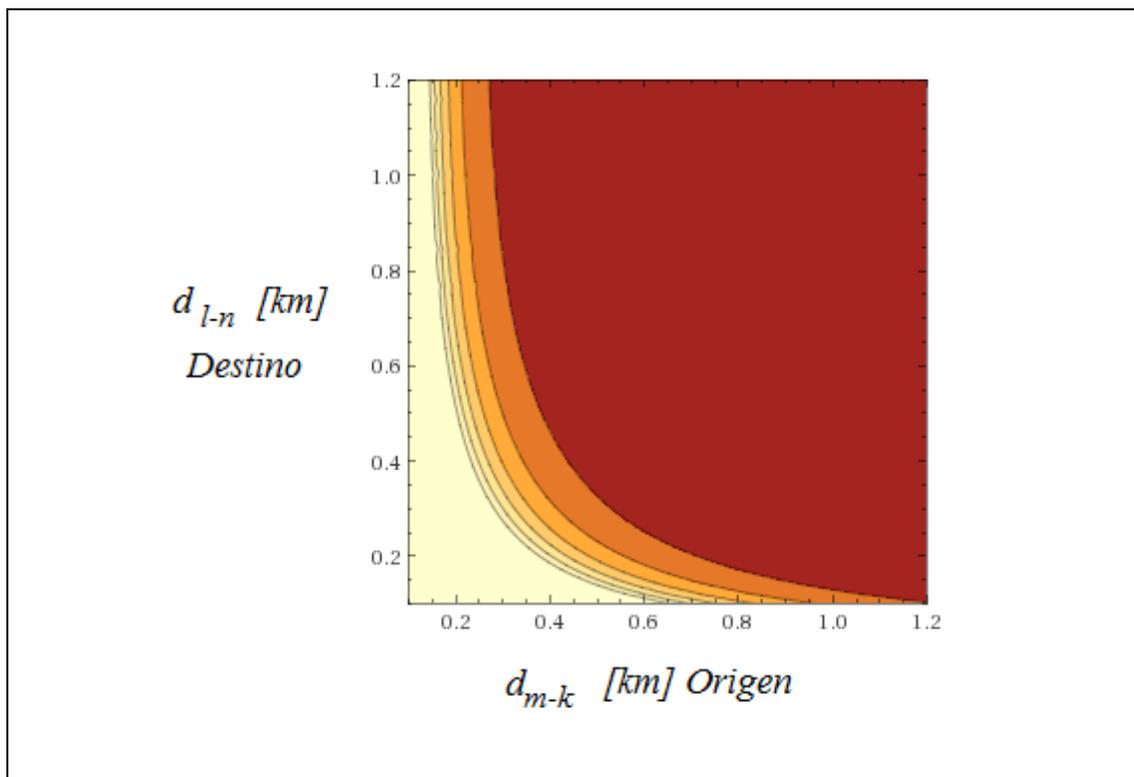


Figura 7-1: Función de costo inversa f_{mn}^{kl} del modelo MNL5

Fuente: Elaboración propia

Se observa que los usuarios encuestados son aparentemente más sensibles a largas distancias de caminata en el segmento inicial de sus viajes, tolerando mayores tiempos de caminata hacia su destino. Una explicación posible para este resultado es que en la mañana los usuarios suelen comenzar sus viajes cerca del hogar, donde tienen un mejor conocimiento de las alternativas cercanas de transporte público, pudiendo evitar tiempos de caminata mayores. Además, en el período de análisis la mayoría de los viajes son atraídos al centro de la ciudad, que generalmente es más seguro que los orígenes y tiene un uso de suelo más diverso, por lo que puede resultar más atractivo durante la caminata. Es importante destacar que estos resultados son consistentes con lo obtenido en los modelos de elección de ruta presentados en el estudio de Abud (2015), donde también se observó que mientras más avanzado el viaje la caminata producía una des-utilidad menor.

Usando este modelo, el Anexo D presenta ejemplos de aplicación en que se calculan asignaciones de viajes a zonas aledañas a algunas estaciones de la red de metro. Específicamente, presenta en porcentajes cuántos de los viajes iniciados (y terminados) en Tobalaba y San Joaquín tienen como origen (y destino) cada una de las zonas conectadas por arcos de acceso con estas estaciones.

7.3 Formulaciones alternativas

Como parte de la tesis, se exploraron distintas maneras de formular el modelo, considerando nuevas estructuras, variables e interacciones, que se detallan a continuación.

7.3.1 Diferenciación de tarjetas de estudiante

En primer lugar, como se recomienda en la literatura especializada, se analizó la posibilidad de diferenciar viajeros que son atraídos de distinta manera a distintas alternativas de destino.

En esta línea, considerando también que los estudiantes de Santiago tienen tarjetas inteligentes diferenciadas (para permitir el cobro de una menor tarifa) y que en la encuesta se preguntó qué tipo de tarifa se pagaba, se estudiaron modelos en que se daba un tratamiento especial a quienes poseían tarjetas de estudiante. Para ello, se incluyó una

interacción entre la cantidad de metros cuadrados educacionales de una zona y el tipo de tarjeta, utilizando una variable dicotómica que toma valor uno para las tarjetas de estudiantes. Los resultados de este modelo se presentan en la Tabla 7-4.

Tabla 7-4: Modelo diferenciando estudiantes (MNL5e)

Asociado a	Parámetro	Valor	Test t
Generación	θ_O	1	Fija
	$\theta_{m^2_{HO}}$	1	Fija
Atracción	θ_D	1	Fija
	$\theta_{m^2_{CD}}$ - Comercial m^2	0,228	2,56
	$\theta_{m^2_{ED}}$ - Educacional m^2	0,240	1,97
	$\theta_{m^2_{Est-ED}}$ - Educacional m^2 estudiantes	2,93	2,03
	$\theta_{m^2_{HD}}$ - Habitacional m^2	0,0515	2,35
	$\theta_{m^2_{ID}}$ - Industrial m^2	0,0332	1,07*
	$\theta_{m^2_{OD}}$ - Oficinas m^2	0,196	1,92
	$\theta_{m^2_{SD}}$ - Salud m^2	1	Fija
Costo	γ_O	2,83	14,75
	β_D	1,75	2,00
	γ_D	2,49	5,26
Ajuste	Log-verosimilitud nula (LVN)	-1.612,9	
	Log-verosimilitud final (LVF)	-1.313,6	
	Test de razón de verosimilitud (TRV)	598,5	

Fuente: Elaboración propia

Como es posible observar, esta diferenciación de tarjetas permite separar el efecto de atracción de la superficie educacional construida en cada zona sobre cada tipo de viajero. El parámetro de atracción asociado a la interacción entre tarjeta de estudiante y metros cuadrados educacionales tiene un valor comparativamente muy alto, mucho mayor que la

atracción de este mismo tipo de uso de suelo sobre tarjetas tipo adulto. De esta manera, se confirma que el uso de suelo educacional atrae a quienes poseen tarjetas de estudiantes con una intensidad mucho mayor.

Cabe destacar que este modelo se presenta con un carácter exploratorio y actualmente no puede ser aplicado en la herramienta computacional, puesto que la información de viajes incluida no diferencia entre tarjetas de distintos tipos. Sin embargo, el sistema de pago de Transantiago sí es capaz de diferenciar entre tarjetas, por lo que existiría la posibilidad de separar los viajes a futuro y aplicar un modelo de este tipo.

Por último, se destaca el potencial de profundizar con la distinción de los distintos tipos de estudiantes (educación básica, media o superior), puesto que presumiblemente son atraídos a distintos tipos de instituciones educativas. Para aprovechar esta condición, se requeriría no solo un mayor nivel de detalle en la información recopilada por las tarjetas, sino también incluir al sistema zonal información sobre el número de matrículas de cada nivel de enseñanza en cada zona, similar a la tarea realizada en el estudio MIDEPLAN (2010).

7.3.2 Constantes modales para metro

En vista de que los usuarios de transporte público de Santiago tienen distintas tolerancias a la caminata dependiendo del modo que escogen, se decidió estimar modelos que incluyen diferenciaciones de costo al caminar hacia o desde estaciones de metro.

Luego de probar distintas especificaciones, el mejor modelo obtenido mantiene los poderes de atracción y generación del modelo MNL5, pero utiliza la siguiente función de costo inverso:

$$f_{mn}^{kl} = \left(\frac{e^{(\beta_O * d_{m-k} + \beta_{MO} * M_o * d_{m-k} + \beta_D * d_{l-n})}}{(d_{m-k})^{\gamma_O} * (d_{l-n})^{\gamma_D}} \right) \quad (7.4)$$

donde M_o representa una variable dicotómica que toma valor uno cuando el paradero de inicio es una estación de metro.

Esta especificación, indica que resultó parcialmente significativa la interacción entre la distancia caminada y el tipo de estación al inicio del viaje (entre la zona de origen m y el paradero inicial k), en la parte de la función de costo asociada a la exponencial. La base de calibración no permitió capturar efectos de interacción en el denominador ni en la caminata de destino (entre el paradero final l y la zona de destino n). Los resultados de este modelo se detallan en la Tabla 7-5.

Tabla 7-5: Modelo con costos diferenciados entre modos (MNL5c)

Asociado a	Parámetro	Valor	Test t
Generación	θ_O	1	Fija
	$\theta_{m2_{H_O}}$	1	Fija
Atracción	θ_D	1	Fija
	$\theta_{m2_{C_D}}$ - Comercial m^2	0,228	2,53
	$\theta_{m2_{E_D}}$ - Educacional m^2	0,689	2,59
	$\theta_{m2_{H_D}}$ - Habitacional m^2	0,0623	2,39
	$\theta_{m2_{I_D}}$ - Industrial m^2	0,0315	1,04*
	$\theta_{m2_{O_D}}$ - Oficinas m^2	0,198	1,93
	$\theta_{m2_{S_D}}$ - Salud m^2	1	Fija
Costo	β_O	-2,28	-1,20*
	β_{MO}	1,51	1,75*
	γ_O	2,83	14,75
	β_D	1,71	1,96
	γ_D	2,48	5,24
	Log-verosimilitud nula (LVN)	-1.612,9	
Ajuste	Log-verosimilitud final (LVF)	-1.325,8	
	Test de razón de verosimilitud (TRV)	574,2	

Fuente: Elaboración propia

Como es posible constatar, el único efecto capturado no resulta ser significativo al 95%. Adicionalmente, se realizó un test de razón de verosimilitud para comparar los modelos MNL5c y MNL5, que resultó en la aceptación del modelo restringido (en este caso, correspondiente al modelo MNL5). En vista de lo anterior y considerando el criterio de parsimonia, las constantes modales no fueron incluidas en el modelo definitivo.

7.3.3 Formulación Logit Jerárquica

Se detectó que las distintas alternativas del modelo de elección Logit (es decir, los pares origen-destino zonales del viaje) podrían presentar algún grado de correlación entre ellas. En efecto, un viaje observado entre dos paraderos de transporte público tiene un conjunto limitado de potenciales zonas de origen m y destino n , por lo que existen alternativas de pares que tienen una misma zona de origen (o destino).

En vista de lo anterior, se estimaron formulaciones de tipo Logit Jerárquico en que los distintos nidos estaban formados por agrupaciones de pares que compartían una zona en el origen o destino. Puesto que las agrupaciones de zonas son arbitrarias (y aleatorias), no existen diferencias intrínsecas entre los distintos nidos, por lo que se forzó que sus parámetros estructurales fuesen iguales.

Sin embargo, en ninguna de las formulaciones estimadas se logró rescatar este efecto y los parámetros asociados a la jerarquía no resultaron significativamente distintos de uno, lo que implica que la formulación jerárquica colapsaba al modelo Logit Multinomial.

7.3.4 Estimación diferenciando datos de punta mañana

Una de las limitaciones del modelo estimado para Santiago es que el horario en que se tomaron las encuestas excedía al de punta mañana, para el que se pretendía reconstruir las matrices origen-destino de viajes entre zonas. Esta consideración resulta preocupante, puesto que los patrones de viajes en distintos períodos del día no son iguales y podrían obtenerse modelos imprecisos.

En consecuencia, se estimó modelos que permitían distintos potenciales de atracción para cada uso de suelo según el horario en que fue registrada la observación. Específicamente, se incluyó una interacción entre la cantidad de metros cuadrados construidos de los distintos usos de una zona y el horario en que se tomó la encuesta, utilizando una variable dicotómica que toma valor uno para aquellas observaciones que fueron tomadas en el período punta mañana (antes de las 9:00 am). Los resultados del mejor modelo estimado con esta consideración se presentan en la Tabla 7-6.

Tabla 7-6: Modelo diferenciando punta mañana (MNL5am)

Asociado a	Parámetro	Valor	Test t
Generación	θ_O	1	Fija
	θ_{m2H_O}	1	Fija
Atracción	θ_D	1	Fija
	θ_{m2C_D} - Comercial m^2	0,393	2,49
	$\theta_{m2\text{9am}-C_D}$ - Comercial m^2 punta mañana	-0,341	-2,25
	θ_{m2E_D} - Educacional m^2	0,697	2,66
	θ_{m2H_D} - Habitacional m^2	0,0573	2,40
	θ_{m2I_D} - Industrial m^2	0,0354	1,09*
	$\theta_{m2\text{9am}-O_D}$ - Oficinas m^2 punta mañana	0,359	1,87
	θ_{m2S_D} - Salud m^2	1	Fija
Costo	γ_O	2,83	14,75
	β_D	1,76	1,99
	γ_D	2,51	5,24
Ajuste	Log-verosimilitud nula (LVN)	-1.612,9	
	Log-verosimilitud final (LVF)	-1.321,5	
	Test de razón de verosimilitud (TRV)	582,7	

Fuente: Elaboración propia

De la Tabla 7-6 se puede notar que el modelo con diferenciación de observaciones de punta mañana presenta algunos parámetros muy distintos a los obtenidos en el modelo MNL5. En particular, los efectos que resultaron significativos fueron los asociados al comercio y oficinas, que poseen potenciales de atracción significativamente menores y mayores para el período de punta mañana, respectivamente. Además, cabe destacar que la muestra no permitió capturar diferencias horarias estadísticamente significativas para los potenciales de atracción de los usos educacionales, habitacionales, industriales y de salud. Por su parte, el parámetro asociado al potencial de atracción de oficinas en el período fuera de punta obtuvo un test t de 0,70 y no se presenta en el modelo.

Finalmente, se realizó un test de razón de verosimilitud para comparar los modelos MNL5am y MNL5, que resultó en el rechazo del modelo restringido (es decir, el modelo con diferenciaciones horarias explica mejor la muestra). En vista de lo anterior, se estudiará la inclusión del modelo MNL5am en la herramienta FONDEF, lo que será discutido en reuniones próximas del proyecto y, en caso de ser aceptado, requerirá la recalibración de los arcos de acceso considerando los potenciales asociados al período punta.

7.3.5 Estimación de modelos considerando variables socioeconómicas

Finalmente, cabe destacar que por un criterio de aplicabilidad, se decidió no utilizar variables socioeconómicas (tales como sexo, edad, ingreso y otras) en la estimación de los modelos, a pesar de que muchas de estas variables sí estaban disponibles en la encuesta; la razón es que el modelo de inferencia resultante tiene un carácter agregado y se aplicará para asignar a zonas viajes observados de tarjetas Bip!, las que por el momento no cuentan con información socioeconómica asociada.

7.4 Reconstrucción de matriz origen-destino zonal

Una vez estimado el modelo de inferencia zonal definitivo y conforme a los objetivos de esta tesis, este fue utilizado para reconstruir la matriz origen-destino zonal de transporte público para la punta mañana de la ciudad de Santiago.

Recordando el capítulo 3, podemos encontrar los viajes entre una zona i y j calculando:

$$T_{ij} = \sum_k \sum_l Prob(ij/kl) * T_{kl} \quad \forall(i,j) \quad (7.5)$$

donde $Prob(ij/kl)$ se calcula utilizando el modelo Logit de inferencia estimado y T_{kl} corresponde a los viajes entre el par de paraderos k y l , conectados a las zonas i y j , respectivamente.

De esta manera, el procedimiento a seguir puede ser resumido en el siguiente algoritmo de asignación de viajes:

- (a) Para cada par de paraderos que tenga viajes en la matriz T_{kl} , seleccionar las posibles zonas de origen m y destino n , provenientes de los arcos de acceso y egreso asociados a k y l , respectivamente.
- (b) Usando la información de uso de suelo y costos, calcular las probabilidades de origen y destino para cada par de zonas, según el modelo de inferencia zonal.
- (c) Asignar los viajes T_{kl} a sus distintos pares origen-destino zonales según las probabilidades obtenidas en (b).
- (d) Repetir (a) hasta que se hayan asignado zonas de origen y destino a todos los viajes observados en paraderos.

Considerando este procedimiento, es importante recordar que la matriz de viajes entre paraderos que se esté asignando es fundamental en los resultados de la metodología; tanto sus bondades como sus limitaciones serán extendidas a la matriz de viajes zonales. En la corrida que se presenta a continuación, se utilizó la información de viajes entre paraderos correspondientes a abril 2013 para la punta mañana (6:30 a 8:30 am), calculada según lo presentado en Munizaga y Palma (2012), Devillaine (2012) y Munizaga et al. (2014).

Cabe destacar que un 1,77% de los viajes de la matriz de viajes entre paraderos tienen como origen o destino un paradero que no se encuentra en la red del proyecto FONDEF,

por lo que resulta imposible asignarlos a zonas. Este problema debiese solucionarse tras la actualización de la red de transporte público del proyecto, proceso que permitiría un calce completo con la matriz de viajes.

Una alternativa de solución temporal sería expandir las matrices que se cuentan en la actualidad en esta misma proporción, para así al menos mantener el total de viajes real de la ciudad en el período de análisis. Sin embargo, puesto que en corridas previas del proyecto se omitió estos viajes, se decidió en esta tesis omitirlos también, para mantener la comparabilidad entre ambos tipos de asignaciones de flujos (asignación de viajes entre paraderos y asignación de viajes entre zonas).

Las Figuras 7-2 y 7-3 exhiben los resultados del algoritmo de asignación de viajes previamente descrito, presentando gráficamente los totales de viajes generados y atraídos (respectivamente) por cada zona de la ciudad, provenientes de la matriz origen-destino zonal estimada para el período de punta mañana de la ciudad de Santiago. En ambos mapas, se encuentran coloreadas en tonos más oscuros las zonas con mayor cantidad de viajes (clasificadas según deciles).

La Figura 7-2 permite observar algunos focos de generación de la ciudad, como Maipú, Quilicura, Puente Alto, Peñalolén y Santiago; sin embargo, la concentración no es tan marcada como la producida por los focos de atracción de viajes (Figura 7-3).

Además, se presenta una clara concentración de viajes originados en zonas aledañas a los ejes de Metro. Lo anterior se puede explicar por la mayor densidad habitacional en torno a estaciones de la red, pero es posible conjeturar que parte de este efecto proviene de viajes en que no se está detectando la primera etapa, ya sea porque esta se realizó en un modo no integrado con Transantiago (automóvil, colectivo, etc.) o porque el usuario evadió el pasaje en su primera etapa de bus.

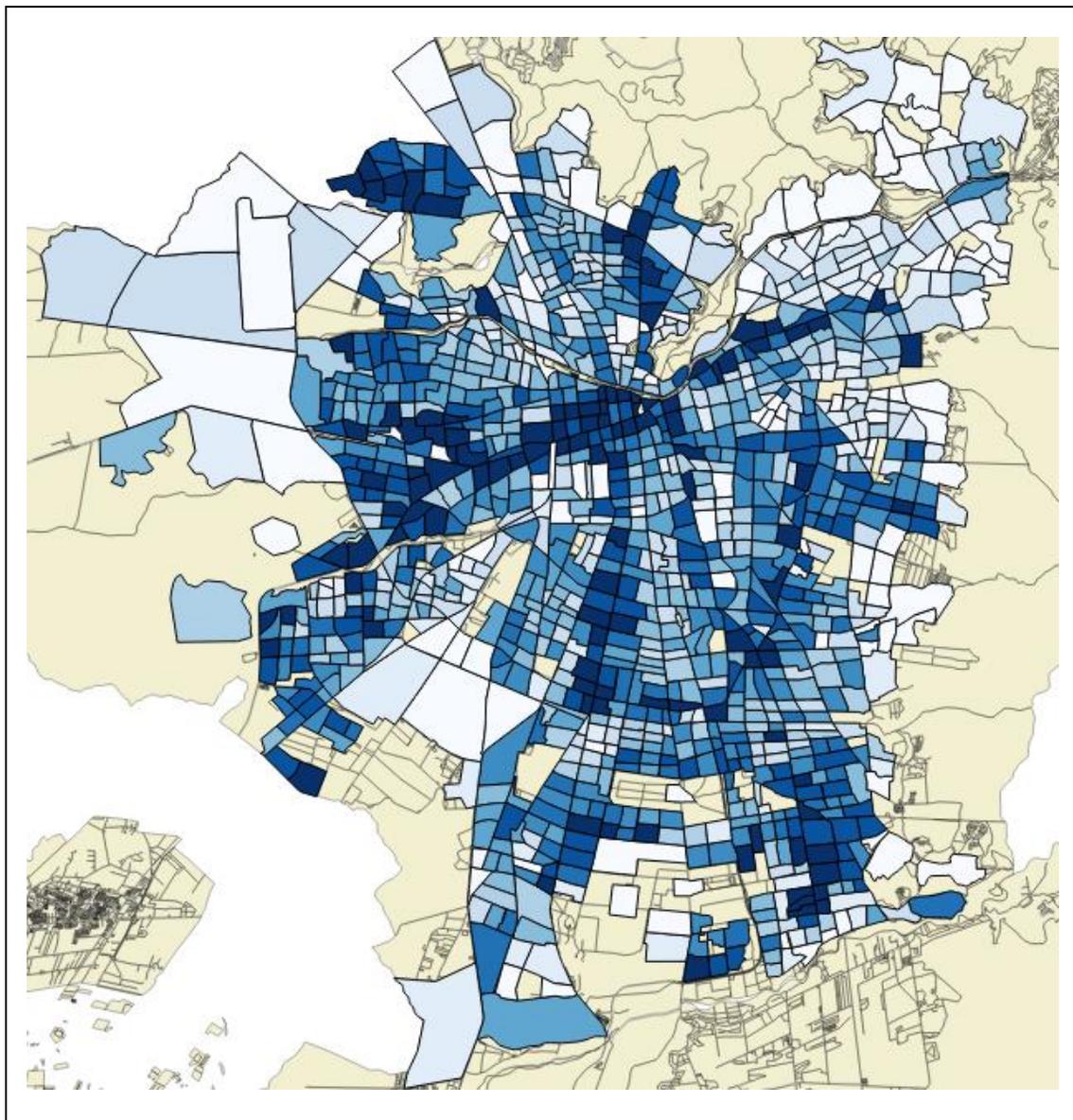


Figura 7-2: Totales de viajes generados por zona, según el modelo MNL5

Fuente: Elaboración propia en base a *QGIS*

Por su parte, la Figura 7-3 muestra un patrón espacial mucho más claro. En ella, observamos como las zonas del sector centro y oriente de la ciudad son las que atraen la mayor cantidad de viajes. Por el contrario, otros sectores periféricos más centrados en usos de suelo habitacionales presentan menor número de viajes (salvo algunas zonas puntuales

que concentran centros de atracción importantes como hospitales o centros comerciales o de oficinas).

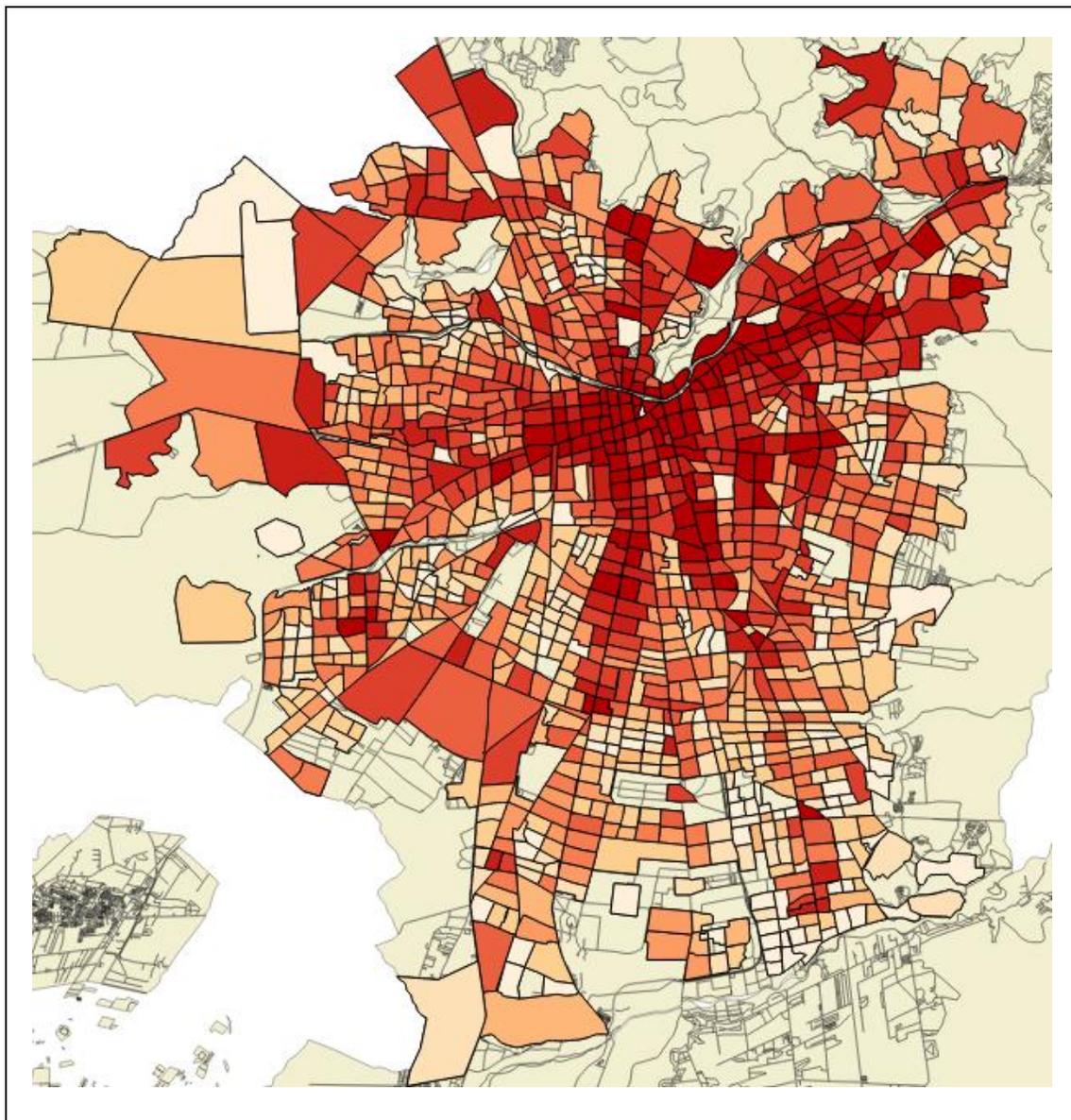


Figura 7-3: Totales de viajes atraídos por zona, según el modelo MNL5

Fuente: Elaboración propia en base a *QGIS*

Usando la clasificación de macro-zonas (según lo definido en la Figura C-2 del Anexo C), los totales de viajes generados y atraídos por cada una de ellas se presentan en la Figura 7-4, para el período de punta mañana (de 6:30 a 8:30 am).

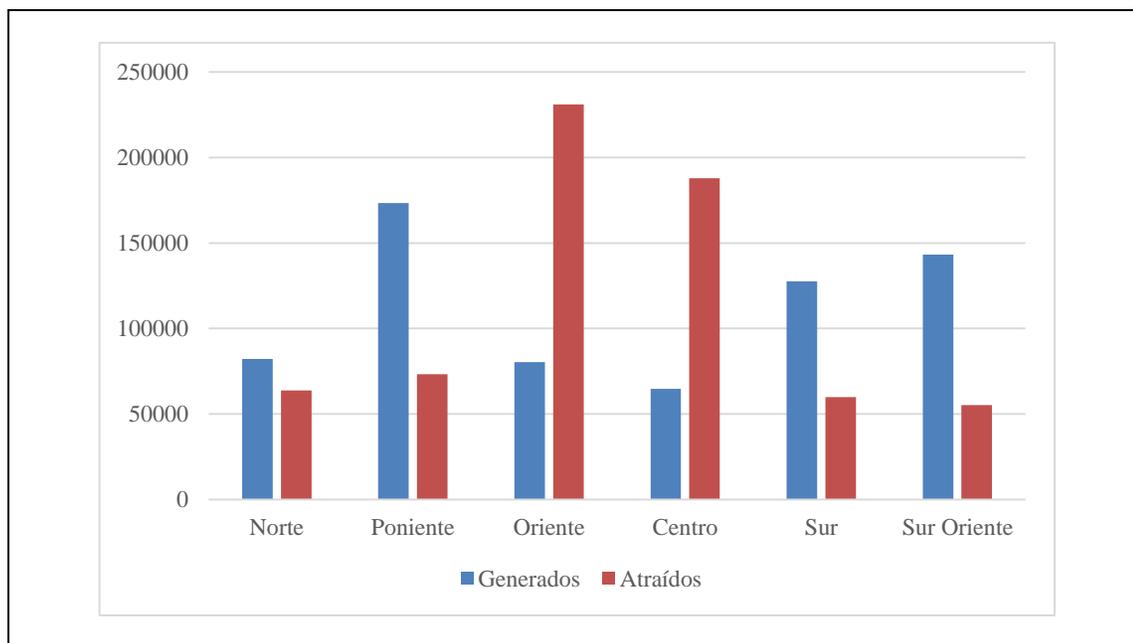


Figura 7-4: Número de viajes generados y atraídos por macro-zona, punta mañana

Fuente: Elaboración propia

Esta visualización confirma los resultados anteriores y la intuición general sobre los patrones de viajes en la ciudad, mostrando que los sectores Centro y Oriente atraen la mayor cantidad de viajes de la ciudad. Por su parte, los sectores Poniente, Sur y Sur Oriente son principalmente residenciales y generan la mayor parte de los viajes de Santiago durante la mañana. Finalmente, el sector Norte tiene usos de suelo menos definidos, presentando totales similares entre generación y atracción de viajes.

8 CONCLUSIONES Y RECOMENDACIONES

La presente investigación expone una metodología que permite estimar una matriz origen-destino zonal, a partir de una matriz de viajes entre paraderos de transporte público. Esta metodología es completamente general y podría ser aplicada a cualquier ciudad y período del día que cuente con la información suficiente.

Específicamente, para los alcances de esta tesis el modelo fue estimado para la red de Transantiago en el período de punta mañana. Si bien fue necesario unificar grandes cantidades de datos de distintas fuentes, todas ellas se actualizan regularmente pues son utilizadas con otros propósitos también. Por esto, la formulación presentada es significativamente más eficiente y económica para estimar matrices origen-destino zonales, que una encuesta origen-destino (EOD) a gran escala. Además, puede actualizarse con una mayor periodicidad a muy bajo costo.

Otra contribución relevante de este trabajo fue el desarrollo de una zonificación actualizada para la ciudad. En esta línea, cabe destacar la importancia de contar con suficiente información sobre la ciudad cuya zonificación se desee actualizar, lo que permite identificar automáticamente las zonas que requieren más atención para evaluar su división. En el caso de este estudio, se incluyó información de uso de suelo y relacionada con la red de transporte público, ayudando a detectar zonas con grandes cantidades de viajes y/o con muchas alternativas para realizarlos. Asimismo, la tecnología de imágenes satelitales permitió revisar y actualizar bordes zonales en que la mancha urbana se expandió considerablemente. Finalmente, es importante recalcar que para el desarrollo de estas tareas fue fundamental el apoyo interdisciplinario y manejo de *software* geográficos.

Por otro lado, este estudio contribuyó generando centroides y arcos de acceso para la zonificación propuesta, permitiendo estudiar los viajes tomando en consideración tiempos de caminata iniciales y finales. Esta tarea fue abordada con dos metodologías novedosas, ambas aprovechando información del comportamiento de los usuarios recopilada en la encuesta de preferencias reveladas. La primera de ellas, permitió definir qué paraderos de

bus y estaciones de metro se conectarían con qué zonas, para lo que se observaron los paraderos y estaciones de metro que usaban los encuestados. La segunda metodología, correspondiente a la determinación de la distancia asociada a estos arcos de acceso, se obtuvo con un procedimiento iterativo. Este algoritmo permitió aprovechar información sobre los potenciales de atracción y generación de cada tipo de uso de suelo (obtenidos de estimaciones preliminares del modelo de inferencia zonal) para encontrar tiempos de acceso que se condijeran con la distribución espacial de los distintos focos de viajes dentro de cada zona.

La inclusión de arcos de acceso, permite que las decisiones de los usuarios se puedan modelar de manera más precisa y posibilita obtener indicadores de desempeño del sistema de transporte para usar en evaluación social de proyectos. Antes del desarrollo de la red ficticia, se tornaba imposible valorar el impacto en los tiempos de caminata de acceso y egreso de alguna política de transporte. Por último, la metodología de cálculo de distancias de acceso podría extenderse para optimizar la posición de nuevos paraderos que se quieran agregar a la red, escogiendo ubicaciones que minimicen las caminatas de los usuarios.

Respecto a los resultados del modelo Logit definitivo de inferencia zonal, es posible afirmar que, *ceteris paribus*, las zonas residenciales tienen probabilidades más altas de ser el origen de un viaje en punta mañana. Similarmente, centros educacionales, comerciales, de salud y oficinas tienen poderes de atracción significativos en este período. En este sentido, el modelo propuesto permite capturar los efectos esperados del uso de suelo en la generación y atracción de viajes, concordando a su vez con lo reportado en la literatura.

Respecto a la función de costos del modelo, se destaca que los usuarios resultaron más sensibles a largas caminatas cuando estas ocurren cerca del origen de su viaje. Este fenómeno también fue capturado por el modelo de elección de rutas propuesto en Abud (2015), lo que muestra consistencia entre los resultados de ambas metodologías.

En cuanto a mejoras potenciales del modelo, se destaca la oportunidad de incluir diferenciaciones entre distintos tipos de tarjeta (regular o estudiante) en los viajes entre

paraderos cargados en la base de datos. Lo anterior sería beneficioso puesto que, como se demostró con modelos exploratorios, los usuarios de tarjetas de estudiantes son atraídos más específicamente a usos de suelo de tipo educacional (como colegios y universidades), diferencia que se podría aprovechar en las predicciones.

Adicionalmente, es esperable que la inclusión del modelo de elección de ruta y paradero en la función de costos tenga efectos determinantes en el proceso de asignación de viajes a zonas, asegurando que los paraderos cercanos atractivos sean predominantemente utilizados por los viajes originados en cada centroide. La inclusión de estas variables es un trabajo necesario y se encuentra actualmente en proceso, pero se espera que ayude a mejorar considerablemente al modelo de inferencia zonal.

En cuanto a las limitaciones de este estudio, es importante recalcar que no incluye explícitamente otros modos no integrados con la red de Transantiago (como taxis colectivos o automóviles privados). Por esto, el modelo no es capaz de rescatar el origen o destino real de viajes que usan estos modos en una o más etapas, en cambio, los asigna a zonas cercanas al lugar donde se conectaron con la red de transporte público.

De estos modos, probablemente el más urgente de incluir en la modelación es el taxi colectivo, puesto que tiene una partición modal importante que además se concentra en algunos sectores de la capital. Para incluirlo, se haría necesario recopilar e integrar a la herramienta computacional gran cantidad de información sobre los recorridos, frecuencias y demandas de viajes en la red de taxi colectivos. Lo anterior tendría que obtenerse de estudios origen-destino previos o podría llegar a ser extremadamente costoso, puesto que no existe suficiente tecnología implementada en los vehículos para posibilitar una recolección automática de información.

Existen también limitaciones relativas a los datos de encuestas usados para la estimación del modelo de inferencia zonal. En primer lugar, la encuesta fue realizada en un período más largo que la punta mañana propiamente tal, lo que puede llevar a peores estimadores en los parámetros del modelo puesto que los patrones de viajes difieren en el período fuera

de punta; sin embargo, la estimación de modelos diferenciando viajes realizados fuera de punta presenta resultados prometedores, por lo que su inclusión en la herramienta FONDEF será discutida en reuniones próximas del proyecto. Además, es posible que centros de salud y educacionales se encuentren sobrerrepresentados como destinos de viajes, puesto que fue más fácil encontrarlos en el mapa y asignarlos unívocamente a una zona, por lo que se perdió una menor proporción de ellos.

En este sentido, una encuesta especialmente diseñada para los propósitos de este estudio podría mejorar la estimación del modelo de inferencia zonal. En primer lugar, la encuesta debiese ser realizada exclusivamente en el período para el que interesa calibrar el modelo. Además, los lugares de realización de la encuestas podrían ser definidos considerando la zonificación que se usará para la estimación, para así capacitar específicamente a los encuestadores sobre los límites zonales cercanos y que se preocupen de indagar información suficiente para encontrar las zonas de origen de manera unívoca. Asimismo, conocer la zonificación de los principales focos de atracción de viajes (sector centro y oriente para el caso de Santiago) permitiría a los encuestadores enfocar mejor sus preguntas para obtener destinos más precisos. Por último, el cuestionario de la encuesta incluía un gran número de preguntas que no resultan necesarias para los efectos de este modelo, aumentando el tiempo de respuesta y la cantidad de datos perdidos.

No obstante lo anterior, la metodología propuesta demostró ser capaz de estimar un modelo estadísticamente significativo que rescata los efectos esperados de las distintas variables, concordando con lo reportado en la literatura de distribución de viajes y permitiendo la estimación de una matriz origen-destino zonal de transporte público para Santiago.

Por último, destacan las numerosas implicancias de política de esta investigación. Como se mencionó previamente, la matriz origen-destino zonal resultante refleja directamente las necesidades de transporte público de los usuarios, permitiendo una mejor planificación de la red de la ciudad.

En el caso de Santiago, la inclusión de la matriz zonal a la herramienta de planificación del proyecto FONDEF permite evaluar modificaciones mayores al sistema de transporte, tales como la creación de nuevas líneas de Metro, corredores de buses y otros. Previamente, proyectos cuyos impactos modificasen los paraderos o estaciones escogidas por los usuarios para empezar y terminar sus viajes no podían ser evaluados correctamente, puesto que la asignación de flujos se realizaba desde una matriz de viajes entre paraderos y, por ejemplo, no se sabría cuántos viajes empezarían y terminarían en cada nueva estación de Metro.

De esta manera, la metodología y modelos propuestos en esta tesis se convierten en una pieza clave del *software* de planificación para la ciudad de Santiago, posibilitando la realización de análisis de viajes en la ciudad en un nuevo nivel, entregando flexibilidad y mejorando las predicciones.

BIBLIOGRAFIA

Abud, I., (2015). *Modelos de Elección de Paradero, Modo y Ruta para Herramientas de Planificación de Transporte Público*. Tesis de Magíster en Ciencias de la Ingeniería, Pontificia Universidad Católica de Chile.

Amaya, M., (2013). *Análisis de Patrones de Viaje Utilizando Datos Masivos de Transporte Público*. Tesis de Magíster en Ciencias de la Ingeniería, mención Transporte, Universidad de Chile.

Amaya, M., & Munizaga, M. (2013). Estimación de zona de residencia en base a sistemas tecnológicos de transporte público. *Actas del XVI Congreso Chileno de Ingeniería de Transporte, Santiago, Chile*.

Bell, M. G. (1983). The estimation of an origin-destination matrix from traffic counts. *Transportation Science*, 17(2), 198-217.

Ben-Akiva, M. E. (1987). Methods to combine different data sources and estimate origin-destination matrices. *International Symposium on Transportation and Traffic Theory*, Berkeley, California.

Bierlaire, M., & Toint, P. L. (1995). Meuse: an origin-destination matrix estimator that exploits structure. *Transportation Research Part B: Methodological*, 29(1), 47-60.

Bierlaire, M. (2003). *BIOGEME*: A free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Suiza.

Bohannon, R. W. (1997). Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. *Age and Ageing*, 26(1), 15-19.

Chriqui, C., & Robillard, P. (1975). Common bus lines. *Transportation Science*, 9(2), 115-121.

Cochrane, R. A. (1975). A possible economic basis for the gravity model. *Journal of Transport Economics and Policy*, 9, 34-49.

Cools, M., Moons, E., & Wets, G. (2010). Assessing the quality of origin-destination matrices derived from activity travel surveys. *Transportation Research Record: Journal of the Transportation Research Board*, 2183, 49-59.

Daly, A. J. (1979). Some developments in transport demand modelling. In *Behavioural Travel Modelling* (Edited by D.A. Hensher y P.R. Stopher). Croom Helm, Londres.

Daly, A. J. (1982). Estimating choice models containing attraction variables. *Transportation Research Part B: Methodological*, 16(1), 5-15.

Devillaine, F., (2012). *Estimación de Viajes y Actividades en Base a Sistemas Tecnológicos de Transporte Público*. Tesis de Magíster en Ciencias de la Ingeniería, mención Transporte, Universidad de Chile.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(3), 269-271.

DTPM (2014) *Informe de Gestión 2013*. Directorio de Transporte Público Metropolitano. Recuperado: 10 de febrero, 2015. [http://www.dtpm.cl/archivos/Memoria%202013-Web_Final%20\(1\).pdf](http://www.dtpm.cl/archivos/Memoria%202013-Web_Final%20(1).pdf)

Domencich, T. & McFadden, D. (1975). *Urban Travel Demand: A Behavioural Analysis*. North Holland, Amsterdam.

Erlander, S., & Stewart, N. F. (1990). *The Gravity Model in Transportation Analysis: Theory and Extensions*. Brill Academic Publishers, Utrech.

Fernández, J. E. & De Cea, J. (2005). *Análisis y actualización del modelo ESTR AUS, Informe Ejecutivo*. Secretaría de Planificación de Transporte, Santiago.

Fisk, C. S. (1989). Trip matrix estimation from link traffic counts: the congested network case. *Transportation Research Part B: Methodological*, 23(5), 331-336.

Florian, M., & Chen, Y. (1995). A coordinate descent method for the bi-level O-D matrix adjustment problem. *International Transactions in Operational Research*, 2(2), 165-179.

Gordon, J. B., Koutsopoulos, H. N., Wilson, N. H. M. & Attanucci, J. P. (2013). Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. *Transportation Research Record: Journal of the Transportation Research Board*, 2343 (1), 17-24.

INE (2002). *Informe Final de la Comisión Externa Revisora del Censo 2002*. Instituto Nacional de Estadísticas, Santiago.

INE (2014). *Población País y Regiones: Actualización 2002-2012 y Proyecciones 2013-2020*. Instituto Nacional de Estadísticas, Santiago.

Jara-Díaz, S., Munizaga, M., & Olgún, J. (2013). The role of gender, age and location in the values of work behind time use patterns in Santiago, Chile. *Papers in Regional Science*, 92(1), 87-102.

- Lam, W. H., Wu, Z. X., & Chan, K. S. (2003). Estimation of transit origin–destination matrices from passenger counts using a frequency-based approach. *Journal of Mathematical Modelling and Algorithms*, 2(4), 329-348.
- Maher, M. J., Zhang, X., & Vliet, D. V. (2001). A bi-level programming approach for trip matrix estimation and traffic control problems with stochastic user equilibrium link flows. *Transportation Research Part B: Methodological*, 35(1), 23-40.
- Mayberry, J. P. (1973). Structural requirements for abstract-mode models of passenger transportation. En R.E. Quandt (ed.), *The Demand for Travel: Theory and Measurement*. D.C. Heath and Co., Lexington, Mass.
- McFadden, D. (1979). Quantative methods for analysing travel behaviour of individuals. In *Behavioural Travel Modelling* (Edited by D.A. Hensher and P.R. Stopher). Croom Helm, London.
- Metro (2011). “*Encuesta Origen Destino de Viajes Geo-referenciada en Metro, año 2011*”. Metro S.A., Santiago.
- MIDEPLAN (2010). *Análisis y Formulación de Nuevos Modelos de Generación y Atracción de viajes*. Ministerio de Planificación y Cooperación, Santiago.
- Munizaga, M., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.
- Munizaga, M., Devillaine, F., Navarrete, C., & Silva, D. (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, 44, 70-79.
- Muñoz, J. C. & Gschwender, A. (2008). Transantiago: a tale of two cities. *Research in Transportation Economics*, 22(1), 45-53.
- Ortúzar, J. de D. (2000). *Modelos Económicos de Elección Discreta*. Ediciones Universidad Católica de Chile, Santiago.
- Ortúzar, J. de D. & Willumsen, L. G. (2011). *Modelling Transport* (Fourth Edi.). John Wiley & Sons, Chichester.
- Raveau, S., Muñoz, J. C., & De Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A: Policy and Practice*, 45(2), 138-147.

SECTRA (2008). *Informe Final del estudio “Actualización y Recolección de Información del Sistema de Transporte Urbano, III Etapa”*. Secretaría Interministerial de Planificación de Transporte, Santiago.

SII (2014). *Roles de la Región Metropolitana de la Segunda Serie No Agrícola y Superficie Construida*. Datos obtenidos en virtud de la Ley Nro. 20.285. Servicio de Impuestos Internos, Santiago.

Van Zuylen, H. J., & Willumsen, L. G. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, 14(3), 281-293.

Williams, H.C.W.L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning*, 9A, 285–344.

Yang, H., Meng, Q., & Bell, M. G. (2001). Simultaneous estimation of the origin-destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium. *Transportation Science*, 35(2), 107-123.

ANEXOS

ANEXO A: ESTIMACIÓN POR MINIMIZACIÓN DE ERRORES CUADRÁTICOS

Una aproximación alternativa para calibrar el modelo de inferencia zonal es combinar la información de viajes observada T_{kl} con un modelo de elección de rutas, con el objetivo de identificar restricciones que la matriz origen-destino buscada debiese cumplir y estimar el modelo de inferencia zonal en concordancia. Este procedimiento de estimación está relacionado a Lam et al. (2003), puesto que en ambos casos el comportamiento observado de los usuarios es combinado con modelos de elección para estimar matrices O-D. Sin embargo, la metodología aquí expuesta se diferencia puesto que usa los datos para calibrar los parámetros del modelo de inferencia zonal en vez de encontrar directamente una matriz O-D, lo que permitiría obtener un modelo predictivo que puede ser utilizado para planificación futura.

Adicionalmente, existen algunas diferencias clave en los datos disponibles, que justifican esta nueva formulación. Para esta tesis, asumimos en primer lugar que no se cuenta con una matriz O-D *a priori* confiable. Por otro lado, asumimos que se tiene información pasiva basada en validaciones de tarjetas inteligentes, lo que representa una oportunidad de explotar los conteos de flujo de una manera más detallada. Por último, el modelo de elección de rutas usado en Lam et al. (2003) es relativamente simple y puede ser mejorado con la inclusión de variables no tradicionales.

Para deducir esta metodología alternativa, en primer lugar debemos definir la probabilidad de que un viaje originado en la zona i con destino en la zona j use los paraderos k y l (como sus puntos de subida y bajada, respectivamente) como $Prob\left(\frac{kl}{ij}\right)$.

Esta probabilidad se encuentra implícita en cualquier modelo de elección de rutas entre zonas, como muestra la Ecuación (A.1):

$$Prob(kl/ij) = \sum_{\forall r:(k=i(r) \& l=f(r)) \in R_{ij}} Prob(r/ij) \quad (A.1)$$

Donde $Prob(r/ij)$ representa la probabilidad de que un pasajero que viaja de la zona i a la j use la ruta r (del conjunto de rutas R_{ij}); mientras que $i(r)$ y $l(r)$ simbolizan las estaciones de acceso iniciales y finales para la ruta r , respectivamente.

Luego, siguiendo las definiciones del capítulo 3, cada elemento T_{ij} de la matriz O-D buscada debe cumplir:

$$T_{ij}^{kl} = Prob(kl/ij) * T_{ij} \quad \forall (i,j) \wedge (k,l) \quad (A.2)$$

Así, si sumamos para todo i y j obtenemos:

$$\sum_i \sum_j T_{ij}^{kl} = \sum_i \sum_j Prob(kl/ij) * T_{ij} \quad \forall (k,l) \quad (A.3)$$

Y reemplazando las Ecuaciones (3.2) en (A.3):

$$T_{kl} = \sum_i \sum_j Prob(kl/ij) * T_{ij} \quad \forall (k,l) \quad (A.4)$$

De esta manera, dedujimos un sistema de v ecuaciones que la matriz buscada debiese cumplir, donde v equivale al cuadrado de los paraderos de transporte público para toda la red. Ahora, si asumimos que T_{kl} y $Prob(kl/ij)$ son conocidos (del comportamiento de viaje observado por las tarjetas inteligentes y el modelo de elección de rutas, respectivamente) y reemplazamos T_{ij} con la formulación del modelo de inferencia zonal

(de las Ecuaciones (3.7)) con parámetros desconocidos θ , obtenemos un sistema de v ecuaciones con w incógnitas (con w igual al número de parámetros θ).

Por supuesto, en cualquier formulación del modelo v será mucho mayor que w . Esto resultaría en un sistema de ecuaciones sobredeterminado sin soluciones factibles, puesto que ni la información pasiva, ni el modelo de elección de ruta, ni el modelo de inferencia zonal son perfectamente consistentes entre sí y podrían contener errores.

Sin embargo, podemos utilizar este sistema de ecuaciones para estimar el modelo y encontrar el conjunto de parámetros θ que minimizan un término de error ε_{kl} agregado a las Ecuaciones (A.4), resolviendo el problema de optimización definido por las Ecuaciones (A.5) y (A.6), descrito a continuación:

$$\underset{\{\theta, \varepsilon_{kl}\}}{\text{minimizar}} \sum_{k,l} \varepsilon_{kl}^2 \quad (\text{A.5})$$

Sujeto a:

$$T_{kl} = \sum_i \sum_j \text{Prob}(kl/ij) * T_{ij}(\theta) + \varepsilon_{kl} \quad \forall(k, l) \quad (\text{A.6})$$

Otra alternativa es considerar solo un subconjunto de paraderos como restricciones en las Ecuaciones (A.6) o crear algún tipo de agregación entre ellas; con el objetivo de reducir el número de restricciones y la variabilidad diaria o estacional de las observaciones T_{kl} .

ANEXO B: ENCUESTA DE PREFERENCIAS REVELADAS

En la siguiente sección, se presenta el formulario utilizado para la encuesta en paraderos y estaciones (usada en la estimación del modelo de inferencia zonal), así como mapas detallados de los sectores seleccionados para su realización.

a. Cuestionario definitivo



Pregunta filtro: ¿Este va a ser su primer bip?

1. ¿Llegó hasta aquí caminando o usó locomoción?

Caminando Transantiago

Auto/Colectivo

¿Cuál servicio/línea? _____

Folio _____ Fecha _____

Estación _____ Hora _____

Encuestador _____

2. ¿Viene de su casa? Sí No **¿De dónde?** *(utilizar en pregunta 3)*

3. ¿Dónde queda (lugar inicio)? *(dirección, intersección, edificio, hito, otro)*

Calle _____

Con calle _____ Comuna _____

4. ¿Cuál es el motivo de su viaje?

Trabajo

Estudio

Trámites

Compras

Salud

Otro, ¿Cuál? _____

5. ¿Dónde queda (lugar destino)? *(dirección, intersección, edificio, hito, otro)*

Calle _____

Con calle _____ Comuna _____

6. ¿Qué tan seguido hace este viaje?

Todos los días Cada dos semanas

Dos a tres veces por semana Una vez al mes

Una vez por semana Excepcionalmente

(Omitir preguntas 7, 11: sentado y deja pasar y 17)

7. ¿Cuánto demoró en su viaje completo la última vez que lo hizo? _____ horas _____ minutos

8. ¿Por qué eligió este paradero/estación? *(Puede ser más de una)*

Cercanía Frecuencia

Oferta de servicios/líneas Otro, ¿Cuál? _____

9. ¿En este paradero, le sirve otra línea o servicio? No Sí ¿Cuáles? _____

10. De las líneas/servicios que le sirven ¿Espera uno específico? ¿Toma el primero que pasa?

11. ¿Por favor, describame su viaje completo? La última vez:

Servicio(s):	hasta:	Sentado	Dejó pasar buses			
_____	_____	<input type="checkbox"/> Sí	<input type="checkbox"/> Sí	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
ó	_____	<input type="checkbox"/> No				
Servicio(s):	hasta:	<input type="checkbox"/> Sí	<input type="checkbox"/> Sí	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
ó	_____	<input type="checkbox"/> No				
Servicio(s):	hasta:	<input type="checkbox"/> Sí	<input type="checkbox"/> Sí	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
ó	_____	<input type="checkbox"/> No				
Servicio(s):	hasta:	<input type="checkbox"/> Sí	<input type="checkbox"/> Sí	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3
ó	_____	<input type="checkbox"/> No				

12. Tipo de Pasaje

Normal Punta

Estudiante Válie

3ª Edad

13. Edad _____ *(Si no responde, estimar edad)*

15. ¿Puede indicarme su rango de ingreso personal?

No trabaja Part time

No responde Jubilado

14. Género Hombre Mujer *(No se pregunta)*

16. ¿Me indicaría el número de su tarjeta bip!? _____ No responde

Figura A-1: Cuestionario final encuesta en paraderos, hoja 1

Fuente: Abud (2015)

17. ¿La última vez que hizo este viaje lo hizo de la misma manera? No Sí

18. ¿Le sirven otros paraderos del sector? No Sí ¿Cuáles?

19. Si ya no existiera este paradero ¿qué haría?

¿por qué descartó esta alternativa?

20. Si ya no existiera la(s) ...*(línea(s) que usa)*... ¿cómo haría su viaje?

¿por qué descartó esta alternativa?

21. ¿De qué otra forma podría realizar el viaje?

¿por qué descartó esta alternativa?

IMPORTANTE
 Si responde cosas como: "iría al paradero de al lado y tomaría la misma micro, o "no sé que haría"
 Inducir al diálogo utilizando preguntas como:
 ¿Y aunque te salga más largo, conoces otra alternativa?
 ¿Y aunque tú no la uses, te serviría ...?
 ¿Qué más haría, qué servicios y hasta dónde?
 ¿Y no has intentado irte por ...?

} Si explica razones de por qué no utiliza alternativas dejarlas registradas

Comentarios

Figura A-2: Cuestionario final encuesta en paraderos, hoja 2

Fuente: Abud (2015)

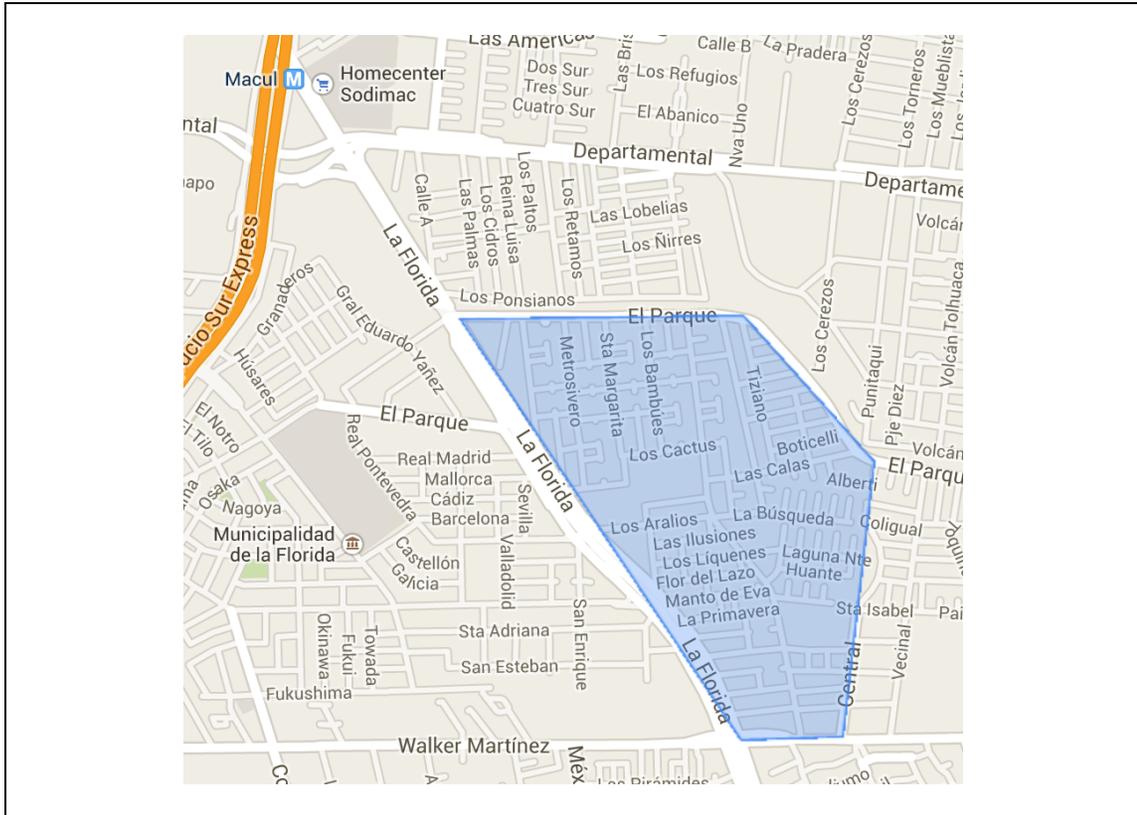


Figura A-5: Mapa de la zona encuestada en La Florida

Fuente: Abud (2015)



Figura A-6: Mapa de la zona encuestada en Ñuñoa

Fuente: Abud (2015)

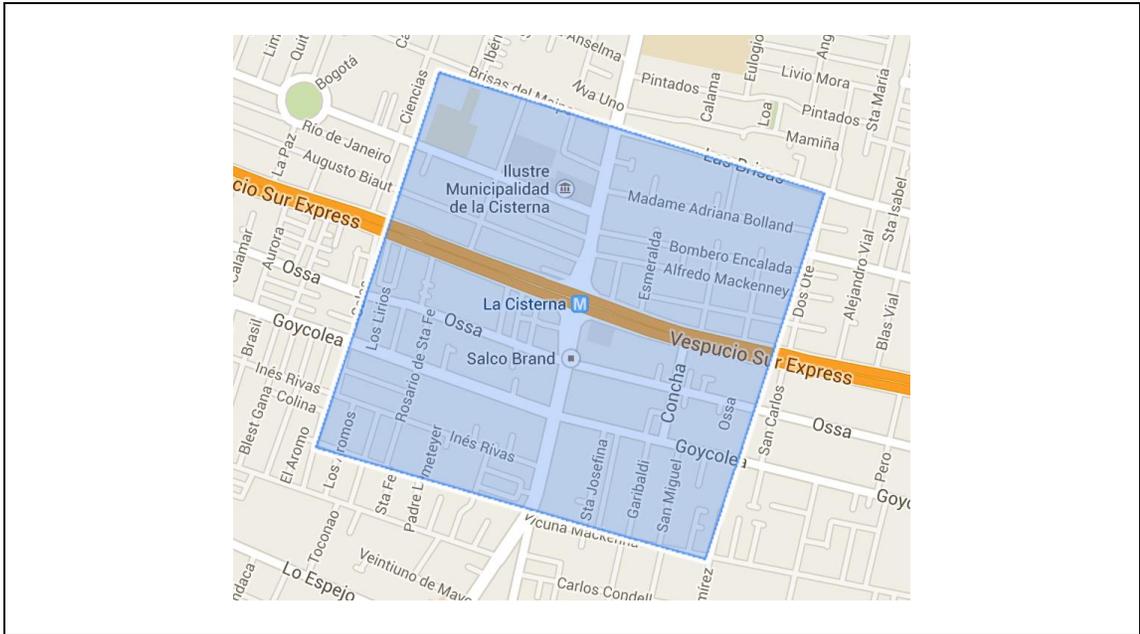


Figura A-7: Mapa de la zona encuestada en Metro La Cisterna

Fuente: Abud (2015)

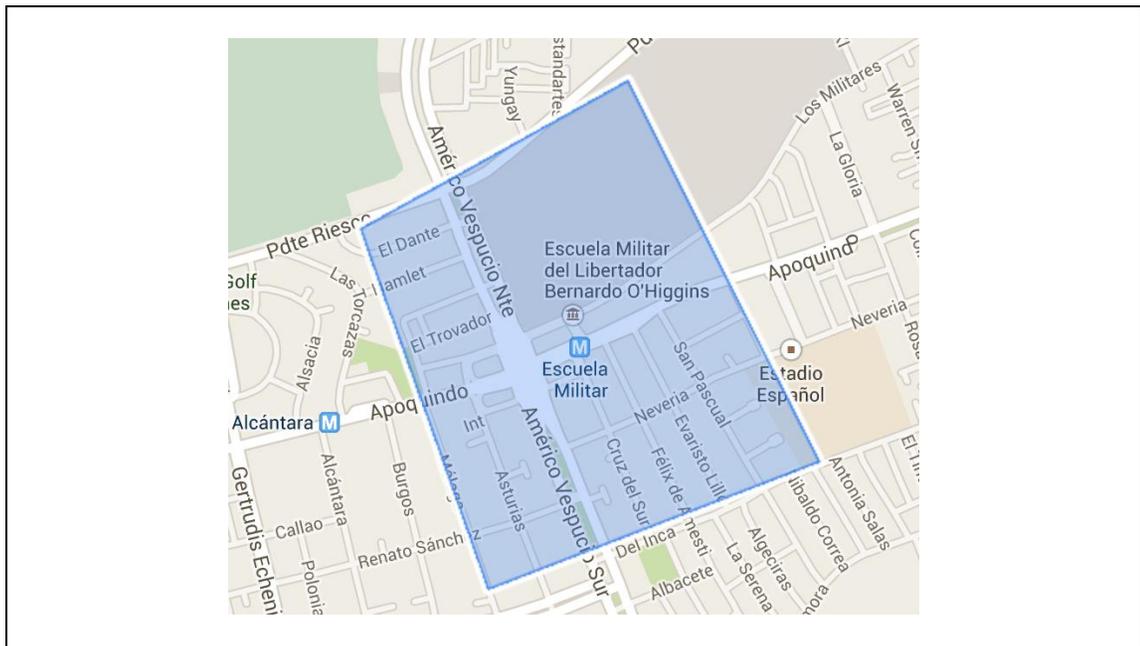


Figura A-8: Mapa de la zona encuestada en Metro Escuela Militar

Fuente: Abud (2015)

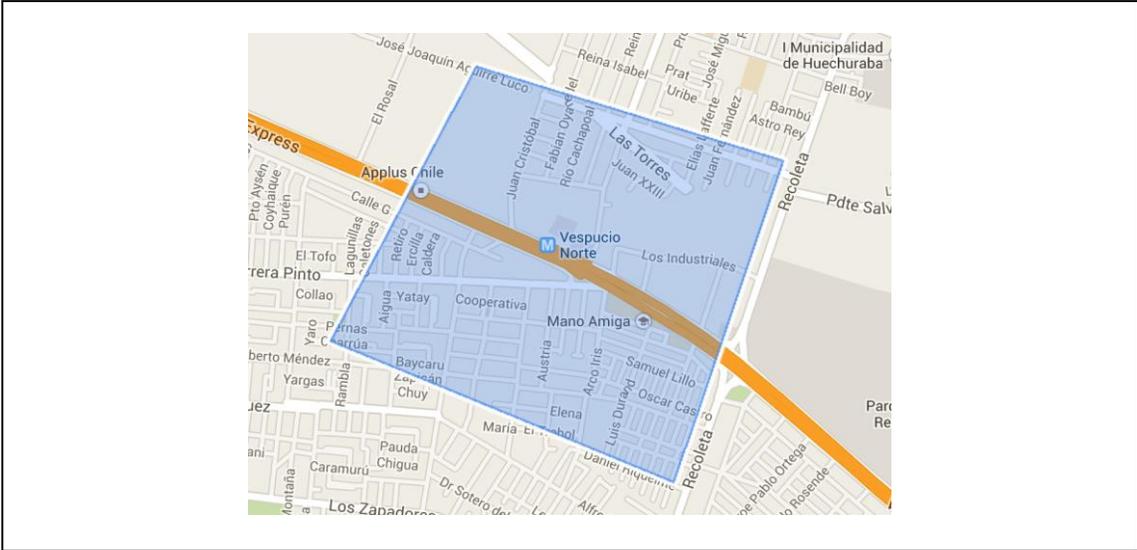


Figura A-9: Mapa de la zona encuestada en Metro Vespucio Norte

Fuente: Abud (2015)

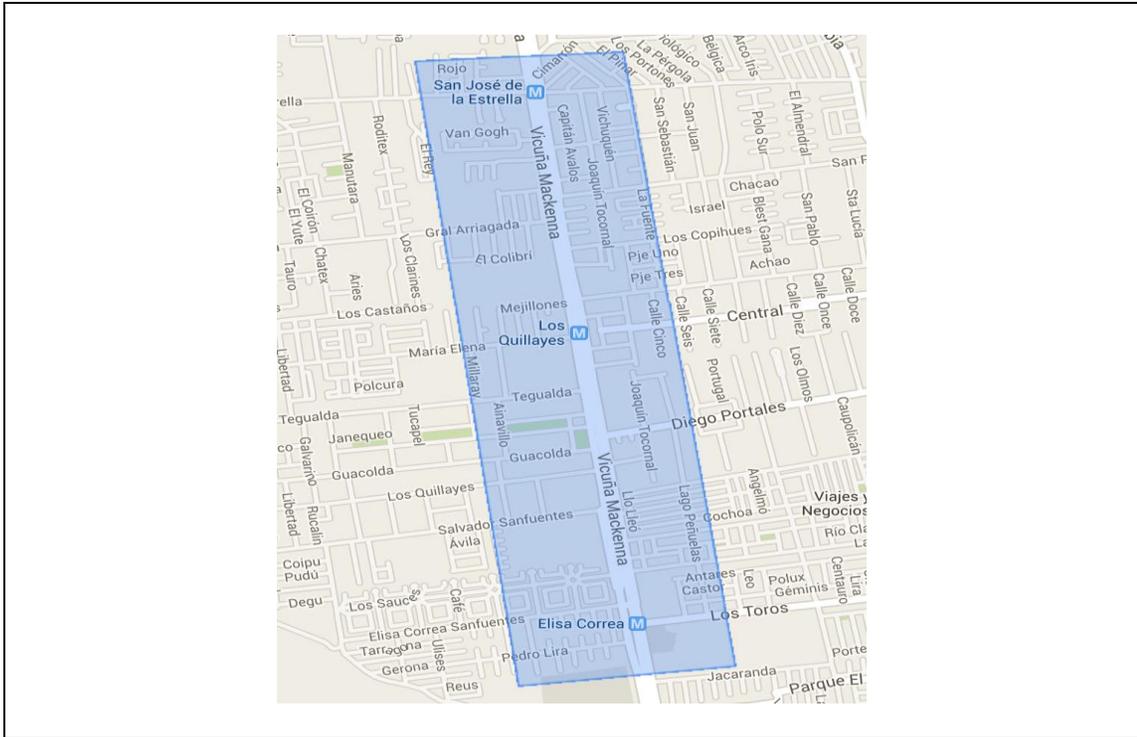


Figura A-10: Mapa de la zona encuestada en Metros Elisa Correa, Los Quillayes y San José de la Estrella

Fuente: Abud (2015)

ANEXO C: ZONIFICACIÓN

En el presente anexo se detalla la zonificación definitiva implementada en la herramienta del proyecto, presentando visualmente sus características y explicando los distintos pasos que llevaron a su finalización.

En primer lugar, cabe recordar que la zonificación propuesta se basa en SECTRA (2008), sistema que contaba con 779 zonas para la Región Metropolitana. Sin embargo, no todas ellas resultaban atinentes al proyecto pues incluyen zonas rurales no conectadas con paraderos de la red de Transantiago. Además, algunos de sus bordes se encontraban desactualizados pues fueron definidos el 2006, y la mancha urbana ha sufrido modificaciones considerables desde entonces, sobre todo en sectores periféricos.

En vista de lo anterior, la primera tarea consistió en eliminar zonas no conectadas con la red de transporte público y actualizar los bordes de zonas en que la mancha urbana haya crecido considerablemente. De esta manera, se aislaron y eliminaron 49 zonas no urbanas, para luego actualizar los bordes de las zonas restantes.

A continuación, dados los requerimientos del proyecto, se realizó una revisión completa del sistema zonal remanente, con el objetivo de densificar las divisiones zonales para permitir un análisis más detallado de los viajes de la ciudad. En esta línea, se siguieron los criterios definidos en el subcapítulo 4.4 y se dividieron zonas atractivas o con un alto poder de generación de viajes, así como también aquellas que poseían una oferta de transporte excesiva.

Un ejemplo de división se presenta en la Figura C-1, donde inicialmente existían tres grandes zonas, definidas por los colores verde, azul y naranja. En este caso, se identificó que las zonas eran excesivamente grandes para su importancia y la oferta de transporte que poseían (en especial las dos zonas superiores), pues por ejemplo la zona del norte quedaría conectada con cuatro estaciones de metro, lo que se considera excesivo pues difícilmente un usuario considera más de dos estaciones de una misma línea como alternativas válidas. Además, esta zona incluía el Club de Golf Los Leones, que

prácticamente no atrae viajes en el período analizado. En vista de lo anterior, estas tres zonas iniciales fueron divididas en 6 zonas de modelación y una zona vacía, correspondiente al parque.

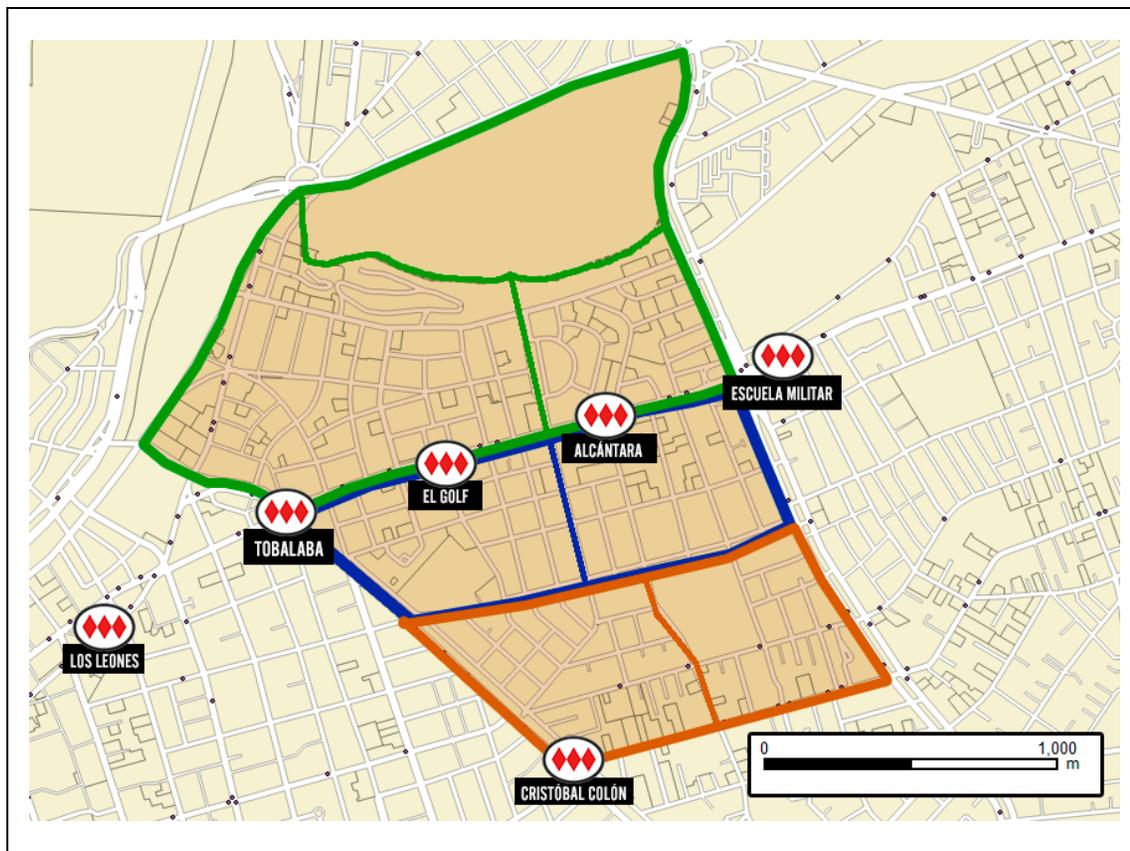


Figura C-1: Ejemplo partición de zonas sector Tobalaba – Escuela Militar

Fuente: Elaboración propia en base a *QGIS*

Una vez terminado este proceso, se fragmentaron 343 zonas generando 485 nuevas zonas, lo que significaba un total de 1.215 zonas. Sin embargo, este número excedía las limitaciones impuestas al número máximo de zonas, por lo que se realizó una nueva revisión con el objetivo de reagrupar zonas excesivamente pequeñas y/o poco atractivas, finalizando con 1.176 zonas.

Finalmente, tras la aplicación de modelos de inferencia zonal iniciales y el análisis de sus resultados, se identificaron nuevas zonas consideradas sobrantes que fueron eliminadas, por lo que la zonificación definitiva del proyecto cuenta con 1.171 zonas.

La Figura C-2, presenta un mapa de la zonificación de Santiago, distinguiendo en colores las seis macro-zonas de la ciudad. Variando desde tonos más claros a más oscuros, se distinguen los sectores Norte, Poniente, Oriente, Centro, Sur y Sur Oriente.

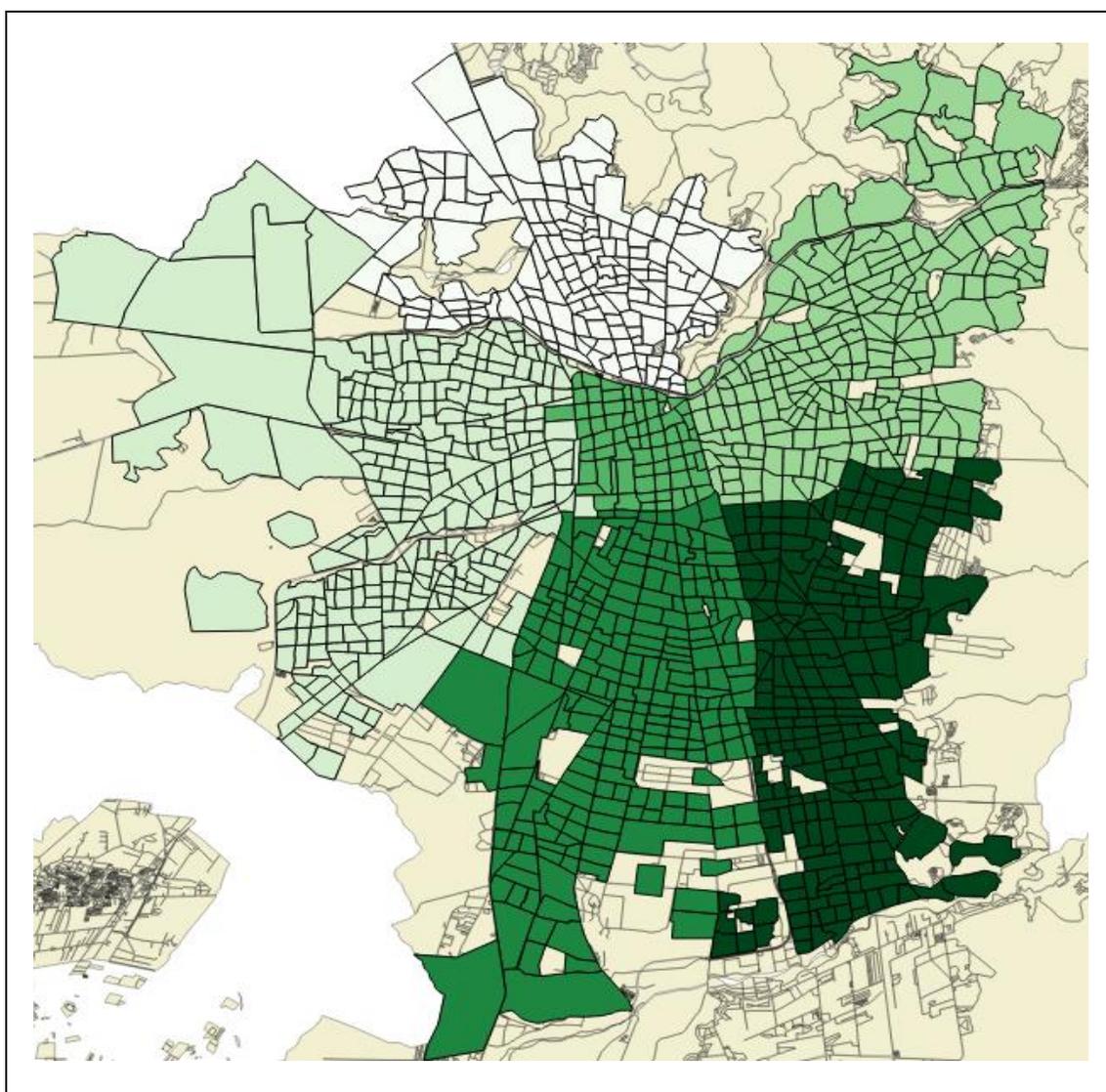


Figura C-2: Zonificación definitiva, distinguiendo en colores las seis macro-zonas

Fuente: Elaboración propia en base a *QGIS*

ANEXO D: EJEMPLOS DE APLICACIÓN DEL MODELO DEFINITIVO

Utilizando el modelo de inferencia zonal definitivo (correspondiente al modelo MNL5, presentado en la Tabla 7-3), esta sección presenta ejemplos de asignaciones de viajes desde algunas estaciones de metro a sus zonas aledañas, para el período de punta mañana.

Para ello, cabe destacar que la forma funcional del modelo permite separar las utilidades asociadas a origen y destino, haciendo posible el análisis individual de generación y atracción para cada estación.

En esta línea, la Figura D-1 presenta las probabilidades de que un viaje que termina en la estación Tobalaba, se dirija a cada una de las zonas conectadas a ella con arcos de acceso (pintadas con color amarillo).

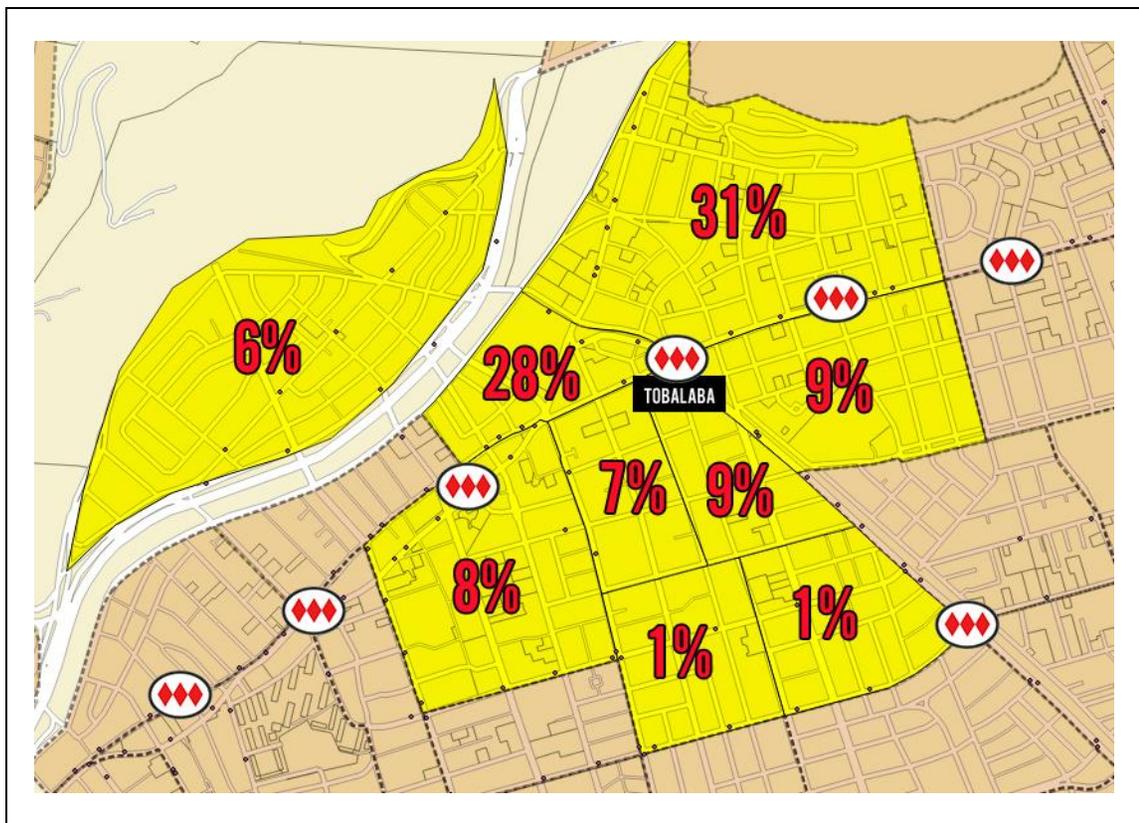


Figura D-1: Ejemplo asignación de viajes a zonas, destino estación Tobalaba

Fuente: Elaboración propia en base a *QGIS*

Es posible observar de este mapa que la mayoría de los viajes fueron asignados al sector norte de las cercanías de la estación, donde se concentran oficinas y sectores comerciales. Asimismo, las zonas del sur poseen una menor cantidad de viajes atraídos, puesto que su uso es mayormente residencial.

Análogamente, la Figura D-2 muestra este análisis pero asignando los viajes generados en la estación a sus potenciales zonas de origen.

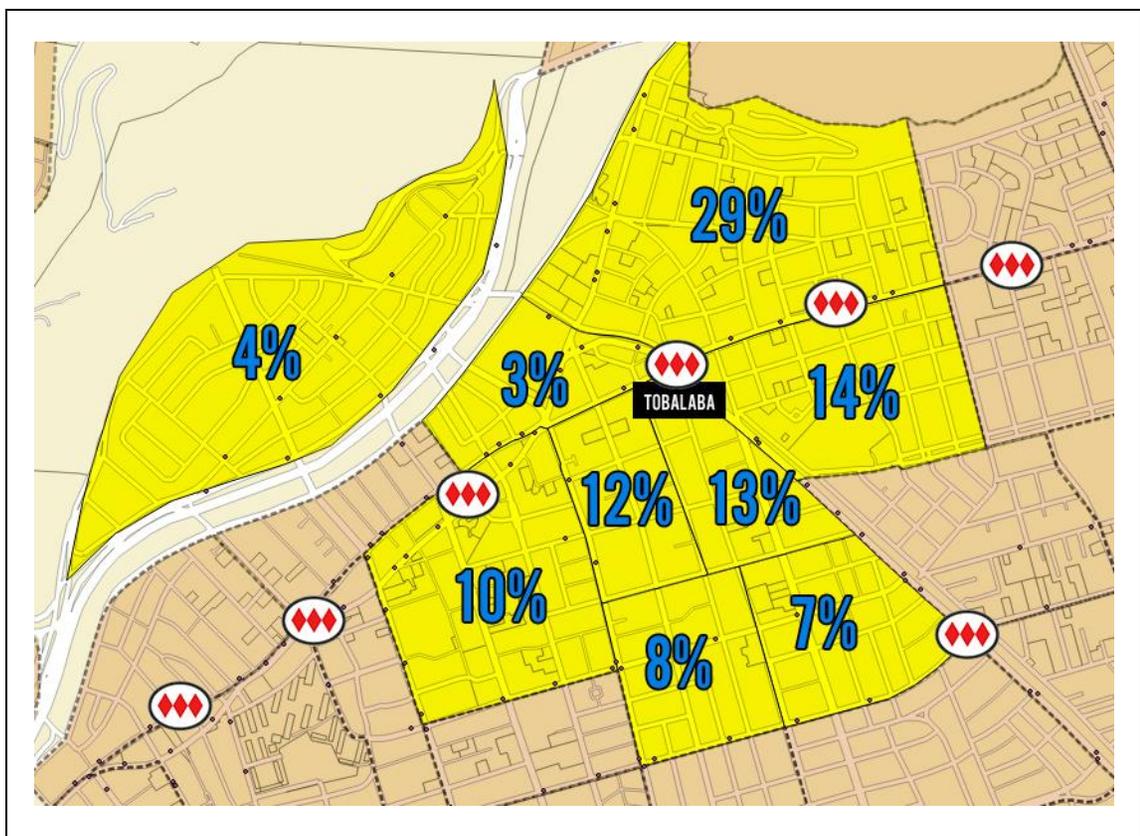


Figura D-2: Ejemplo asignación de viajes a zonas, origen estación Tobalaba

Fuente: Elaboración propia en base a *QGIS*

En cuanto a los viajes originados en la estación, se observa una distribución bastante distinta entre las zonas aledañas. En particular, se observa que las zonas ubicadas al sur del eje Providencia – Apoquindo toman una importancia mayor, concentrando la mayoría de los viajes. Además, resulta especialmente interesante la gran diferencia entre

generación y atracción de la zona ubicada inmediatamente al noroeste de la estación, explicable por la baja superficie habitacional construida en ella.

Siguiendo la línea anterior, se aplicó también el modelo de inferencia zonal a los viajes que terminan en la estación de metro San Joaquín, obteniéndose los resultados que presenta la Figura D-3.

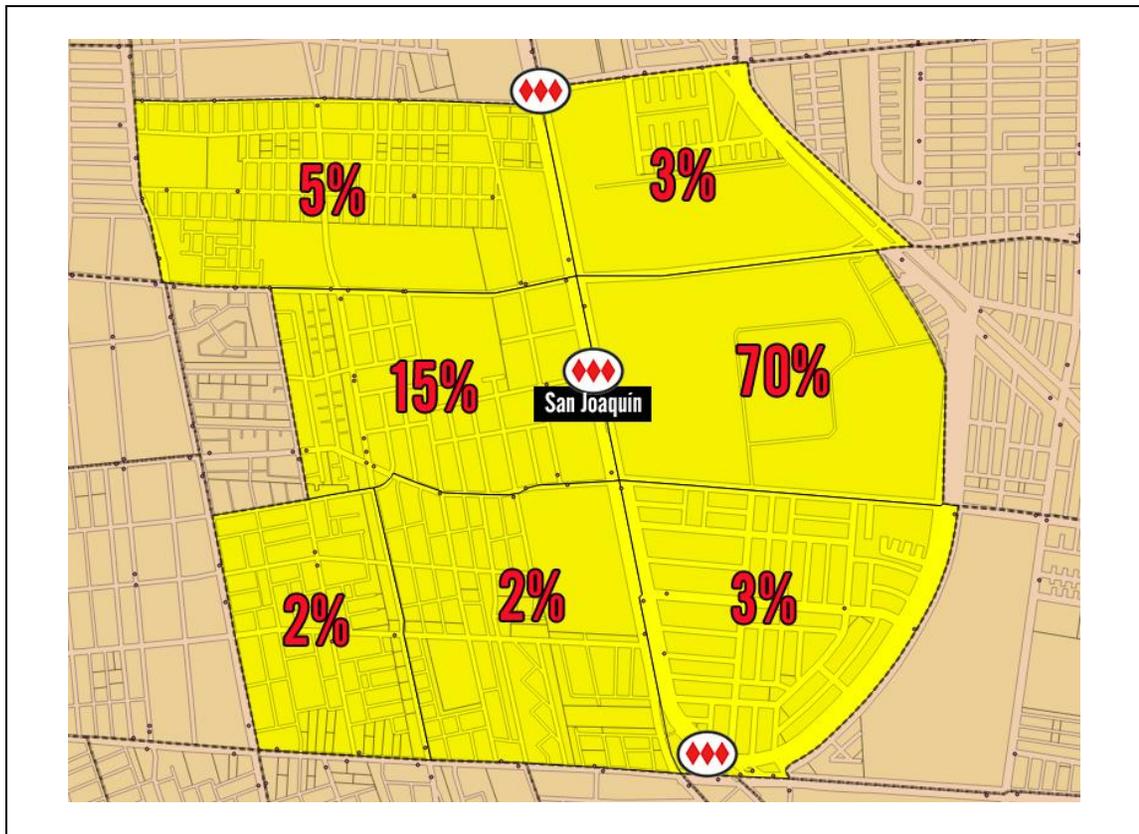


Figura D-3: Ejemplo asignación de viajes a zonas, destino estación San Joaquín

Fuente: Elaboración propia en base a *QGIS*

En este caso, se observa que el modelo predice que el campus de la Universidad Católica (que incluye además el Centro Médico San Joaquín), atrae cerca del 70% de los viajes que terminan en esta estación. Por su parte, la zona ubicada al lado opuesto de Vicuña Mackenna también tiene edificios de educación superior, pero en una extensión

significativamente menor y atrae solo el 15% de los viajes. Los viajes restantes se dividen entre el resto de las zonas, que resultan menos atractivas según el modelo de inferencia.

Por último, se realizó la asignación de viajes originados en la estación San Joaquín, resultando los siguientes porcentajes de repartición, exhibidos en la Figura D-4.

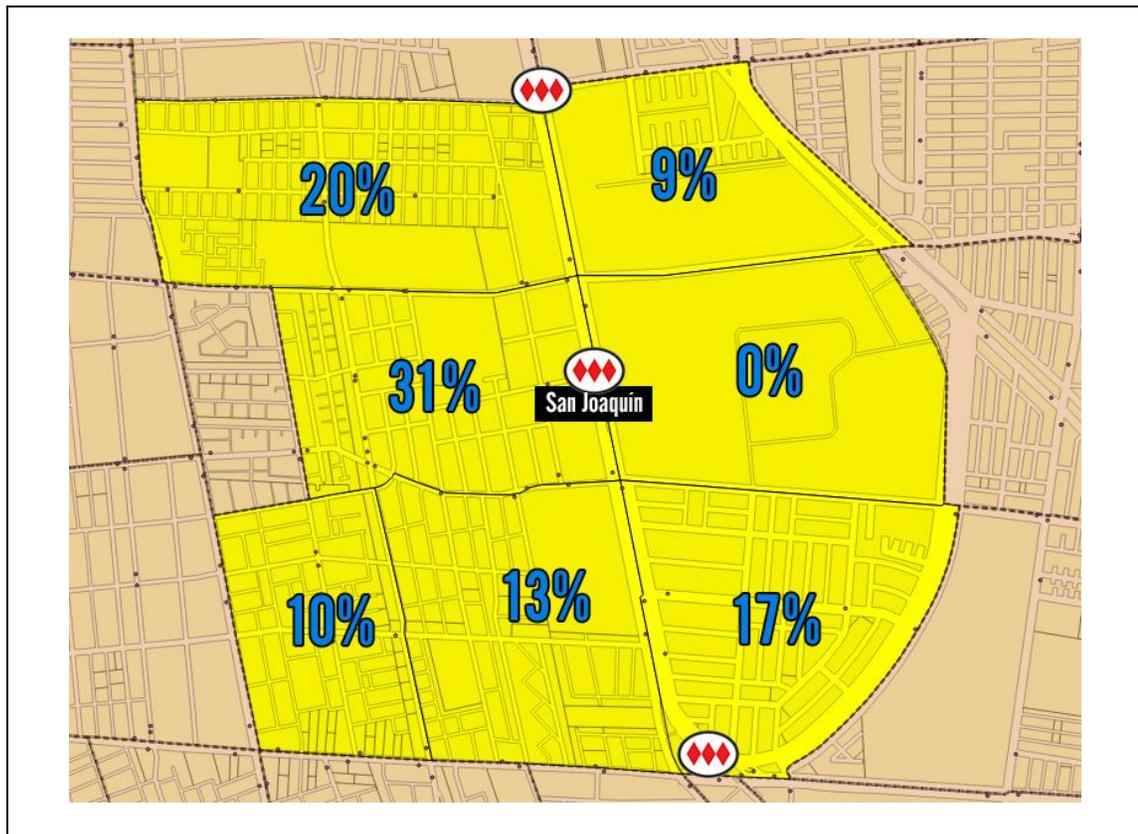


Figura D-4: Ejemplo asignación de viajes a zonas, origen estación San Joaquín

Fuente: Elaboración propia en base a *QGIS*

El modelo de inferencia zonal indica una repartición radicalmente distinta de los viajes en las zonas aledañas. Concretamente, la zona que contiene el campus San Joaquín prácticamente no genera viajes en el período punta mañana, pues no posee edificios residenciales. En cambio, zonas que como destinos eran poco atractivas obtienen ahora utilidades más altas por su gran cantidad de viviendas y generan la mayor cantidad de los viajes de la estación de metro.