

PERIODIZACIÓN DE UNA RED A PARTIR DE LA CLUSTERIZACIÓN DE LA DISTRIBUCIÓN DE VELOCIDADES DE OPERACIÓN DE BUSES DE TRANSPORTE PÚBLICOS

Martín López-Pavez, Pontificia Universidad Católica de Chile, milopez7@uc.cl
Jaime Soza-Parra, University of California at Davis, jsozaparra@ucdavis.edu
Juan Carlos Herrera, Pontificia Universidad Católica de Chile, jch@ing.puc.cl

RESUMEN

La periodización de una red de transporte consiste en la agrupación de intervalos de tiempo que presentan patrones o estados de tráfico similares. Para realizarla, es necesario reunir datos y efectuar mediciones en terreno. Este estudio tiene como objetivo generar una periodización a partir de los perfiles de velocidad provenientes del análisis de información GPS de los buses de transporte público en una zona específica en Santiago. Se clusterizan las velocidades en intervalos de media hora para cada día de agosto 2019 mediante el método *k-means*. Se identifican 4 períodos que resumen el comportamiento de una “semana tipo”. Esta metodología aporta para generar una periodización a bajo costo en una zona delimitada mediante el uso exclusivo de información GPS y de manera semi automatizada.

Palabras claves: periodización, transporte público, patrones de velocidad

ABSTRACT

A network periodization looks for time periods with similar traffic states. For this, it is necessary to collect data and make measurements in the streets. This study identifies patterns in the operational speeds of public transport buses for a specific area in Santiago. The speeds added by blocks of half an hour for each day of August 2019 are clustered using *k-means*. The replicability of the blocks of the month is shown in a "typical week" with 4 clusters. This methodology contributes to generate a periodization at low cost in a delimited area through the exclusive use of GPS information and in a semi-automated way.

Keywords: periodization, public transport, speed patterns

1. INTRODUCCIÓN

La velocidad comercial u operacional de un servicio de buses urbanos se entiende como la velocidad promedio sobre el tramo recorrido incluyendo detenciones en paraderos, semáforos y terminales, y es una variable clave en el funcionamiento de los sistemas de transporte público (Cortés et. al, 2010).

Existe evidencia que la velocidad operacional varía su distribución entre distintos días de una semana y dentro de un mismo día, más allá de la mera estocasticidad presente en un sistema de transporte público y es, además, una aproximación del estado de tráfico que enfrenta cada bus. Estas variaciones son posibles de ver en distintos niveles de agregación, como por ejemplo a través del avance de cada ruta o por segmentación por tramos (Cortés *et al.*, 2011). Además existen diferencias entre los días de una semana laboral y los de los fines de semana, así como entre los horarios de punta y fuera de punta (Rakha y Van Aerde, 1995; Weijermars y Van Berkum, 2005). Esta categorización en distintos bloques horarios con características similares se conoce como periodización y es útil para distintas instancias. Por ejemplo, se utiliza para definir planes operacionales de servicios de buses, sincronizaciones y conformaciones de redes de semáforos (Ministerio de Transporte y Telecomunicaciones, 2017), definir rangos horarios representativos para la medición de variables de tráfico en terreno para proyectos que necesiten una evaluación de impacto vial (SECTRA, 2008), entre otros aspectos que impactan directamente el sistema de transporte.

Por otro lado, el auge de los sistemas de información pasivos, tales como los sistemas de conteo de pasajeros y localización de vehículos permiten entender cada día mejor la compleja dinámica de un sistema de transporte público masivo. En particular, los sistemas de localización, típicamente analizados mediante información de tipo GPS, permiten comprender de mejor manera la variación de los perfiles de velocidad a lo largo de una ruta y del tiempo. Estas propuestas favorecen dos elementos: por un lado, mejorar la calidad del nivel de servicio del transporte público, de interés para las agencias y los planificadores del transporte en la ciudad (Ma y Wang, 2014) y por otro lado, entender con mayor profundidad la reproducibilidad de variables del transporte a través del tiempo, como insumo para la realización de mediciones en terreno.

Este artículo propone una manera de analizar esta gran cantidad de datos de localización para determinar distintos tramos horarios cuyos estados de tráfico sean similares sin proponer una estructura de categorización de manera previa. Con ello, se podrán responder las preguntas relacionadas a la existencia o no de un día “normal” para realizar mediciones de variables de tráfico, o si existen diferencias significativas en la distribución de velocidades entre distintos días en un mismo horario. Sumado a lo anterior, se mostrará una metodología basada en el análisis de información GPS que permitirá realizar una periodización de una zona sin la necesidad de realizar mediciones en terreno.

Este estudio se realiza analizando los datos del sistema de transporte público de Santiago de Chile. Así, asumiendo que el comportamiento de los buses de transporte público representa de buena forma aquel del tráfico general, los estados de tráfico serán caracterizados a través de la distribución de velocidades de operación que presenten los buses. El artículo se estructura de la siguiente forma: en primer lugar se describen los datos y la metodología utilizada para realizar este análisis. Posteriormente, se presentan los resultados principales de este trabajo. Finalmente,

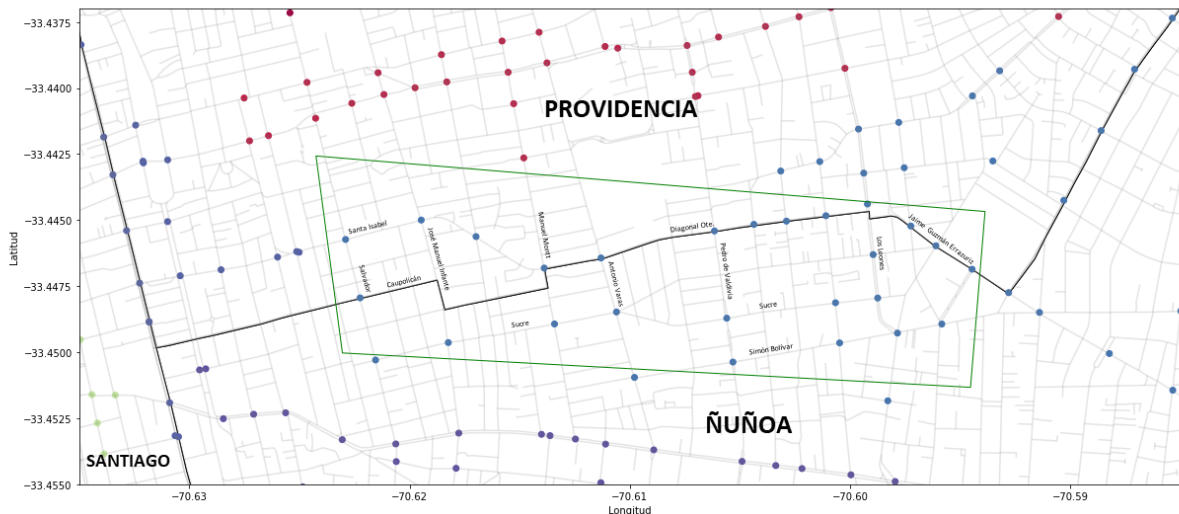
se presentan las conclusiones y recomendaciones generadas a partir del análisis de los resultados obtenidos.

2. METODOLOGÍA

Los datos utilizados corresponden a las observaciones GPS de los buses de Santiago durante el mes de agosto del año 2019. Estas observaciones son generadas cada 30 segundos e incluyen la posición (latitud, longitud), el tiempo, la velocidad instantánea, la identificación del vehículo, y la identificación del servicio. En lugar de analizar la ciudad como un todo, se determinó un área de estudio, en un sector transitado de esta. De esta forma es posible aislar el análisis de aquellos efectos relacionados con la ubicación geográfica del servicio.

La zona de estudio que se muestra en la Figura 1 tiene un ancho y largo aproximado de 1 y 2,7 kilómetros respectivamente. Es importante destacar que en el interior hay presencia de pistas sólo buses (pistas sólo para la circulación de vehículos del Transporte Público ubicadas en el lado derecho de la vía) y vías exclusivas (operan de Lunes a Viernes, excepto festivos, entre las 07:30 a 10:00 hrs. y entre las 17:00 a 21:00 hrs y son controladas en su mayoría por cámaras). Los datos fueron extraídos de las observaciones de las emisiones GPS del mes de Agosto 2019 para todos los servicios de buses de la Región Metropolitana. Para el filtro se utilizó PostgreSQL, obteniendo una base de datos de 2.416.000 observaciones GPS (considerando contar con observaciones de cada expedición antes de entrar y después de salir de la zona de estudio). Esto se traduce en 165.150 expediciones de buses que corresponden a 28 servicios que cruzan, en algún momento dada su programación, la zona de estudio. Los datos son analizados a lo largo de las 24 horas del día, rango que posteriormente se acota de 05:00 a 23:00.

Figura 1. Zona de estudio

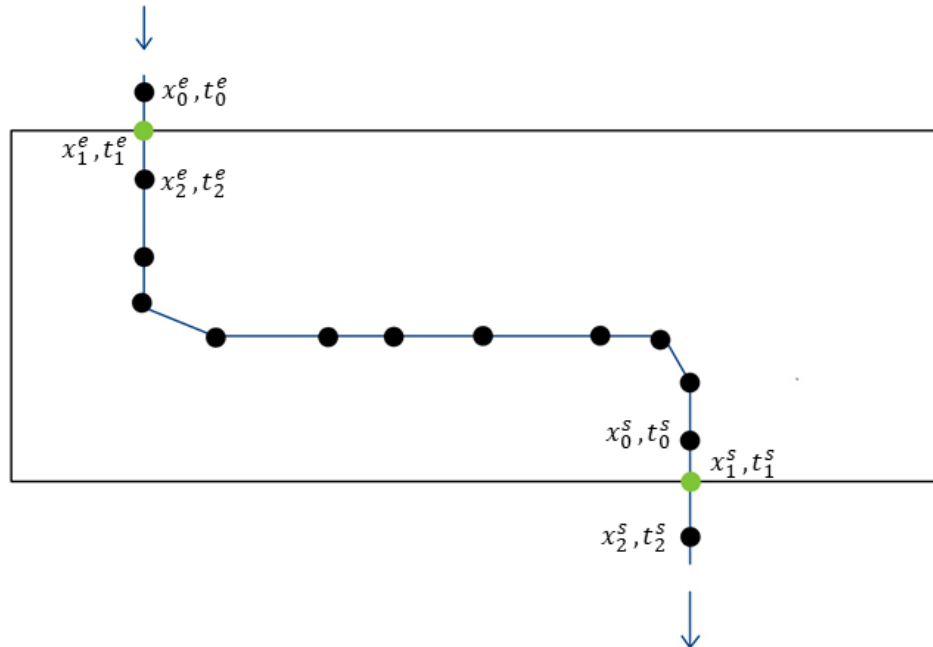


Se consideró expediciones distintas si un bus sale y vuelve a entrar a la zona o si deja de emitir observaciones por más de 60 segundos. Para el análisis de los datos, se filtró el total de expediciones con el fin de que los cálculos tuvieran mayor consistencia y no se generaran errores por falta de observaciones, obteniendo alrededor de 75.000 expediciones. Los criterios aplicados a cada una de las expediciones son:

- Presentar al menos 3 observaciones dentro de la zona de estudio
- Presentar una observación antes de entrar y una observación después de salir de la zona de estudio

Para calcular las velocidades operacionales de cada expedición, se determinó el momento en que cada bus ingresó y salió de la zona de estudio, interpolando dicho instante en base a las observaciones más cercanas en tiempo y espacio a dichos límites como muestra la Figura 2. Los puntos verdes muestran las observaciones interpoladas en base a datos previo y posterior al ingreso y salida de cada bus a la zona de estudio.

Figura 2. Representación de observaciones GPS de una expedición que atraviesa la zona



Los puntos verdes representan las observaciones interpoladas en base a los datos cuando un bus ingresa o sale de la zona de estudio. La notación e se refiere a observaciones de entrada y la s a observaciones de salida del bus de la zona

Para interpolar la posición geográfica, se utilizó el punto de intersección entre los bordes de la zona de estudio y la recta que une las observaciones previo y posterior al ingreso o salida de cada bus a la zona de estudio. Para interpolar los tiempos, se asumió una velocidad uniforme entre las observaciones previo y posterior al ingreso o salida de cada bus a la zona de estudio. Así, los tiempos de las observaciones interpoladas en los bordes de la zona de estudio son calculados con (1).

$$t_1 = t_0 + \frac{t_2 - t_0}{\text{dist}(x_2, x_0)} \cdot \text{dist}(x_1, x_0) \quad (1)$$

Con esto fue posible obtener la distancia recorrida por cada expedición, como la suma de las distancias entre cada observación GPS (considerando los puntos interpolados de los extremos) y

el tiempo empleado por cada expedición, como el tiempo total desde que cada bus entró y salió de la zona. Esta información permitió obtener la velocidad operacional para cada expedición como el cociente entre la distancia recorrida y el tiempo empleado en recorrer dicho trayecto dentro de la zona de estudio, como lo indica (2).

$$\frac{\sum d_i}{\sum t_i}, \text{ para todo } i \text{ (tramos entre observaciones GPS) de una expedición} \quad (2)$$

Posteriormente, se eliminaron expediciones con velocidades operacionales consideradas *outliers* en base al criterio del rango intercuartil, en que se consideran *outliers* los valores que son mayores que 1.5 veces el valor del rango intercuartil más allá de los cuartiles Q1 y Q3. Con lo anterior se obtuvo un total de 75.000 expediciones con sus respectivas velocidades de operación.

Posteriormente, se construyó una grilla de celdas de media hora para cada día del mes. Para el caso de agosto (31 días) corresponde a un total de 1.488 celdas. Cada una de las velocidades operacionales por expedición fue asignada a uno de los bloques según el instante de ingreso a la zona de estudio. Del total de celdas, 14 no registraron ningún tipo de observación. Luego, se procedió a obtener 9 estadísticos representativos de cada bloque, descritos en la Tabla 1.

Tabla 1. Indicadores de cada media hora usados para la clusterización mediante k-means.

Media	Mínimo	1er cuartil	Mediana (m)	3er Cuartil	Máximo	IQR/Mediana	MAD	MAD/Mediana
$\frac{\sum d_i}{\sum t_i}$	$\min(X)$	$Q_1(X)$	$Q_2(X)$	$Q_3(X)$	$\max(X)$	$\frac{Q_3(X) - Q_1(X)}{m}$	$m x_i - m $	$\frac{m x_i - m }{m}$

El subíndice i representa cada expedición

La media se obtuvo como el cociente de la suma de todas las distancias recorridas y la suma de todos los tiempos empleados de todas las expediciones de cada bloque de 30 minutos. Los estadísticos mínimo, 1er cuartil, 2do cuartil, mediana, 3er cuartil y máximo se obtienen considerando todas las velocidades de las expediciones de un mismo bloque. El IQR/m representa el cociente entre la diferencia del 3er cuartil y el 1er cuartil (rango inter cuartil) sobre la mediana. El MAD (*median absolute deviation*) representa la mediana de los valores absolutos de las diferencias entre cada velocidad y la mediana dentro de cada bloque. El MAD/m representa el cociente entre el MAD y la mediana.

Finalmente, los estadísticos para cada tramo de media hora fueron utilizados para realizar una clusterización mediante k-means. Este procedimiento corresponde al tipo de aprendizaje no-supervisado dentro del aprendizaje de máquinas (machine learning). Esto es relevante pues la categorización surge entonces de manera espontánea sin necesidad de imponer restricciones. En palabras sencillas, dado un número determinado de grupos (o clusters), se agrupan las observaciones de tal forma que la diferencia dentro del grupo sea mínima y las diferencias entre grupos sea máxima. Esta diferencia se determina a partir de una medida de distancia considerando diferentes atributos. En este caso se realizó con los 9 indicadores, descritos en la Tabla 1, que representan la distribución de las velocidades comprendidas en cada bloque de 30 minutos.

Este análisis se realizó para dos distintos rangos horarios: a) De 00:00 a 23:59 horas, con un total de 1.474 bloques de media hora, con el fin de estudiar el día en su totalidad y b) De 05:00 a 22:59 horas, con un total de 1.110 bloques de medias horas, con el fin de estudiar las horas en que comúnmente hay una operación continua de los servicios de buses. Para los resultados posteriores, es importante considerar que agosto tiene 31 días, que en 2019 tuvo un feriado el día jueves 15, y que no se observaron días con fuertes lluvias o eventos disruptivos que hayan afectado significativamente la operación del sistema de buses para la zona de estudio.

3. RESULTADOS

En primer lugar, se realizó un análisis para definir el número final de clusters a utilizar (ya que es una variable que se debe entregar al modelo) a través de 3 métodos. En primer lugar, un análisis gráfico de “*Elbow method*”, “*Silhouette method*”, “*Gap statistic method*”. Luego se realizó un procedimiento computacional, ocupando el paquete *NbClust*, el cual sugiere el número óptimo de clusters bajo 30 indicadores. Adicionalmente se decidió complementar el análisis tradicional de clusterización con las grillas presentadas en la Figura 3. Estas figuras presentan en distintos colores el cluster asignado para cada bloque de media hora (eje Y) para cada día del mes (eje X). En la Tabla 2 es posible ver el número de clúster óptimo según cada metodología y para cada rango horario de estudio. El análisis de clusters se realizó con las variables de la Tabla 1 (en total 9). De esta forma, es posible ver que el número sugerido de clusters oscila entre 1 y 5.

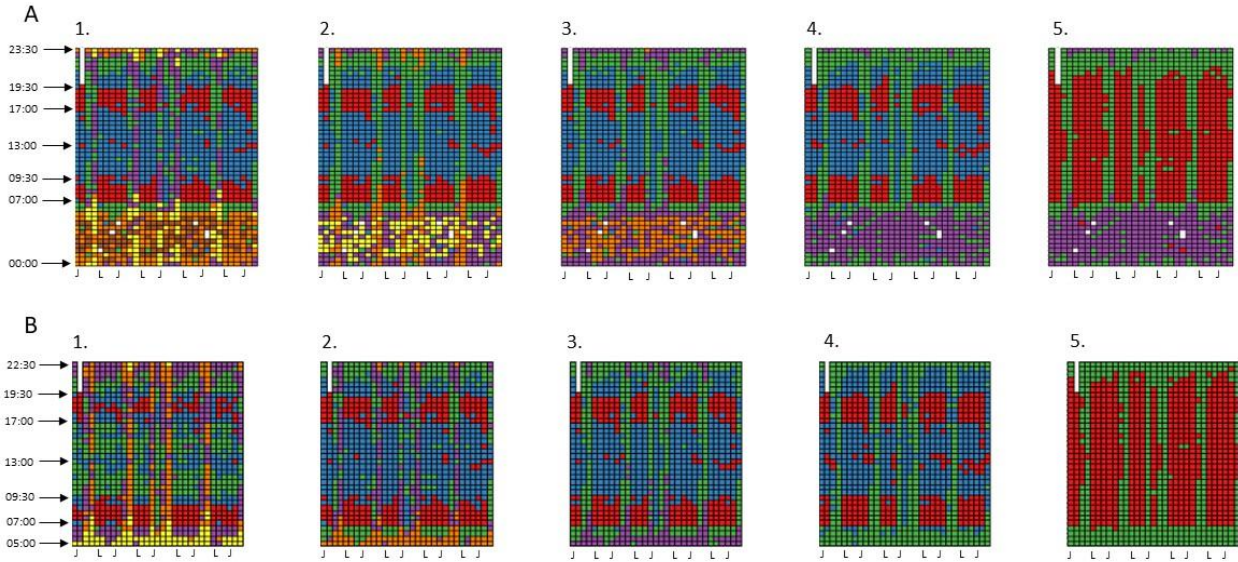
Tabla 2. Número de clusters óptimo por metodología

Rango Horario	Nº Variables	Elbow	Silhouette	Gap	NbClust	Grillas
00:00 - 23:59	9	3	2	3	3	4
	8	3	2	5	2	4
05:00 - 23:00	9	3	2	1	2	3
	8	3	2	1	2	3

Se muestran métodos gráficos (Elbow, Silhouette, Gap), computacionales (NbClust) y de representación gráfica (Grillas). Los resultados son para dos rangos horarios de estudio y considerando 9 y 8 variables estadísticas.

En la imagen A de la Figura 3 se distingue que entre las 07:00 y las 20:00 el patrón es similar para todos los números de clusters, salvo cuando este es 3, donde desaparece el grupo observado entre las 10:00 y las 17:00. Por otro lado, a mayor número de clusters, se diversifica la clusterización observada entre las 00:00 a las 07:00 y entre las 20:00 a las 23:59. En la imagen B ocurre algo similar. La clusterización generada entre las 07:00 y las 20:00 horas se mantiene prácticamente inalterada al cambiar el número de clusters, con la excepción de cuando el número de clusters es 2, donde desaparece el cluster entre las 10:00 a las 17:00. Por otro lado, a mayor número de clusters, se diversifica la clusterización entre las 05:00 y las 07:00 y entre las 20:00 y las 23:59. Dados los argumentos anteriores es que se define que el número de clusters a utilizar es 4 para el rango horario 00:00 - 23:59 (Figura 1, imagen A-4.) y 3 para el rango de 05:00 - 23:00 (Figura 1, imagen B-4.). De esta forma, se tiene el número mínimo de clusters sin perder información relevante dentro del horario habitual de operación, que se muestra en la Figura 4.

Figura 3. Grillas de mes completo con distinto número de clusters para cada media hora del mes



La imagen A representa los clusters para el procedimiento de 00:00 a 23:59 horas, desde 7 a 3 clusters (de izquierda a derecha). La imagen B representa los clusters para el procedimiento de 05:00 a 23:00 horas, desde 6 a 2 clusters (de izquierda a derecha).

A partir de esta clusterización, es posible indicar que, aparentemente, existe una diferencia significativa en las velocidades experimentadas entre un día de semana y un fin de semana. Sin embargo, no existen mayores diferencias entre un mismo tramo horario para los distintos días de una semana laboral. La única salvedad ocurre para los días jueves 15 y viernes 16, producto del feriado del día jueves. En las grillas es posible observar que el jueves feriado se asemeja a un día domingo y que el viernes inter feriado presenta un comportamiento anómalo al no distinguirse velocidades bajas en horas punta a diferencia de los otros viernes. Observamos que en un día laboral normal, se observan al menos tres períodos: horarios punta (mañana y tarde), horarios fuera de punta (o valle) y horarios nocturnos.

Con el fin de resumir el análisis de las observaciones anteriores, se generó una grilla para una semana representativa, como se ve en la Figura 4. El eje X representa los días de una semana y el eje Y los horarios del día cada media hora. Cada celda es asociada a un cluster en base a la cantidad de veces que se observó un cluster para una media hora y día de la semana en la grilla de mes completo, exigiendo que ese horario y día de la semana perteneciera al menos el 75% de las veces a un mismo cluster. Lo anterior está basado en la cantidad de veces que se observó cada día en el mes de agosto como se muestra en la Tabla 3.

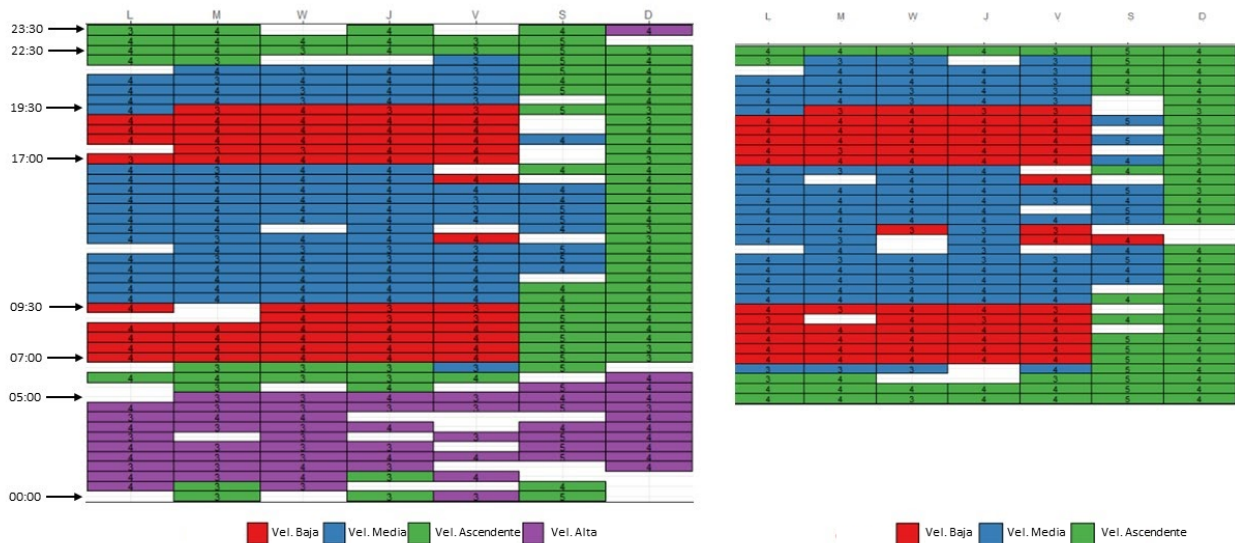
Tabla 3. Frecuencias exigidas y observadas por día de la semana en Agosto 2019

Frecuencias	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Observadas	4	4	4	5*	5*	5	4
Exigidas	3	3	3	3	3	4	3

Los días Jueves y Viernes de Agosto 2019 tienen una frecuencia observada de 5 ocurrencias, sin embargo, el jueves 15 corresponde a un feriado y viernes 16 a un interferiado, por lo que se asumió que no se cuenta con 5 de estos días, si no con 4 de cada uno.

Las celdas en blanco corresponden a medias horas que no fue posible asignar a ningún cluster con las características mencionadas (no alcanzaron el mínimo de observaciones para un mismo cluster). Los números al interior de las celdas representan la cantidad de veces que se observó (en el mes) el cluster asignado para cada bloque de 30 minutos.

Figura 4. Grillas de semanas representativas



Cada color representa el cluster al que pertenece cada media hora. Imagen A. Grilla para el horario de 00:00 a 23:59 horas. Imagen B. Grilla para el horario de 05:00 a 23:00 horas.

Es posible ver en la Figura 4 que mediante esta heurística es posible asignar la gran mayoría de celdas a un cluster. Dado lo anterior, es posible hablar de una “semana tipo” construida en base a la repetición de patrones observados durante el mes de agosto. Finalmente se generó una tabla descriptiva de cada estadístico de la Tabla 1 utilizado para la clusterización, a nivel de los grupos generados. La tabla incluyó la media, percentil 5 y 95 de cada estadístico, como se ilustra en la Figura 5.

Figura 5. Tablas de intervalos de confianza para los estadísticos de la clusterización

A

Clúster	Media	Mínimo	1er Cuartil	Mediana	3er Cuartil	Máximo	IQR/mediana	MAD	MAD/mediana
Vel. Baja	12,51 (10,32; 14,04)	7,09 (5,18; 9,56)	10,96 (8,78; 12,36)	13,02 (10,85; 14,52)	15,3 (13,28; 16,92)	22,57 (18,45; 25,61)	0,34 (0,23; 0,49)	3,22 (2,17; 4,52)	3,22 (0,17; 0,38)
Vel. Media	15,73 (14,33; 17,53)	9,79 (7,37; 11,99)	13,9 (12,32; 15,69)	15,94 (14,5; 17,64)	18,27 (16,58; 20,22)	24,84 (22,42; 27,24)	0,27 (0,2; 0,37)	3,23 (2,31; 4,35)	3,23 (0,14; 0,27)
Vel Media-Alta	19,47 (17,69; 21,56)	13,36 (10,61; 16,45)	17,19 (15,51; 19,18)	19,54 (17,64; 22,07)	22,32 (20,06; 24,95)	27,29 (24,2; 31,04)	0,26 (0,17; 0,36)	3,82 (2,46; 5,58)	3,82 (0,13; 0,27)
Vel Alta	23,06 (20,93; 25,54)	19,36 (15,27; 24,34)	21,7 (19,15; 24,74)	23,44 (20,94; 25,86)	24,85 (21,99; 28,38)	26,28 (23,18; 30,93)	0,13 (0; 0,31)	2,43 (0; 5,57)	2,43 (0; 0,23)

B

Clúster	Media	Mínimo	1er Cuartil	Mediana	3er Cuartil	Máximo	IQR/mediana	MAD	MAD/mediana
Vel. Baja	12,65 (10,34; 14,34)	7,22 (5,2; 9,75)	11,09 (8,87; 12,65)	13,15 (10,92; 14,74)	15,43 (13,32; 17,11)	22,65 (18,54; 25,61)	0,33 (0,23; 0,49)	3,21 (2,19; 4,51)	0,25 (0,17; 0,37)
Vel. Media	15,97 (14,5; 17,85)	9,94 (7,5; 12,2)	14,1 (12,45; 15,9)	16,16 (14,62; 18,03)	18,55 (16,78; 20,76)	25,13 (22,76; 27,88)	0,28 (0,2; 0,37)	3,29 (2,38; 4,39)	0,2 (0,15; 0,27)
Vel. Alta	19,89 (18,02; 23,38)	13,66 (10,69; 17,73)	17,69 (15,79; 21,76)	20,01 (17,95; 24,24)	22,76 (20,39; 26,61)	27,96 (24,7; 31,17)	0,26 (0,13; 0,36)	3,75 (2,3; 5,17)	0,19 (0,1; 0,26)

Ambas tablas muestran para cada estadístico usado en la clusterización, la media (izquierda) e intervalo de confianza al 90% (derecha, entre paréntesis). Imagen A. Tabla para el horario de 00:00 a 23:59 horas. Imagen B. Tabla para el horario de 05:00 a 23:00 horas.

En la tabla A, los intervalos de confianza de la media permiten ver que sólo hay una leve superposición entre los grupos “Vel. Media-Alta” y “Vel. Alta”. Similar es en la tabla B, donde los intervalos de confianza de la media no se superponen para ninguno de los grupos.

En base a lo anterior se deben reconocer dos fenómenos probabilísticos. En primer lugar, las grillas son útiles para evidenciar que tan probable es enfrentarse a un cluster dada una media hora y un día de la semana. En segundo lugar, las tablas estadísticas permiten visualizar qué tan compactos son los clusters y verificar que existe baja superposición de valores entre un cluster y otro, siendo finalmente diferentes entre sí.

4. CONCLUSIONES

Este estudio aporta una metodología la cual, a partir de datos provenientes de los GPS de los buses y de una clusterización k-means, permite identificar patrones en las velocidades de los buses, definiendo probabilidades de enfrentarse a distintos estados de tráfico (i.e. clusters), dependiendo del día y la hora. Es decir, determina períodos cuya operación es similar a lo largo del día, semana y mes de manera semi automatizada. Asumiendo que el comportamiento de los buses de transporte público representa de buena forma aquel del tráfico general, esta homologación de periodos (tarea conocida como “periodización”) resulta clave al momento de definir planes operacionales de servicios de buses, sincronizaciones y conformaciones de redes de semáforos, rangos horarios representativos para la medición de variables de tráfico en terreno para proyectos que necesiten una evaluación de impacto vial, entre otros.

Si bien dentro de un cluster existe una variabilidad asociada, es posible identificar sus intervalos de confianza, que permiten definir con una alta probabilidad el rango de velocidades dentro del cluster y, por ende, el estado de tráfico asociado. Los resultados preliminares presentados en este trabajo confirman, para la zona estudiada, la existencia de una “semana tipo”, en la cual es

posible identificar tres períodos de operación con características distintas para el horario de 05:00 a 22:59 y cuatro periodos para el horario de 00:00 a 23:59.

La robustez de esta metodología puede ser mejorada al aumentar la cantidad de días a incluir en la clusterización. Aumentar la cantidad de datos mejoraría la designación de una semana representativa, puesto que la exigencia para la mayoría de los días de semana en este estudio fue una presencia de algún cluster en el 75% de las veces, equivalente a 3 de 4 días. Si se considerara por ejemplo, un año de observaciones, se contaría con aproximadamente 52 observaciones para cada tipo de día, en vez de las 4 consideradas para este estudio. Esto permitiría determinar la existencia de posibles patrones entre semanas y concluir, por ejemplo, si las primeras semanas de cada mes son similares a las segundas semanas. Por otro lado, con el fin de realizar un ajuste más preciso, se propone para futuras investigaciones incluir indicadores para la clusterización que representen variables como precipitación, incidentes en la zona, restricción vehicular por pre-emergencia ambiental y flujos de buses.

Al analizar estos patrones, es importante tener presente que se pueden ver afectados por variables que no somos capaces de ver en los datos, como son las redes programadas semafóricas (que también se ajustan de forma cíclica con el comportamiento de los buses) y la cantidad de pasajeros que espera al bus en cada paradero. Se espera a futuro diseñar un experimento que permita determinar qué implicancia pueden tener estas variables en los resultados.

5. REFERENCIAS

- Cortés, C. E., Gibson, J., Gschwender, A., Munizaga, M., & Zúñiga, M. (2011). Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies*, 19(4), 695–707. <https://doi.org/10.1016/j.trc.2010.12.008>
- Ma, X., & Wang, Y. (2014). Development of a Data-Driven Platform for Transit Performance Measures Using Smart Card and GPS Data. *Journal of Transportation Engineering*, 140(12), 04014063. [https://doi.org/10.1061/\(asce\)te.1943-5436.0000714](https://doi.org/10.1061/(asce)te.1943-5436.0000714)
- Rakha, H. y Van Aerde, M. (1995). Statistical Analysis of Day-to-Day Variations in Real-Time Traffic Flow Data. *Transportation Research Record* 1510, 26-34.
- Weijermars, W., & Van Berkum, E. (2005). Analyzing highway flow patterns using cluster analysis. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2005(October), 308–313. <https://doi.org/10.1109/ITSC.2005.1520157>
- Ministerio de Transporte y Telecomunicaciones, (2017). Manual Programación y Modelación de Semáforos. Ministerio de Transporte y Telecomunicaciones, Chile.
- SECTRA, (2008). Metodología para el Desarrollo, Presentación y Tramitación de los Estudios de Impacto sobre el Sistema de Transporte Urbano - EISTU. Versión 1.1. Ministerio de Transporte y Telecomunicaciones, Chile.