

## IDENTIFICACIÓN DE ARCOS REPRESENTATIVOS PARA LA ESTIMACIÓN DEL DIAGRAMA FUNDAMENTAL MACROSCÓPICO

Guamán Diego, Pontificia Universidad Católica de Chile, dgguaman@uc.cl  
Herrera Juan Carlos, Pontificia Universidad Católica de Chile, jch@ing.puc.cl

### RESUMEN

Presentamos una metodología para reducir los costos de recolección de datos para construir un diagrama fundamental macroscópico. Utilizamos como dato principal la velocidad media espacial de los vehículos en cada arco de la red. Empleamos el algoritmo de Kmeans varias veces para construir una matriz de agrupamiento de arcos. La matriz fue utilizada junto con una heurística de selección basada en la cantidad de veces que un arco fue representativo y la cantidad de arcos que representó para identificar los arcos que se deben monitorear. Los errores de estimación del MFD son mayores que otras metodologías, pero esta requiere menos datos e infraestructura.

*Palabras Claves:* MFD, agrupamiento de arcos, tecnología GPS.

We present a methodology to reduce data collection costs for constructing a macroscopic fundamental diagram. We used the space mean speed of vehicles on each arc of the network as the primary data. We employed the K-means algorithm multiple times to construct an arcs clustering matrix. We used the matrix in conjunction with a selection heuristic based on the number of times an arc was representative and the number of arcs represented to identify the arcs that should be monitored. The errors of the estimated MFD are higher than other methodologies, however this one requires less data and infrastructure.

*Keywords:* MFD, arcs clustering, GPS technology

### 1. INTRODUCCIÓN

La congestión del tráfico es un problema a nivel global (Zhang et al., 2020). Una alternativa para mitigar la congestión es el monitoreo en tiempo real y la implementación de acciones basadas en modelos simples y agregados (Geroliminis y Daganzo, 2007). El diagrama fundamental macroscópico (MFD por sus siglas en inglés) proporciona un modelo parsimonioso para describir el rendimiento del tráfico a nivel de red (Saffari et al, 2020) ya que relaciona la acumulación de vehículos en esa red (densidad) con su productividad (flujo). Si bien medir el flujo y la densidad de cada arco es sencillo, esto no se hace para toda la red debido al costo generado por la instalación y el mantenimiento de los sensores fijos (Saffari et al, 2020). Una alternativa a los sensores fijos son los datos de los observadores móviles, pero estos tienen limitaciones como las condiciones especiales del modo de transporte, bajas y desconocidas tasas de penetración y la privacidad. Así, la falta de datos para caracterizar un MFD y posteriormente usarlo para administrar la red en tiempo real es un problema. Dada la problemática de la recolección de datos y la utilidad de conocer la

MFD de una zona, estudios como el de Ortigosa et al. (2014), Zockaie et al. (2018), entre otros, buscan identificar los arcos más importantes de la red para caracterizar su MFD. Las metodologías existentes requieren menos recursos que los necesarios para la estimación de MFD, pero siguen siendo demasiado costosas para las ciudades poco equipadas.

Por otra parte, una alternativa para reducir los costos de estimación del MFD es utilizar la velocidad media espacial ya que esta es una medida agregada que se asocia a un solo estado de tráfico en el diagrama fundamental de cada arco. Los desarrolladores de software recopilan la velocidad de los observadores móviles equipados con dispositivos GPS y proveen servicios que tienen una calidad aceptable para evaluar escenarios de gestión de tráfico (van den Haak et al., 2018). Sin embargo, la velocidad media espacial no caracteriza por sí sola al MFD, por lo tanto, es necesaria una metodología para su uso con el objetivo de reducir el costo de la estimación del MFD.

Estimar el MFD de una red y mediante él conocer el estado de tráfico de la misma permite a los tomadores de decisiones implementar acciones de gestión operativa de la oferta en tiempo real como el control perimetral (Geroliminis & Daganzo, 2007). Estas acciones aumentan la productividad al interior de la red aminorando los efectos de la congestión.

El objetivo de este trabajo es desarrollar una metodología que utilice la evolución temporal de la velocidad media en cada arco de una red para identificar un subconjunto de ellos que, luego de conocer su flujo y densidad, sirven para estimar un MFD cercano al real, considerando que la red no está completamente equipada con sensores de punto fijo y de la cual se desconoce su MFD. La hipótesis de este trabajo es que los histogramas de velocidad media de los arcos de una red entregan información suficiente para identificar el menor subconjunto de arcos que son necesarios para estimar un MFD.

## 2. REVISIÓN BIBLIOGRÁFICA

Varios estudios se han desarrollado en las dos últimas décadas en torno al MFD. Inicialmente, los trabajos (Daganzo & Geroliminis, 2008; Geroliminis & Daganzo, 2007; Leclercq et al., 2014; Nagle & Gayah, 2014) se centraron en el desarrollo de métodos analíticos y experimentales para construir el MFD. En la última década se han desarrollado un gran número de estudios en torno a las propiedades, factores de influencia y aplicaciones prácticas del MFD (Zhang, 2020).

Los métodos experimentales para construir el MFD se basan en mediciones, lo que impone restricciones presupuestarias. Las mediciones disponibles para construir el MFD pueden ser trayectorias de todos los vehículos en la red (Courbon & Leclercq, 2011; Ji & Geroliminis, 2012), trayectorias de algunos vehículos (Ambühl & Menendez, 2016; Du et al., 2016; A. Nagle & Gayah, 2013; A. S. Nagle & Gayah, 2014), mediciones de sensores de punto fijo en toda la red y mediciones de sensores de punto fijo en una parte de la red (Ambühl & Menendez, 2016; Buisson & Ladier, 2009; Courbon & Leclercq, 2011; Knoop et al., 2018; Leclercq et al., 2014; Saffari et al., 2020). Las metodologías existentes requieren menos recursos que los necesarios para construir el MFD real, pero siguen siendo demasiados para ciudades poco equipadas.

Dadas las restricciones presupuestarias, es necesario elegir de forma óptima un grupo de arcos que sean representativos del estado del tráfico de toda la red. Estudios como el de Keyvan-Ekbatani et al. (2013); Ortigosa et al. (2014) y Zockaie et al. (2018) plantearon un problema de optimización para identificar arcos representativos del estado del tráfico de la red y encontraron que con una pequeña proporción de los arcos es posible construir un MFD reducido. Sin embargo, los resultados son sensibles a las variaciones de la demanda o requieren recursos similares a los necesarios para construir el MFD real. La primera metodología que utiliza datos de algunos observadores móviles para identificar arcos representativos del estado de tráfico de una red fue propuesta por Saffari et al. (2020). En esta metodología la tasa de penetración de observadores móviles debe ser homogénea en todos los arcos. Por lo tanto, es necesaria una gran flota de observadores móviles, lo cual limita su aplicación.

Para solucionar el problema de los costos de estimación del diagrama fundamental macroscópico una alternativa es utilizar la velocidad media espacial ya que esta es una medida agregada que se asocia a un solo estado de tráfico en el diagrama fundamental de cada arco. Estudios como el de Herrera et al. (2010), Gayah & Dixit (2013) y Nagle & Gayah (2014) sugieren que es posible estimar la velocidad media espacial con solo 2% a 3% de las velocidades de los vehículos en una carretera. La tasa de penetración de observadores móviles que recopilan datos de GPS para desarrolladores de software estimada en los Países Bajos supera el 10% (Knoop et al., 2018). Para validar los datos de los desarrolladores de software, van den Haak et al. (2018) utilizaron series de tiempo de velocidad generadas por observadores móviles de un desarrollador de software para estimar el estado del tráfico de 2200 ubicaciones. El estudio concluyó que los datos anónimos y agregados que se recopilaron tienen una calidad aceptable para ser utilizados en escenarios de gestión del tráfico (van den Haak et al., 2018). Estos datos ya fueron utilizados para construir un MFD (Knoop et al., 2018), aunque para este propósito se requirieron supuestos relacionados con la tasa de penetración de vehículos flotantes, el flujo máximo de los arcos de la red y simplificaciones de la geometría de los arcos.

### 3. METODOLOGÍA.

En esta sección se describe la metodología propuesta para identificar arcos representativos para la estimación del diagrama fundamental macroscópico. La metodología propuesta utiliza la velocidad media espacial en cada uno de los arcos que componen la red en estudio durante un periodo para calcular variables de su distribución de velocidad media. Las variables que se deben calcular son: mínimo, cuartil uno, mediana, media, cuartil tres, máximo, rango inter cuartil y desviación absoluta media. Para toda la red las variables calculadas conforman una matriz cuyas filas corresponden a los arcos de la red y las columnas al valor de cada variable.

La matriz construida se utiliza como dato de entrada en el algoritmo de agrupamiento *Kmeans*. Para el algoritmo de agrupamiento, las filas de la matriz (arcos de la red) son elementos que deben ser agrupados en base a sus características. Las características de cada arco son las columnas de la matriz (variables de la distribución de velocidad media). El algoritmo de agrupamiento divide a todos los arcos en grupos y encuentra un arco representante para cada uno.

El algoritmo de agrupamiento debe ser repetido varias veces, utilizando cada vez una matriz de variables de la distribución de velocidad media espacial diferente. Para obtener distintas matrices se debe consultar la velocidad media espacial de los arcos de la red durante varios periodos. La heurística de selección propuesta considera los resultados de todas las repeticiones del algoritmo de agrupamiento y forma un conjunto de arcos para los cuales se debe construir dos indicadores. El primer indicador cuenta el número de repeticiones en las cuales el arco fue representante y el segundo indicador cuenta el número acumulado de arcos que representó cada arco. En base a los indicadores se determina los arcos que se deben monitorear para estimar el MFD de la red. El resultado de la metodología es una lista de arcos que deben ser monitoreados para construir un MFD estimado. La cantidad de arcos que se deben monitorear constituye una restricción presupuestaria. La metodología se representa como diagrama de flujo en la Figura 1. Donde,  $Z$  indica la cantidad de repeticiones del algoritmo de agrupamiento que se debe realizar, lo que es equivalente a la cantidad de periodos de recolección de datos de velocidad media. En las siguientes subsecciones se realiza una descripción resumida de cada uno de los pasos de la metodología.

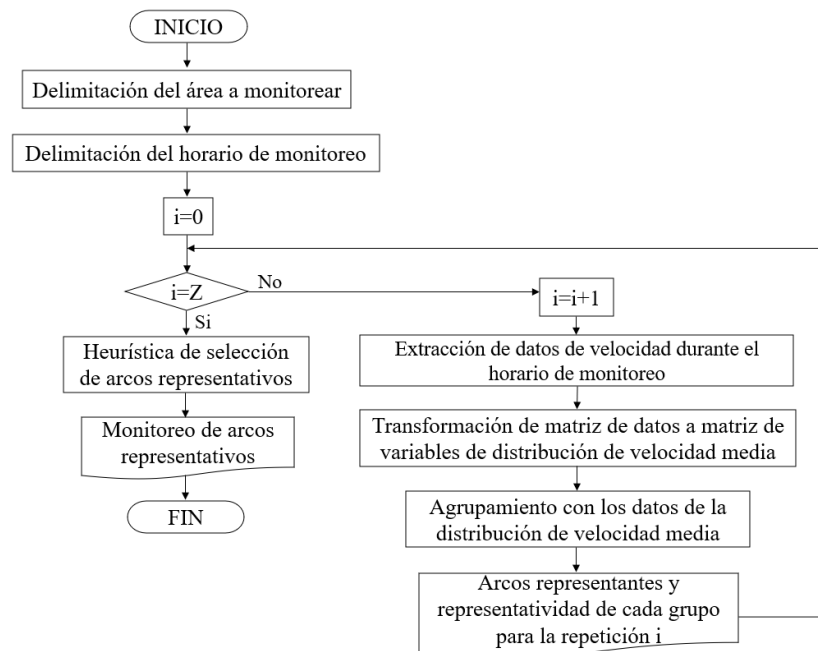


Figura 1 Diagrama de flujo de la metodología de selección de arcos representativos.

### 3.1. Delimitación del área y periodo de recolección de datos.

Se han identificado MFD en áreas de 2.6 km<sup>2</sup> en Zurich (Ortigosa et al., 2014), 6.47 km<sup>2</sup> en San Francisco (Geroliminis & Daganzo, 2007), 7.46 km<sup>2</sup> en Barcelona (Saffari et al., 2020), 219.3 km<sup>2</sup> en Ámsterdam (Knoop et al., 2018) y 41 543 km<sup>2</sup> en Países Bajos (Knoop et al., 2018). Por tanto, el tamaño del área monitoreada y el número  $N$  de arcos dentro de esta no es una limitante para la existencia del MFD. Se debe delimitar un área en la cual no existan cuellos de botella localizados

y separados por arcos de poca congestión para favorecer una baja dispersión en el MFD (Doig et al., 2013).

El periodo de recolección de datos de velocidad media espacial es la ventana de tiempo  $T$  en la cual interesa conocer los arcos representativos del estado de tráfico de la red. Durante este tiempo debe mantenerse constante las características físicas y operacionales de la red ya que el MFD del área es propio de esta oferta (Geroliminis & Daganzo, 2007). Es necesario repetir la recolección de datos durante varios días o distintos horarios de un mismo día para los cuales las características físicas y operacionales son las mismas. Además, durante el periodo  $T$  los cambios de la demanda deben ser lentos (Geroliminis & Daganzo, 2007).

### 3.2. Matriz de datos.

Los datos que se deben recolectar es la velocidad media espacial en cada uno de los  $N$  arcos que componen la red en estudio para cada intervalo de tiempo  $\Delta t$  que es parte del periodo  $T$  de recolección de datos, donde  $\Delta t$  se define discretizando  $T$  en intervalos de tiempo de igual duración. El valor más pequeño de  $\Delta t$  es el largo de ciclo de la red semafórica en operación durante los  $T$  minutos de recolección de datos en el área delimitada. En base a las variables definidas se construye una matriz conformada por  $N$  filas y  $C = \frac{T}{\Delta t}$  columnas, donde cada celda contiene la velocidad media espacial. En caso que se cuente con algunos vehículos equipados con un dispositivo GPS que permite construir su trayectoria, la velocidad media espacial del arco  $n$  durante el intervalo  $c$  ( $v_c^n$ ) se calcula según la Ecuación 1 (Edie, 1963).

$$v_c^n = \frac{\sum_{p=1}^P d_{p,c}^n}{\sum_{p=1}^P t_{p,c}^n} \quad \forall p \text{ que se encuentra en arco } n \text{ durante el intervalo } c \quad (1)$$

$P$  es el conjunto autos que son monitoreados y se encuentran en el arco  $n$  durante el intervalo  $c$ ,  $d_{p,c}^n$  es la distancia recorrida por el auto  $p$  en el arco  $n$  en el intervalo  $c$  y  $t_{p,c}^n$  es el tiempo recorrido por el auto  $p$  en el arco  $n$  en el intervalo  $c$ .

En la matriz de datos cada arco es representado por una fila durante el periodo  $T$ . Los datos de velocidad media espacial de un arco describen una distribución de probabilidad de la velocidad media espacial para ese arco. Para cada arco  $n$ , a partir a las  $C$  observaciones de velocidad, se calculan las siguientes nueve variables: media ( $\bar{v}^n$ ), mínimo, cuartil uno ( $Q_1^n$ ), mediana ( $\tilde{v}^n$ ), cuartil tres ( $Q_3^n$ ), máximo, rango inter cuartil ( $IQR^n$  por sus siglas en inglés) y desviación absoluta media ( $MAD^n$  por sus siglas en inglés). La Tabla 1 muestra la matriz de variables descriptivas que será utilizada por el algoritmo de agrupamiento.

$$\bar{v}^n = \frac{\sum_{c=1}^C v_c^n}{C} \quad (2)$$

$$Q_1^n = v_{(C+1)/4}^n \quad (3)$$

$$\tilde{v}^n = v_{(C+1)/2}^n \quad (4)$$

$$Q_3^n = v_{(3C+3)/4}^n \quad (5)$$

$$IQR^n = Q_3^n - Q_1^n \quad (6)$$

$$D_c^n = |v_c^n - \tilde{v}^n| \quad (7)$$

$$MAD^n = D_{(C+1)/2}^n \quad (8)$$

Tabla 1 Matriz de variables descriptivas estandarizadas para  $N$  arcos

Arco (n)	Variables de la distribución de velocidad								
	$\bar{v}^n$	$Min$	$Q_1^n$	$\tilde{v}^n$	$Q_3^n$	$Max$	$IQR^n$	$MAD^n$	$MAD^n/\bar{v}^n$
1									
2									
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N									

### 3.3. Agrupamiento.

El objetivo de los algoritmos de agrupamiento es descubrir una estructura en los datos o agrupaciones naturales (Jain, 2010). La naturaleza exploratoria de los algoritmos de agrupamiento está en línea con el objetivo de reconocer arcos representativos del estado de tráfico de la red, ya que no se cuenta con patrones de entrenamiento (Duda et al., 1995). Los algoritmos de agrupamiento se dividen en jerárquicos y particionales. Los algoritmos jerárquicos establecen niveles de agrupación anidados, mientras que el agrupamiento particional realiza una sola división de los elementos en varios grupos. Los algoritmos particionales difieren en la elección de la función objetivo, modelos probabilísticos y heurísticas. Uno de los algoritmos particionales más utilizado es *Kmeans*. El éxito de *Kmeans* está en su fácil implementación, simplicidad, eficiencia y éxito empírico (Jain, 2010). El algoritmo de *Kmeans* requiere especificar: el número de  $K$  grupos en los que se desea agrupar todos los datos, una partición inicial de los datos y una métrica de distancia. La métrica de distancia utilizada normalmente es la distancia euclidiana entre dos puntos (Jain, 2010). Los grupos que se generan con *Kmeans* se pueden definir como regiones de alta densidad de objetos ubicados en el espacio de características separados por regiones de baja densidad. Una variación de los algoritmos basados en densidad son los algoritmos subespaciales como CLIQUE o el agrupamiento teórico de grafos (Jain, 2010).

Ji & Geroliminis (2012) aplicaron algoritmos de agrupamiento utilizando datos flujo y densidad de todos los arcos de una red y encontraron que al utilizar *Kmeans* el resultado fue grupos espacialmente dispersos pero homogéneos en su estado de tráfico. Recientemente Saffari et al. (2020) y Ambühl et al. (2021) aplicaron algoritmos de agrupamiento sobre los arcos de una red en base a sus características de tráfico para identificar arcos representativos del estado de tráfico de la red obteniendo resultados favorables sin embargo estas tienen limitaciones relacionadas con los recursos necesarios.

Para aplicar los algoritmos de agrupamiento es recomendable normalizar cada variable que compone el espacio de características con el objetivo de que estas sean comparables (Nisbet et al., 2017). Por tanto, cada valor de las variables de la Tabla 1 deben ser normalizadas según su distancia a la media expresadas en desviaciones estándar. Luego, el agrupamiento de los arcos consiste en utilizar la matriz de variables descriptivas estandarizadas con el algoritmo de *Kmeans* para agrupar los arcos del área delimitada en  $J$  grupos. La cantidad  $J$  de grupos es una restricción presupuestaria, ya que, a mayor número, mayor será la inversión necesaria para monitorear un área. Además, este

número debe ser definido a priori, ya que es un parámetro del algoritmo. Sin considerar la restricción presupuestaria, se puede determinar el número óptimo de grupos con el método gráfico del codo (Coates & Ng, 2012).

Luego de realizar el agrupamiento se debe seleccionar un arco representativo para cada grupo y a este se le asignará un valor de representatividad igual a la cantidad de arcos que conforman su grupo. La selección del arco representativo, se realiza evaluando la distancia euclidiana más corta entre todos los arcos del grupo al centroide del mismo y eligiendo aquel con menor distancia. El resultado de esta etapa es un conjunto de arcos representativos para los cuales se conoce su representatividad  $RP_{n,z}$ .

El proceso descrito en esa subsección se debe repetir  $Z$  veces utilizando para cada repetición una matriz de variables descriptivas estandarizadas distinta. Para esto último, se debe repetir la recolección de datos de velocidad media espacial variando el periodo de recolección de datos  $T$ . Después de repetir el agrupamiento, el resultado es una matriz de agrupamiento conformada por  $N$  filas y  $Z$  columnas, donde cada celda contiene la representatividad del arco  $n$  en la repetición  $z$  ( $RP_{n,z}$ ).

### 3.4. Heurística de selección de arcos representativos.

Debido a la naturaleza estocástica del sistema de transporte una red puede exhibir diferentes distribuciones de velocidad para un mismo arco y patrón de demanda (Cascetta, E, 1989), por lo que los arcos representativos podrían variar dependiendo del periodo  $T$  de recolección de datos. Lo anterior no implica cambios en la forma del MFD ya que este no es sensible a la demanda (Geroliminis & Sun, 2011). Por tanto, la heurística propuesta aborda este problema construyendo una lista de arcos en orden de prioridad, donde la prioridad significa qué tan importante es monitorear el flujo y la densidad de ese arco para estimar un MFD cercano al real utilizando la matriz de agrupamiento. Para decidir la prioridad de un arco, primero se calcula la representatividad acumulada de este arco en todas las réplicas según la Ecuación 9 y el número de veces que el arco fue representativo según la Ecuación 10.

$$RT_n = \sum_{z=1}^Z RP_{n,z} \quad (9) \quad Conteo_n = \sum_{z=1}^Z I_{n,z} \quad (10) \quad I_{n,z} = \begin{cases} 1 & \text{Si } RP_{n,z} > 0 \\ 0 & \text{Si } RP_{n,z} = 0 \end{cases} \quad (11)$$

Donde,  $RT_n$  es la representatividad acumulada del arco  $n$ ,  $Conteo_n$  es el número de veces que el arco fue representativo luego de las  $Z$  repeticiones y  $I_{n,z}$  es una variable dummy que adquiere el valor de uno si el arco  $n$  fue representativo en la repetición  $z$ . Los valores de  $RT_n$  se utilizan para ordenar los arcos de mayor a menor. El arco que tenga un mayor valor de  $RT_n$  es el más importante según este criterio y se le asigna la posición uno. Para todos los arcos de la lista se almacena su posición como una nueva variable la cual se denomina *Ranking de Representatividad*. De forma independiente al proceso anterior,  $Conteo_n$  se utiliza para ordenar los arcos de mayor a menor. El arco que tenga mayor valor de  $Conteo_n$  es el más importante según este criterio y se le asigna la

posición uno. Para todos los arcos de la lista se almacena su posición como una nueva variable la cual se denomina *Ranking de Conteo*. Transformar  $RT_n$  y  $Conteo_n$  en un ranking permite realizar una suma lineal entre ambas variables al hacerlas comparables. La suma de  $RT_n$  y  $Conteo_n$  para cada arco  $n$  da lugar a la variable *Orden*. Mientras menor sea el valor de *Orden* más importante será obtener los datos de flujo y densidad de ese arco para construir el MFD estimado.

La metodología descrita es insensible a la topología de la red, ya que sus propiedades no son consideradas en el algoritmo de agrupación o en la heurística de selección. Sin embargo, es necesario que en el área delimitada exista un MFD bien definido. Esta condición supone una limitante de este trabajo ya que si luego de equipar los arcos representativos con sensores fijos que midan flujo y densidad existen modificaciones importantes en la topología de la red, entonces existirá un nuevo MFD.

#### 4. RESULTADOS Y DISCUSIÓN

La metodología propuesta se evaluó en una red ficticia de 6x6 vías unidireccionales de dos pistas cada una, en donde las 36 intersecciones se encuentran reguladas por un semáforo. La operación en esta red se simuló utilizando AIMSUN, en la cual se ingresaron 18 escenarios con diferentes patrones y niveles de demanda (Guaman, 2022). Para calcular los errores de estimación del MFD utilizando los arcos representativos identificados con la metodología propuesta, se construyó el MFD real y el estimado para cada escenario. El MFD real es construido con todas las trayectorias de los vehículos de la microsimulación y el MFD estimado es construido a partir del flujo y densidad del subconjunto de arcos definidos al aplicar la metodología propuesta y el proceso descrito en el apéndice A (Figura 2). El MFD estimado representa, de manera agregada, el comportamiento de todos los arcos de la red ya que el valor de flujo y densidad de cada arco representativo es multiplicado por un ponderador descrito en el apéndice A.

Los errores en el MFD estimado obtenidos con la metodología propuesta son mayores que los errores de otras metodologías. Por ejemplo, Ortigosa et al. (2014) informan que monitoreando el 5% de la red se logra un 3% de error en la estimación del nivel de congestión, mientras que esta metodología, monitoreando el 7% de la red, logra un 10% de error. Saffari et al. (2020) informa que monitoreando el 10% de la red logra un error cuadrático medio (*RMSE* por sus siglas en inglés) de 0.07, mientras que esta metodología, monitoreando el 7% de la red, logra un *RMSE* de 0.12. Sin embargo, los resultados son prometedores ya que los datos utilizados se obtienen a bajo costo y sin necesidad de infraestructura propia para la identificación de los arcos representativos, lo cual no ocurre en los otros trabajos. La metodología propuesta es más robusta que la de Ortigosa et al. (2014) ya que la magnitud del error varía poco al modificar los patrones de demanda. La heurística propuesta relaja los requisitos de Saffari et al. (2020) ya que no requiere una flota de observadores móviles propios distribuidos uniformemente por la red para cada intervalo de tiempo.



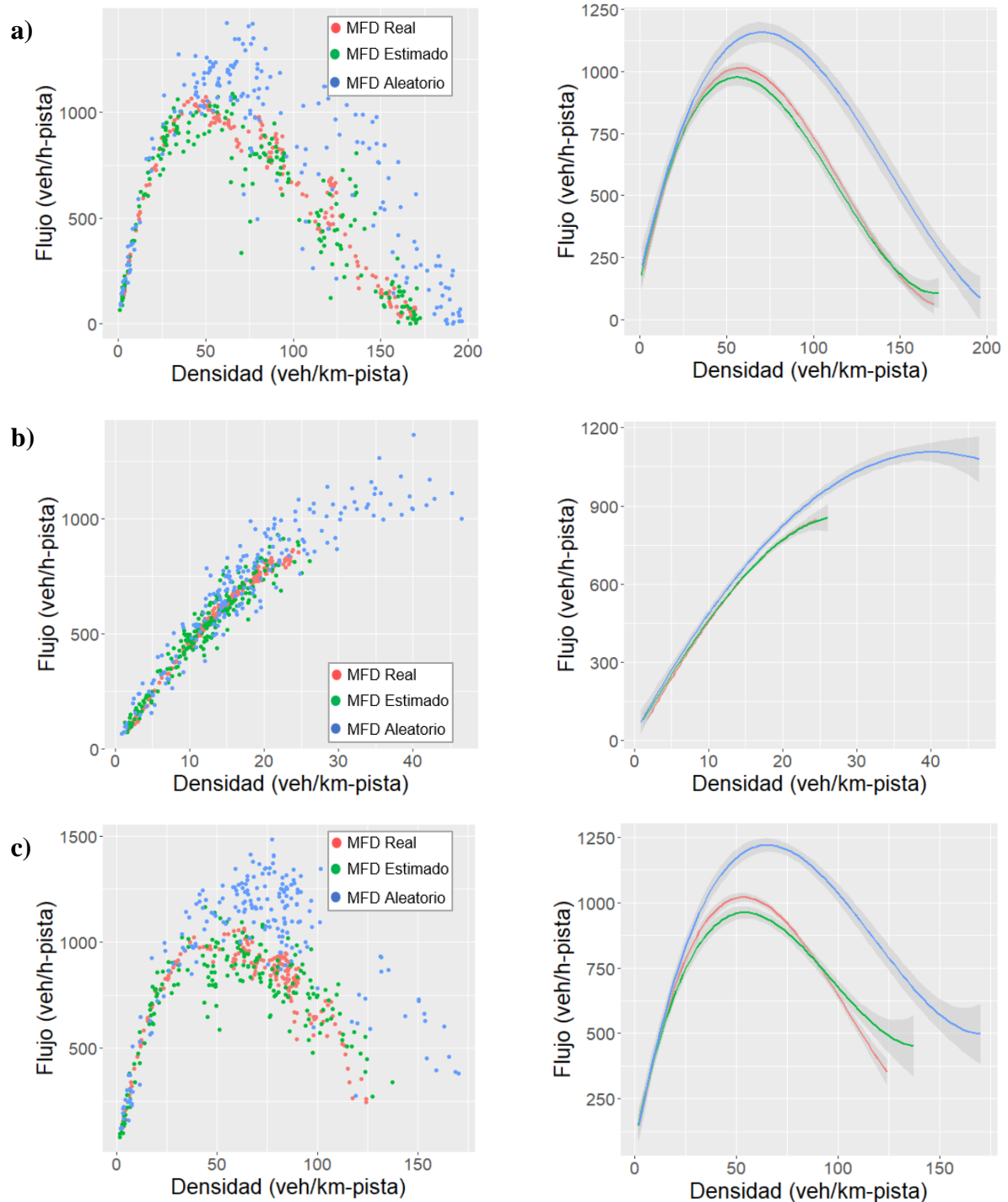


Figura 2. A la izquierda puntos del MFD real, MFD estimado y MFD aleatorio de las simulaciones de set de validación. A la derecha ajuste polinomial grado 3 de los MFD. a) Simulación 1 (Escenario base), b) Simulación 2 (Escenario nivel de demanda) y c) Simulación 3 (Escenario de variación de patrón)

Por otro lado, la metodología propuesta es sensible a la variación del algoritmo de agrupamiento. Kmeans generó los errores de estimación más bajos en comparación con algoritmos como Spectral,

Clique y Optics. Adicionalmente, se estudió la sensibilidad de la metodología a la variación en el número de arcos representativos. En la Figura 3 se muestran los errores calculados, y en base a ello se determina que a medida que aumenta el número de arcos representativos monitoreados, el error promedio disminuye. Para la red simulada, sobre el 14% de la red monitoreada el beneficio de incrementar el número de arcos es menor. Finalmente, se estudió la sensibilidad de la metodología a la diversidad de los datos utilizados. Para el análisis de esta sensibilidad se modificó la cantidad y diversidad de simulaciones. Los resultados muestran que el *RMSE* disminuye cuando el conjunto de entrenamiento incluye escenarios de variación del nivel de demanda.

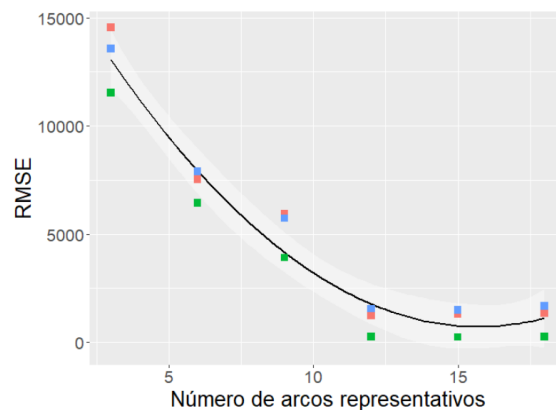


Figura 3: Comparación del RMSE según el número de arcos representativos.

Para comprobar que la metodología propuesta se puede aplicar con datos de un desarrollador de software, se realizó una aplicación parcial de la metodología. La metodología se aplica con una sola repetición del algoritmo de agrupamiento debido al costo de consultar el tiempo y la distancia de viaje en Google Directions API. El área de monitoreo consta de 71 arcos, los cuales fueron monitoreados durante 170 minutos. El costo del monitoreo fue de 127,80 USD.

El resultado de la aplicación parcial se muestra en la Figura 4, donde se puede apreciar que el número de arcos por grupo ( $RP_{n,z}$ ) es razonable y los arcos representativos de cada grupo parecen adecuados ya que sus características físicas y operativas son diferentes. La aplicación parcial muestra que con datos reales es posible realizar un buen agrupamiento en función de la velocidad de los vehículos en la red y que los arcos representativos resultantes tienen características diferentes. Monitorear los seis arcos seleccionados significa instalar sensores en solo el 8,5% de la red, lo que reduce considerablemente el costo de monitoreo.

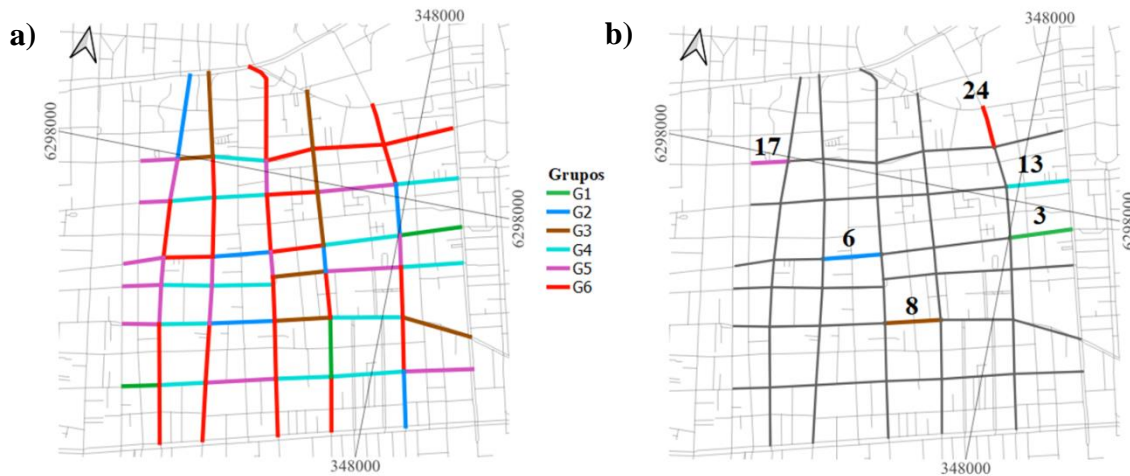


Figura 3 Resultados del agrupamiento: a) Grupos, b) Arcos representativos de cada grupo y cantidad de arcos representados.

## 5. CONCLUSIONES E INVESTIGACIONES FUTURAS

La metodología propuesta permite identificar un grupo reducido de arcos que, una vez conocido su flujo y densidad, pueden ser utilizados para estimar un MFD cercano al real. La metodología propuesta no requiere sensores de punto fijo o una flota de observadores móviles propia para identificar los arcos representativos.

La heurística propuesta considera tanto la cantidad de veces que un arco fue representativo como la cantidad de arcos que representó. Los criterios utilizados permiten considerar diferentes escenarios generados debido a la variación de la demanda. Si bien los errores de estimación del MFD obtenidos con la heurística propuesta son superiores a los de otras metodologías, tiene la ventaja de requerir menos datos y requerimientos de infraestructura que ellas, siendo esto último un aspecto clave en muchas ciudades.

Durante este trabajo, surgieron interrogantes a abordar en trabajos futuros, como evaluar el desempeño de la metodología al variar la topología y el tamaño de la red que se está monitoreando. Finalmente, la metodología propuesta tiene limitantes relativas a la representatividad en el tiempo ante lo cual surgen las preguntas ¿Es posible mejorar la estimación del estado del tráfico de la red actualizando la representatividad de cada arco monitoreado en tiempo real? ¿Es posible estimar el estado del tráfico de la red en periodos con diferentes MFD modificando la representatividad de cada arco monitoreado?

## AGRADECIMIENTOS

A la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) de Ecuador por el financiamiento recibido para la obtención del grado de magister de Diego Guamán.

**REFERENCIAS**

- Ambühl, L., & Menendez, M. (2016). Data fusion algorithm for macroscopic fundamental diagram estimation. *Transportation Research Part C: Emerging Technologies*, 71, 184–197. <https://doi.org/10.1016/j.trc.2016.07.013>
- Buisson, C., & Ladier, C. (2009). Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record*, 2124, 127–136. <https://doi.org/10.3141/2124-12>
- Coates, A., & Ng, A. Y. (2012). *Neural Networks: Tricks of the Trade* (G. Montavon, G. B. Orr, & K.-R. Müller, Eds.; Vol. 7700). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-35289-8>
- Courbon, T., & Leclercq, L. (2011). Cross-comparison of macroscopic fundamental diagram estimation methods. *Procedia - Social and Behavioral Sciences*, 20, 417–426. <https://doi.org/10.1016/j.sbspro.2011.08.048>
- Daganzo, C. F., & Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, 42(9), 771–781. <https://doi.org/10.1016/j.trb.2008.06.008>
- Doig, J., Gayah, V., & Cassidy, M. (2013). Inhomogeneous flow patterns in undersaturated road networks. *Transportation Research Record*, 2390, 68–75. <https://doi.org/10.3141/2390-08>
- Du, J., Rakha, H., & Gayah, V. v. (2016). Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. *Transportation Research Part C: Emerging Technologies*, 66, 136–149. <https://doi.org/10.1016/j.trc.2015.08.015>
- Duda, R. O., Hart, P. E., & Stork, D. G. (1995). *Pattern Classification and Scene Analysis* 2nd ed. Part 1: Pattern Classification.
- Edie, & L. C. (1963). Discussion of traffic stream measurements and definitions. In OECD (Ed.), J. Almond (Ed.), *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow* (pp. 139–154). OECD.
- Gayah, V., & Dixit, V. (2013). Using mobile probe data and the macroscopic fundamental diagram to estimate network densities. *Transportation Research Record*, 2390, 76–86. <https://doi.org/10.3141/2390-09>
- Geroliminis, N., & Daganzo, C. F. (2007). Macroscopic modeling of traffic in cities.

- Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9), 759–770.  
<https://doi.org/10.1016/j.trb.2008.02.002>
- Geroliminis, N., & Sun, J. (2011). Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, 45(3), 605–617.  
<https://doi.org/10.1016/j.trb.2010.11.004>
- Guamán, D. (2022). Identificación de arcos representativos para la estimación del diagrama fundamental macroscópico. Pontificia Universidad Católica de Chile.  
<https://repositorio.uc.cl/handle/11534/63714>
- Herrera, J. C., Work, D. B., Herring, R., Ban, X. (Jeff), Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4), 568–583.  
<https://doi.org/10.1016/j.trc.2009.10.006>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Ji, Y., & Geroliminis, N. (2012). On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological*, 46(10), 1639–1656.  
<https://doi.org/10.1016/j.trb.2012.08.005>
- Keyvan-Ekbatani, M., Papageorgiou, M., & Papamichail, I. (2013). Urban congestion gating control based on reduced operational network fundamental diagrams. *Transportation Research Part C: Emerging Technologies*, 33, 74–87. <https://doi.org/10.1016/j.trc.2013.04.010>
- Knoop, V., van Erp, P. B. C., Leclercq, L., & Hoogendoorn, S. P. (2018). Empirical MFDs using Google Traffic Data. 2018 21st International Conference on Intelligent Transportation Systems (ITSC).  
[https://doi.org/10.0/Linux-x86\\_64](https://doi.org/10.0/Linux-x86_64)
- Leclercq, L., Chiabaut, N., & Trinquier, B. (2014). Macroscopic Fundamental Diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological*, 62, 1–12.  
<https://doi.org/10.1016/j.trb.2014.01.007>
- Nagle, A., & Gayah, V. (2013, October 6). A method to estimate the macroscopic fundamental diagram using limited mobile probe data. *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013)*.

- Nagle, A. S., & Gayah, V. v. (2014). Accuracy of Networkwide Traffic States Estimated from Mobile Probe Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2421(1), 1–11. <https://doi.org/10.3141/2421-01>
- Nisbet, R., Miner, G., & Yale, K. (2017). Handbook of statistical analysis and data mining applications. In *Handbook of Statistical Analysis and Data Mining Applications*. <https://doi.org/10.1016/c2012-0-06451-4>
- Ortigosa, J., & Menendez, M. (2014). Traffic performance on quasi-grid urban structures. *Cities*, 36, 18–27. <https://doi.org/10.1016/j.cities.2013.08.006>
- Ortigosa, J., Menendez, M., & Tapia, H. (2014). Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich. *EURO Journal on Transportation and Logistics*, 3(3–4), 245–266. <https://doi.org/10.1007/s13676-013-0034-0>
- Saffari, E., Yildirimoglu, M., & Hickman, M. (2020). A methodology for identifying critical links and estimating macroscopic fundamental diagram in large-scale urban networks. *Transportation Research Part C: Emerging Technologies*, 119. <https://doi.org/10.1016/j.trc.2020.102743>
- van den Haak, P., Bakri, T., van Katwijk, R., Emde, M., & Snelder, M. (2018, January 7). Validation of Google floating car data for applications in traffic management. *Transportation Research Board 97th Annual Meeting*.
- Zhang L, Yuan Z, Yang L. & Liu Z. (2020) Recent developments in traffic flow modelling using macroscopic fundamental diagram, *Transport Reviews*, 40:6, 689-710. DOI: 10.1080/01441647.2020.1738588.
- Zockaie, A., Saberi, M., & Saedi, R. (2018). A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. *Transportation Research Part C: Emerging Technologies*, 86, 245–262. <https://doi.org/10.1016/j.trc.2017.11.017>

## APÉNDICE

### A. Construcción del MFD estimado.

Para construir el MFD estimado se deben recolectar los datos de flujo y densidad para los  $J$  arcos representativos cada  $\Delta t$  unidades de tiempo. Los  $J$  arcos representativos son los reconocidos utilizando la metodología propuesta. Es necesario ponderar los valores de flujo y densidad correspondientes a cada periodo  $\Delta t$  de cada arco representativo con un valor que indique la cantidad de arcos que representa. El ponderador se calcula utilizando la representatividad acumulada de los arcos representativos y las Ecuaciones 1 y 2.

$$RS = \sum_{j=1}^J RT_j \quad (1) \quad P_j = \frac{RT_j}{RS} \cdot N \quad (2)$$

Donde,  $j$  es un arco representativo,  $RT_j$  es la representatividad acumulada de un arco  $j$ ,  $RS$  es la sumatoria de la representatividad acumulada de todos los arcos representativos,  $N$  es el número de arcos de la red y  $P_j$  es la ponderación que se debe utilizar para la construcción del MFD estimado para cada arco representativo. El flujo promedio ponderado y la densidad promedio ponderada quedan definidas por las Ecuaciones 3 y 4. La construcción del MFD mediante el procedimiento descrito es válido si todos los arcos de la red tienen las mismas propiedades físicas.

$$q_{\Delta t}^w = \frac{\sum_j P_j \cdot q_{j,\Delta t} \cdot l_j}{\sum_j l_j} \quad (3) \quad k_{\Delta t}^w = \frac{\sum_j P_j \cdot k_{j,\Delta t} \cdot l_j}{\sum_j l_j} \quad (4)$$