



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

IDENTIFICACIÓN DE ARCOS REPRESENTATIVOS PARA LA ESTIMACIÓN DEL DIAGRAMA FUNDAMENTAL MACROSCÓPICO

DIEGO GERMAN GUAMAN MATUTE

Tesis para optar al grado de
Magister en Ciencias de la Ingeniería

Profesor Supervisor:
JUAN CARLOS HERRERA MALDONADO

Santiago de Chile, abril, 2022



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

IDENTIFICACIÓN DE ARCOS REPRESENTATIVOS PARA LA ESTIMACIÓN DEL DIAGRAMA FUNDAMENTAL MACROSCÓPICO

DIEGO GERMAN GUAMAN MATUTE

Tesis presentada a la Comisión integrada por los profesores:

JUAN CARLOS HERRERA MALDONADO

LUIS IGNACIO RIZZI CAMPANELLA

RODRIGO EDUARDO FERNANDEZ AGUILERA

VALERIA PAZ HERSKOVIC MAIDA

Para completar las exigencias del grado de
Magister en Ciencias de la Ingeniería

Santiago de Chile, abril, 2022

A mi esposa quien me acompañó en esta aventura y que con el soporte de su amor me ha ayudado a cumplir un sueño más.

AGRADECIMIENTOS

Este trabajo es el cierre del primer paso de una vida dedicada a la ciencia y la enseñanza, por tanto, quiero agradecer en primer lugar a quienes me guiaron en esta vocación. Agradezco a mi mamá por siempre recordarme la capacidad que Dios me dio y que ella cultivó. También agradezco a mi cuñado y amigo Juan por escuchar mis dudas y transmitirme tranquilidad.

Quiero agradecer a quienes han velado y ayudado en nuestro bienestar en Chile. A mi papá y a mis suegros por su respaldo respecto a las decisiones tomadas en estos años. Agradezco además a nuestros nuevos amigos por estrecharnos su mano en Santiago.

Quiero agradecer a Juan Carlos Herrera por su guía durante todo el programa y por permitirme explorar durante esta investigación, así como por sus recomendaciones en la misma.

Además, agradezco a la "Secretaría de Educación Superior, Ciencia, Tecnología e Innovación" (SENESCYT) de Ecuador por financiar mis estudios de maestría.

INDICE GENERAL

AGRADECIMIENTOS	iv
INDICE GENERAL.....	v
INDICE DE TABLAS	viii
INDICE DE FIGURAS.....	ix
RESUMEN.....	xi
ABSTRACT.....	xiii
1. INTRODUCCIÓN	1
1.1. Hipótesis.....	2
1.2. Objetivos	3
1.3. Contenido y estructura de la tesis.....	3
2. REVISIÓN BIBLIOGRÁFICA	6
2.1. Diagrama fundamental macroscópico	6
2.2. Inhomogeneidad en la estimación del MFD.....	9
2.3. Métodos para construir un MFD	12
2.3.1. Métodos experimentales.....	12
2.4. Datos de velocidad y tiempo de viaje de <i>Google</i>	18
2.5. Agrupamiento de arcos en redes.....	22

2.6.	Cálculo de errores de estimación del MFD	24
2.7.	Algoritmos de agrupamiento	26
2.8.	Resumen	29
3.	METODOLOGÍA	31
3.1.	Delimitación del área y periodo de recolección de datos.	33
3.2.	Matriz de datos	34
3.3.	Agrupamiento de los arcos	37
3.4.	Heurística de selección de arcos representativos.....	39
3.5.	Construcción del MFD estimado.....	43
4.	VALIDACIÓN DE LA METODOLOGÍA	47
4.1.	Simulación.....	48
4.1.1.	Escenarios de simulación	50
4.1.1.1.	Escenario base	50
4.1.1.2.	Variación de escenarios.....	53
4.2.	Velocidad media en cada arco	55
4.3.	Resultados de la heurística de selección.....	56
4.4.	Evaluación de otros algoritmos de agrupamiento.	64
4.5.	Sensibilidad del método ante la variación del número de arcos monitoreados. 66	
4.6.	Sensibilidad del método ante la cantidad y diversidad de los datos.....	69

5.	APLICACIÓN PARCIAL DE LA METODOLOGÍA EN SANTIAGO	71
5.1.	Delimitación del área y periodo de recolección de datos	71
5.2.	Extracción de datos	72
5.3.	Agrupamiento de los arcos	74
6.	CONCLUSIONES	77
	REFERENCIAS	80
	REFERENCIAS DIGITALES	86

INDICE DE TABLAS

Tabla 1 Matriz de extracción de datos para un área de N arcos, T minutos	35
Tabla 2 Matriz de datos para el algoritmo de agrupamiento para N arcos	36
Tabla 3 Tabulación de resultados del agrupamiento.....	37
Tabla 4 Matriz de entrenamiento de Z repeticiones de un área de N arcos	39
Tabla 5 Distribución de los viajes entre todos los pares origen destino	51
Tabla 6 Heurística de selección de arcos representativos	57
Tabla 7 Errores en la estimación del MFD para cada día del set de validación	60
Tabla 8 Errores en la estimación del MFD con el escenario base para todos los métodos de agrupación	65
Tabla 9 Ponderación P_k de los arcos representativos para cada K arcos representativos	67
Tabla 10 Ponderación P_k de los arcos representativos para cada set de entrenamiento ..	70

INDICE DE FIGURAS

Figura 1 Relación trapezoidal entre producción y acumulación en un arco	7
Figura 2 Histéresis del MFD en una red de autopistas durante 3 días	10
Figura 3 Función de ocupación de la red en 5 instantes de tiempo	11
Figura 4 Relación entre la velocidad media, flujo medio y densidad media de una red mediante el MFD.....	13
Figura 5 Flujo promedio vs. Densidad promedio en dos días con 4 niveles de demanda.	17
Figura 6 MFD obtenidos mediante selección aleatoria de 100 arcos de una red de 1260 arcos. Cada color representa una selección aleatoria de arcos.....	22
Figura 7 Diagrama de flujo de la metodología de selección de arcos representativos. ...	32
Figura 8 Estructura y comportamiento de la red recomendada para aplicar la metodología	33
Figura 9 División de la red de un área en 6 grupos.....	38
Figura 10 Ejemplo del conteo del número de veces que los arcos son representativos...41	
Figura 11 Ejemplo de la representatividad acumulada de los arcos, Conteo.....	41
Figura 12 Diagrama de flujo de la heurística de selección	42
Figura 13 Ejemplo de arcos representativos del estado de tráfico de la red.	43
Figura 14 Representación de la red simulada	49
Figura 15 División temporal del total de viajes entre cada par origen destino	51
Figura 16 MFD de la primera réplica del escenario base.	52
Figura 17 MFD real de la zona monitoreada considerando las 15 simulaciones del set de entrenamiento	54

Figura 18 a) Conteo y b) representatividad de los arcos del área monitoreada. Fuente: Elaboración propia	57
Figura 19 MFD de los días del set de validación	58
Figura 20 A la izquierda puntos del MFD real y estimados del 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)	61
Figura 21 Desviaciones del flujo y la acumulación estimada. a) día 1 (Escenario base), b) día 2 (Escenario nivel de demanda) y c) día 3 (Escenario de variación de patrón).	63
Figura 22 MFD real y eMFD construidos con diferentes métodos de agrupamiento para un día del escenario base.....	65
Figura 23 Grupos de arcos resultantes con datos de tabla 1. a)Kmeans, b)Spectral.....	66
Figura 24 Comparación del número de arcos representativos. Fuente: Elaboración propia	68
Figura 25 Comparación según el set de entrenamiento. Fuente: Elaboración propia.....	70
Figura 26 Área de monitoreo y arcos monitoreados en la comuna de Santiago	72
Figura 27 Campo de velocidades del área monitoreada.....	74
Figura 28 Método gráfico del codo. Fuente: Elaboración propia	74
Figura 29 Resultado del agrupamiento: a) Grupo de pertenencia de cada arco, b) Representatividad de cada arco	75

RESUMEN

El problema de la congestión de tráfico puede ser abordado mediante la implementación de acciones en tiempo real como el cambio en la programación semafórica. Para esto, es necesario un modelo simple y agregado de la operación de la red que permita conocer su estado de tráfico, como por ejemplo el diagrama fundamental macroscópico (MFD por sus siglas en inglés). El MFD es una relación funcional simple entre la tasa a la que terminan los viajes al interior de una zona y la cantidad de autos en la misma. El MFD requiere para su construcción datos del flujo y la densidad de todos los arcos (vías) de la red. La recolección de los datos no representa mayor esfuerzo, sin embargo, su costo es alto. Por tanto, es importante desarrollar una metodología que permita reducir costos, por ejemplo, al disminuir el número de arcos para los cuales se debe conocer los datos de flujo densidad.

La metodología propuesta emplea el algoritmo de *Kmeans* para agrupar arcos y seleccionar el más representativo de cada grupo. El agrupamiento se realiza utilizando algunas variables descriptivas de la distribución de la velocidad de cada arco. El agrupamiento se repite varias veces y cada repetición utiliza la distribución de velocidad de cada arco de un periodo distinto. Para seleccionar los arcos más representativos, se construye una matriz con el resultado de todas las repeticiones del agrupamiento. La matriz resultante constituye el dato de entrada para una heurística de selección basada en la cantidad de veces que un arco resultó representativo y la cantidad de arcos que representó.

La metodología fue validada mediante la comparación del MFD real de una red simulada en el software Aimsun y el MFD estimado con los datos de los arcos más representativos reconocidos con la metodología propuesta. La diferencia entre el MFD real y el MFD estimado es mayor a la obtenida con otras metodologías que tienen el mismo objetivo, sin embargo, la diferencia adicional se justifica dado que la metodología propuesta requiere datos que son distintos a los de las otras metodologías y pueden obtenerse a un costo

menor. Además, se verificó la sensibilidad de la metodología ante la variación de tres parámetros: el algoritmo de agrupamiento, la cantidad de arcos representativos y la diversidad del set de entrenamiento. Se encontró que la diferencia entre los MFD se reduce a medida que aumenta la proporción de la red monitoreada y que esta reducción no es lineal. Además, la diferencia se reduce cuando los datos provienen de periodos con distintos niveles y patrones de demanda.

Otras metodologías que tienen el mismo objetivo se basan en información que no se encuentra disponible en la mayoría de las ciudades de países en vías de desarrollo. Este trabajo presenta un método para identificar los arcos representativos para la estimación del MFD que requiere solo la velocidad promedio de cada arco de la red. Esta velocidad es un dato de tráfico que se encuentra fácilmente disponible dada el creciente uso de apps como Google Maps, Waze, y Uber, entre otros.

La principal conclusión de este trabajo es que empleando la metodología propuesta la velocidad promedio de cada arco de la red (como la consultada en *Google Directions API*) permite identificar un grupo acotado de arcos que luego de medir su flujo y densidad sirven para estimar un MFD cercano al real. Así, la construcción y uso del MFD se vuelve factible con una inversión económica asequible.

Palabras Claves: MFD, diagrama fundamental macroscópico, velocidad media espacial, agrupamiento de arcos, tecnología GPS.

ABSTRACT

Traffic congestion problem can be addressed through the implementation of real-time actions such as traffic light programming changes. For this, it is necessary a simple model of the network operation which allows being updated on the traffic levels, for instance the macroscopic fundamental diagram (MDF). An MDF is a simple functional relation between the rate into which the inland travel ends in a specific zone and the quantity of cars within it. The MDF requires for its construction the flow data and the density of all arcs (track) of the network. The data collection does not represent major efforts; however, its cost is high. Therefore, it is important to develop a methodology which allows reducing costs, for instance, when diminishing the number of arcs from which flow-density data needs to be known. The suggested methodology employs the Kmeans algorithm in order to cluster arcs and select the most representative of each cluster. The clustering is conducted by using some descriptive variables of speed distribution of each arc. The clustering is repeated several times and each repetition uses the speed distribution of each arc from a different period. In order to select the most representative arcs a matrix is built with the results of all the clustering repetitions. The resulting matrix constitutes the input data for a heuristic selection based on how many times a single arc was representative and the number of arcs it represented.

The methodology was validated by comparing both a real MDF of a simulated network on Aimsun software and the estimated MDF with the data of the most representative arcs recognized by the proposed methodology. The difference between the real MDF and the estimated MDF is bigger than the obtained using other methodologies with the same goal, nevertheless, the additional difference is justified given that the proposed methodology requires distinct data from other methodologies and they can be obtained at a minor cost. Besides, it was verified the sensibility of this methodology towards the variation of three parameters: the clustering algorithm, the quantity of representative arcs and the diversity of the training group. It was found that the difference between MDFs is reduced to the

increase of the arcs monitored and that this reduction is not lineal. Also, the difference is reduced when the data is collected from periods of different levels and demand patterns.

Other methodologies that have the same objective are based on information that is not available in most of the cities of developing countries. This work introduces a method to identify the representative arcs for the MFD estimation that requires only the average speed of each arc of the network. This speed is traffic data that is easily available given the crescent use of apps such as Google Maps, Waze, Uber, among others.

The main conclusion of this work is that employing the proposed methodology, the average speed of each network's arc (as the consulted on Google Directions API) permits to identify a delimited group of arcs that after measuring their flow and density can be used to estimate a MDF. This manner the construction and use of MFD turns feasible with an affordable economic investment.

Key words: MFD, macroscopic fundamental diagram, mean speed, arc clustering, GPS technology.

1. INTRODUCCIÓN

La congestión de tráfico es un problema a nivel global. Según Yang et al. (2020), esta situación es especialmente importante en países de ingresos medios como los mercados emergentes, debido a que la infraestructura vial y las políticas públicas no avanzan al mismo ritmo que el incremento del número de vehículos (Goodwin, 1996). Es por tanto necesario desarrollar herramientas en estos países para abordar el problema de la congestión de tráfico.

Los modelos que buscan predecir la operación de la red pueden aportar a la disminución de la congestión de tráfico. Sin embargo, esta estrategia es sensible a errores de los datos de entrada (Daganzo, 1998). Una alternativa a los modelos predictivos es el monitoreo e implementación de acciones en tiempo real (Geroliminis y Daganzo, 2007). Sin embargo, es necesario un modelo simple y agregado de la operación de la red que permita monitorear y ejecutar acciones en tiempo real.

El diagrama fundamental macroscópico (MFD por su sigla en inglés) proporciona un modelo parsimonioso para describir el rendimiento de tráfico a nivel de red (Saffari et al, 2020), pues relaciona la acumulación de vehículos en la red (densidad) con su productividad (flujo).

Pese a que medir el flujo y densidad de cada arco es sencillo, esto no se realiza en toda la red por el costo generado por la instalación y mantenimiento de los sensores necesarios. Además, existen costos asociados al almacenaje y procesamiento de la gran cantidad de los datos recolectados (Saffari et al, 2020). Una alternativa a los sensores son los datos de observadores móviles. Sin embargo, para ser empleados se deben asumir algunas limitaciones como las condiciones especiales del modo de transporte (datos provistos por flotas de vehículos comerciales o vehículos de transporte público), baja tasa de penetración, privacidad, entre otros. Dada la problemática de la recolección de datos y la utilidad de conocer el MFD de una zona, estudios como el de Ortigosa et al. (2014),

Zockaie et al. (2018) y Fei et al. (2007) buscan identificar los arcos más importantes de la red para caracterizar su MFD. Sin embargo, las metodologías existentes tienen impedimentos relacionados con la cantidad de sensores de punto fijo y el tamaño y propiedades de los observadores móviles necesarios para sus metodologías.

La velocidad media espacial de los vehículos que circulan en un arco es una medida agregada para la cual no se necesita conocer el valor de la velocidad de cada vehículo. Este valor puede ser recolectado a través de plataformas como *Google* que disponen de productos comerciales como *Google Directions API* que entregan el tiempo de viaje entre dos puntos. La ventaja de la API es la disponibilidad inmediata y a bajo costo con la cual se pueden consultar datos.

La falta de datos para caracterizar un MFD y posteriormente utilizarlo para gestionar la red en tiempo real es un problema. Servicios como *Google Directions API*, que proveen el tiempo de viaje entre dos puntos, abren la posibilidad de usar sus productos para reducir los costos asociados a la estimación del MFD en ciudades poco equipadas. Debido a que la velocidad media en un arco no es directamente una de las variables que caracterizan el MFD, es necesario elaborar una metodología para su uso correcto con el objetivo de reducir el costo en la estimación del MFD.

La pregunta de investigación que aborda esta tesis es ¿Cómo reducir la cantidad de arcos a monitorear para construir un MFD en una zona poco equipada a partir de los datos de velocidad media en cada arco?

1.1. Hipótesis

La hipótesis de este trabajo es que la velocidad media de cada arco de una red, a lo largo de un periodo, permite identificar un grupo acotado de arcos que, luego de conocer su flujo y densidad, sirven para estimar un MFD cercano al real.

1.2. Objetivos

El objetivo general de esta tesis consiste en desarrollar una metodología que utilice la velocidad de los vehículos en cada arco de una red para identificar aquellos arcos que, luego de conocer su flujo y densidad, sirven para estimar un MFD cercano al real, considerando que la red no se encuentra equipada en su totalidad con sensores de punto fijo y de la cual se desconoce a priori el MFD real. Dentro de los objetivos específicos se encuentran los siguientes:

- a) Determinar una forma o transformación adecuada para utilizar la velocidad de los autos en un arco como dato básico para un algoritmo de agrupamiento.
- b) Encontrar un algoritmo de agrupamiento que se adapte a los datos disponibles y entregue grupos de arcos adecuados.
- c) Definir una heurística que permita determinar los arcos que deben ser monitoreados para construir un MFD cercano al real, utilizando los resultados del agrupamiento.
- d) Estimar la sensibilidad de la metodología frente a la variación de sus parámetros.
- e) Aplicar parcialmente la metodología al caso de una zona en Santiago de Chile.

1.3. Contenido y estructura de la tesis

La bibliografía revisada durante este trabajo muestra que la estimación del MFD de un área es factible si se conocen los datos de flujo y densidad de todos los arcos de una red. La recolección de estos con equipos adecuados, como los sensores de punto fijo, no se realiza con frecuencia debido a su elevado costo derivado de la necesidad de implementar numerosos puntos de medición para alcanzar resultados de buena calidad. En la actualidad se han desarrollado métodos alternativos de estimación, basados en la reducción de los puntos de medición o en el uso de observadores móviles, sin embargo, aquellas requieren

recursos comparables al primer método mencionado lo que no permite emplearlos en ciudades que cuentan con recursos limitados.

La presente tesis utiliza datos de velocidad de los vehículos que circulan por una red durante un periodo para calcular, variables de la distribución de velocidad media para cada arco evaluado. Las variables que se deben calcular son: mínimo, cuartil uno, mediana, media, cuartil tres, máximo, rango intercuartil 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 de *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). 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 diferente. Para obtener distintas matrices se debe consultar la velocidad media de los arcos 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 construyen 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 es posible colocar en orden de prioridad los arcos que se deben monitorear para estimar el MFD de la red estudiada. El resultado de la metodología es por tanto una lista ordenada por prioridad de los arcos que se deben monitorear para luego en base a el monitoreo del flujo y la densidad de estos construir un MFD estimado. La cantidad de arcos que se deben monitorear constituye una restricción presupuestaria.

Para validar la metodología descrita se elaboró un modelo de micro simulación de una red tipo Manhattan, para la cual se construyó el MFD real y el MFD estimado en base a la metodología propuesta. Luego, se comparó la diferencia entre el MFD real y el MFD estimado y se concluyó que la diferencia es mayor a la obtenida con otras metodologías que tienen el mismo objetivo, sin embargo, la diferencia adicional se justifica dado que la metodología propuesta requiere datos que son distintos a los de las otras metodologías y pueden obtenerse a un costo menor. Además, se verificó la sensibilidad de la metodología ante la variación de tres parámetros: el algoritmo de agrupamiento, la cantidad de arcos representativos y la diversidad del set de entrenamiento. Se encontró que la diferencia entre los MFD se reduce a medida que aumenta la proporción de la red monitoreada y que esta reducción no es lineal. Además, la diferencia se reduce cuando los datos provienen de periodos con distintos niveles y patrones de demanda.

Por último, se presenta un ejemplo de aplicación parcial de la metodología en Santiago de Chile, utilizando como fuente de datos *Google Directions API*. El objetivo de este ejemplo es que la metodología pueda ser reproducida para otras zonas de Santiago de Chile y otras ciudades con características similares.

Las siguientes secciones de este documento se estructuran en cinco capítulos. En el capítulo dos se presenta una revisión bibliográfica que establece el marco teórico de esta investigación. En el capítulo tres se expone la metodología propuesta. En el capítulo cuatro se describe la validación de la metodología y se estudia la sensibilidad de esta ante la variación de sus parámetros. En el capítulo cinco se describe un ejemplo de aplicación parcial de la metodología para una zona de Santiago. Por último, en el capítulo seis se resumen los hallazgos y se revisan las conclusiones derivadas de los objetivos planteados junto con la exposición de las futuras líneas de investigación.

2. REVISIÓN BIBLIOGRÁFICA

En este capítulo se presenta una revisión bibliográfica que establece el marco teórico de esta investigación. Primero se introduce el concepto de MFD, algunas de sus propiedades, métodos utilizados para su estimación y cálculo de errores. Luego, se describe la velocidad como una variable agregada del tráfico. Además, se resume los trabajos realizados con objetivos semejantes al de esta tesis. Al final de este capítulo se presenta el marco teórico referente al método de agrupamiento que fue utilizado en la metodología.

2.1. Diagrama fundamental macroscópico

Una alternativa de solución a los problemas de congestión vehicular es utilizar un mecanismo basado en el monitoreo e implementación de acciones en tiempo real (Geroliminis & Daganzo, 2007). Para lo anterior, es necesario contar con un modelo sencillo y que reproduzca de forma robusta el estado de tráfico de una red. Las necesidades expuestas indican que un modelo macroscópico de tráfico es el adecuado ya que utiliza indicadores de un grupo de vehículos en lugar de modelar cada uno de ellos.

Las variables de tráfico a considerar en un modelo macroscópico deben ser relevantes en la representación de las limitaciones físicas que causa el hacinamiento, así como estar correlacionadas con el desempeño de tráfico de la zona (Geroliminis & Daganzo, 2007). Las variables de un modelo macroscópico deben poder medirse de forma directa en campo (Geroliminis & Daganzo, 2007). Bajo las consideraciones mencionadas, el modelo gráfico propuesto por Godfrey (1969) que relaciona la velocidad promedio de viaje en una red con la densidad de la misma, sería el adecuado.

El modelo gráfico de Godfrey (1969) fue reintroducido por Daganzo (2007), el cual propuso una relación macroscópica unimodal no negativa, $G(\mathbf{m})$, entre la tasa a la cual terminan los viajes en una zona y la cantidad de vehículos en la misma (\mathbf{m}). La existencia de $G(\mathbf{m})$ supone condiciones de estado estacionario. A la función $G(\mathbf{m})$ se le conoce como

diagrama fundamental macroscópico (MFD por su sigla en inglés). Conocer el MFD de una zona es importante porque la tasa a la que terminan los viajes está relacionada con la accesibilidad de esta (Geroliminis & Daganzo, 2008).

El MFD se fundamenta en que para un arco (enlace o calle) la teoría del tráfico de ondas cinemáticas muestra que la producción de viajes (distancia total recorrida por todos los vehículos que viajan en un arco en un periodo) está relacionada con la acumulación (cantidad de vehículos que viajan en un arco al inicio de un periodo) mediante una función trapezoidal como se muestra en la Figura 1. La forma trapezoidal cóncava del MFD es producto de la relación entre el diagrama fundamental (DF) de los arcos que componen el área y la configuración de las intersecciones.

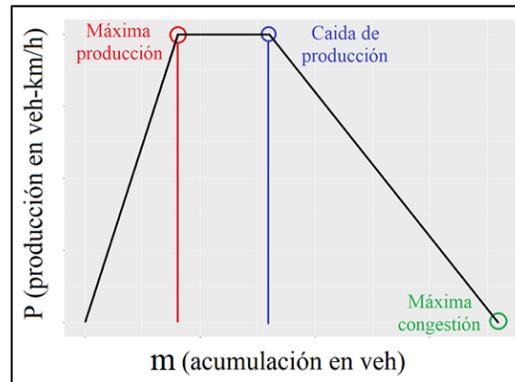


Figura 1 Relación trapezoidal entre producción y acumulación en un arco

Fuente: Elaboración propia

Considerando que una red es el conjunto de arcos interconectados por nodos (intersecciones), Geroliminis & Daganzo (2007) postularon que si en una red los vehículos llegan a todos los enlaces a la misma velocidad (carga homogénea) y si la congestión se distribuye uniformemente entonces la producción de toda la red es la suma de la producción individual de cada arco y esta puede expresarse como una función de la acumulación en toda la red. La relación descrita se muestra en la Ecuación 1.

$$P = \sum_n P_n = \sum_n Q_n(m_n) \cong Q\left(\sum_n m_n\right) \quad (1)$$

Donde,

$P = P(t, t + \Delta t)$ es la producción o distancia total recorrida por todos los vehículos que viajan entre t y Δt en la red. P tiene unidades $\frac{veh-km}{h}$.

$m_n = m_n(t)$ es el número de vehículos que viajan en un arco n en el momento t .

Q_n es la función de producción de viajes para cada arco n .

Q es la función de producción de viajes para toda la red.

La función $Q(\mathbf{m})$ derivada en la Ecuación 1 es semejante a la función $G(\mathbf{m})$ ya que son dependientes del número de vehículos en una zona. Sin embargo, la primera representa la producción de viajes en la zona en tanto que la segunda es la tasa de finalización de viajes en la zona. Geroliminis & Daganzo (2007) demostraron que bajo condiciones de estacionalidad la relación entre producción y tasa de finalización de viajes viene dada por el inverso de la distancia media de los viajes según se muestra en la Ecuación 2.

$$\frac{O}{P} = \frac{1}{\bar{L}} \quad (2)$$

Donde,

\bar{L} es la distancia media de los viajes considerando todos los pares origen destino.

O es la tasa de finalización de viajes en la zona.

P es la producción de viajes en la zona.

Con la Ecuación 1 y 2 es posible construir el MFD de una zona conociendo la acumulación y producción de la misma. Para validar las ecuaciones presentadas, Geroliminis & Daganzo (2007) simularon una arteria vial de Los Ángeles, California y una zona del centro de San Francisco. La validación, además de ratificar los hallazgos, encontró robustez bajo diferentes patrones de demanda y concluyeron que mientras más grande el

sistema analizado es mejor la estimación. Adicionalmente, un experimento en Yokohama con datos reales reveló la existencia de un MFD en un área urbana (Geroliminis & Daganzo, 2008). 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). El MFD es ampliamente ocupado debido a que es una herramienta semejante al diagrama fundamental (DF) de un arco. El MFD refleja el estado de tráfico de un grupo de arcos en el cual existe flujo interrumpido.

La existencia del MFD sugiere que las redes urbanas grandes se comportan de manera predecible e independiente de los patrones de origen y destino lo cual permite ejecutar políticas públicas relacionadas a la gestión de tráfico en tiempo real como el control perimetral con el objetivo de mejorar la accesibilidad al interior de la red. Por ejemplo, en una red semaforizada el MFD permitiría modificar la configuración de los semáforos perimetrales, con el objetivo de reducir el flujo de ingreso de vehículos a la zona evitando de esta forma que la productividad al interior se reduzca.

2.2. Inhomogeneidad en la estimación del MFD

La forma del MFD es una propiedad de la red, por lo tanto, tiene un máximo bien definido e insensible a la demanda (Geroliminis & Sun, 2011). Para que el MFD exista como una curva bien definida y reproducible, Geroliminis & Daganzo (2007) establecieron que es necesario que la congestión esté distribuida de forma uniforme en la red, que la red se cargue de forma homogénea y que los patrones de demanda sean constantes o varíen de forma lenta. Sin embargo, las condiciones óptimas para encontrar un MFD bien definido rara vez suelen estar presentes en la realidad, por lo que estudios empíricos basados en datos medidos en campo encuentran varios valores de flujo medio para la misma densidad media. Este fenómeno se llama histéresis.

La histéresis se puede generar debido a que el tráfico es dinámico (Mariotte et al., 2017), a la heterogeneidad en la distribución espacial de los vehículos (Daganzo et al., 2012) o a

la operación en los sistemas semafóricos como prioridad al transporte público (Castrillon & Laval, 2017). Teniendo en cuenta la histéresis, Ambühl et al. (2020) indicaron que el MFD descrito por Geroliminis & Daganzo (2007) es el límite máximo que se puede observar para el par flujo medio; densidad media de la red. Estudios como Buisson & Ladier (2009), Tilg et al. (2020), Ambühl et al. (2021) reconocen que pueden existir estados de tráfico bajo el MFD. Una de las formas para que los estados de tráfico se ubiquen sobre el MFD es eliminando la histéresis mediante el enrutamiento dinámico de los vehículos, es decir, que la congestión se distribuya de forma homogénea por toda la red y que los aumentos en la demanda sean lo suficientemente lentos como para que el enrutamiento dinámico pueda generarse de forma adecuada. El enrutamiento dinámico se refiere a que los vehículos actualizan su ruta a medida que avanzan en su viaje.

En redes de autopistas los cuellos de botella y la ausencia de rutas alternas generan congestión solo en algunos arcos y no en otros, por lo que la falta de homogeneidad en este tipo de redes se traduce en un MFD con histéresis (Doig et al., 2013), como el que se muestra en la Figura 2. En una red urbana donde existen rutas alternas los conductores potencialmente pueden evitar los cuellos de botella mediante el enrutamiento dinámico, obteniendo como resultado la propagación de la congestión de forma homogénea (Doig et al., 2013).

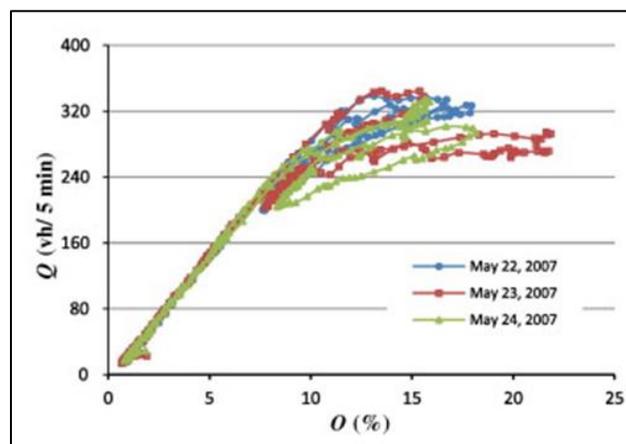


Figura 2 Histéresis del MFD en una red de autopistas durante 3 días

Fuente: Geroliminis & Sun, 2011

Geroliminis & Sun (2011) demostraron que no es necesario una estricta homogeneidad de los distintos estados de tráfico en cada arco de la red para observar un MFD bien definido. En su lugar proponen que, si la distribución espacial de la densidad de vehículos en la red es la misma en dos intervalos de tiempo diferentes que tienen una densidad promedio igual, entonces ambos periodos tienen el mismo flujo medio. Por ejemplo, la Figura 3 representa un día en el cual se reconocieron 5 instantes de tiempo en los cuales la ocupación promedio de los sensores de punto fijo es del 35%, entonces un MFD bien definido se reconoce si la densidad en cada uno de los cinco instantes de tiempo se distribuye como una única función. En la Figura 3 se puede ver que para cada instante de tiempo el gráfico de barras tiene la misma forma, por tanto, existe un MFD bien definido.

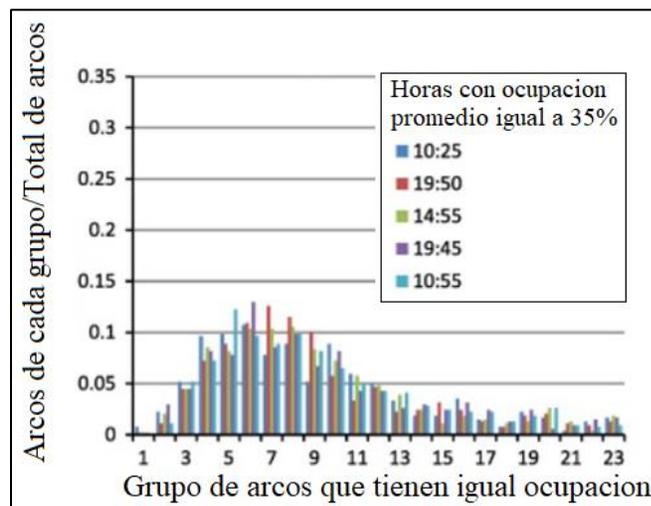


Figura 3 Función de ocupación de la red en 5 instantes de tiempo

Fuente: Geroliminis & Sun, 2011

En redes del mundo real la estricta homogeneidad es difícil de encontrar, incluso construir una simulación que cumpla esta condición es difícil. Por tanto, la existencia de un MFD bajo condiciones de inhomogeneidad es importante.

2.3. Métodos para construir un MFD

Los métodos para construir un MFD se dividen en analíticos y experimentales. Los métodos analíticos se basan en la teoría de onda cinemática LWR (Lighthill & Whitham, 1955; Richards, 1956) y en el diagrama fundamental triangular. Los métodos analíticos proporcionan una estimación del MFD sin requerir mediciones o simulaciones. Sin embargo, se limitan a situaciones con valores de flujo medio iguales en todos los enlaces (Leclercq et al., 2014). Por otra parte, los métodos experimentales se basan en mediciones, lo cual supone restricciones presupuestarias, errores provenientes de la ubicación de observaciones eulerianas (observaciones en un punto fijo del espacio) o baja representatividad de observaciones lagrangianas (observaciones de una secuencia de puntos que se mueven en el espacio).

2.3.1. Métodos experimentales

Los métodos experimentales para la construcción del MFD varían de acuerdo al tipo de medición disponible. Las mediciones disponibles pueden ser: trayectorias de todos los vehículos de la red, trayectorias de algunos vehículos (observadores móviles), mediciones de sensores de punto fijo en toda la red y mediciones de sensores de punto fijo en una porción de arcos de la red.

Para cualquier medición disponible, un punto del MFD representa un estado de tráfico de la red para un intervalo de tiempo Δt . El intervalo Δt debe ser mayor o igual a un ciclo semafórico para garantizar que el estado de tráfico estimado sea representativo del arco y no de una fase del semáforo.

Los estudios basados en micro simuladores (Courbon & Leclercq, 2011; Ji & Geroliminis, 2012) permiten conocer todas las trayectorias de los vehículos, por tanto, es posible conocer la tasa de finalización de viajes (\mathbf{O}) y la cantidad de vehículos (\mathbf{m}) mediante el análisis de curvas acumuladas de inicio y finalización de viajes en la zona para cada

intervalo de tiempo. Conociendo \mathbf{O} y \mathbf{m} se puede construir el MFD de la zona. Utilizando los mismos datos de trayectorias de todos los vehículos y las definiciones generalizadas de Edie (1963) es posible construir el MFD en función de la producción (\mathbf{P}) y cantidad de vehículos (\mathbf{m}) de la zona para cada intervalo de tiempo. Esta última forma de construir el MFD toma en cuenta la dificultad de medir la tasa de finalización de viajes en la vida real. Sin embargo, aún requiere conocer todas las trayectorias de los vehículos.

Cuando las mediciones disponibles son solamente las trayectorias de algunos vehículos (observadores móviles) es posible estimar los estados de tráfico de una red en base a los datos del tiempo y distancia que viajan estos observadores móviles. Varios estudios (Ambühl & Menendez, 2016; Du et al., 2016; A. Nagle & Gayah, 2013; A. S. Nagle & Gayah, 2014) estimaron el MFD utilizando las trayectorias de los observadores móviles y las definiciones generalizadas de Edie (1963). Este método asume que es conocida la tasa de penetración de los observadores móviles con respecto al total de vehículos. Cuando se dispone de este tipo de datos es posible calcular la acumulación de vehículos, el flujo medio, densidad media y velocidad media de la red según las Ecuaciones 3, 4, 5 y 6.

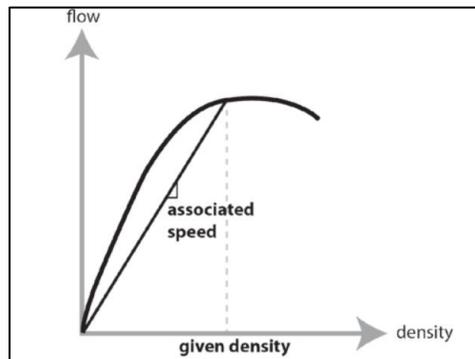


Figura 4 Relación entre la velocidad media, flujo medio y densidad media de una red mediante el MFD

Fuente: Gayah & Dixit (2013)

$$\hat{q} = \frac{M_p \bar{d}_p}{\rho L \Delta t} \quad (3)$$

$$\hat{k} = \frac{M_p \bar{t}_p}{\rho L \Delta t} \quad (4)$$

$$\hat{v} = \frac{\bar{d}_p}{\bar{t}_p} \quad (5)$$

$$\hat{m} = \frac{M_p \bar{t}_p}{\rho \Delta t} \quad (6)$$

Donde,

\hat{q} es el flujo medio.

\hat{k} es la densidad media.

\hat{v} es la densidad media.

\hat{m} es la acumulación de vehículos en la zona.

ρ es la tasa de penetración de los observadores móviles con respecto al total de vehículos.

Δt es el periodo para el cual se calcula el flujo y la densidad media de la red. Su rango tiene una cota inferior igual al tiempo de ciclo de la red semafórica y una cota superior definida por la mínima resolución temporal deseada.

M_p es el número de observadores móviles en la red durante Δt .

\bar{t}_p es la media del tiempo viajado por los observadores móviles en la red durante Δt .

\bar{d}_p es la media de la distancia viajada por los observadores móviles en la red durante Δt .

L es la longitud total de pistas-km de la red.

La estimación del MFD mediante observadores móviles según la metodología revisada requiere que la tasa de penetración y la longitud media del viaje de todos los vehículos de la red sea constante en el tiempo y en todos los pares origen destino. Las condiciones mencionadas limitan la aplicación de esta metodología ya que la longitud media de viaje de cada par O D es distinta. Para relajar los requerimientos mencionados se puede estimar una tasa de penetración equivalente para el cálculo de la densidad media de la red. La tasa de penetración equivalente es la media armónica de las tasas de penetración de los observadores móviles entre cada par origen destino (O D) ponderadas por el tiempo de viaje del observador móvil entre cada par O D (Du et al., 2016).

$$\rho_{eq}^k = \frac{M_p \bar{t}_p}{\sum_{o,d} \frac{t'_{o,d}}{\rho_{o,d}}} \quad (7)$$

$$\rho_{eq}^q = \frac{M_p \bar{d}_p}{\sum_{o,d} \frac{d'_{o,d}}{\rho_{o,d}}} \quad (8)$$

Donde,

$\rho_{eq}^{k,q}$ es la tasa de penetración equivalente para el cálculo de la densidad o flujo medio.

$t'_{o,d}$ es el tiempo de viaje total de todos los observadores móviles del par O, D durante Δt .

$d'_{o,d}$ es la distancia total de todos los observadores móviles del par O, D durante Δt .

ρ_{eq}^q y ρ_{eq}^k sustituyen a ρ en la Ecuación 3 y Ecuación 4 respectivamente. La inclusión de estas tasas de penetración equivalentes mejora las estimaciones del MFD cuando la matriz O, D es heterogénea (Du et al., 2016).

La tasa de penetración (ρ) es un factor importante cuando se construye un MFD con trayectorias de un número limitado de observadores móviles. Contar con una flota de observadores móviles representativa de los viajes que se realizan en la red y con una alta tasa de penetración es uno de los retos de los estudios experimentales.

Pese a que ρ debe ser conocido a priori, es posible estimar un valor $\hat{\rho}$ mediante la razón entre el conteo de vehículos de sensores de punto fijo (M^d) y el número de observadores móviles que pasaron por el sensor (M_p^d) para cada Δt . Estimar ρ mediante sensores de punto fijo es una limitante de este método ya que implica que la red debe estar equipada de forma adecuada con estos sensores de punto fijo y además contar con las trayectorias de la flota de observadores móviles.

Si solo se considera las mediciones de sensores de punto fijo es posible estimar el MFD de una zona que está equipada en todos sus arcos. Varios estudios (Ambühl & Menendez, 2016; Buisson & Ladier, 2009; Courbon & Leclercq, 2011; Knoop et al., 2018; Leclercq

et al., 2014; Saffari et al., 2020) utilizaron las definiciones del promedio ponderado del flujo y de la densidad de toda la red descritas por Daganzo & Geroliminis (2008) para construir el MFD de un área. El promedio ponderado del flujo de toda la red en un periodo Δt y el promedio ponderado de la densidad de toda la red en un periodo Δt se calculan según las Ecuaciones 9 y 10.

$$q_{\Delta t}^w = \frac{\sum_n q_{n,\Delta t} * l_n}{\sum_n l_n} \quad (9)$$

$$k_{\Delta t}^w = \frac{\sum_n k_{n,\Delta t} * l_n}{\sum_n l_n} \quad (10)$$

Donde,

$q_{\Delta t}^w$ es el flujo promedio ponderado.

$k_{\Delta t}^w$ es la densidad promedio ponderada.

n es un arco de la red.

l_n es la longitud del arco n .

Δt es el periodo para el cual se calcula el flujo y la densidad promedio ponderada.

Utilizar la Ecuación 9 es razonable debido a que puede ser vista como una normalización de la producción ya que su numerador es la producción. Por otra parte, considerando las definiciones generalizadas (Edie, 1963) la Ecuación 9 es el promedio espacial del flujo.

Utilizando las ecuaciones 9 y 10 Geroliminis & Daganzo (2008) construyeron un MFD bien definido del área de Yokohama el cual se muestra en la Figura 5. No todos los arcos de Yokohama fueron monitoreados con sensores de punto fijo capaces de recolectar flujo y densidad, por lo que los promedios ($q_{\Delta t}^w$ y $k_{\Delta t}^w$) fueron representativos de aquellos arcos monitoreados. El MFD elaborado con los datos de algunos arcos se conoce como

MFD reducido, ya que solo considera los arcos monitoreados. Otros estudios (Keyvan-Ekbatani et al., 2013; Ortigosa & Menendez, 2014) también han utilizado el MFD reducido ya que este es suficiente para aplicaciones del MFD como la reducción de la tasa de ingreso de vehículos a una red para evitar el descenso de su productividad (control perimetral).

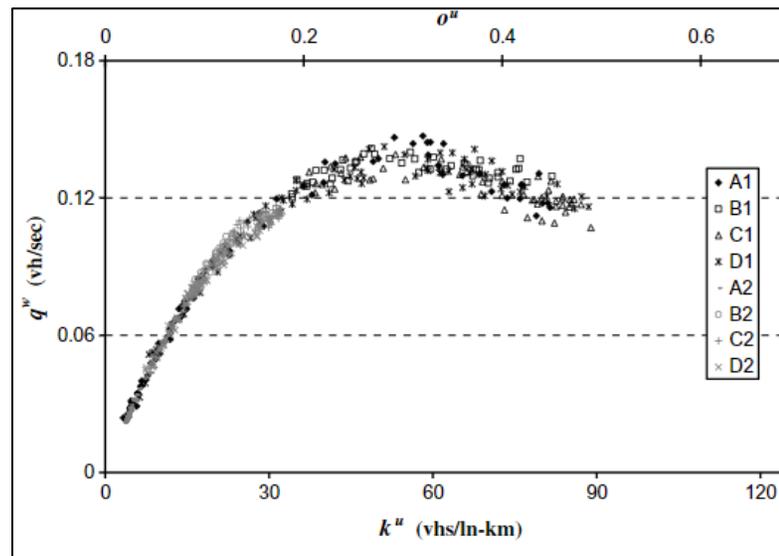


Figura 5 Flujo promedio vs. Densidad promedio en dos días con 4 niveles de demanda.
Fuente: (Geroliminis & Daganzo, 2008)

El problema del método anterior es que la velocidad, flujo y densidad observadas siempre serán sensibles a la ubicación del sensor de punto fijo debido a los efectos de la cola en la intersección. Por tanto, este método no puede capturar los distintos estados de tráfico que suceden de forma simultánea en distintos lugares de un arco al mismo tiempo (comportamiento dinámico del tráfico en un arco).

Para capturar el comportamiento dinámico del tráfico en un arco observado se puede sustituir la medición de los sensores de punto fijo por las trayectorias de los vehículos que circulan por ese mismo arco. Utilizando las definiciones generalizadas de Edie (1963) y las trayectorias de los vehículos en cada arco, $q_{\Delta t}^w$ y $k_{\Delta t}^w$ quedan definidas por las

ecuaciones 11, 12 y 13. El método descrito elimina la necesidad de conocer la media de la distancia viajada por los observadores móviles y la tasa de penetración para cada par origen destino ya que las trayectorias se discretizan para cada arco.

$$q_{\Delta t}^w = \frac{\sum_n(\hat{d}_{\Delta t,n})}{\Delta t * L} \quad (11)$$

$$k_{\Delta t}^w = \frac{\sum_n(\hat{t}_{\Delta t,n})}{\Delta t * L} \quad (12)$$

$$L = \sum_n l_n \quad (13)$$

Donde,

$\hat{d}_{\Delta t,n}$ es la distancia viajada por todos los vehículos en el arco n durante Δt .

$\hat{t}_{\Delta t,n}$ es el tiempo total viajado por todos los vehículos en el arco n durante Δt .

l_n es la longitud del arco n.

Este método fue utilizado por Paipuri et al. (2020) para construir un MFD reducido. El método puede ser utilizado si se cuenta con las trayectorias de los vehículos en pocos arcos.

2.4. Datos de velocidad y tiempo de viaje de Google

Los datos recolectados por observadores móviles que utilizan un sistema GPS son secuencias de coordenadas registradas a lo largo de su recorrido en intervalos de tiempo establecidos. El uso de datos de observadores móviles es algo cada vez más común, ya que estos son utilizados con diferentes objetivos como los sistemas de ayuda de explotación de flota de transporte público (SAEF). El problema de los observadores móviles es su representatividad con respecto al resto de vehículos.

A partir de 2009 *Google Maps* muestra las condiciones de tráfico en tiempo real de los arcos de algunas ciudades del mundo. Las condiciones de tráfico se muestran en una escala de colores que indica la velocidad media en cada arco (*Google, 2009_1*). Lo anterior indica que *Google* en 2009 contaba con una base de datos de observadores móviles lo suficientemente robusta como para lanzar al mercado un producto comercial como *Google Maps* con información de tráfico en tiempo real. Los datos recolectados son de propiedad de los usuarios de las aplicaciones de *Google* y estos son almacenados después de que los usuarios hayan aceptado compartirlos.

Para estimar las condiciones de tráfico en tiempo real, *Google* recibe datos anónimos generados por un equipo con GPS que describen qué tan rápido se está moviendo un usuario. Luego, se combina la velocidad de todos los usuarios del arco. Los miles de usuarios que aportan sus datos permiten generar la información que se muestra de forma gratuita en *Google Maps* (*Google, 2009_2*).

La velocidad promedio de los vehículos que recorren un arco es una medida agregada que tiene asociada un único estado de tráfico. La velocidad por sí sola no permite conocer cuál es este estado; para ello es necesario contar con el diagrama fundamental del arco o alguna de las otras dos variables fundamentales del tráfico. Estudios como Gayah & Dixit (2013) sugieren que se necesita una gran cantidad de vehículos de prueba para determinar con precisión las condiciones del tráfico basadas en flujo y densidad en toda la red, pero para estimar las condiciones del tráfico en base a la velocidad se necesita una tasa de penetración menor. Los resultados de Herrera et al (2010) sugieren que una penetración de 2 a 3% de teléfonos celulares en la población de conductores es suficiente para proporcionar mediciones precisas de la velocidad del flujo de tráfico en autopistas. Nagle & Gayah (2014) apoyan la conclusión anterior al encontrar que un error del 10% en la estimación de las condiciones del tráfico en base a la velocidad en la red se puede lograr con tasas de penetración de observadores móviles menores al 5%. La tasa de penetración

de los observadores móviles de *Google* estimada en Los Países Bajos está muy por encima del 10% (Knoop et al., 2018).

La recolección y procesamiento de datos que realiza *Google* en base a observadores móviles generan productos comerciales como *Google Maps app*, *Google Routes API*, entre otros. *Google Directions API* permite extraer información del tráfico en tiempo real. La API permite realizar consultas de la distancia y tiempo de viaje entre dos puntos, el tiempo de viaje puede ser transformado a velocidad utilizando la distancia. La ventaja de la API es la facilidad con la cual se pueden consultar datos, aunque por ahora su resolución es cercana al minuto y cada consulta tiene un costo de 0.01 USD por cada par origen destino consultado (Developers, 2021).

El uso de los productos comerciales de *Google* brinda, desde el lado de los usuarios, la posibilidad de conocer la ruta más corta entre un par origen destino considerando la congestión en tiempo real y la congestión pronosticada. Desde el lado del planificador, conocer la velocidad del tráfico en toda la red permite aplicar políticas públicas para aminorar los efectos de la congestión, como el control perimetral (Geroliminis & Daganzo, 2007). Además, dado que *Google Maps* actualiza constantemente la mejor ruta entre un par O D los usuarios de la APP pueden modificar su recorrido según las condiciones del tráfico de la red lo cual ayuda a que la congestión se distribuya uniformemente, siendo esta una condición favorable para la existencia de un MFD (Geroliminis & Daganzo, 2007).

Para validar los datos de *Google*, van den Haak et al. (2018) utilizaron series de tiempo de velocidad que generan los observadores móviles de *Google* para estimar el estado de tráfico de 2200 ubicaciones. El estudio concluyó que los datos anónimos y agregados que recolecta *Google* son de calidad aceptable con un impacto limitado y pueden conducir a reducciones sustanciales de costos, ya que la diferencia en la estimación de los estados de tráfico con respecto a sensores de punto fijo era adecuada para escenarios de gestión de tráfico (van den Haak et al., 2018)

Dado el potencial de los datos de *Google* estos han sido utilizados para la planificación y control de la gestión en corredores de autopistas en San Francisco (Sana et al., 2017). Los datos también han sido utilizados para derivar un MFD en Ámsterdam (Knoop et al., 2018), aunque para ello se requirió supuestos relacionados con la tasa de penetración de los usuarios de *Google Maps*, el flujo máximo de los arcos de la red y simplificaciones de la geometría de los arcos.

Los datos de *Google* para ambos casos fueron entregados a los investigadores mediante el programa *Better Cities* de *Google* el cual mantiene un convenio con La Autoridad de Transporte del Condado de San Francisco y El Instituto de Soluciones Metropolitanas Avanzadas de Ámsterdam, en ambos casos se busca minimizar la congestión de tráfico, mejorar la seguridad vial y reducir la cantidad de dinero invertido en infraestructura.

Los datos que se utilizaron en los estudios anteriores no son parte de los productos comerciales de *Google* por lo que disponer de ellos para aplicaciones de gestión de tráfico en ciudades que no poseen convenios de desarrollo o comerciales con *Google* resulta muy complicado. Durante esta investigación se intentó acceder a datos semejantes a los utilizados por Knoop et al. (2018) mediante el Vínculo electrónico 1 pero no se recibió una respuesta favorable o una forma alternativa de conseguirlos. Por el contrario, el tiempo de viaje entre dos puntos si es un producto comercial de *Google Maps* y por tanto accesible a cualquier consultor, municipalidad o grupo de investigación.

2.5. Agrupamiento de arcos en redes

El agrupamiento de arcos en redes en el contexto del MFD se ha utilizado con dos objetivos distintos. El agrupamiento fue utilizado en primera instancia para delimitar vecindarios que tengan un MFD observable y, más recientemente, es utilizado para conformar grupos de arcos que permitan reducir el costo de monitoreo.

Pese a que medir el flujo y densidad de cada arco es sencillo, esto no se realiza en toda la red debido al costo que significa instalar y mantener los sensores necesarios. Además, existen costos asociados al almacenaje y procesamiento de la gran cantidad de datos recolectados (Saffari et al., 2020). Dadas las restricciones presupuestarias, es necesario escoger los arcos a monitorear de tal forma que estos sean representativos del estado de tráfico de toda la red. Si se monitorea un porcentaje de los arcos de una red de forma aleatoria los MFD resultantes no se alinearán en una única curva según se puede ver en la Figura 6.

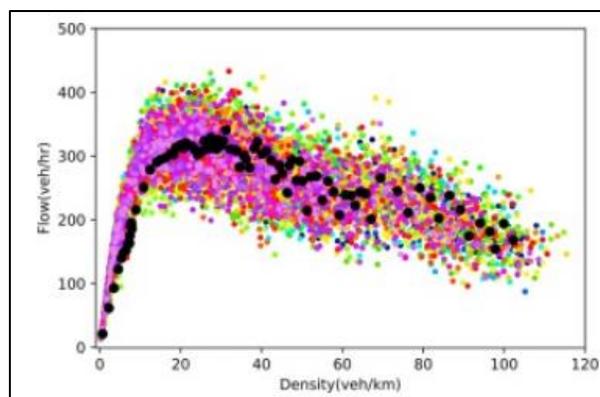


Figura 6 MFD obtenidos mediante selección aleatoria de 100 arcos de una red de 1260 arcos. Cada color representa una selección aleatoria de arcos.

Fuente: Saffari et al. (2020)

Ji & Geroliminis (2012) aplicaron algoritmos de agrupamiento a los arcos de una red con el objetivo de delimitar áreas con condiciones de tráfico semejantes (vecindarios). Sin embargo, al utilizar *Kmeans* el resultado fue grupos espacialmente dispersos los cuales no formaban un vecindario, pero si eran homogéneos en su estado de tráfico. Debido a que

su objetivo no era el reducir la cantidad de arcos a monitorear para construir un MFD, no existe información respecto a la calidad del agrupamiento. Ji & Geroliminis (2012) no tomaron en cuenta la variación del flujo y densidad en el tiempo de cada arco en la aplicación del algoritmo de agrupamiento ya que lo realizaron para un instante de tiempo específico. Para realizar el agrupamiento la metodología requería conocer el flujo y la densidad de todos los arcos de la red por lo que esta tiene limitantes semejantes a la estimación completa del MFD.

Keyvan-Ekbatani et al. (2013) elaboraron un modelo en Aimsun de una zona de la red urbana de Chania, Grecia, con el objetivo de realizar un control perimetral para reducir los efectos de la congestión en base a un MFD reducido. Para seleccionar los arcos que debían equiparse con sensores realizaron una inspección visual del modelo para identificar los arcos con los que iniciaba la congestión y otros que muestren un estado de congestión promedio. Keyvan-Ekbatani et al (2013) concluyeron que incluso con solo el 5% de arcos monitoreados es posible realizar una estrategia de control perimetral. En el estudio no se calculó los errores de estimación del MFD reducido respecto al MFD real y tampoco se describió una heurística de selección de arcos replicable en otras redes.

Ortigosa et al. (2014), por su parte, establecen que el 5% de los arcos de una red deben ser monitoreados para obtener un error menor al 3% en la estimación de densidad promedio. La selección de los arcos a monitorear se basó en un problema de optimización que busca maximizar la precisión del MFD reducido resultante del monitoreo de p arcos. La metodología utilizada por Ortigosa et al. (2014) es sensible a los distintos patrones de demanda que fueron probados en la validación de sus resultados, por lo que la selección de los p arcos es válida solo para el patrón de demanda bajo el cual se estimó. La metodología propuesta requiere conocer a priori el MFD de la zona por lo que esta tiene limitantes semejantes a la estimación completa del MFD. De forma semejante Zockaie et al. (2018) plantea un problema de optimización para encontrar los mejores arcos para instalar sensores de punto fijo y las mejores trayectorias de observadores móviles para

estimar el MFD de una red. La metodología de Zockaie et al. (2018) también requiere para su aplicación conocer el flujo y densidad de todos los arcos de la red.

La primera metodología que utiliza solamente datos de algunos observadores móviles para identificar arcos representativos del estado de tráfico de una red fue propuesta por Saffari et al. (2020). Para encontrar los enlaces representativos se aplica un análisis de componentes principales (PCA) sobre el flujo y la densidad de observadores móviles en cada arco para cada periodo Δt . El PCA reconoce N componentes principales que luego son comparadas con los arcos de la red para encontrar aquellos arcos más parecidos mediante un algoritmo de búsqueda (Saffari et al., 2020). Una vez que los arcos representativos fueron reconocidos, en ellos se colocaron medidores de punto fijo para obtener los datos de flujo y densidad mediante los cuales se puede reconstruir el MFD real utilizando los vectores de características del PCA. En la metodología de Saffari et al. (2020), es necesario que la tasa de penetración de los observadores móviles sea homogénea en todos los arcos. Si no se cumple la condición anterior los flujos y densidades relativas tendría como referencia distintos flujos y densidades reales. Esta metodología requiere una flota de observadores móviles de gran tamaño y con una distribución en la red que satisfaga la tasa de penetración constante, limitando de esta forma su aplicación.

2.6. Cálculo de errores de estimación del MFD

Las métricas para cuantificar la precisión en la estimación del MFD se han dividido en aquellas que comparan el MFD real de la zona con el MFD estimado por alguna metodología y aquellas que comparan el MFD real con el MFD reducido de la misma. En ambos casos se asume disponible el MFD real.

Las métricas utilizadas en estudios previos (Nagle & Gayah, 2014, Saffari et al., 2020 y Du et al., 2016) para evaluar la precisión del MFD estimado son el error cuadrático medio (RMSE) de las estimaciones de flujo medio y densidad media para cada intervalo de

estudio Δt . Estas métricas indican qué tan parecido es el MFD estimado al real mediante las Ecuaciones 14 y 15.

$$RMSE(q) = \sqrt{\frac{\sum_{\Delta t} (\hat{q}_{\Delta t} - q_{\Delta t})^2}{I}} \quad (14)$$

$$RMSE(k) = \sqrt{\frac{\sum_{\Delta t} (\hat{k}_{\Delta t} - k_{\Delta t})^2}{I}} \quad (15)$$

Donde,

I es el número de intervalos Δt .

$\hat{q}_{\Delta t}$ y $\hat{k}_{\Delta t}$ son el flujo y la densidad media estimada para cada Δt .

$q_{\Delta t}$ y $k_{\Delta t}$ son el flujo y la densidad media real para cada Δt .

Una tercera métrica permite cuantificar la precisión conjunta en la estimación del flujo medio y la densidad media teniendo en cuenta los valores extremos del MFD. Esta métrica se calcula mediante la Ecuación 16.

$$RMSE(q_c, k_j) = \sqrt{\frac{\sum_{\Delta t} \left[\left(\frac{\hat{q}_{\Delta t}}{q_c} - \frac{q_{\Delta t}}{q_c} \right)^2 + \left(\frac{\hat{k}_{\Delta t}}{k_j} - \frac{k_{\Delta t}}{k_j} \right)^2 \right]}{I}} \quad (16)$$

Donde,

q_c es flujo medio máximo.

k_j es la densidad máxima.

El MFD reducido, por construcción, no es comparable con el MFD real de la red utilizando las métricas anteriores, ya que este último considera todos los arcos. Sin embargo, existen aplicaciones, como el control perimetral, en las cuales conocer el MFD reducido es suficiente. Ortigosa et al. (2014) utilizaron una metodología para evaluar la precisión del

MFD reducido en base a la relación entre la densidad en un cierto intervalo de tiempo ($k_{\Delta t}$) y la densidad crítica (k_c). La relación entre las dos variables indica qué tan cerca está el sistema de llegar a un estado de congestión. Un MFD reducido idéntico al MFD real bajo este criterio será aquel que presente las mismas relaciones entre $k_{\Delta t}$ y k_c para un intervalo de tiempo dado. La relación de densidades para el MFD real cuando la red no está congestionada se define por la Ecuación 17 y cuando la red esta congestionada por la Ecuación 18.

$$R_{un} = \frac{k_{\Delta t} - k_c}{k_c} \quad \forall k_{\Delta t} \leq k_c \quad (17)$$

$$R_{con} = \frac{k_{\Delta t} - k_c}{k_j - k_c} \quad \forall k_{\Delta t} > k_c \quad (18)$$

Donde, k_c se obtiene promediando la densidad correspondiente al 1% de las medidas con mayor flujo medio y k_j se obtiene promediando el 1% de las medidas con mayor densidad media. Las relaciones de densidad para el MFD reducido son idénticas a las Ecuaciones 17 y 18 pero utilizando las densidades $k'_{\Delta t}$, k'_c y k'_j propias del MFD reducido, obteniendo de esta forma R'_{un} y R'_{con} . El error en la relación de densidades para un intervalo Δt se muestra en la Ecuación 19. Considerando I intervalos de tiempo Δt el error medio de la relación de densidades se muestra en la Ecuación 20. El error medio tiene unidades de puntos porcentuales ya que se basa en una relación de densidades.

$$\Delta R_{\Delta t}(MFD \text{ real}, MFD \text{ reducido}) = \begin{cases} R_{un} - R'_{un} & \text{si } k_{\Delta t} < k_c \\ R_{con} - R'_{con} & \text{si } k_{\Delta t} > k_c \end{cases} \quad (19)$$

$$\Delta R = \frac{1}{I} \sum_t |\Delta R_{\Delta t}(MFD \text{ real}, MFD \text{ reducido})| \quad (20)$$

2.7. Algoritmos de agrupamiento

El objetivo de los algoritmos de agrupamiento es descubrir 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 algoritmos jerárquicos y particionales. Los algoritmos jerárquicos establecen niveles de agrupación anidados que pueden iniciar desde N grupos con un elemento cada grupo o un grupo con N elementos y a partir de estos se unen varios grupos para formar un grupo con más elementos o se subdivide el grupo principal respectivamente. El proceso de agrupamiento y subdivisión se realiza varias veces para establecer una jerarquía dada por la cantidad de elementos agrupados en cada nivel. Por otra parte, el agrupamiento particional realiza una sola división de los elementos en varios grupos. Los dos tipos de algoritmos pueden aplicarse por separado, pero también puede utilizarse de forma conjunta. Por ejemplo, el algoritmo jerárquico permite al investigador definir el número K de grupos en los cuales se debe agrupar los elementos en base a un diagrama de témpanos o un dendograma de aglomeración y luego el algoritmo particional, por ejemplo, *Kmeans*, utiliza el número de grupos como entrada para su ejecución (Forina et al., 2002).

Los algoritmos particionales difieren en la elección de la función objetivo, modelos probabilísticos y heurísticas. El algoritmo particional más antiguo y uno de los más utilizados es *Kmeans*. Este algoritmo es casi tan antiguo como el agrupamiento de datos en sí. *Kmeans* fue aplicado desde 1956 en varias áreas del conocimiento y sigue siendo uno de los algoritmos más utilizados para la agrupación. 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 definir una métrica de distancia. Debido a que los datos de entrada influyen de forma importante en el resultado de la clasificación el algoritmo, tal como está planteado, tiende a converger a un mínimo local. Para incrementar la probabilidad de encontrar un mínimo global se debe aplicar el algoritmo en múltiples ocasiones con diferente partición inicial y elegir aquella partición

que entregue el menor error cuadrático. La métrica de distancia utilizada normalmente es la distancia euclidiana entre dos puntos, pero el algoritmo puede ser adaptado a otras (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. Siguiendo la definición anterior existe un grupo de algoritmos de agrupamiento particionales que se basan en la búsqueda de estas regiones densas. Algunos otros de estos algoritmos son BDSCAN y Optics. La ventaja de los algoritmos basados en densidad es su capacidad para tratar grupos que describen varias formas en el espacio de características y su desventaja es que tiene limitaciones en el manejo de datos de alta dimensión (Jain, 2010) porque cuando existen varias dimensiones los espacios de baja densidad son escasos.

Una variación de los algoritmos basados en densidad son los algoritmos subespaciales. Estos superan la limitante antes descrita ya que encuentran espacios de baja densidad en subespacios del espacio de alta dimensión (Jain, 2010). Un ejemplo de este grupo es el algoritmo CLIQUE.

Otra variante a los algoritmos basados en la densidad es el agrupamiento teórico de grafos o agrupamiento espectral. Estos algoritmos representan los puntos de datos como nodos en un grafo ponderado y los arcos del grafo se encuentran ponderados por la similitud entre nodos (Jain, 2010). Estos algoritmos dividen los nodos en dos subconjuntos (grupos) de manera que se minimice el tamaño del corte, donde el tamaño del corte es la suma de los pesos asignados a los bordes que conectan los grupos. Un ejemplo de este grupo es el algoritmo Spectral.

Una vez reconocidos los grupos se debe validar los resultados y determinar si los datos tienen una tendencia natural a la agrupación. Los algoritmos encontrarán agrupaciones en los datos independientemente si hay o no grupos claros (Smith & Jain, 1984).

Estudios como Ji y Geroliminis (2012), Saffari et al. (2020), Ambühl et al. (2021) han aplicado algoritmos de agrupamiento sobre los arcos de una red en base a sus características de tráfico. Sin embargo, no se ha utilizado el tiempo de viaje o la velocidad media espacial como una variable de agrupación para determinar grupos heterogéneos entre sí, pero homogéneos en su interior.

2.8. Resumen

Luego de que Daganzo (2005) reintrodujera la relación entre la tasa de finalización de viajes con la acumulación de vehículos en una zona, varios estudios se han desarrollado en las dos últimas décadas en torno al MFD. Inicialmente, los trabajos (Daganzo & Geroliminis, 2008; Geroliminis & Daganzo, 2007, 2008; Leclercq et al., 2014; A. S. Nagle & Gayah, 2014) se centraron en el desarrollo y estimación de la precisión de los métodos analíticos y experimentales para construir el MFD. Estudios más recientes buscan reducir la inversión necesaria para construir el MFD (Ortigosa & Menendez, 2014; Saffari et al., 2020; Zockaie et al., 2018). Sin embargo, estos trabajos necesitan información que no está disponible en ciudades poco equipadas.

Los datos de observadores móviles que son muestreados a través de dispositivos GPS, se han vuelto relevantes en el desarrollo de las ciudades, ya que el beneficio que obtiene la flota de observadores móviles se traduce en servicios comerciales de utilidad como *Google Maps*. El beneficio que obtienen los observadores móviles ha contribuido a superar barreras relacionadas con la privacidad de los datos. Sin embargo, esta privacidad ha pasado del observador móvil a los desarrolladores de las apps. Por tanto, la nueva fuente de información solo dispone de algunas características de los datos originales como la velocidad media de los arcos de una red.

La relación unívoca entre la velocidad media de los arcos de una red y su estado de tráfico podría ser aprovechada para reducir el esfuerzo en la estimación de un MFD. Por si sola la velocidad de los arcos de una red no define un único estado de tráfico.

Los trabajos indicados en la sección 2.5. buscan reducir la inversión para estimar un MFD y en todos se obtienen resultados alentadores, sin embargo, para ello utilizan datos que no se encuentran disponibles en la mayoría de ciudades. Por tanto, la motivación de este trabajo es que las ciudades poco equipadas pueden conocer velocidad media en cada arco a un bajo costo y por tanto pueden utilizar esta información para reducir la inversión necesaria para estimar un MFD cercano al real.

La metodología para la identificación de arcos representativos para la estimación del MFD que se presenta en el siguiente capítulo utiliza como dato base la velocidad media de los arcos de una red. Luego, se transforman estos datos en un conjunto de variables que describen para cada arco la distribución de la velocidad media observada en un periodo de monitoreo. El conjunto de variables es utilizado por el algoritmo de agrupamiento *Kmeans* para dividir los arcos de la red en grupos y encontrar un representante para cada uno. El agrupamiento se repite varias veces utilizando datos de varios periodos de monitoreo con el objetivo de conformar un conjunto de arcos para los cuales se construyen 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 es posible colocar en orden de prioridad los arcos y seleccionar tantos como el presupuesto disponible permita monitorear. Finalmente, se puede estimar el MFD en base al flujo y densidad de los arcos monitoreados.

La principal innovación de la metodología propuesta es que solo se requiere la velocidad media de los arcos durante varios periodos para encontrar los arcos representativos y estas velocidades pueden ser obtenidas fácilmente a bajo costo mediante *Google Directions API*. La principal diferencia con otras metodologías que tienen un objetivo similar y fueron expuestas en la sección 2.5 es que antes del monitoreo de arcos representativos no se requiere sensores de punto fijo o una flota de observadores móviles propios.

3. METODOLOGÍA

En este capítulo se describe la metodología propuesta que busca identificar arcos representativos para la estación del diagrama fundamental macroscópico. Esta metodología utiliza datos de velocidad de los vehículos que circulan por una red 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 diferente. Para obtener distintas matrices se debe consultar la velocidad media 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 construyen 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 es posible colocar en orden de prioridad los arcos que se deben monitorear para estimar el MFD de la red estudiada. El resultado de la metodología es por tanto una lista ordenada por prioridad de los arcos que se deben monitorear para luego en base a su flujo y densidad 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 7.

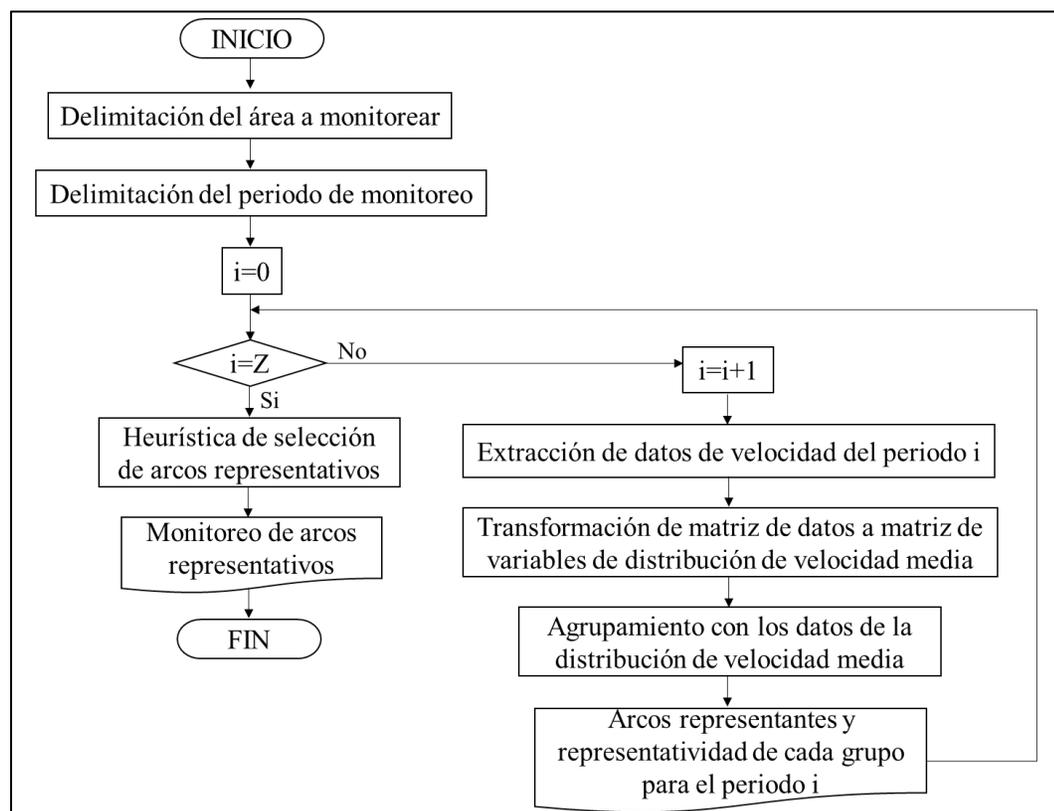


Figura 7 Diagrama de flujo de la metodología de selección de arcos representativos.
Fuente: Elaboración propia

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.

El resto de este capítulo explica en su primera sección algunas consideraciones para la delimitación del área que se desea monitorear, así como el periodo de monitoreo. La segunda sección describe la matriz de datos de velocidad media que se debe recolectar y la transformación de esta matriz en una matriz diferente que contiene las variables de la distribución de velocidad media. La tercera sección corresponde a la aplicación del algoritmo de agrupamiento *Kmeans* sobre la matriz de variables de la distribución de

velocidad media. La cuarta sección corresponde a la heurística de selección de los arcos representativos del área monitoreada en base a los resultados de la repetición del algoritmo de agrupamiento. La última sección de este capítulo detalla cómo construir un MFD estimado con el resultado de la metodología.

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

Se han identificado MFD en áreas de 2.6 km² en Zurich (Ortigosa et al., 2014), 6.47 km² en San Francisco (Geroliminis & Daganzo, 2007), 7.46 km² en Barcelona (Saffari et al., 2020), 219.3 km² en Ámsterdam (Knoop et al., 2018) y 41 543 km² 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. Sin embargo, es conveniente que en el área exista un importante número de usuarios provenientes de otras áreas que deseen interactuar con la delimitada (Knoop et al., 2018) con el objetivo de utilizar el MFD para formular estrategias como el control perimetral. Por otra parte, se debe delimitar un área en la cual no existan cuellos de botella localizados y separados por arcos de poca congestión. Una buena delimitación es semejante al ejemplo de la Figura 8. Se debe evitar autopistas que no formen parte de una gran red ya que la ausencia de rutas alternas genera un MFD disperso (Doig et al., 2013).



Figura 8 Estructura y comportamiento de la red recomendada para aplicar la metodología

Fuente: Elaboración propia

El periodo de recolección de datos de velocidad media 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 configuración (Geroliminis & Daganzo, 2007). Es decir, se debe evitar cambios en la oferta de la red. Además, durante el periodo de recolección de velocidad media no deben existir cambios bruscos en la demanda, como restricciones de movilidad, ya que la existencia del MFD está condicionada a cambios lentos de los patrones de demanda (Geroliminis & Daganzo, 2007).

Debido a que la metodología necesita varios periodos de recolección de datos de velocidad media para repetir el algoritmo de agrupamiento, es necesario repetir la recolección de estos datos durante varios días o distintos horarios de un mismo día para los cuales las características físicas y operacionales son las mismas.

3.2. Matriz de datos

Los datos iniciales para la metodología es la velocidad media de los vehículos que circulan por cada arco del área delimitada para cada intervalo de tiempo Δt que es parte del periodo T de recolección de datos, donde Δt se define discretizando T en intervalos de tiempo de igual duración. El valor más pequeño de Δt es el largo de ciclo de la red semafórica en operación durante los T minutos de recolección de datos. Escoger un Δt grande tiene como consecuencia una menor dispersión del MFD estimado del área, pero una menor cantidad de observaciones.

Conociendo la cantidad de arcos de la red (N), el periodo de recolección de datos (T) y la discretización del periodo de recolección de datos (Δt), la matriz indicada en la Tabla 1 se conforma con N filas y $C = \frac{T}{\Delta t}$ columnas, donde cada celda contiene la velocidad media de cada arco para cada intervalo Δt .

Tabla 1 Matriz de extracción de datos para un área de N arcos, T minutos

Arco (n)	Intervalo (c)									
	1	2	3	4	.	.	.	C-2	C-1	C
1										
2										
3										
4										
.					.					.
.						.				.
.							.			.
N-2										
N-1										
N					.	.	.			

El valor en cada celda corresponde a la velocidad media espacial de cada arco según Edie (1963). Esta velocidad media se calcula mediante la Ecuación 21.

$$V_{S_c}^n = \frac{\sum_{p=1}^P d_{p,c}^n}{\sum_{p=1}^P t_{p,c}^n} \quad \forall p \text{ que se encuentra en arco } n \text{ durante el intervalo } c \quad (21)$$

Donde,

$V_{S_c}^n$ es la velocidad media del arco n durante el intervalo c .

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 .

$t_{p,c}^n$ es la distancia recorrida por el auto p en el arco n en el intervalo c .

Las filas de la Tabla 1 representan los arcos del área de delimitada y las columnas los intervalos de tiempo. La Tabla 1 debe ser construida varias veces para distintos periodos T , con el objetivo de contar con datos que sean independientes entre ellos. A la Tabla 1 se le denomina matriz de extracción de datos.

Cada arco del área, fila de la Tabla 1, tiene asociada una distribución de velocidad media definida por las columnas de la tabla 1 para cada arco, por tanto, es posible calcular para cada fila variables que describan su distribución de velocidad media. Las variables que se deben calcular son: media, mínimo, cuartil uno, mediana, cuartil tres, máximo, rango inter cuartil (IQR por sus siglas en inglés) y desviación absoluta media (MAD por sus siglas en inglés). Este proceso es una transformación de variables que describen a cada arco.

Las variables de la distribución de velocidad se calculan para cada arco y luego deben ser normalizadas para toda la red según la distancia de cada variable a su media expresadas en desviaciones estándar. La normalización sirve para lograr comparabilidad entre las variables (Nisbet et al., 2017). La Tabla 2 muestra la matriz de datos estandarizados que será utilizada por el algoritmo de agrupamiento.

Tabla 2 Matriz de datos para el algoritmo de agrupamiento para N arcos

Arco (n)	Variables de la distribución de velocidad								
	Media	Mínimo	cuartil 1	Mediana	cuartil 3	Máximo	IQR/Mediana	MAD	MAD/Mediana
1									
2									
3									
4									
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N-2									
N-1									
N									

En esta sección se construyó la matriz que será utilizada por el algoritmo de agrupamiento a partir de los datos de velocidad media de cada arco durante el periodo T. La transformación de datos de velocidad media en variables de la distribución de velocidad media es necesaria para mejorar el rendimiento del algoritmo de agrupamiento en arcos. Debido a que el algoritmo de agrupamiento se repetirá varias veces se debe construir la matriz de la Tabla 2 en base a la Tabla 1 para distintos periodos de recolección de datos. Las distintas matrices construidas conforman un set de entrenamiento.

3.3. Agrupamiento de los arcos

El agrupamiento de los arcos consiste en utilizar la matriz de la Tabla 2 como dato de entrada para ejecutar el algoritmo de *Kmeans*, con el objetivo de obtener un arco representativo para cada uno de los \mathbf{K} grupos, en los cuales se dividirá los arcos del área delimitada. Además, el número de arcos que se encuentren en cada grupo será una medida de la representatividad del arco seleccionado.

La cantidad \mathbf{K} de grupos que se formarán con *Kmeans* 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).

El algoritmo de *Kmeans* entrega como resultado un arco representante y el conjunto de arcos que representa para cada uno de los \mathbf{K} grupos. La selección del arco representante de cada grupo, se realiza evaluando la distancia euclidiana más corta entre todos los arcos del grupo al centroide del grupo. El arco con menor distancia será el representante de su grupo y su representatividad será el número de arcos que pertenecen a ese grupo.

El resultado de esta etapa es una división de la red. Por ejemplo, los 84 arcos de la Figura 9 fueron clasificados en seis grupos que se muestran en distintos colores. Además, se obtiene una tabla con el arco representante de cada grupo junto con su representatividad, como la que se muestra en la Tabla 3.

Tabla 3 Tabulación de resultados del agrupamiento

Arco	Representatividad ($RP_{n,z}$)
4	7
37	19
49	11
59	8
79	21
84	18

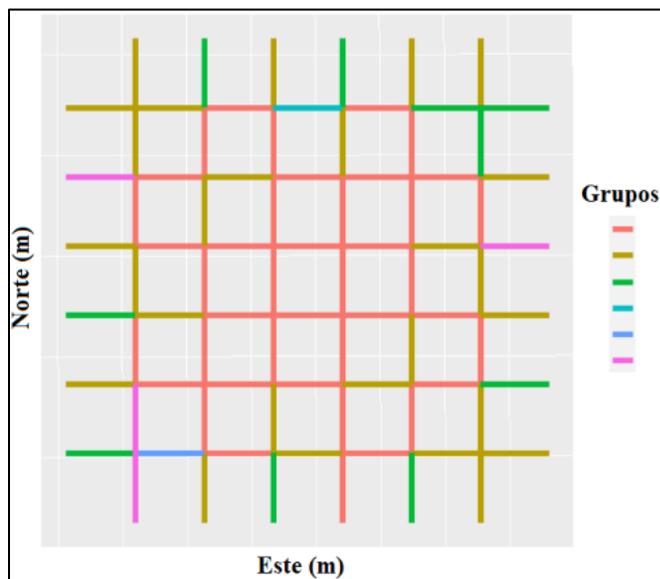


Figura 9 División de la red de un área en 6 grupos

Fuente: Elaboración propia

En la sección 3.2. se construyó un set de entrenamiento de matrices. Cada matriz del set contiene para cada arco las variables de la distribución de velocidad. En base a cada matriz del set de entrenamiento se debe aplicar el algoritmo de agrupamiento. Por tanto, se generará una Tabla 3 y una Figura 9 para cada una de las matrices del set de entrenamiento. Entonces la representatividad indicada en la Tabla 3 se nombra como $RP_{n,z}$, donde $RP_{n,z}$ es la representatividad del arco n para la repetición z .

El resultado de la repetición del algoritmo de agrupamiento sobre el set de entrenamiento forma una matriz de entrenamiento como la indicada en la Tabla 4, donde cada celda indica la representatividad de cada arco para cada repetición ($RP_{n,z}$).

Por ejemplo, en la Tabla 4, el arco N-2 es representativo en la repetición 2 y en la repetición Z-2. Por tanto, tiene una representatividad $RP_{N-2,2} = 54$ y $RP_{N-2,Z-2} = 2$, respectivamente.

Tabla 4 Matriz de entrenamiento de Z repeticiones de un área de N arcos

Arco	Repeticiones del agrupamiento								
	1	2	3	4	...	Z-2	Z-1	Z	
1	0	0	0	0		0	0	0	
2	0	18	0	0		0	0	0	
3	0	0	0	0		0	0	0	
4	1	0	0	0		0	0	23	
5	0	0	0	0		0	1	0	
⋮									
⋮									
⋮									
N-3	0	0	0	0		0	0	0	
N-2	0	54	0	0		2	0	0	
N-1	0	0	0	0		0	0	0	
N	27	0	0	0	...	1	0	0	

3.4. Heurística de selección de arcos representativos

En las secciones anteriores se realizaron recomendaciones sobre como delimitar el área y el periodo de recolección de datos de la velocidad media. Luego, se explicó cómo transformar los datos de velocidad media a variables que describen la distribución de velocidad media. Por último, en la sección anterior se explicó cómo utilizar el algoritmo de agrupamiento para construir una matriz de entrenamiento que contiene los arcos representativos y la representatividad de los mismos para cada una de las repeticiones del algoritmo de agrupamiento.

En esta sección se detalla la heurística propuesta que se debe aplicar sobre la matriz de entrenamiento para identificar los arcos representativos para la estimación del MFD. Es necesario aplicar una heurística de selección de arcos ya que el resultado del agrupamiento solo indica los arcos representativos de una única distribución de velocidad en la red. Sin embargo, una red puede exhibir diferentes distribuciones de velocidad incluso para un mismo patrón de demanda, por lo que los arcos representativos podrían variar. La heurística propuesta tiene como resultado una lista de arcos ordenados en orden de

prioridad, donde la prioridad significa que tan importante es monitorear el flujo y la densidad de ese arco para estimar un MFD cercano al real.

Para decidir la prioridad de un arco, primero calcula la representatividad acumulada de este arco en todas las réplicas según la Ecuación 22 y el número de veces que el arco fue representativo según la Ecuación 23. Los resultados de las ecuaciones 22 y 23 se ejemplifican en las Figuras 10 y 11.

$$RT_n = \sum_{z=1}^Z RP_{n,z} \quad (22)$$

$$Conteo_n = \sum_{z=1}^Z I_{n,z} \quad (23) \quad I_{n,z} = \begin{cases} 1 & \text{Si } RP_{n,z} > 0 \\ 0 & \text{Si } RP_{n,z} = 0 \end{cases} \quad (24)$$

Donde,

RT_n es la representatividad acumulada del arco n.

$RP_{n,z}$ es la representatividad del arco n para la repetición z.

Z es la cantidad de repeticiones del algoritmo de agrupamiento que es igual a la cantidad de periodos de recolección de datos de velocidad media.

$Conteo_n$ es el número de veces que el arco fue representativo luego de las Z repeticiones.

$I_{n,z}$ es una variable dummy que adquiere el valor de uno si el arco **n** fue representativo en la repetición **z**

Los valores de la representatividad acumulada de cada arco **n** (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**

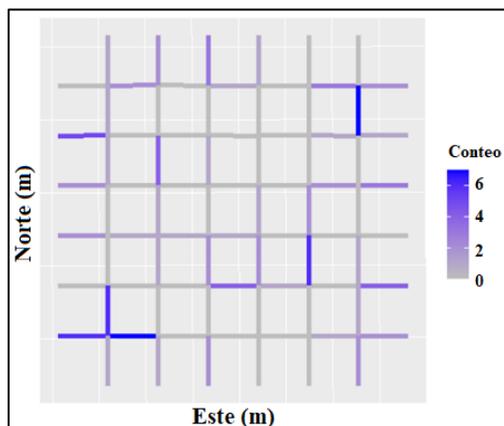


Figura 10 Ejemplo del conteo del número de veces que los arcos son representativos

Fuente: Elaboración propia

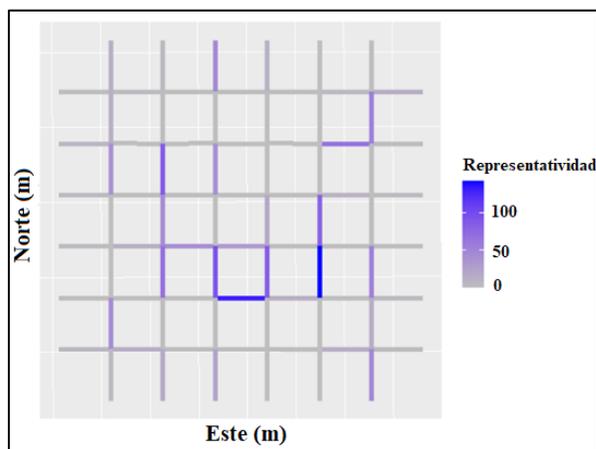


Figura 11 Ejemplo de la representatividad acumulada de los arcos, Conteo

Fuente: Elaboración propia

De forma independiente al proceso anterior, el número de veces que el arco n fue representativo ($Conteo_n$) 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.

Según lo descrito, cada arco de la red tiene asociado un valor de **Ranking de representatividad** y un valor de **Ranking de conteo**. La suma de estas variables se denomina **Orden**. **Orden** indica que arcos son más representativos para la estimación del MFD. Mientras menor sea el valor de **Orden** más importante será obtener los datos de flujo y densidad de ese arco. La heurística propuesta se muestra como diagrama de flujo en la Figura 13.

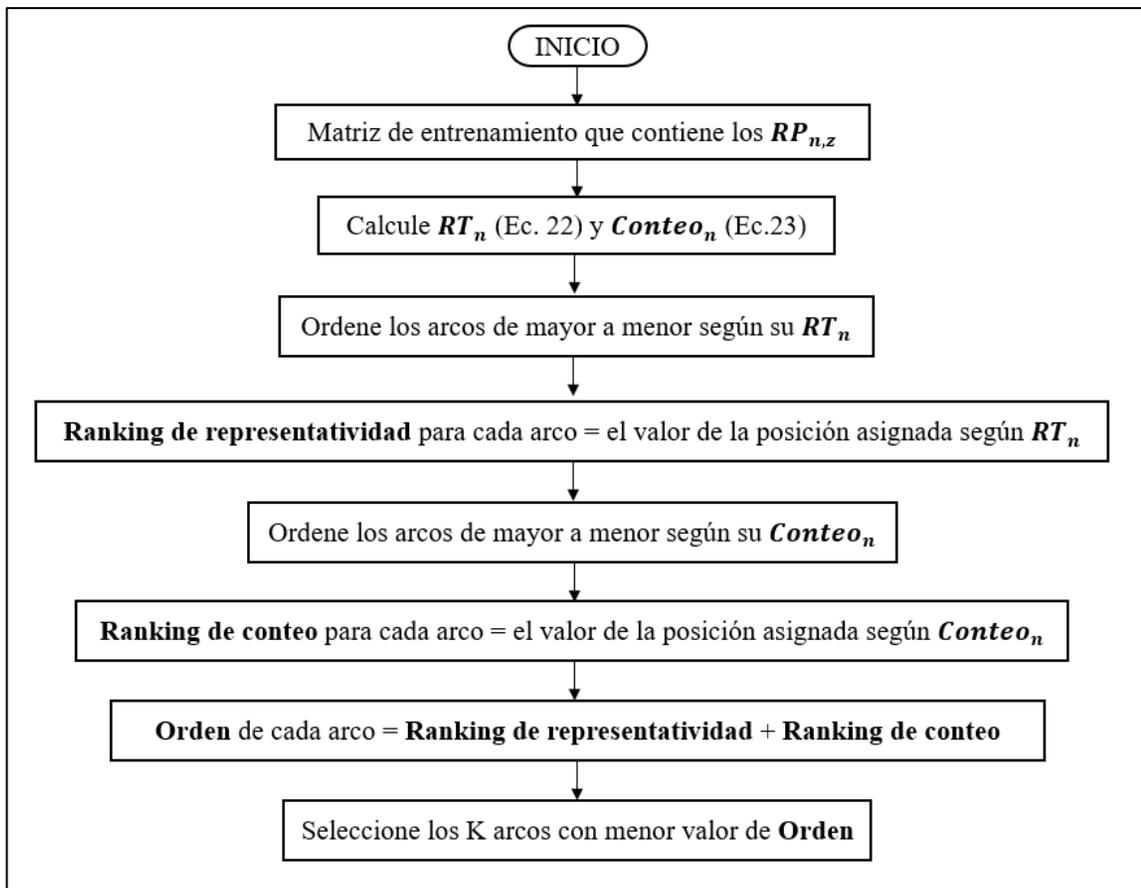


Figura 12 Diagrama de flujo de la heurística de selección

Fuente: Elaboración propia

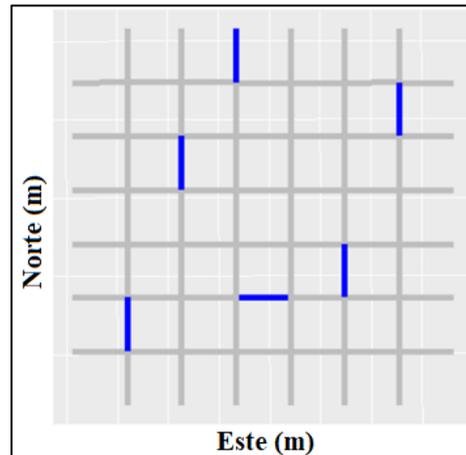


Figura 13 Ejemplo de arcos representativos del estado de tráfico de la red.
Fuente: Elaboración propia

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 se cumplan las condiciones descritas en la sección 2.1 y 2.2 para la existencia de un único MFD sin histéresis.

La condición anterior supone una limitante de este trabajo ya que si luego de equipar los arcos representativos con sensores que midan flujo y densidad existen modificaciones en la topología de la red, entonces existirá un nuevo MFD. Ante esta limitante es necesario realizar una nueva investigación que permita determinar si es posible mantener los arcos equipados como representativos o se deben identificar nuevamente.

3.5. Construcción del MFD estimado

En esta sección se explica un procedimiento que mediante los resultados de la metodología propuesta permite construir un MFD estimado. Los supuestos para estimar el MFD son: todos los arcos de la red tienen las mismas propiedades físicas y operacionales y es posible medir los flujos y densidades de los arcos más representativos (arcos seleccionados).

Luego de aplicar la metodología propuesta, se puede medir los datos de flujo y densidad en los \mathbf{K} arcos seleccionados en varios instantes de tiempo Δt . Utilizando la Formula 9 y 10 de la sección 2.3.1 para cada instante de tiempo se obtiene el flujo promedio y la densidad promedio del subconjunto de arcos \mathbf{K} de la red. Si se grafican los resultados de todos los instantes de tiempo se construye un MFD reducido. El MFD es reducido ya que solo se utiliza los \mathbf{K} arcos seleccionados.

Para construir un MFD estimado es necesario ponderar los valores de flujo y densidad de cada instante de tiempo Δt de cada arco seleccionado con un valor que indique la cantidad de arcos que representa. El valor de RT_k es la representatividad acumulada de un arco que pertenece al conjunto de arcos seleccionados y este valor puede ser obtenido de la metodología propuesta. Sin embargo, RT_k no es la ponderación buscada ya que la suma de todos estos valores para los \mathbf{K} arcos seleccionados no es la cantidad total de arcos de la red. Para transformar RT_k en un ponderador adecuado se utiliza las ecuaciones 25 y 26.

$$RS = \sum_{k=1}^K RT_k \quad (25)$$

$$P_k = \frac{RT_k}{RS} * N \quad (26)$$

Donde,

k es un arco seleccionado con la metodología propuesta.

RT_k es la representatividad acumulada de un arco k seleccionado.

RS es la sumatoria de la representatividad acumulada de todos los arcos seleccionados.

N es el número de arcos de la red.

P_k es la ponderación para la construcción del MFD estimado para cada arco seleccionado.

Utilizando P_k como ponderador para los datos de flujo y densidad de los arcos seleccionados se puede reconstruir un MFD estimado con las Ecuaciones 9 y 10. Por ejemplo, si los arcos seleccionados son los números 4, 37, 49, 59, 79 y 84 ($K=6$) y considerando las Ecuaciones 9 y 10 los valores para el par flujo y densidad promedio de la red para cada Δt serían:

$$q_{\Delta t}^w = P_4 * q_{\Delta t,4} + P_{37} * q_{\Delta t,37} + P_{49} * q_{\Delta t,49} + P_{59} * q_{\Delta t,59} + P_{79} * q_{\Delta t,79} + P_{84} * q_{\Delta t,84} \quad (27)$$

$$k_{\Delta t}^w = P_4 * k_{\Delta t,4} + P_{37} * k_{\Delta t,37} + P_{49} * k_{\Delta t,49} + P_{59} * k_{\Delta t,59} + P_{79} * k_{\Delta t,79} + P_{84} * k_{\Delta t,84} \quad (28)$$

En las Ecuaciones 27, 28, los valores de $q_{\Delta t,n}$, $k_{\Delta t,n}$ se obtienen de sensores de punto fijo como espiras o cámaras de tráfico ubicadas en los arcos representativos luego de que estos son reconocidos. Si en lugar de sensores de punto fijo se cuenta con datos de trayectorias de observadores móviles en los arcos seleccionados los mismos valores de flujo y densidad promedio de la red para cada Δt se calculan con las ecuaciones 29 y 30.

$$q_{\Delta t}^w = \frac{P_4 * \hat{d}_{\Delta t,4} + P_{37} * \hat{d}_{\Delta t,37} + P_{49} * \hat{d}_{\Delta t,49} + P_{59} * \hat{d}_{\Delta t,59} + P_{79} * \hat{d}_{\Delta t,79} + P_{84} * \hat{d}_{\Delta t,84}}{L * N * \Delta t} \quad (29)$$

$$k_{\Delta t}^w = \frac{P_4 * \hat{t}_{\Delta t,4} + P_{37} * \hat{t}_{\Delta t,37} + P_{49} * \hat{t}_{\Delta t,49} + P_{59} * \hat{t}_{\Delta t,59} + P_{79} * \hat{t}_{\Delta t,79} + P_{84} * \hat{t}_{\Delta t,84}}{L * N * \Delta t} \quad (30)$$

Donde,

L es el largo promedio de cada arco.

N es la cantidad de arcos.

\hat{d} es la distancia recorrida por todos los observadores móviles en el instante Δt en un arco en particular.

\hat{t} es el tiempo de permanencia de todos los observadores móviles en el instante de tiempo Δt en un arco en particular.

4. VALIDACIÓN DE LA METODOLOGÍA

En este capítulo se busca verificar mediante un modelo de simulación, si el MFD estimado que se construye a partir de los arcos representativos seleccionados con la metodología propuesta en el capítulo 3 se ajusta al MFD real. Para ello, se elaboró un modelo de microsimulación en el software Aimsun el cual permite conocer todas las trayectorias de los vehículos de la red. En base a las trayectorias de los vehículos se mide la velocidad media de los vehículos, el flujo y la densidad en cada arco para cada instante de tiempo Δt .

El flujo y la densidad en cada arco para un instante de tiempo Δt se utilizan junto con las ecuaciones 9 y 10 para calcular el flujo y la densidad promedio de la red. Con esos promedios se obtiene un punto del MFD real. Al repetir el procedimiento para todos los Δt se construye el MFD real de la red simulada.

La velocidad media en cada arco para un instante de tiempo Δt es utilizada para generar la Tabla 2 indicada en la sección 3.2. El modelo de microsimulación es ejecutado varias veces modificando cada vez los patrones y niveles de demanda (escenarios de simulación) con el objetivo de generar varios ejemplos semejantes a la Tabla 2. Todos los ejemplos semejantes a la Tabla 2 forman los datos iniciales del set de entrenamiento.

Luego de aplicar la metodología propuesta con los datos del set de entrenamiento, se seleccionan los arcos representativos de la red simulada y se calculan los ponderadores P_k . Utilizando los ponderadores, los flujos y densidades para cada instante Δt del set de validación para los arcos seleccionados y las Ecuaciones 9 y 10 se construye el MFD estimado. Para determinar si el MFD real y el MFD estimado se ajustan, se calculan los errores de estimación del MFD indicados en la sección 2.6.

El resto de este capítulo describe primero el modelo de microsimulación (oferta y demanda) y en las últimas secciones se estudia la sensibilidad de la metodología ante

variaciones del algoritmo de agrupamiento, los datos base para el agrupamiento, el número de arcos representativos y las simulaciones del set de entrenamiento.

4.1. Simulación

Aimsun es un software que permite la construcción de modelos microscópicos de tráfico. El entorno de este software es amigable y fácil de utilizar. Además, los resultados numéricos de la simulación se pueden procesar en un gestor de base de datos. Aimsun es un software de uso frecuente en la región, por lo que el modelo construido para la validación puede ser entendido por estudiantes, profesionales e investigadores del área.

El área total simulada es una red tipo Manhattan de 11x11 de 1.74 km². La red está conformada por vías unidireccionales intercaladas de dos pistas cada una. Las intersecciones son controladas por un plan semafórico que tiene un ciclo de 60 segundos y dos fases de verde de 26 segundos acompañados de 3 segundos de amarillo y 1 segundo de todo rojo. No existe desfase entre los semáforos por lo que la fase de verde de cada semáforo en una misma dirección de circulación se activa al mismo tiempo. La demanda es asignada mediante una matriz de viajes y esta se incorpora al modelo mediante los centroides azules (generación), verdes (atracción) y amarillos (generación y atracción) indicados en la Figura 14. En total el modelo tiene 45 centroides de origen y 45 centroides de destino de los cuales 20 de origen y 20 de destino se encuentran en la periferia. La oferta de la red descrita fue modelada por García-Huidobro (2018).

Para esta investigación se utilizan solo vehículos livianos de los cuales el 30% tienen enrutamiento dinámico. Es decir, al 70% de vehículos se les asigna una ruta al inicio de la simulación la cual se mantiene constante hasta que terminan su viaje.

El área a monitorear, y para la cual se desea estimar un MFD es la delimitada por líneas rojas entrecortadas en la Figura 14, y que corresponde a una red tipo Manhattan de 7x7 de

0.7 km² que se encuentra en la parte central de la red simulada. El área a monitorear está compuesta por 84 arcos y 36 intersecciones.

La oferta vial y operacional descrita tiene asociada un único MFD. Si cambia la oferta, el MFD también lo hará. Por tanto, la oferta vial y operacional deben mantenerse constantes durante la recolección de datos.

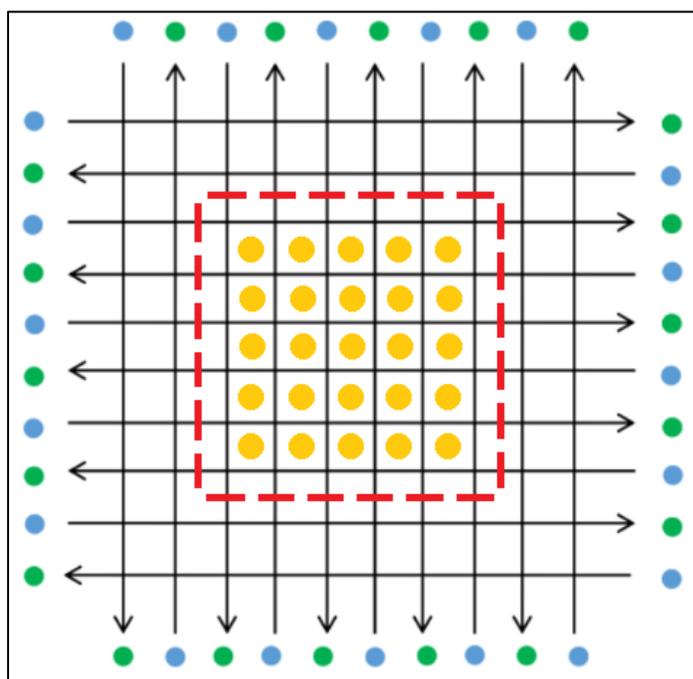


Figura 14 Representación de la red simulada

Fuente: Elaboración propia en base a García – Huidobro (2018)

El modelo de microsimulación mantiene los parámetros por defecto del software Aimsun ya que no se busca simular una red en particular. Para garantizar la existencia de un MFD sin histéresis en el área monitoreada se deben cumplir condiciones estacionarias. La oferta modelada que cumple estas condiciones se explica con detalle en las siguientes secciones.

Es necesario ejecutar el modelo de microsimulación bajo diferentes escenarios de simulación ya que cada ejecución del modelo reemplazará a un periodo de recolección de datos. Los escenarios se explican a continuación.

4.1.1. Escenarios de simulación

Los distintos escenarios de simulación que se describirán en las siguientes secciones se diferencian entre ellos de acuerdo a su demanda de tráfico. Los escenarios son utilizados para representar días con comportamientos de tráfico diferentes teniendo en cuenta la aleatoriedad del tráfico, variaciones en el nivel de demanda o nivel de congestión y la variación interna de un escenario de demanda.

Todos los escenarios de demanda se basan en tres tipos de viajes:

- Viajes que tienen origen en la periferia y destino en la zona monitoreada (P-M).
- Viajes que tienen origen en la zona monitoreada y destino en la periferia (M-P).
- Viajes que tienen origen y destino en la zona monitoreada (M-M).

Todos los escenarios modelan el comportamiento del área en punta mañana. Debido a que el área a monitorear representa un modelo de ciudad mono céntrica, los viajes P-M son los de mayor cantidad mientras que los viajes M-P son los de menor cantidad. Asimismo, se asume que los viajes que cruzan el área (P-P) no existen. Esta configuración se mantiene para todos los escenarios de simulación.

Se construyeron tres escenarios de demanda. El primer escenario fue denominado “base”, el segundo escenario fue denominado “nivel de demanda” y el tercer escenario fue denominado “variación de patrón”. Para cada escenario se realizaron 6 simulaciones. A continuación, se presenta con mayor detalle los escenarios utilizados.

4.1.1.1. Escenario base

El escenario base carga a la red una demanda total de 43 000 viajes en un periodo de 195 minutos. Los viajes que ingresan a la red en todo el periodo de monitoreo se dividen según la Tabla 5. Los viajes P-M y M-P se dividen de forma homogénea entre los 500 pares origen destino posibles para cada tipo de viaje. Los viajes M-M se dividen de forma

homogénea entre los 600 pares origen destino posibles. La Tabla 5 indica el patrón de demanda utilizado para el escenario base.

Tabla 5 Distribución de los viajes entre todos los pares origen destino

		Destino	
		Centroides en la zona de monitoreo	Centroides en la periferia
Origen	Centroides en la zona de monitoreo	17 500 viajes	5 000 viajes
	Centroides en la periferia	20 500 viajes	0 viajes

La tasa a la cual ingresan los viajes de la Tabla 5 al modelo es variable en el tiempo con el objetivo de generar un patrón de demanda adecuado para la situación modelada. La tasa de ingreso cambia cada 5 minutos con el objetivo de realizar cambios lentos en la demanda según se muestra en la Figura 15.

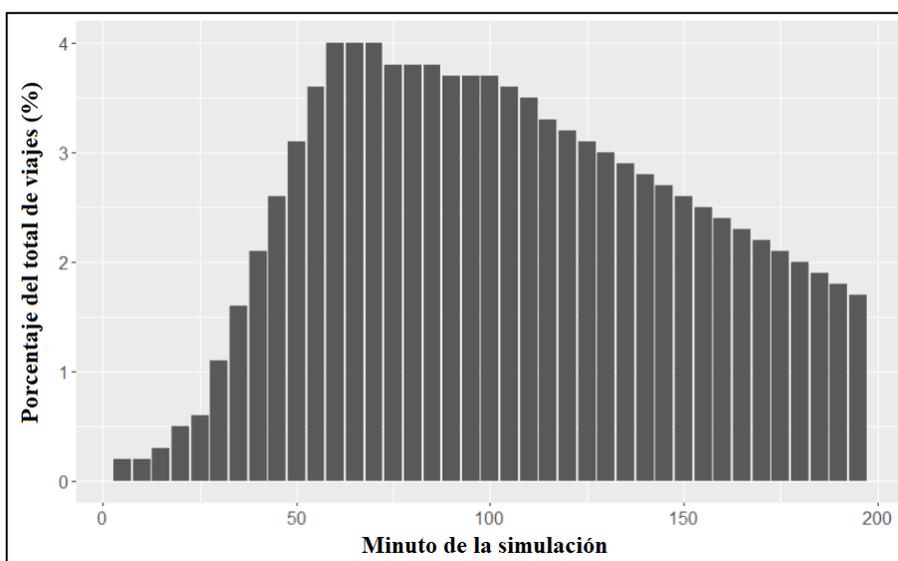


Figura 15 División temporal del total de viajes entre cada par origen destino

Fuente: Elaboración propia

El escenario base fue calibrado para que la demanda sea suficiente para congestionar la red sin llegar a la acumulación máxima. El objetivo de encontrar estados de tráfico en este rango fue reconocer al menos un ciclo de carga e inicio de descarga de la red para asegurar la existencia del MFD sin histéresis.

Después de ejecutar una simulación del escenario descrito se extraen del modelo de microsimulación las trayectorias de todos los vehículos en base a las cuales se calcula, el flujo y la densidad en cada arco para cada instante de tiempo Δt . El flujo y la densidad en cada arco para un instante de tiempo Δt se utilizan junto con las ecuaciones 9 y 10 para calcular el flujo y la densidad promedio de la red. Con esos promedios se obtiene un punto del MFD real de la Figura 16. Al repetir el procedimiento para los 195 minutos que dura la simulación, con $\Delta t = 1$ minuto, se construye el MFD real de la figura 16.

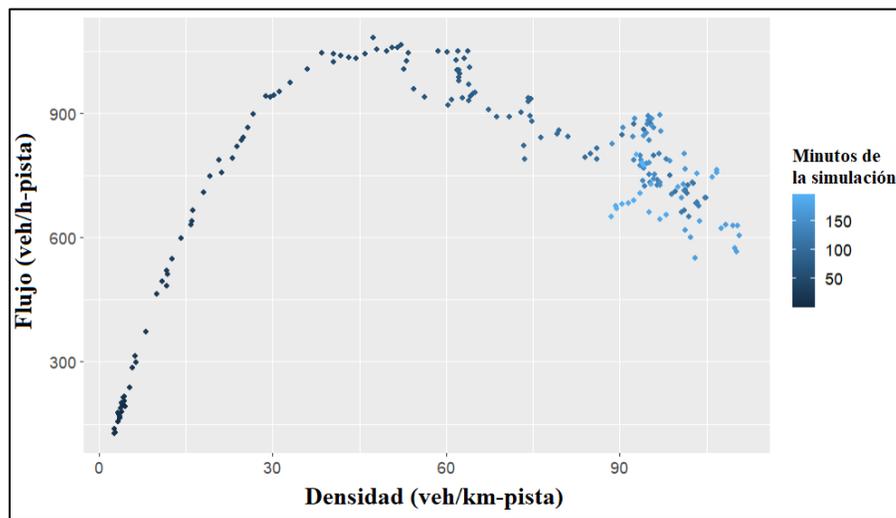


Figura 16 MFD de la primera réplica del escenario base.
Fuente: Elaboración propia

En la Figura 16 los puntos más claros son aquellos que se calculan para el final de la simulación y estos corresponden a un descenso de la demanda. Los puntos más claros se ajustan a los puntos más oscuros que corresponden a otro intervalo de tiempo. Por tanto, el MFD no tiene histéresis. El obtener un MFD sin histéresis requirió un proceso de calibración de la demanda para cumplir las condiciones de estacionalidad.

Los vehículos simulados ingresan a la red desde los nodos de generación según la demanda calibrada. Para cumplir la demanda calibrada, los nodos de generación envían vehículos, donde el tiempo entre vehículos es distribuye exponencial. La aleatoriedad que

incluye la distribución exponencial fue aprovechada para modelar seis simulaciones diferentes para este escenario. Cada simulación tiene una semilla aleatoria distinta.

4.1.1.2. Variación de escenarios

Como se mencionó al final de la sección 4.1.1 se construyeron tres escenarios de demanda. El primer escenario fue denominado “base” y este se explicó en la sección anterior. El segundo escenario fue denominado “nivel de demanda” y el tercer escenario fue denominado “variación de patrón”. Los escenarios “nivel de demanda” y “variación de patrón” son independientes entre ellos y se construyen a partir del escenario “base”.

Las simulaciones del escenario del “nivel de demanda” se modelaron multiplicando la matriz viajes del escenario “base” por un coeficiente obtenido mediante una extracción aleatoria de la distribución uniforme $U(0.5, 1.5)$. Multiplicando la Tabla 5 por el coeficiente extraído y manteniendo el mismo modelo del escenario base, se obtiene una simulación del escenario del “nivel de demanda”. Se generaron seis valores para este coeficiente de modo de contar con seis simulaciones distintas con distinto nivel de demanda cada una. Por tanto, cada simulación puede tener estados de flujo libre, llegar a congestión y regresar a flujo libre o llegar al bloqueo completo.

Las simulaciones del escenario de “variación de patrón” se modelaron multiplicando cada celda de la matriz de viajes del escenario “base” por un coeficiente distinto obtenido mediante una extracción aleatoria de la distribución $U(0.8, 1.2)$. Multiplicando cada celda de la Tabla 5 por un coeficiente distinto, se obtiene una simulación del escenario “variación de patrón”. Al multiplicar las celdas de la Tabla 5 por un coeficiente distinto se modifica la cantidad de viajes que van al centro o que van a la periferia en magnitudes distintas, generando de esta forma un patrón de demanda distinto para cada simulación. Para generar seis días en base a este escenario se realizaron nuevas extracciones de la distribución $U(0.8, 1.2)$ para cada celda de la matriz y para cada simulación. Por tanto, cada simulación de este escenario tiene un patrón de demanda distinto.

Con todo lo anterior, se cuenta con 18 simulaciones. Seis de ellas corresponden al escenario base, otras seis consideran distintos niveles de demanda y las últimas seis consideran distintos patrones de demanda. La metodología se aplicó solamente con $Z=15$ simulaciones (cinco simulaciones de cada escenario), es decir, hay 15 simulaciones que entregan datos de velocidad media y estas conforman el set de entrenamiento. La sexta simulación de cada escenario fue utilizada para el cálculo de los errores de estimación detallados en la sección 2.6. Es necesario calcular los errores de estimación con un conjunto de simulaciones distintas a las usadas para el set de entrenamiento con el objetivo de detectar si la metodología propuesta tiene un sesgo asociado al set de entrenamiento.

El MFD real es una propiedad de la oferta y no de las variaciones de demanda (Geroliminis & Daganzo, 2007), por tanto, utilizando cualquier escenario de simulación el MFD construido debe ser el mismo. Lo anterior se verifica realizando una inspección visual de la Figura 17.

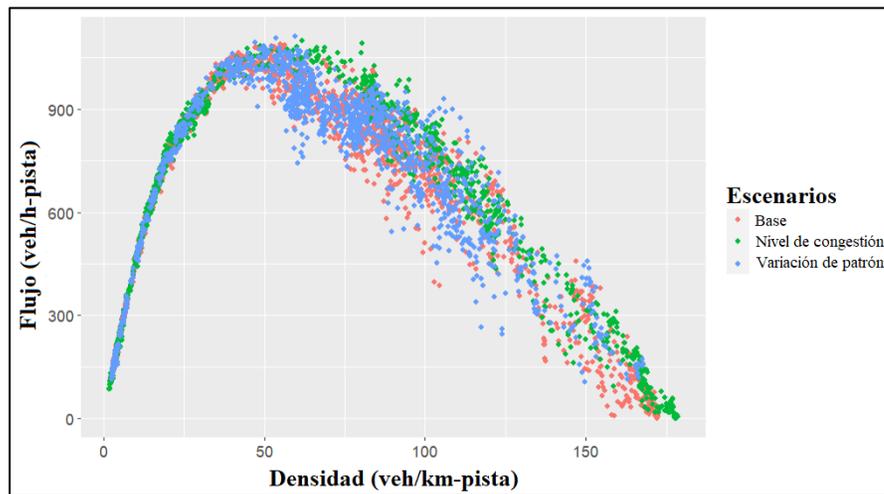


Figura 17 MFD real de la zona monitoreada considerando las 15 simulaciones del set de entrenamiento

Fuente: Elaboración propia

Después de ejecutar una simulación de cada escenario se extraen del modelo de microsimulación las trayectorias de todos los vehículos en base a las cuales se calcula, el flujo y la densidad en cada arco para cada instante de tiempo Δt . El flujo y la densidad en

cada arco para un instante de tiempo Δt se utilizan junto con las Ecuaciones 9 y 10 para calcular el flujo y la densidad promedio de la red. Con esos promedios se obtiene un punto del MFD real de la Figura 17. Al repetir el procedimiento para los 195 minutos que dura una la simulación, con $\Delta t = 1$ minuto, se construye una secuencia de puntos del MFD real. Al repetir el proceso descrito de forma independiente para todas las simulaciones de un escenario y para todos los escenarios el resultado se puede ver en la Figura 17.

4.2. Velocidad media en cada arco

Para aplicar la metodología se requiere, para cada intervalo Δt , la velocidad media en cada arco de la red a monitorear. Como se mencionó al inicio de este capítulo, el micro simulador Aimsun permite conocer la trayectoria de todos los vehículos, por lo que es conocido el total de kilómetros recorridos por todos los vehículos en cada arco y el tiempo que permanecieron todos los vehículos en cada arco para cada intervalo de tiempo Δt . Los datos del total de kilómetros recorridos y el tiempo de permanencia de los vehículos en cada arco se contabiliza desde que un vehículo ingresa al arco hasta que sale. Utilizado las definiciones generalizadas de Edie (1963), la velocidad media espacial de cada arco para cada intervalo de tiempo ($\Delta t=1$ minuto) se calcula mediante la Ecuación 31.

$$Vs_c^n = \frac{travel_{c,n}}{\frac{traveltime_{c,n}}{3600}} \quad (31)$$

Donde,

Vs_c^n es la velocidad media del arco **n** en el intervalo de tiempo **c**.

$travel_{c,n}$ es el total de kilómetros recorridos durante el intervalo **c** en el arco **n**.

$traveltime_{c,n}$ es el tiempo que permanecieron todos los vehículos durante el intervalo **c** en el arco **n**.

En todas las simulaciones el tiempo simulado es de 195 minutos y el intervalo de tiempo para la recolección de datos se estableció en un minuto. Por tanto, cada arco cuenta con 195 datos de velocidad media espacial para cada simulación.

Para el caso de la red simulada, la Tabla 1 descrita en la sección 3.2 tiene 84 filas (arcos) y 195 columnas (intervalos de monitoreo). Las celdas de la Tabla 1 tienen el valor de Vs_c^n descrito en esta sección. Para cada simulación se construye la Tabla 1 y se separan 15 simulaciones que conforman el set de entrenamiento y tres simulaciones que conforman el set de prueba. Las tres simulaciones del set de prueba están conformadas por una simulación de cada escenario.

4.3. Resultados de la heurística de selección

En las secciones anteriores se describió la red simulada y los escenarios de demanda que se utilizaron para generar 18 simulaciones. Luego, se indicó como conformar el set de entrenamiento y validación con los datos de velocidad media de las 18 simulaciones. Al inicio de esta sección se muestran los resultados de realizar el agrupamiento de arcos con todas las simulaciones del set de entrenamiento (15 simulaciones) y los resultados de la heurística de selección propuesta. Al final de esta sección se muestran los errores de estimación entre el MFD real y el MFD estimado construido con el resultado de la metodología.

En la Tabla 6 y Figura 18 se muestra el resultado del agrupamiento mediante los valores de RT_n y $Conteo_n$. La heurística de selección utiliza estas variables para calcular el **Ranking de representatividad**, **Ranking de conteo** y **Orden** que también son mostradas en la Tabla 6.

En esta validación de la metodología se monitorearon seis arcos ($K=6$) que corresponden al 7.14% de la red. Para definir K se utilizó el método gráfico del codo (Coates & Ng, 2012). Aplicando la heurística los arcos seleccionados son: 6, 55, 44, 66, 20 y 22.

Tabla 6 Heurística de selección de arcos representativos

ID Arco	RT _n	Conteo _n	Ranking de representatividad	Ranking de conteo	Orden	RS	P _k
6	52	6	3	1	4	320	13.7
55	76	3	1	4	5		20
44	70	3	2	5	7		18.4
66	50	3	5	6	11		13.1
20	40	3	8	7	15		10.5
22	32	6	15	2	17		8.4
33	51	2	4	13	17		
82	45	2	6	14	20		
5	41	2	7	15	22		
77	28	4	21	3	24		
52	37	2	10	16	26		
63	28	3	20	8	28		
59	33	2	12	17	29		
84	33	2	13	18	31		
54	38	1	9	23	32		
1	32	2	14	19	33		
26	35	1	11	24	35		
4	21	3	29	9	38		
45	32	1	16	25	41		
27	31	1	17	26	43		
37	17	3	35	10	45		
47	29	1	18	27	45		
78	22	2	27	20	47		

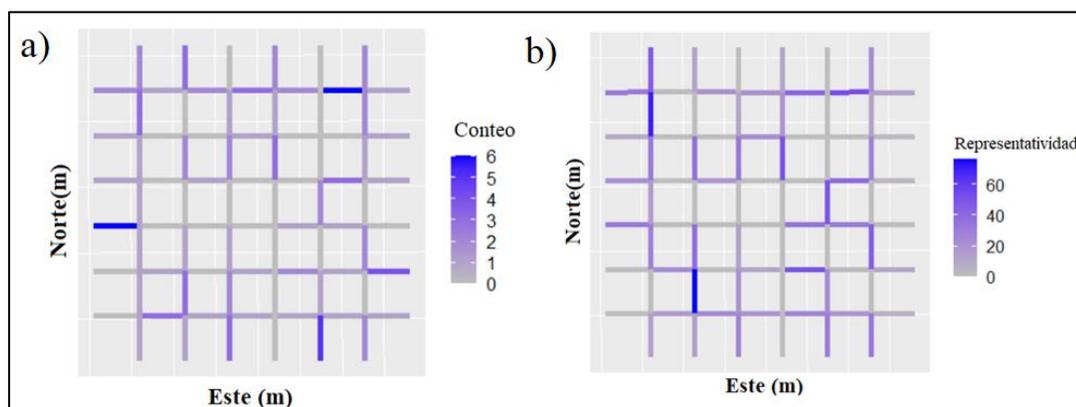


Figura 18 a) Conteo y b) representatividad de los arcos del área monitoreada. Fuente: Elaboración propia

Para calcular los errores de estimación del MFD se utilizaron las tres simulaciones que no formaron parte del set de entrenamiento. Las tres simulaciones forman el set de validación.

Cada simulación corresponde a un nivel y patrón de demanda distinta. El MFD real para estos días se presenta en la Figura 19.

Después de ejecutar una simulación que forma parte el set de validación se extraen del modelo de microsimulación las trayectorias de todos los vehículos en base a las cuales se calcula, el flujo y la densidad en cada arco para cada instante de tiempo Δt . El flujo y la densidad en cada arco para un instante de tiempo Δt se utilizan junto con las Ecuaciones 9 y 10 para calcular el flujo y la densidad promedio de la red. Con esos promedios se obtiene un punto del MFD real de la Figura 19. Al repetir el procedimiento para los 195 minutos que dura una simulación, con $\Delta t = 1$ minuto, se construye el MFD real para cada simulación del set de validación. Al repetir el proceso descrito de forma independiente para todas las simulaciones del set de validación se conforma la Figura 19.

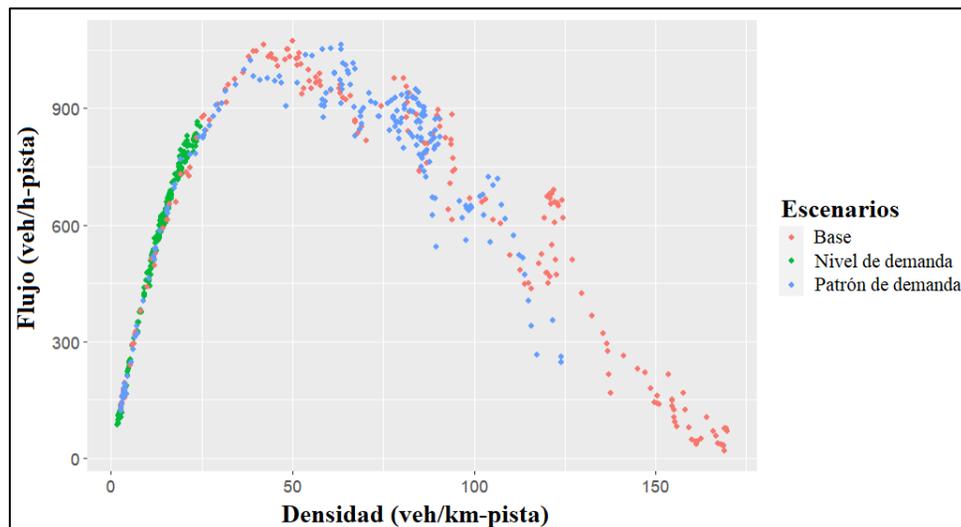


Figura 19 MFD de los días del set de validación
Fuente: Elaboración propia

Utilizando los ponderadores P_k de la Tabla 6, el flujo y la densidad de cada arco seleccionado para cada instante de tiempo Δt correspondiente a una simulación del set de validación junto con las Ecuaciones 9 y 10 se calculó el flujo y la densidad promedio de la red. Con esos promedios se obtiene un punto del MFD estimado. Al repetir el

procedimiento para los 195 minutos que dura una simulación, con $\Delta t = 1$ minuto, se construye el MFD estimado para cada simulación del set de validación. El MFD estimado para cada simulación del set de validación se muestra en la Figura 20. En los MFD estimados de la Figura 20, el flujo máximo y la densidad máxima se encuentra alrededor de 1000 veh/h-pista y 150 veh/km-pista. Estos valores son razonables si se asume que son semejantes a los parámetros del diagrama fundamental de una pista promedio que tiene un semáforo al final.

Además del MFD real y el MFD estimado se construyó para cada simulación del set de validación dos MFD estimados adicionales. El primer MFD estimado adicional se denominó MFD-*Kmeans*. El segundo MFD estimado adicional se denominó MFD-Aleatorio.

Si se aplica el algoritmo de *Kmeans* con los datos de una simulación y $K=6$, el resultado es seis grupos de arcos para los cuales es posible encontrar su arco representativo. El arco representativo de cada grupo es el que tenga la menor distancia euclidiana con respecto al centroide del grupo. Para la estimación del MFD-*Kmeans* se puede sustituir los ponderadores P_k por el número de arcos de cada grupo. Entonces, utilizando los arcos representativos, la cantidad de arcos por cada grupo, el flujo y la densidad de cada arco representativo para cada instante de tiempo Δt junto con las Ecuaciones 9 y 10 se calculó el flujo y la densidad promedio de la red. Con esos promedios se obtiene un punto del MFD-*Kmeans*. Al repetir el procedimiento para los 195 minutos que dura una simulación, con $\Delta t = 1$ minuto, se construye el MFD-*Kmeans* para cada simulación del set de validación.

Para construir el MFD-Aleatorio se escoje 6 arcos de forma aleatoria que serán considerados arcos seleccionados. Luego, se escoje una ponderacion P_k de forma aleatoria, pero considerando que la suma de todos los P_k sea igual al número de arcos. Utilizando los ponderadores P_k recién definidos, el flujo y la densidad de cada arco seleccionado para cada instante de tiempo Δt correspondiente a un día del set de validación junto con

las Ecuaciones 9 y 10 se calculó el flujo y la densidad promedio de la red. Con esos promedios se obtiene un punto del MFD-Aleatorio. Al repetir el procedimiento para los 195 minutos que dura una simulación, con $\Delta t=1$ minuto, se construye el MFD-Aleatorio para cada simulación del set de validación.

En la Figura 20 se puede observar como el MFD estimado en las tres simulaciones de validación es semejante en forma, dominio y rango al MFD real. Para calcular los resultados numéricos de los errores de estimación mediante las ecuaciones de la 14 a la 20, se utiliza por separado los datos de cada simulación del set de validación. Por tanto, para cada simulación del set de validación, se utilizó el flujo promedio ponderado real y la densidad promedio ponderada real junto con el flujo promedio ponderado estimado y la densidad promedio ponderada estimada para calcular cada uno de los datos de la Tabla 7. En la Tabla 7 se muestra que, en dos de las tres simulaciones, el MFD estimado es el que tiene menor error con respecto al MFD real y en ninguno de las tres simulaciones el MFD estimado es el más distante al MFD real. El MFD-Aleatorio es el que se encuentra más distante al MFD.

Tabla 7 Errores en la estimación del MFD para cada día del set de validación

Error de estimación	MFD	Set de validación		
		Simulación 1	Simulación 2	Simulación 3
RMSE	eMFD	7.5.E+03	6.4.E+03	7.9.E+03
	eMFD- <i>Kmeans</i>	7.2.E+03	7.3.E+03	9.7.E+03
	eMFD-Aleatorio	1.8.E+04	1.6.E+04	2.4.E+04
RMSE (qc, kj)	eMFD	0.105	0.135	0.129
	eMFD- <i>Kmeans</i>	0.062	0.150	0.087
	eMFD-Aleatorio	0.230	0.366	0.289
$\Delta R(\text{MFD real, MFD reducido})$	eMFD	0.157	0.087	0.100
	eMFD- <i>Kmeans</i>	0.151	0.144	0.161
	eMFD-Aleatorio	0.144	0.146	0.174

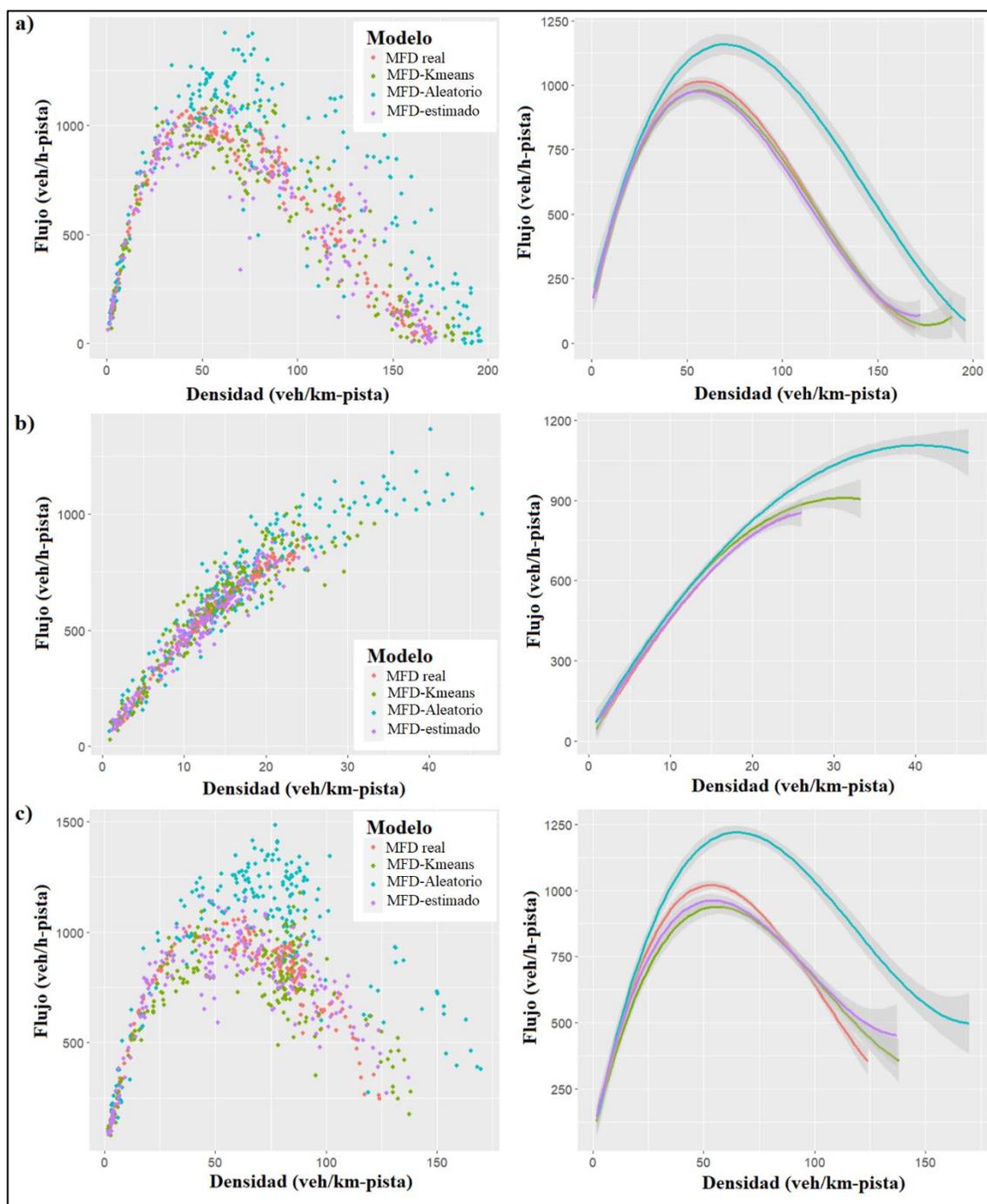


Figura 20 A la izquierda puntos del MFD real y estimados del set de validación. A la derecha ajuste polinomial grado 3 de los MFD. a) Simulación 1 (Escenario base), b) Simulación 2 (Escenario nivel de demanda) y c) Simulación 3 (Escenario de variación de patrón)

Fuente: Elaboración propia

Los errores de estimación consideran tanto el error del flujo promedio como el de la densidad promedio. Para analizar de forma separada el flujo y la densidad se elaboró la Figura 21, en la cual se comparan los valores reales y los estimados de estas variables. No se aprecia un error dominante en uno de sus componentes. Es posible notar que el MFD-*Kmeans* exhibe un dominio mayor que los datos reales. El MFD estimado distribuye de forma homogénea los errores en todo el dominio de los datos.

Los errores del MFD estimado que se obtienen con la metodología propuesta son mayores a los errores de otras metodologías. Por ejemplo: Ortigosa et al. (2014) reporta que monitoreando el 5% de la red logra un ΔR de 3%, mientras que esta metodología monitoreando el 7% de la red logra un ΔR de 10%; Saffari et al. (2020) por su parte reporta que monitoreando el 10% de la red logra un RMSE (qc, kj) de 0.07, mientras que esta metodología monitoreando el 7% de la red logra un RMSE (qc, kj) de 0.12. Pese a que los errores obtenidos con la metodología propuesta son mayores a los de otras metodologías los resultados son prometedores ya que los datos que fueron utilizados se obtienen fácilmente a bajo costo y sin contar con infraestructura propia.

La heurística propuesta presenta ventajas respecto a aquellas que tienen errores de estimación del MFD menores. Por ejemplo: la heurística 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 del set de validación, por lo que los arcos representativos son adecuados para varios patrones de demanda distintos a los ocupados en el entrenamiento; la heurística relaja los requisitos de Saffari et al. (2020) ya que no requiere una flota de observadores móviles propios que se encuentre distribuidos de forma homogénea en toda la red para cada intervalo de tiempo.

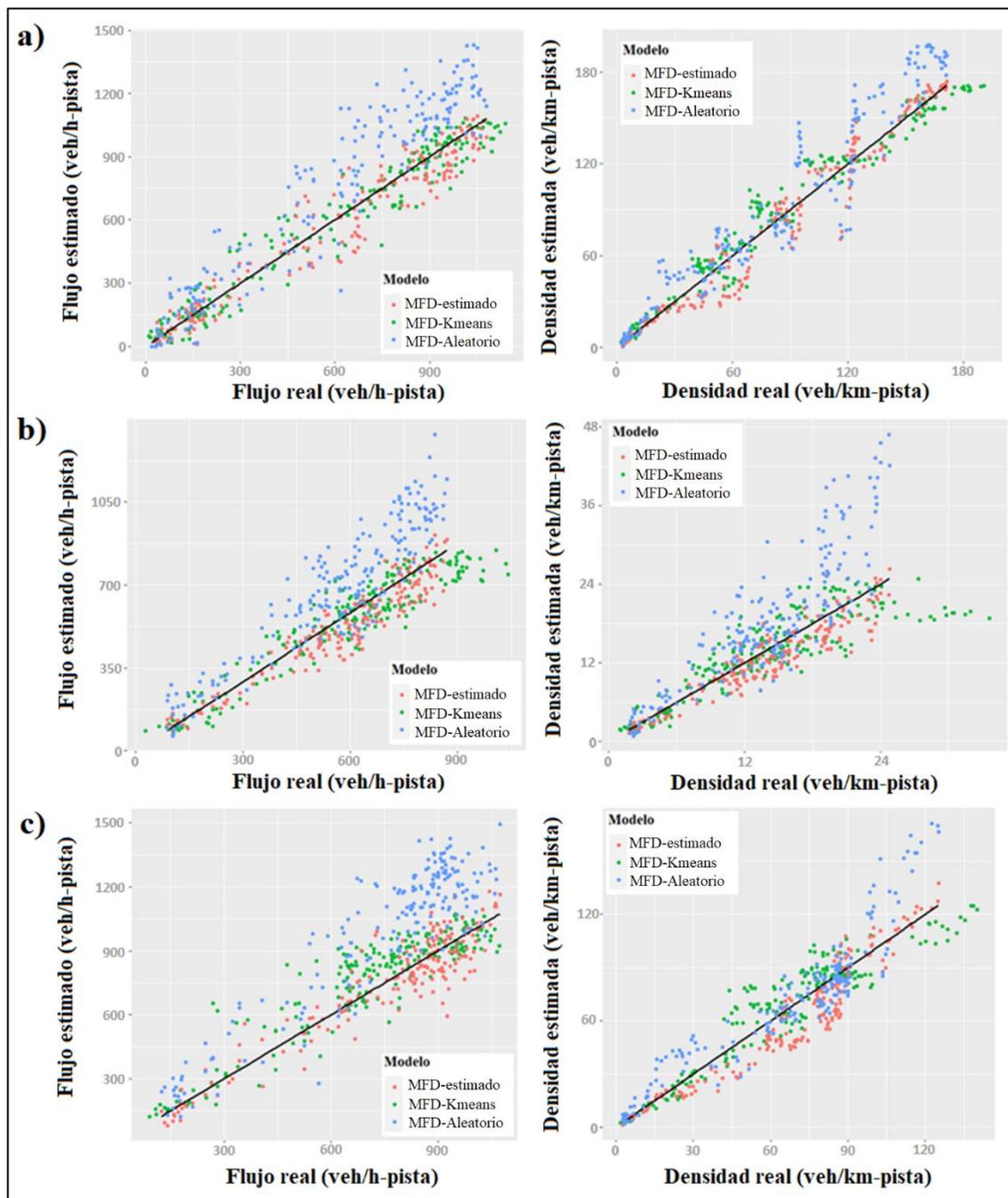


Figura 21 Desviaciones del flujo y la acumulación estimada. a) día 1 (Escenario base), b) día 2 (Escenario nivel de demanda) y c) día 3 (Escenario de variación de patrón).

Fuente: Elaboración propia

4.4. Evaluación de otros algoritmos de agrupamiento.

La variedad de algoritmos de agrupamiento existentes indicados en la sección 2.7. pueden generar diferentes grupos y arcos representativos, modificando de esta forma el resultado final de la metodología. Para verificar que *Kmeans* es un buen algoritmo para agrupar arcos se aplicó la metodología descrita en el Capítulo 3 hasta la sección 3.3 variando el algoritmo de agrupamiento. Los algoritmos de agrupamiento probados fueron *Kmeans*, Spectral, Clique y Optics, y solo se utilizaron las simulaciones del set de entrenamiento correspondientes al escenario base.

Para reconstruir el MFD estimado para una simulación del set de entrenamiento con cada algoritmo de agrupamiento se siguió la metodología indicada en la sección 3.3 y en base a su resultado se modificó el valor de P_k por $RP_{n,z}$. Siguiendo este proceso, se construyeron los MFD estimados para las cinco simulaciones del escenario base con cada uno de los cuatro algoritmos de agrupamiento y luego se calculó los errores de estimación del MFD. Los arcos monitoreados con cada algoritmo de agrupamiento resultaron distintos.

En la Tabla 8 se puede ver cómo los métodos de agrupamiento que tienen un menor error en la estimación del MFD corresponden a *Kmeans* y Spectral. Los resultados de la construcción de los MFD estimados muestran que no todos los algoritmos generan un MFD estimado cercano al real. En la Figura 22, además del MFD real se muestra los MFD estimados construidos con los distintos métodos de agrupamiento. Los MFD estimados de la Figura 22 se construyen con seis arcos monitoreados para una simulación del set de entrenamiento.

Tabla 8 Errores en la estimación del MFD con el escenario base para todos los métodos de agrupación

Error de estimación	Método	Simulaciones del escenario base				
		1	2	3	4	5
RMSE/E04	Clique	1.6	1.5	1.8	1.3	1.4
	<i>Kmeans</i>	1.1	1.2	1.2	1.1	1.2
	Optics	9.8	1.5	1.4	1.1	1.2
	Spectral	1.3	1.1	1.0	1.0	1.3
RMSE(qc,kj)	Clique	0.242	0.199	0.292	0.212	0.191
	<i>Kmeans</i>	0.17	0.159	0.204	0.179	0.168
	Optics	0.142	0.186	0.185	0.182	0.198
	Spectral	0.216	0.153	0.145	0.2	0.151
ΔR	Clique	0.218	0.139	0.373	0.234	0.136
	<i>Kmeans</i>	0.472	0.093	0.216	0.192	0.100
	Optics	0.179	0.173	0.253	0.270	0.205
	Spectral	0.377	0.121	0.158	0.338	0.123

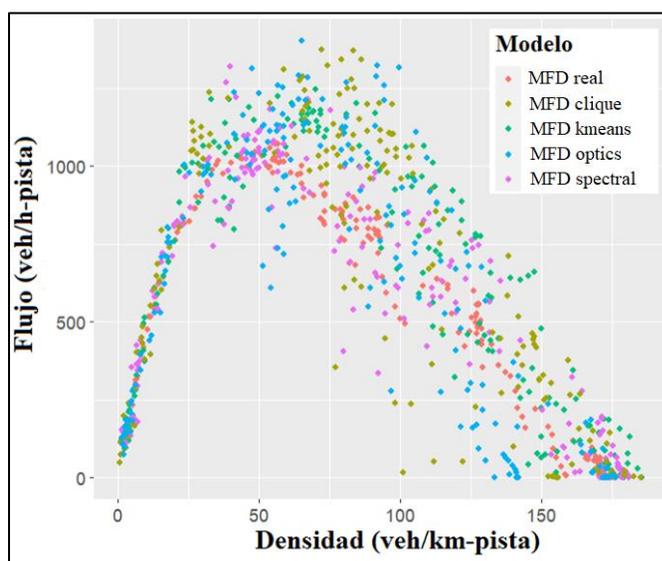


Figura 22 MFD real y eMFD construidos con diferentes métodos de agrupamiento para un día del escenario base

Fuente: Elaboración propia

Para verificar que en todos los escenarios se mantenga el mismo orden de magnitud del error y determinar si es mejor el algoritmo Spectral o *Kmeans* se repitió el proceso descrito en esta sección para las 15 simulaciones del set de entrenamiento. Los resultados numéricos se encuentran en el Anexo 1 y estos indican que el orden de magnitud del error se mantiene para todos los días y que no existe una diferencia numérica evidente en el error generado por ambos algoritmos.

La metodología propone utilizar *Kmeans* como método de agrupamiento con los datos de la Tabla 2 pese a que se dispone de datos más desagregados como los de la Tabla 1. Lo anterior se debe a que la selección del método y los datos a utilizar están relacionados. *Kmeans* es un método particional basado en densidad por lo que tiene limitaciones en cuanto al número de variables a utilizar para cada arco. Además, la complejidad computacional se reduce considerablemente al reducir el número de variables. Los errores de estimación del MFD obtenidos al utilizar los datos de la Tabla 1 y la Tabla 2 son comparables como se muestra en la Tabla 4 del Anexo 1. Por otra parte, al aplicar el algoritmo Spectral a los datos de la Tabla 1 los grupos que se obtuvieron eran de difícil contextualización a diferencia de *Kmeans* con el cual se lograban grupos de arcos lógicos teniendo en cuenta la estructura de oferta (intersecciones con control semafórico, vías unidireccionales de circulación interna, vías unidireccionales de entrada y de salida del área monitoreada) y demanda del modelo de simulación según se muestra en la Figura 23.

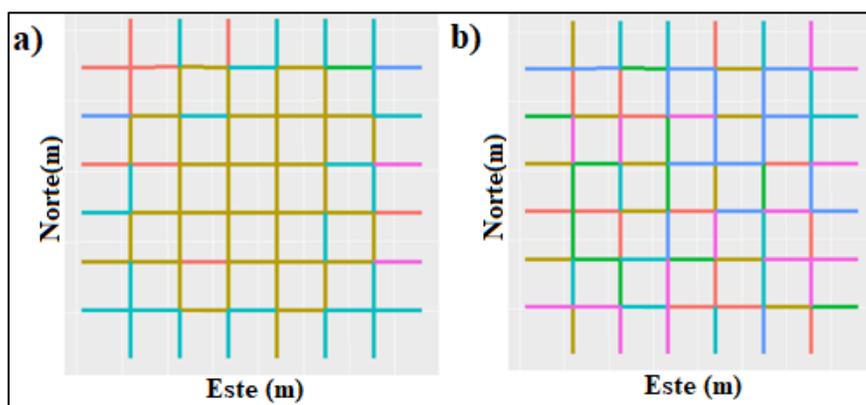


Figura 23 Grupos de arcos resultantes con datos de tabla 1. a) *Kmeans*, b) *Spectral*

Fuente: Elaboración propia

4.5. Sensibilidad del método ante la variación del número de arcos monitoreados.

Para el análisis de esta sección se utilizaron las 15 simulaciones del set de entrenamiento y se modificó la cantidad de arcos monitoreados que pueden ser seleccionados a partir de la Tabla 6. Para estudiar la sensibilidad del método ante la variación del número de arcos

monitoreados se calcularon los errores de estimación para los MFD estimados construidos con el monitoreo de 3, 6, 9, 12, 15 y 18 arcos que corresponden al 3.57%, 7.14%, 10.71%, 14.28%, 17.85% y 21.42% de la red respectivamente. Los arcos representativos resultantes y su ponderación se muestran en la Tabla 9.

Tabla 9 Ponderación P_k de los arcos representativos para cada K arcos representativos

Arco	Pk					
	3 arcos	6 arcos	9 arcos	12 arcos	15 arcos	18 arcos
6	22	14	10	8	7	6
44	30	18	13	10	9	9
55	32	20	14	11	10	9
20		11	7	6	5	5
22		8	6	5	4	4
66		13	9	7	6	6
5			8	6	5	5
33			9	8	7	6
82			8	7	6	6
52				6	5	4
59				5	4	4
84				5	4	4
1					4	4
63					4	3
77					4	3
4						3
37						2
50						1

La Figura 24 presenta los errores de estimación de los MFD estimados con los arcos y ponderaciones indicados en la Tabla 9 para las tres simulaciones del set de validación. A medida que aumenta el número de arcos representativos monitoreados se reduce el rango de error de los indicadores. Al monitorear más de 12 arcos representativos no existe una reducción del rango de error considerable. La Figura 24 muestra que a medida que aumenta el número de arcos representativos monitoreados el error medio disminuye para todos los indicadores. Para la situación modelada, sobre el 14% de la red monitoreada el beneficio de incrementar el número de arcos es menor.

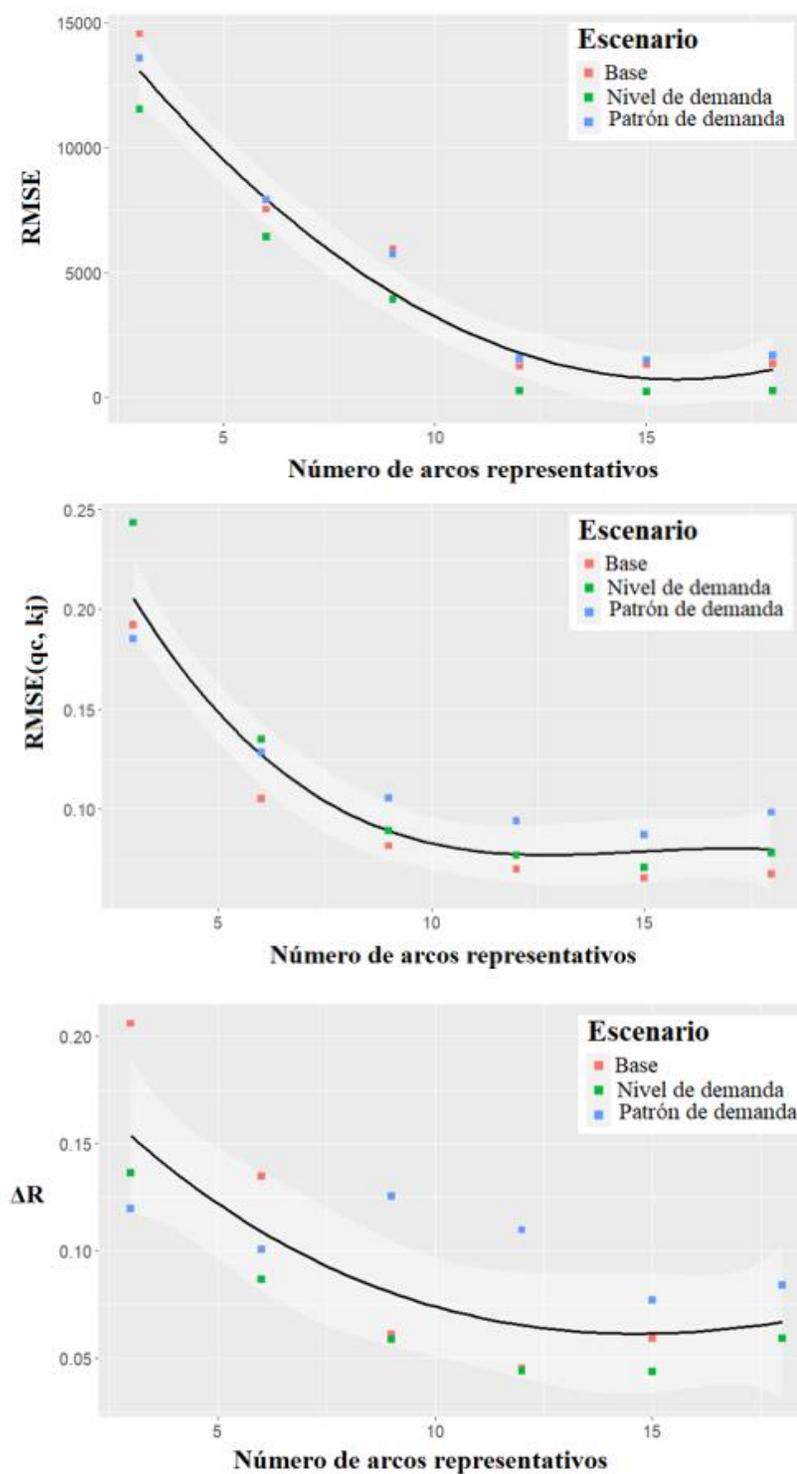


Figura 24 Comparación del número de arcos representativos. Fuente: Elaboración propia

4.6. Sensibilidad del método ante la cantidad y diversidad de los datos.

Para el análisis de esta sección se seleccionaron seis arcos representativos y se modificó la cantidad y diversidad de las simulaciones del set de entrenamiento. Las simulaciones que conforman los sets de entrenamiento son:

- A) Tres simulaciones (una simulación del escenario base, una simulación del escenario de nivel de demanda y una simulación del escenario de variación de patrón)
- B) Cinco simulaciones (cinco simulaciones del escenario base).
- C) Cinco simulaciones (cinco simulaciones del escenario de nivel de demanda).
- D) Cinco simulaciones (cinco simulaciones del escenario de variación de patrón).
- E) Diez simulaciones (cinco simulaciones del escenario base y cinco simulaciones del escenario de nivel de demanda).
- F) Diez simulaciones (cinco simulaciones del escenario base y cinco simulaciones del escenario de variación de patrón).
- G) Diez simulaciones (cinco simulaciones del escenario de nivel de demanda y cinco simulaciones del escenario de variación de patrón).

Para estudiar la sensibilidad del método ante la variación del número y diversidad de los datos se calcularon los errores de estimación para los MFD estimados que son construidos para cada set de simulaciones de entrenamiento. Los arcos representativos resultantes con cada grupo y su ponderación se muestran en la Tabla 10.

La Figura 25 presenta los errores de estimación de cada set de entrenamiento para las tres simulaciones del set de validación. El indicador ΔR es insensible a la variación del set de entrenamiento. RMSE y RMSE (qc, kj) se reducen cuando en el set de entrenamiento se incluyen escenarios de variación del nivel de demanda. La diversidad de escenarios en el set de entrenamiento representa una reducción del 50% de los errores RMSE y RMSE (qc, kj), por lo cual se concluye que la diversidad de escenarios favorece la metodología.

Tabla 10 Ponderación Pk de los arcos representativos para cada set de entrenamiento

Arco	Set de entrenamiento							Original
	A	B	C	D	E	F	G	
1						11		
3	9							
5			13				12	
6			11		19	8		14
20							11	11
22								8
26				16				
27	17							
33			13				15	
40	16	11						
41				9				
43	10							
44			19		18		21	18

Arco	Set de entrenamiento							Original
	A	B	C	D	E	F	G	
45		13						
47				13				
50					4			
52		14			13	12		
54			15					
55				23		25	15	20
59			13				10	
62				10				
66	16	16			14	17		13
80	16			13				
82		17			16			
84		13				11		

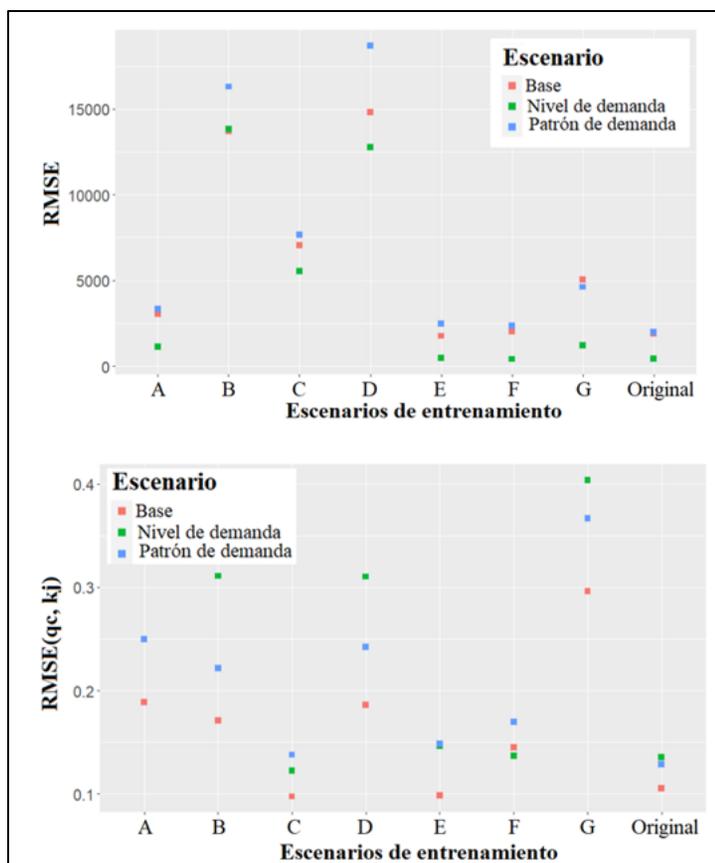


Figura 25 Comparación según el set de entrenamiento. Fuente: Elaboración propia

5. APLICACIÓN PARCIAL DE LA METODOLOGÍA EN SANTIAGO

En este capítulo se presenta un ejemplo de aplicación de la metodología propuesta. La metodología se aplica solo con una repetición del algoritmo de agrupamiento debido al costo que representa las consultas de tiempo y distancia de viaje en *Google Directions API*. El objetivo de esta aplicación es verificar que con los datos disponibles es posible aplicar la metodología y reducir la cantidad de arcos a monitorear sin contar con infraestructura propia, así como identificar futuras líneas de investigación en cuanto a las limitaciones de los datos disponibles.

El resto de este capítulo explica cómo se realizó la delimitación del área de monitoreo y como se escogió el periodo de consulta de datos. Luego, se detallan los datos recolectados, algunas de sus propiedades y su costo. Finalmente, se realiza un agrupamiento de los arcos del área de monitoreo y se establecen los arcos representativos.

No se construyó un MFD estimado para el área de monitoreo ya que el método requiere varias repeticiones del algoritmo de agrupamiento y posteriormente la medición del flujo y densidad en campo.

5.1. Delimitación del área y periodo de recolección de datos

Para delimitar el área de monitoreo se observaron las redes semaforizadas del Gran Santiago con el fin de reconocer una zona que tenga cruces semaforizados pertenecientes a una misma red. El objetivo de que el área monitoreada pertenezca a una sola red semaforizada fue facilitar el reconocimiento de los posibles periodos de monitoreo.

En la Figura 26 se muestra en gris el área de monitoreo seleccionada, la cual está cubierta por una sola red semaforizada y delimita al norte con Av. Libertador Bernardo O'Higgins, al sur con Av. Manuel Antonio Matta, al este con Av. Vicuña Mackenna y al oeste con las calles Serrano y Arturo Pratt. El área tiene 1.49 km² y cuenta con 120 arcos de los cuales

71 conforman la red estudiada. Los otros 49 arcos no se incluyen en la red estudiada por que son de tipo residencial o pasajes.

El día 21 de diciembre de 2021 se verificó que la red semafórica del área tiene una programación constante entre las 10:00 y las 13:00, por lo que se estableció esa franja horaria para realizar la extracción de datos de velocidad. La red semafórica tiene un ciclo de 104 segundos. Los semáforos de la red se dividen en dos grupos con diferente reparto de verde. El reparto de un grupo es 70 segundos de verde para la fase uno, tres segundos de amarillo, un segundo de todo rojo, 26 segundos de verde para la fase dos, tres segundos de amarillo y un segundo de todo rojo. El reparto de verde para el segundo grupo es de 52 segundos de verde para la fase uno, tres segundos de amarillo, un segundo de todo rojo, 44 segundos de verde para la fase dos, tres segundos de amarillo y un segundo de todo rojo.

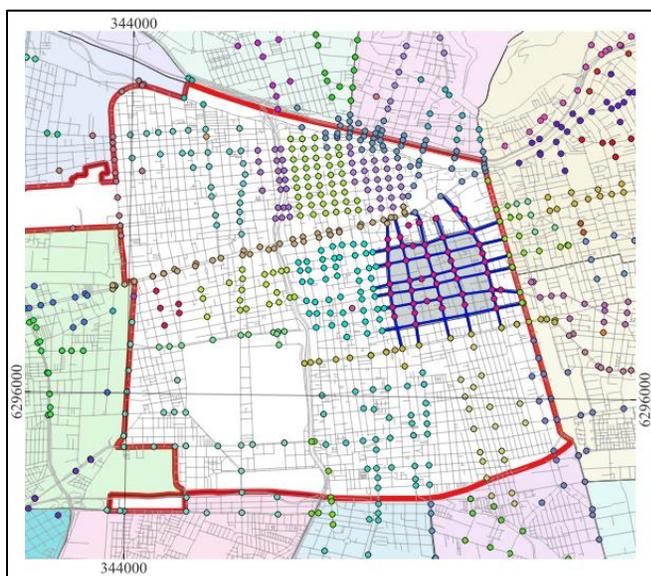


Figura 26 Área de monitoreo y arcos monitoreados en la comuna de Santiago
Fuente: Elaboración propia

5.2. Extracción de datos

La velocidad promedio de los autos en cada arco fue obtenida mediante *Google Directions API* a través de dos cuentas privadas de *Google Cloud Platform*. Se realizó una consulta

del tiempo y la distancia de viaje para recorrer cada arco de la red. El viaje consultado inicia aguas arriba de la intersección inicial y termina aguas abajo de la intersección final del arco monitoreado. La consulta para cada arco se repitió durante 170 minutos, una vez por minuto y los datos consultados corresponden al 23 de diciembre de 2021. Cada intersección se recorre en dos ocasiones, una vez con el arco aguas arriba y otra vez con el arco aguas abajo. Lo anterior entrega una estimación de velocidad correcta al considerar tanto la distancia como el tiempo en recorrer ambas intersecciones.

Para realizar las consultas fueron necesarias dos cuentas de *Google Cloud Platform* debido a que *Google Directions API* tiene un límite del número de consultas por segundo para cada cuenta, el cual está alrededor de una consulta por segundo. El costo de cada consulta fue de 0.01 USD, por tanto, el costo de la extracción de datos de un día para el área y periodo determinado es de 127.80 USD. Para obtener un set de entrenamiento que permita aplicar la metodología completa se recomienda la consulta de varios días de datos en los cuales se reconozca distintos niveles de demanda según los resultados de la validación. El costo total de la extracción de datos sería de 1278.00 USD si se consultan datos por 10 días.

Se asume que la velocidad consultada en *Google Directions API* es la velocidad media espacial utilizada en la metodología. Como resultado de la consulta se muestra en la Figura 27 una matriz de 71x170, donde cada fila representa un arco y cada columna un periodo c de ancho Δt igual a un minuto. Los datos extraídos muestran una variabilidad menor a la obtenida con el modelo de microsimulación del Capítulo 4. El rango de velocidades detectadas es menor. En la Figura 27 se puede notar que existen arcos en los cuales la velocidad se mantiene constante durante toda la consulta de datos. Lo anterior genera que el rango intercuartil y la desviación absoluta media de la Tabla 2 sean distintos de 0 solo en el 12% de los arcos por lo que estas variables no aportan mayor información para el agrupamiento en este caso.

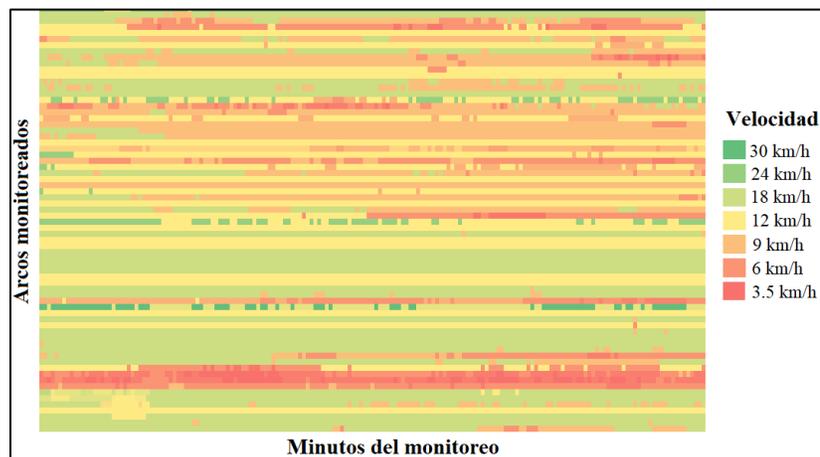


Figura 27 Campo de velocidades del área monitoreada
Fuente: Elaboración propia

5.3. Agrupamiento de los arcos

Para determinar el número de grupos que se utilizó como parámetro en el algoritmo de *Kmeans* se aplicó el método gráfico del codo (Coates & Ng, 2012). En la Figura 28 se puede ver como el negativo de la distancia media interna de los grupos (varianza intra grupal) comienza a disminuir conforme aumenta el número de grupos. Sin embargo, luego de seis grupos la distancia media interna no se reduce de forma sustancial con un grupo más. Por tanto, seis es un buen número de grupos para el conjunto de datos disponible.

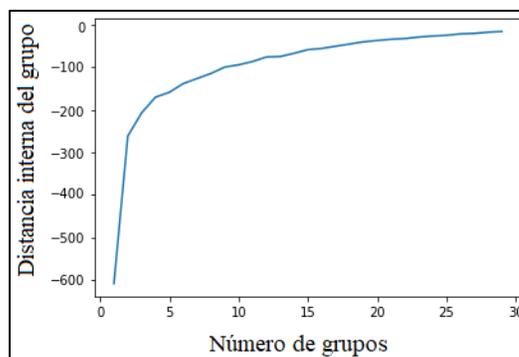


Figura 28 Método gráfico del codo. Fuente: Elaboración propia

El resultado del agrupamiento con seis grupos utilizando *Kmeans* sobre la matriz de variables de la distribución de velocidad se muestra en la Figura 29.

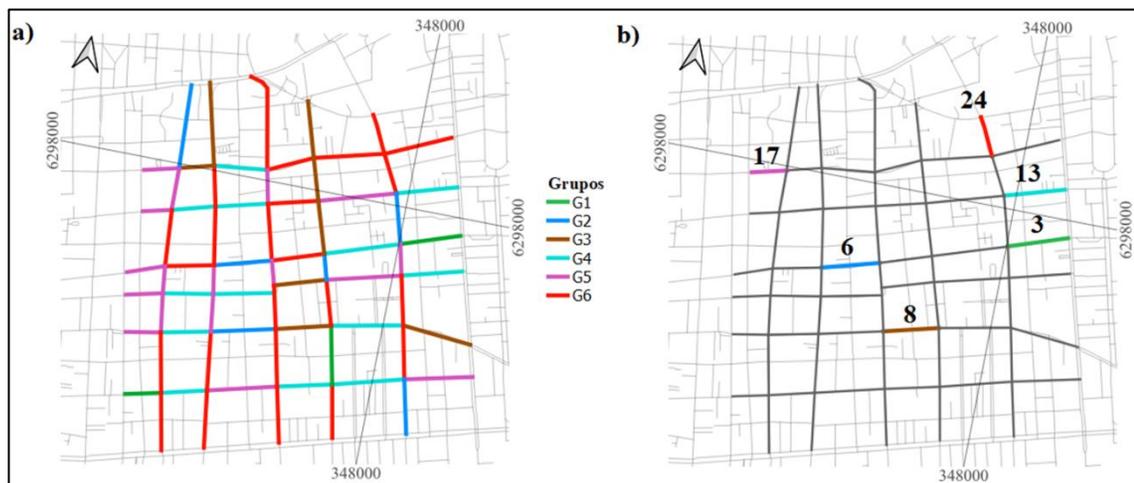


Figura 29 Resultado del agrupamiento: a) Grupo de pertenencia de cada arco, b) Representatividad de cada arco
Fuente: Elaboración propia

La cantidad de arcos por grupo ($RP_{n,z}$) es razonable teniendo en cuenta el número de grupos utilizados y está acorde a los resultados obtenidos en el modelo de microsimulación del Capítulo 4.

En el aprendizaje no supervisado, el poder describir los grupos con alguna de sus características da razón de un buen resultado. Teniendo en cuenta esto, los grupos 5 y 6 son arcos generalmente de una pista y poca congestión, mientras que los grupos 2 y 3 son arcos congestionados de varias pistas por sentido. Los arcos representativos de cada grupo parecen adecuados. En el caso de los grupos 2 y 3; estos se encuentran en dos de las principales avenidas que cruzan el área de estudio y tienen sentido de circulación contrario. De igual forma, los representantes del grupo 1, 4, 5 y 6 son arcos que tienen propiedades físicas y operacionales distintas, como por ejemplo la categoría de vía, la cantidad de pistas y el sentido de circulación.

Los hallazgos de esta aplicación muestran, con datos reales, que es posible realizar un buen agrupamiento de arcos en función de la velocidad de los vehículos de la red y que los arcos representativos resultantes tienen características diferentes entre ellos. Monitorear los seis arcos seleccionados significa instalar sensores solo en el 8.5% de la

red lo cual reduce considerablemente el costo de monitoreo. El costo de los datos para aplicar la metodología propuesta es reducido si se compara con obtener la misma información con mediciones en campo. Para recolectar la información de la velocidad de los vehículos en los 71 arcos del área estudiada durante 170 minutos el costo diario sería de 754.38 USD, considerando que el pago por hora de trabajo es de 3.75 USD. Por tanto, el costo utilizando *Google Directions API* es 5.9 veces menor.

6. CONCLUSIONES

Este trabajo propone una nueva metodología para la identificación de arcos representativos para la estimación del MFD. La metodología propuesta permite identificar un grupo acotado de arcos que, luego de conocer su flujo y densidad, sirven para estimar un MFD cercano al real. La metodología propuesta no requiere de sensores de punto fijo o de una flota de observadores móviles para identificar los arcos representativos.

Estudios anteriores que tienen objetivos semejantes a los de este trabajo, utilizan datos de flujo y la densidad de los arcos de la red, lo cual supone una limitante difícil de superar al momento de replicar esas metodologías ya que ambas variables fundamentales son aditivas. En contraposición, la velocidad media espacial no es aditiva por tanto solo se necesita conocer la velocidad de algunos de los autos que circulan por cada arco. El problema con la velocidad media espacial es que esta no es directamente una de las variables utilizadas para la construcción de un MFD. Sin embargo, la metodología propuesta solo utiliza la velocidad media espacial para identificar semejanzas del estado de tráfico de los arcos dejando la construcción del MFD estimado a cargo del flujo y densidad que se deben medir solo sobre los arcos representativos.

Los resultados de la validación muestran que una buena forma de utilizar la velocidad de los autos en un arco, es transformando esta información para obtener variables descriptivas de la distribución de velocidad de los autos en el arco. El uso de estas variables reduce la complejidad computacional del algoritmo de agrupamiento y presenta resultados semejantes a los que se pueden lograr con los datos más desagregados. Además, el uso de variables descriptivas de la distribución de velocidad de los autos en un arco genera agrupamientos que son consistentes con lo que se esperaría obtener conociendo la configuración de oferta y demanda de la red.

Los hallazgos de este estudio muestran que *Kmeans* es un algoritmo adecuado para realizar el agrupamiento de arcos. Para validar su uso se evaluaron varios algoritmos que se

diferencian de *Kmeans* por su formulación, complejidad y función objetivo. Al evaluar los errores de estimación del MFD obtenidos con *Kmeans* se puede ver que no siempre su uso genera el menor error de estimación, sin embargo, es competitivo en la mayoría de casos. Los grupos de arcos que se generaron con *Kmeans* fueron siempre consistentes con lo que se esperaría obtener si se conoce la configuración de oferta y demanda de la red.

La heurística propuesta para seleccionar los arcos representativos toma en consideración tanto la cantidad de días que un arco fue representativo como la cantidad de arcos que representó. Los criterios utilizados permiten tomar en cuenta los diferentes agrupamientos y arcos representativos que se obtienen debido a la variación de la demanda, a la vez que los pondera de acuerdo a su frecuencia. Los errores de estimación del MFD que se obtienen con la heurística propuesta son mayores a los de otras metodologías, pero siguen siendo alentadores considerando los datos necesarios.

A medida que aumenta el número de arcos representativos monitoreados el método es más preciso y más confiable. Sin embargo, para la red simulada, sobre el 14% de la red monitoreada el beneficio de incrementar el número de arcos es menor. Por otra parte, la metodología mejora su precisión y confiabilidad mientras mayor diversidad de escenarios de demanda son considerados. Incrementar el número de arcos monitoreados o el número de escenarios constituye una restricción presupuestaria de este problema.

La aplicación de la metodología en un área de Santiago muestra que los datos consultados mediante *Google Directions API* permiten realizar un buen agrupamiento de arcos en una red real en base a la velocidad promedio y los arcos representativos son de buena calidad. Además, el costo de las consultas en *Google Directions API* es de aproximadamente 5.9 veces menor si se compara con el costo de obtener los mismos datos con mediciones en campo.

Finalmente, durante este desarrollo surgieron interrogantes para ser resueltas en trabajos futuros como evaluar el desempeño de la metodología al variar la topología y tamaño de

la red a monitorear. Por otra parte, como se mencionó en la sección 4.1.1.1. fue necesario calibrar la demanda del modelo de microsimulación para obtener un MFD sin histéresis por lo cual surge la pregunta ¿Cómo determinar si existe histéresis del MFD en un área con los datos de velocidad de sus arcos? En el modelo de microsimulación se asignó al 30% de los vehículos el enrutamiento dinámico, sin embargo, este porcentaje es desconocido en la realidad y como se mencionó en la sección 2.2. el enrutamiento dinámico es un parámetro importante para el MFD ya que ayuda en la distribución homogénea de la congestión. Por tanto, el conocimiento del porcentaje de vehículos en Santiago que utiliza aplicaciones de enrutamiento dinámico y los incentivos para su uso se vuelven preguntas relevantes. Por otra parte, como se mencionó en la sección 3.4 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 de tráfico de la red si se actualiza la representatividad de cada arco monitoreado en tiempo real? ¿Es posible estimar el estado de tráfico de la red en periodos con distinto MFD modificando la representatividad de cada arco monitoreado? Por último, en las últimas secciones del capítulo 4 se ha determinado que la metodología propuesta es sensible a la variación de sus parámetros, por tanto, resulta atractivo modificar la metodología propuesta en base a herramientas basadas en datos como el aprendizaje profundo de máquinas.

REFERENCIAS

Ambühl, L., Loder, A., Bliemer, M. C. J., Menendez, M., & Axhausen, K. W. (2020). A functional form with a physical meaning for the macroscopic fundamental diagram. *Transportation Research Part B: Methodological*, 137, 119–132. <https://doi.org/10.1016/j.trb.2018.10.013>

Ambühl, L., Loder, A., Leclercq, L., & Menendez, M. (2021). Disentangling the city traffic rhythms: A longitudinal analysis of MFD patterns over a year. *Transportation Research Part C: Emerging Technologies*, 126. <https://doi.org/10.1016/j.trc.2021.103065>

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>

Castrillon, F., & Laval, J. (2017). Impact of buses on the macroscopic fundamental diagram of homogeneous arterial corridors. *Transportmetrica B*, 1–16. <https://doi.org/10.1080/21680566.2017.1314203>

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. (2005). A variational formulation of kinematic waves: Basic theory and complex boundary conditions. *Transportation Research Part B: Methodological*, 39(2), 187–196. <https://doi.org/10.1016/j.trb.2004.04.003>
- Daganzo, C. F. (2007). Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1), 49–62. <https://doi.org/10.1016/j.trb.2006.03.001>
- Daganzo, C. F., Gayah, V. v., & Gonzales, E. J. (2012). The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions. *EURO Journal on Transportation and Logistics*, 1(1–2), 47–65. <https://doi.org/10.1007/s13676-012-0003-z>
- 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.

Forina, M., Armanino, C., & Raggio, V. (2002). Clustering with dendrograms on interpretation variables. In *Analytica Chimica Acta* (Vol. 454).

Garcia-Huidobro. (2018). Estimación del diagrama fundamental macroscópico mediante datos de bluetooth. Pontificia Universidad Católica de Chile. <https://repositorio.uc.cl/handle/11534/23043>

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>

Godfrey, J. W. (1969). The mechanism of a road network. *Traffic Engineering & Control*, 8(8).

Goodwin, P. B. (1996). Empirical evidence on induced traffic A review and synthesis. In *Transportation* (Vol. 23).

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

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>

Lighthill Michael James, & Whitham Gerald Beresford. (1955). On kinematic waves II. A theory of traffic flow on long crowded roads. <https://doi.org/10.1098/rspa.1955.0089>

Mariotte, G., Leclercq, L., & Laval, J. A. (2017). Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. *Transportation Research Part B: Methodological*, 101, 245–267. <https://doi.org/10.1016/j.trb.2017.04.002>

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>

Paipuri, M., Xu, Y., González, M. C., & Leclercq, L. (2020). Estimating MFDs, trip lengths and path flow distributions in a multi-region setting using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 118. <https://doi.org/10.1016/j.trc.2020.102709>

Richards, P. I. (1956). Shock Waves on the Highway. *Operations Research*, 4(1), 42–51. <https://doi.org/10.1287/opre.4.1.42>

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>

Sana, B., Castiglione, J., Cooper, D., & Tischler, D. (2017). Using google's aggregated and anonymized trip data to support freeway corridor management planning in San

Francisco, California. *Transportation Research Record*, 2643(1), 65–73.
<https://doi.org/10.3141/2643-08>

Smith, S. P., & Jain, A. K. (1984). Testing for Uniformity in Multidimensional Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(1), 73–81.
<https://doi.org/10.1109/TPAMI.1984.4767477>

Tilg, G., Amini, S., & Busch, F. (2020). Evaluation of analytical approximation methods for the macroscopic fundamental diagram. *Transportation Research Part C: Emerging Technologies*, 114, 1–19. <https://doi.org/10.1016/j.trc.2020.02.003>

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*.

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>

REFERENCIAS DIGITALES

Google (2015). Abordar la movilidad urbana con tecnología, 18 de noviembre de 2015

<https://europe.googleblog.com/2015/11/tackling-urban-mobility-with-technology.html>

Google (2009_1). Arterial traffic available on Google, 25 de agosto de 2009

<http://google-latlong.blogspot.com/2009/08/arterial-trafficavailable-on-google.html>

Google (2009_2). The bright side of sitting in traffic: Crowdsourcing road congestion data

<https://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html>

Google (2020). How AI helps predict traffic and determine routes, 03 de septiembre de 2020

<https://blog.google/products/maps/google-maps-101-how-ai-helps-predict-traffic-and-determine-routes/>

Developers (2021). Distance Matrix API Usage and Billing

<https://developers.google.com/maps/documentation/distance-matrix/usage-and-billing>

Deepmind (2021)

<https://deepmind.com/blog/article/traffic-prediction-with-advanced-graph-neural-networks>

ANEXO

Tabla 1: Errores en la estimación del MFD con los días del escenario base

Error de estimación	Método	Días del escenario base				
		1	2	3	4	5
RMSE/E04	Kmeans	1.1	1.2	1.2	1.1	1.2
	Spectral	1.3	1.1	1	1	1.3
RMSE(qc,kj)	Kmeans	0.17	0.159	0.204	0.179	0.168
	Spectral	0.216	0.153	0.145	0.2	0.151
ΔR	Kmeans	0.472	0.093	0.216	0.192	0.100
	Spectral	0.377	0.121	0.158	0.338	0.123

Tabla 2: Errores en la estimación del MFD con los días del escenario de nivel de demanda

Error de estimación	Método	Días del escenario de nivel de demanda				
		1	2	3	4	5
RMSE/E04	Kmeans	1.5	0.9	9.1	1.1	1.1
	Spectral	1.2	1.1	7.5	1.1	1.3
RMSE(qc,kj)	Kmeans	0.14	0.2	0.06	0.12	0.1
	Spectral	0.35	0.38	0.23	0.27	0.26
ΔR	Kmeans	0.307	0.219	0.086	0.078	0.076
	Spectral	0.207	0.135	0.076	0.085	0.067

Tabla 3: Errores en la estimación del MFD con los días del escenario de variación de patrón

Error de estimación	Método	Días del escenario de variación de patrón				
		1	2	3	4	5
RMSE/E04	Kmeans	1	1.4	7.3	7.8	9.3
	Spectral	1.6	1.1	9.7	1.5	8.7
RMSE(qc,kj)	Kmeans	0.145	0.174	0.122	0.154	0.136
	Spectral	0.215	0.148	0.168	0.264	0.148
ΔR	Kmeans	0.108	0.121	0.111	0.265	0.136
	Spectral	0.212	0.142	0.182	0.382	0.168

Tabla 4: Errores en la estimación del MFD con los días del escenario base con datos de la tabla 1 y 2

Error de estimación	Datos*	Días del escenario base				
		1	2	3	4	5
RMSE * E04	Tabla 1	1.7	1.4	1.4	1.7	2.4
	Tabla 2	1.1	1.2	1.2	1.1	1.2
RMSE(qc,kj)	Tabla 1	0.16	0.106	0.155	0.166	0.119
	Tabla 2	0.17	0.159	0.204	0.179	0.168
ΔR	Tabla 1	0.181	0.087	0.223	0.231	0.167
	Tabla 2	0.472	0.093	0.216	0.192	0.100

* Tabla 1: 195 datos de velocidad por arco. Tabla 2: datos de la distribución de velocidad