



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

INFERENCIA BAYESIANA DE ELECCIÓN DE RUTAS A PARTIR DE TECNOLOGÍA BLUETOOTH

FRANCISCO ORLANDO GARRIDO VALENZUELA

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

Profesores Supervisores:
JUAN CARLOS HERRERA MALDONADO
SEBASTIÁN RAVEAU FELIÚ

Santiago de Chile, octubre, 2018

© MMXVIII, FRANCISCO ORLANDO GARRIDO VALENZUELA



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

INFERENCIA BAYESIANA DE ELECCIÓN DE RUTAS A PARTIR DE TECNOLOGÍA BLUETOOTH

FRANCISCO ORLANDO GARRIDO VALENZUELA

Miembros del Comité:

JUAN CARLOS HERRERA MALDONADO

SEBASTIÁN RAVEAU FELIÚ

HOMERO LARRAIN IZQUIERDO

RAFAEL DELPIANO COSTABAL

RICARDO JORGE RAINERI BERNAIN

Tesis para optar al grado de

Magíster en Ciencias de la Ingeniería

Santiago de Chile, octubre, 2018

© MMXVIII, FRANCISCO ORLANDO GARRIDO VALENZUELA

*A mi abuelo Francisco, mi madre
Eliana, familia y amigos que fueron
el pilar fundamental en el
desarrollo de esta etapa.*

AGRADECIMIENTOS

Este documento implica el cierre de un capítulo tanto en mi vida personal como en mi formación profesional. Pero este logro no hubiera sido posible sin el apoyo de mi familia, amigos y profesores, a los que quisiera agradecer en este apartado.

Quisiera comenzar por mi madre Eliana quien ha sido el pilar fundamental en el desarrollo de mi formación universitaria, gracias por el continuo apoyo y las ganas de superación que sembraste en mi. Gracias por estar presente en toda circunstancia y nunca dejarme de lado pese a las frustraciones que implicó el desarrollo de este proceso. También quiero agradecer a mis abuelos Francisco Valenzuela y Adriana Gálvez por estar presentes en mi formación temprana, inculcarme valores y a ser responsable con mis quehaceres lo que me ayudó a ser la persona que soy hoy en día. Agradezco además a mi hermano Matias por soportar mi ánimos y a mi padre Ramón.

Esta tesis no hubiera podido ser realizada sin la intachable supervisión de los profesores Juan Carlos Herrera y Sebastián Raveau. Ellos fueron los encargados de mantenerme motivado en todo momento y no tirar la toalla en las situaciones de frustración. Ambos son profesores de muy alto nivel y su experiencia fue crucial en este trabajo. Quisiera añadir que Sebastián me motivó a presentar este trabajo en inglés en dos conferencias internacionales. Hecho que me llenó de experiencias y nuevas habilidades que me entregan la confianza que tengo hoy. Muchas gracias amigo y ahora colega.

Quiero agradecer a Rafa Delpiano por motivarme en el mundo de la programación, ser su ayudante e incluso tener el honor de ayudarlo en otra universidad. En esta última se mantuvieron las primeras discusiones de este trabajo. Gracias también a los profesores Homero Larrain, Ricardo Olea y Ricardo Aravena por ayudarme en temas teóricos en el desarrollo de esta investigación. Por otro lado quisiera agradecer al profesor David

Painequeo por su motivación temprana y convertirme en una persona crítica, hacerme notar que soy capaz de hacer las cosas que me propongo y que el esfuerzo y perseverancia es el camino al éxito.

También quisiera agradecer a mis amigos, sin ellos, el paso por postgrado hubiera sido solitario. La motivación y ayuda recibida por ellos queda plasmada en la calidad de este trabajo. Agradesco a Ignacio Tiznado, Jaime Soza, Guillermo Soto, Sebastián Muñoz, Ignacio Arismendi, Andra Vera, Muriel Manriquez, Owen Bull, Nicolás Badiola, Gonzalo Suazo, Luis Cantillo, Felipe González, Mariana Valle, Sebastián García-Huidobro, Sebastián Tamblay, Jaime Moya, Alejandro Schmidt y Marina Dragicevic.

En la última etapa de este trabajo fue súper importante el apoyo de Constanza Gutiérrez. Gracias por acompañarme todas esas noches pasando de largo, por ayudarme a construir esos gráficos, por tu motivación y por creer que podía. Muchas gracias por eso y por lo ricos sentimientos que acompañaron esta etapa.

Finalmente quiero agradecer a la Pontificia Universidad Católica de Chile por la formación entregada, a la Escuela de Ingeniería por la calidad de sus profesores y en especial al Departamento de Ingeniería de Transporte y Logística por ser una comunidad cercana, preocupada de sus alumnos y que cuenta con un equipo humano de primer nivel. Por último a CONICYT que por medio del FONDECYT 1160943 se financió parte de esta investigación.

ÍNDICE GENERAL

AGRADECIMIENTOS	iv
Índice de figuras	ix
Índice de cuadros	xii
RESUMEN	xiii
ABSTRACT	xv
Capítulo 1. INTRODUCCIÓN	1
1.1. Hipótesis de investigación	2
1.2. Objetivo general de investigación	3
1.3. Objetivos específicos	3
1.4. Alcances de investigación	3
1.5. Contenidos	4
Capítulo 2. MARCO TEÓRICO	5
2.1. Tecnología Bluetooth	5
2.1.1. El Bluetooth y su clasificación según tipos de dispositivo	5
2.1.2. Funcionamiento de la comunicación Bluetooth entre dispositivos	7
2.1.3. Sensores Bluetooth y detecciones	8
2.1.4. Limitaciones propias de la tecnología	9
2.2. Usos de tecnologías inalámbricas en transporte	10
2.2.1. Posicionamiento de entidades en entornos cerrados	12
2.2.2. Predicción de tiempos de viaje	14
2.2.3. Generación de matrices origen destino	16
2.2.4. Reconstrucción de rutas de vehículos	17
2.2.5. Consideraciones del uso de datos Bluetooth en transporte	20

2.3. Resumen del capítulo	21
Capítulo 3. METODOLOGÍA	23
3.1. Etapa 1: Definición de la red	25
3.2. Etapa 2: Precarga de información	27
3.2.1. Tiempo de permanencia en nodos (DT_i^h)	28
3.2.2. Tiempo de viaje en arcos (TT_j^h)	31
3.3. Etapa 3: Reconstrucción de rutas	34
3.3.1. Inferencia Bayesiana	34
3.3.2. Cálculo de probabilidades de ruta	47
Capítulo 4. VALIDACIÓN	49
4.1. Definición del experimento	49
4.1.1. Sensores Bluetooth	50
4.1.2. Red	51
4.1.3. Demanda	52
4.1.4. Escenarios a evaluar	55
4.2. Medidas de desempeño	60
4.2.1. Análisis de ajuste predictivo	60
4.2.2. Análisis de flujos obtenidos	62
4.2.3. Análisis de predicción dependiendo de la velocidad	63
Capítulo 5. RESULTADOS	64
5.1. Comparaciones entre escenarios	64
5.1.1. Estimación de flujos por escenario	64
5.1.2. Impactos del uso de los modelo <i>a priori</i>	66
5.1.3. Capacidad de predicción en la reconstrucción de rutas	67
5.1.4. Cantidad de detecciones producidas en cada escenario	69
5.2. Resultados de escenarios evaluados	71
5.2.1. Escenario 36:1:C	72

5.2.2. Escenario 36:8:C	74
5.2.3. Escenario 18h:1:C	77
5.2.4. Escenario 18u:1:C	80
5.2.5. Escenario 36:1:E	82
5.3. Comparación con metodología existente en literatura	85
Capítulo 6. CONCLUSIONES	89
REFERENCIAS	93
ANEXOS	99
A. ERRORES DE ESTIMACIÓN	100
A.1. Error cuadrático medio (MSE)	100
A.2. Error porcentual ponderado absoluto medio (WMAPE)	100
B. DESARROLLO Y PROGRAMACIÓN DE LA METODOLOGÍA	102
B.1. Archivo de base de datos de detecciones Bluetooth	102
B.2. Archivo de codificación de nodos	103
B.3. Archivo de codificación de arcos	104
C. RESULTADOS GENERALES DE ESCENARIOS	105

ÍNDICE DE FIGURAS

2.1. Geometrías de zonas de detección para sensores Bluetooth direccionales y omnidireccionales	9
2.2. Porcentaje de celulares con Bluetooth desde 2001 a 2010	12
2.3. Vehículos con redes Bluetooth y tasa de penetración de la red en Europa desde 2001 a 2010	14
2.4. Ejemplo de una situación cuando no es posible conocer la ruta que realizó un vehículo producto de la falla de un sensor Bluetooth	19
3.1. Etapas de la metodología	25
3.2. Ejemplo de una red vial y una posible configuración de sensores Bluetooth . .	26
3.3. Grafo $G(N, L)$ construido a partir de la red vial de la Figura 3.2	27
3.4. Tiempo de permanencia en nodo para un vehículo del caso (iii)	30
3.5. Tiempo de viaje en arcos para un vehículo del caso (iii)	33
3.6. Distribuciones de tiempo de permanencia en nodos y tiempo de viaje en arcos para un red ejemplo al finalizar la etapa 2 de la metodología	34
3.7. Detecciones de un vehículo en las que se puede realizar una inferencia de ruta	35
3.8. Definición de costos utilizados en el algoritmo de k rutas mínimas	37
3.9. Error de medición de la variable T_{obs}	44
3.10. Ejemplo esquemático de la aplicación de la metodología sobre la ruta completa	48
4.1. Probabilidad de detección en función de la velocidad de circulación	51

4.2.	Red vial utilizada en la simulación	52
4.3.	Distribución de la demanda sobre la red vial	53
4.4.	Clasificación de la demanda respecto a las detecciones realizadas por los SB .	53
4.5.	Matriz OD de flujos en vehículos por hora para la red vial	54
4.6.	Distribución de SB en configuración 18 <i>h</i>	57
4.7.	Distribución de SB en configuración 18 <i>u</i>	58
4.8.	Distribución de SB en configuración 27	59
4.9.	Distribución de SB en configuración 36	60
5.1.	Indicador de error WMAPE determinado para todos los escenarios	65
5.2.	Error cuadrático medio para comparar modelos <i>a priori</i> entre escenarios de igual demanda	67
5.3.	Valores del indicador FPR analizado por secciones	68
5.4.	Valores del indicador FPR analizado por viajes	69
5.5.	Total de detecciones producidas por escenario en función de las características de cada uno	70
5.6.	Comparación de flujos reales e inferidos para Escenario 36:1:C	73
5.7.	Histograma de proporción de aciertos en secciones por velocidades de vehículos para Escenario 36:1:C	74
5.8.	Comparación de flujos reales e inferidos para Escenario 36:8:C	76
5.9.	Histograma de proporción de aciertos en secciones por velocidades de vehículos para Escenario 36:8:C	77

5.10. Comparación de flujos reales e inferidos para Escenario 18h:1:C	79
5.11. Histograma de proporción de aciertos de secciones por velocidades de vehículos para Escenario 18h:1:C	80
5.12. Histograma de proporción de aciertos de secciones por velocidades de vehículos para Escenario 18u:1:C	82
5.13. Comparación de flujos reales e inferidos para Escenario 36:1:E	84
5.14. Histograma de proporción de aciertos de secciones por velocidades de vehículos para Escenario 36:1:E	85
5.15. Valores de FPR de secciones entre el escenario de conteos en la BD y la literatura escenarios evaluados	86
5.16. Valores de FPR de viajes entre el escenario de conteos en la BD y la literatura escenarios evaluados	86
5.17. Error WMAPE entre los escenarios de conteos en BD y la literatura (vista 2) .	88
B.1. Ejemplo de un extracto del archivo de base de datos detecciones Bluetooth . .	103
B.2. Ejemplo de un extracto del archivo de codificación de nodos de la red	104
B.3. Ejemplo de un extracto del archivo de codificación de arcos de la red	104

ÍNDICE DE CUADROS

2.1. Clasificación de dispositivos Bluetooth según energía de salida y distancia máxima de conexión	6
3.1. Ejemplo de un registro de la base de datos de detecciones Bluetooth	24
4.1. Resumen de escenarios en función de los segundos de interdetección, modelo de probabilidad <i>a priori</i> y cantidad de sensores Bluetooth	56
5.1. Resultados análisis predictivo Escenario 36:1:C	72
5.2. Indicadores de error en flujos inferidos de viajes Escenario 36:1:C	73
5.3. Resultados análisis predictivo Escenario 36:8:C	75
5.4. Indicadores de error en flujos inferidos de viajes Escenario 36:8:C	76
5.5. Resultados análisis predictivo Escenario 18h:1:C	78
5.6. Indicadores de error en flujos inferidos de viajes Escenario 18h:1:C	79
5.7. Resultados análisis predictivo Escenario 18u:1:C	81
5.8. Indicadores de error en flujos inferidos de viajes Escenario 18u:1:C	81
5.9. Resultados análisis predictivo Escenario 36:1:E	83
5.10. Indicadores de error en flujos inferidos Escenario 36:1:E	84
C.1. Resumen resultados generales de escenarios	106

RESUMEN

La tecnología Bluetooth puede ser utilizada para seguir vehículos equipados mediante la dirección MAC. Esto es posible por medio de la instalación de sensores Bluetooth en un conjunto específico de intersecciones de un área de estudio. Sin embargo, un vehículo equipado con esta tecnología no necesariamente es detectado por todos los sensores por los que este transita. Esto se debe a que la probabilidad de detección depende de varios factores tales como el clima, la infraestructura cercana o la velocidad de los vehículos. Dada esta falta de información, es necesario inferir la ruta más probable que utilizó el vehículo.

Este estudio presenta una metodología capaz inferir las rutas elegidas por vehículos en una red. La inferencia se realiza a partir de la reconstrucción del recorrido realizado por el vehículo entre dos detecciones sucesivas donde se estiman probabilidades de uso para cada ruta. Una vez que las probabilidades son obtenidas, es posible inferir la ruta completa desde el origen a su destino. El predecir los patrones de viaje es esencial para planificación de transporte y la aplicación de diferentes políticas como gestión de tráfico y tarificación vial.

La metodología cuenta con tres etapas. En la primera de ellas, se construye un grafo a partir de la red vial y la localización de sensores Bluetooth. En la segunda etapa, con la información Bluetooth recolectada se calibran distribuciones del tiempo de permanencia en cada uno de los nodos y del tiempo de viaje para cada arco. El tercer y último paso consiste en convolucionar las distribuciones de los tiempos de nodos y arcos entre un par de detecciones sucesivas con el fin de obtener una distribución agregada para cada ruta potencial. Basado en el tiempo observado de viaje de cada vehículo y el conjunto de detecciones que no se realizaron, se realiza una inferencia Bayesiana para obtener las probabilidades de cada ruta.

La metodología es testeada mediante microsimulación y presenta buenos resultados a la hora de inferir las rutas elegidas por cada vehículo. Además, es comparada con otra metodología presente en la literatura. Ambas poseen similar porcentaje de predicción, pero la metodología propuesta muestra mejores resultados cuando la red presenta menor densidad de sensores y falta más información.

Palabras Claves: Tecnología Bluetooth, Reconstrucción de rutas, Inferencia Bayesiana, Dirección MAC.

ABSTRACT

Bluetooth technologies can be used to track vehicles by their MAC address; this can be done by installing Bluetooth sensors in selected intersections of a study area. However, a vehicle equipped with this technology is not necessarily detected by every sensor because detection probabilities depend on many factors such as weather, nearby infrastructure or vehicles' speeds. Given the lack of perfect information, it is necessary to infer the most likely route chosen by the vehicles, based on the information provided by the Bluetooth sensors.

This study presents a methodology to infer route choices in a network by reconstructing paths between two successive detection of a vehicle. Choice probabilities for each route between these detections' points are estimated. Once these probabilities are obtained, it is possible to infer the entire route from the vehicle's origin to its destination. Predicting these travel patterns is essential for transport planning and analysing different policies such as traffic management and route pricing.

The methodology has three steps. In the first step, a graph is developed with the network configuration and the sensors' location. In the second step, based on the Bluetooth data, distributions of the time vehicles spend at each intersection and the travel time between intersections are obtained. The third step consists on convolving the previous distributions between two successive detections to obtain aggregated distributions for each potential route. Based on the observed travel time and missed detections, Bayesian inference is performed to obtain the choice probabilities of each route.

The methodology is tested with microsimulation with high performance to obtain the route chosen by vehicles. It is also compared with another approach proposed in the literature. Although the proposed methodology has similar prediction capabilities, it shows better results in cases when the sensors data is incomplete.

Keywords: Bluetooth technology, Route reconstruction, Bayesian inference, MAC address.

CAPÍTULO 1. INTRODUCCIÓN

En las últimas décadas los niveles de congestión vehicular han ido en un constante aumento en varias ciudades del mundo. Lindsey y Verhoef (2001) mencionan que la congestión impone varios costos a los viajeros, tales como la reducción en las velocidades de circulación (por ende, mayores tiempos de viaje), mayor incertidumbre en la duración de los viajes, mayor consumo de combustible e incluso es capaz de impactar en cómo las personas escogen su lugar de vivienda y trabajo. Este problema ha generado la necesidad de comprender cómo se comportan las personas y construir modelos que permitan anticipar su reacción ante congestión u otros eventos, tales como, efectos de la lluvia sobre los patrones de comportamiento, desvíos en caso de accidentes, entre otros.

Uno de los aspectos necesarios de estudiar para comprender la movilidad de la población es la forma en que se realizan los viajes. Es decir, conocer desde y hacia dónde se mueven las personas, en qué modo lo hacen y qué ruta eligen. En concreto, conocer en detalle la ruta escogida por cada vehículo permite estimar flujos vehiculares en cada vía de la red, estudiar cambios en los patrones de viaje frente a modificaciones viales y conocer qué rutas se utilizan cuando ocurre un accidente o frente a algún evento importante que pueda afectar la decisión de las personas en la vialidad. Por otro lado, es posible anticipar fenómenos de comportamiento y aplicar políticas de carácter preventivo a mediano–largo plazo para reducir los impactos.

El auge de tecnologías como tarjetas inteligentes de pago en transporte público, dispositivos GPS, señales Bluetooth y/o Wi-Fi han dado origen a nuevos mecanismos para recolectar, agrupar y filtrar información de movilidad con mayor precisión, mayor frecuencia de actualización y menores costos, lo que ha ayudado a tener una mejor comprensión de los viajes. Algunas de estas tecnologías no solo permiten tener herramientas eficientes de recopilación de datos a menores costos, sino que también son capaces de entregar más

información de la habitual (Bhaskar y Chung, 2013). Es esperable que estos nuevos mecanismos traigan consigo nuevos desafíos, principalmente respecto a fallas en los equipos, inconsistencias de detección y privacidad de personas (Michau, Nantes, Chung, Abry, y Borgnat, 2014).

Esta investigación desarrolla una metodología para reconstruir las rutas de vehículos que circulan por una ciudad o entorno urbano equipado parcial o totalmente con sensores Bluetooth. Estos sensores son capaces de detectar vehículos que circulen con algún dispositivo Bluetooth compatible. Los sensores buscan los dispositivos circundantes cada cierta unidad de tiempo, lo que se conoce como tiempo de interdetección. La información entregada por los sensores permite realizar el seguimiento del vehículo a través de la red, con lo que se puede identificar su ruta. La metodología propuesta además es capaz de inferir la ruta realizada por los vehículos que no han podido ser detectados en tramos intermedios por los sensores, que de acuerdo a Michau *et al.* (2014) es una de las fallas más recurrentes en este tipo de tecnologías. Es por eso que, cuando fallan los sensores de una ruta elegida por un vehículo, no es posible determinar qué ruta fue seleccionada. De este modo se hace necesario contar con un modelo que sea capaz de usar la información proporcionada por las señales Bluetooth para discriminar entre las opciones posibles.

1.1. Hipótesis de investigación

Una red equipada con sensores Bluetooth en algunas de sus intersecciones entrega datos suficientes para reconstruir el detalle de las rutas elegidas por los vehículos de manera confiable.

1.2. Objetivo general de investigación

El objetivo de esta investigación es desarrollar una metodología de inferencia y reconstrucción de rutas de vehículos en un contexto urbano a partir de datos obtenidos exclusivamente mediante sensores Bluetooth u otra señal inalámbrica de transferencia de datos (*e.g.* Wi-Fi).

1.3. Objetivos específicos

1. Analizar los niveles de predicción de la metodología propuesta frente a:
 - Distinto número y distribución de sensores.
 - Distintos tiempos de interdetección del sensor Bluetooth.
 - Distintas formas de asignar una probabilidad *a priori* de elección de rutas, necesaria para la aplicación de la metodología.
2. Crear una herramienta computacional que determine las probabilidades de uso de rutas para cada vehículo y que pueda ser utilizada en planificación de transporte.

1.4. Alcances de investigación

1. La cantidad y distribución de sensores Bluetooth dentro de la red es un dato de entrada para la metodología y no se realiza una optimización para conocer la distribución más eficiente.
2. El análisis se realiza en base a resultados de una microsimulación. Esto permite conocer las rutas realmente escogidas por los vehículos para validar la metodología, lo que es difícil de conocer en la práctica con datos reales.

1.5. Contenidos

El resto de este documento se estructura en cinco capítulos. En el segundo Capítulo se presenta el marco teórico de esta investigación donde se introduce qué es la tecnología Bluetooth, sus características, funcionamiento, fallas y otros aspectos técnicos. Además, se desarrolla una revisión de la literatura donde se dan a conocer los avances en investigación del uso de tecnologías inalámbricas aplicadas a transporte y a la reconstrucción de rutas en particular. El tercer Capítulo describe la metodología propuesta para la reconstrucción de rutas y los aspectos de su modelación, desarrollo y validación. En el cuarto Capítulo se detallan los mecanismos utilizados para evaluar la metodología y se describen las herramientas y metodologías de validación. El quinto Capítulo presenta los resultados de la metodología frente a los seis escenarios más concluyentes de los 24 testeados. Además, presenta los resultados de un *benchmark* realizado con una metodología de reconstrucción de rutas ya propuesta en la literatura. Finalmente en el sexto Capítulo de conclusiones, se resumen y cierran los objetivos propuestos en esta investigación y se dejan abiertas futuras líneas de investigación.

CAPÍTULO 2. MARCO TEÓRICO

La propuesta de reconstrucción de rutas de vehículos descrita en esta investigación se realiza mediante el uso de detecciones Bluetooth, por lo que es importante entender el real y completo funcionamiento de esta tecnología. Este capítulo describe detalladamente la tecnología Bluetooth, sus características, confiabilidad, limitaciones y posibles fallas a la hora de su uso. También se presentan y describen diferentes usos de este tipo de tecnologías en el área de transporte, con un especial énfasis en reconstrucción de rutas de vehículos y/o entidades en general.

2.1. Tecnología Bluetooth

Esta Sección comienza con una definición y descripción de las comunicaciones Bluetooth. Se describe el funcionamiento de una conexión y los tipos de dispositivos que se encuentran involucrados en ella. Desde un punto de vista del uso en transporte, se discute sobre qué infraestructura se utiliza, cuál es la potencia y posibles limitaciones de trabajar con estas tecnologías.

2.1.1. El Bluetooth y su clasificación según tipos de dispositivo

El Bluetooth (BT) es un protocolo de comunicación inalámbrica de corto alcance y baja energía que permite el envío y recepción de información entre dos o más dispositivos. Esta conexión se realiza entre un dispositivo maestro y uno o más dispositivos esclavos. El dispositivo maestro es el que gestiona la red, mientras que los dispositivos esclavos se someten a las instrucciones del dispositivo maestro. El BT se inventó con el fin de eliminar los cables en las conexiones entre dispositivos y de este modo facilitar el uso de diversos dispositivos a la vez. El rango de conexión depende de varios factores, mayoritariamente de la potencia del equipo emisor, la sensibilidad del equipo receptor y la tasa de absorción del medio en donde se propaga la señal (Bhaskar y Chung, 2013).

De acuerdo a la web de Bluetooth (2017) existen dos protocolos de tecnologías BT, los de alta velocidad (BR/EDR) y los de baja energía (LE). Dentro de la alta velocidad, están las conexiones punto a punto (P2P) que permite enlaces inalámbricos continuos entre únicamente dos dispositivos. Este tipo de tecnología es utilizada para el envío de audio en altavoces, audífonos, manos libres y radios de vehículos. Por otro lado, la conexión de baja energía permite comunicación de frecuencia corta y tiene la capacidad de realizar conexiones múltiples. De igual forma se permiten las conexiones P2P, pero además incluye los enlaces de tipo *Broadcast*, donde un dispositivo es capaz de comunicarse con múltiples dispositivos, y los enlaces *Mesh* que permiten la comunicación de múltiples dispositivos entre sí. En general, los dispositivos de baja energía son utilizados en monitores de salud o seguimiento deportivo (P2P), puntos de información o soluciones con *beacons* (*Broadcast*) y automatización de procesos o sensores en red (*Mesh*).

Existe una segunda clasificación de los dispositivos en función de su energía de salida y su distancia máxima de conexión. Cada dispositivo puede ser clase 1, clase 2 o clase 3. En la Tabla 2.1 se muestran los rangos de conexión en metros y la máxima potencia de salida para cada una de las clases mencionadas.

Cuadro 2.1. Clasificación de dispositivos Bluetooth según energía de salida y distancia máxima de conexión

Clase	Rango (m)	Potencia de salida (dBm)
Clase 1	100	20
Clase 2	10	40
Clase 3	1	0

Fuente: Bhaskar y Chung (2013)

Murphy, Welsh, y Frantz (2002) demostraron experimentalmente que los rangos de conexión teóricos eran sobrepasados en la práctica, alcanzando rangos de 250 metros y 122 metros para las clases 1 y 2, respectivamente.

2.1.2. Funcionamiento de la comunicación Bluetooth entre dispositivos

Las señales BT operan bajo las bandas industrial científica y médica en el rango de frecuencia de 2,4-2,485 GHz (IEEE, 2002), que también son utilizadas por otras tecnologías inalámbricas tales como Wi-Fi, NFC o teléfonos sin cables (Bhaskar y Chung, 2013). Un dispositivo BT emite y transmite señales sobre 79 canales diferentes de la banda con el fin de disminuir las interferencias entre unidades. Chen y Hung (2011) realizan un estudio de este comportamiento de cambios de canal que en inglés se conoce como *frequency hopping*. El estudio reveló que ocurre a una tasa de 1.600 cambios por segundo ordenados por el dispositivo maestro lo que genera intervalos de 625 μ s por canal.

La comunicación existe gracias a que cada dispositivo posee un identificador único e irrepetible que permite la identificación entre aparatos. Esta identidad se conoce como *Media Access Control Address* (Dirección MAC) y es el código que se envía en una señal de comunicación para que cada dispositivo sea identificado por los demás.

Un dispositivo BT se puede encontrar en dos estados principales: espera y conexión. En el estado “espera” no existe conexión con ningún dispositivo, pero la señal está presente, mientras que en estado “conexión” existe transmisión de datos (Bhaskar y Chung, 2013). Bhaskar y Chung (2013) también mencionan que el proceso de conexión entre un dispositivo controlador maestro y un dispositivo esclavo puede simplificarse con el siguiente proceso:

- (i) Al encender la conexión en un dispositivo maestro, este busca dispositivos esclavos en el entorno (los que se encuentran esperando por una conexión). Para ello

el dispositivo maestro envía paquetes de información que contienen la dirección MAC y la hora.

- (ii) Cuando uno o más de los dispositivos esclavos recibe el paquete enviado por el dispositivo maestro, el esclavo responde con su propia información (MAC y hora).
- (iii) Una vez que la respuesta llega al dispositivo maestro y es validada, este avisa al dispositivo esclavo que la conexión ha sido aceptada.
- (iv) Finalmente ambos dispositivos se enlazan y se habilita la transferencia de información.

2.1.3. Sensores Bluetooth y detecciones

Un sensor Bluetooth (SB) es un artefacto capaz de reconocer las direcciones MAC de otros dispositivos BT que se encuentren dentro de un radio de detección. Su funcionamiento se basa en tratar de establecer conexión con los dispositivos pero sin llevarla a cabo (etapas (i) y (ii) del proceso descrito en la Sección 2.1.2). Por lo tanto, cada vez que intenta establecer conexión con los dispositivos cercanos, se registra la hora y la dirección MAC de cada uno.

De este modo, cuando un dispositivo BT pasa cerca de más de un SB, podrá ser identificado por la dirección MAC y la hora de pasada en cada uno. Dentro de la clasificación de conexiones y dispositivos presentes en la Sección 2.1.1, un SB corresponde a un dispositivo de baja energía (LE) pero solo puede detectar dispositivos de alta velocidad (BD/EDR) (Michau *et al.*, 2017), además pertenece a la clase 1 según la Tabla 2.1.

El *hardware* del sensor utilizado en la fabricación de un SB determina la geometría de la zona de detección. En este sentido, existen dos tipos de SB, los omnidireccionales y los direccionales. La Figura 2.1 muestra la geometría que tienen las zonas de detección en cada tipo de SB. Los omnidireccionales generan ondas radiales (circunferencias) mientras que los direccionales solo son capaces de detectar en la dirección a la que apunta el sensor.

El alcance de los SB (tamaño de las zonas de detección) depende de varios factores tales como: la potencia de ambos dispositivos (emisor y receptor), redes aledañas que puedan interferir en la capacidad de detección, materiales de muros y construcciones cercanas, condiciones climáticas, entre otros (Bhaskar y Chung, 2013).

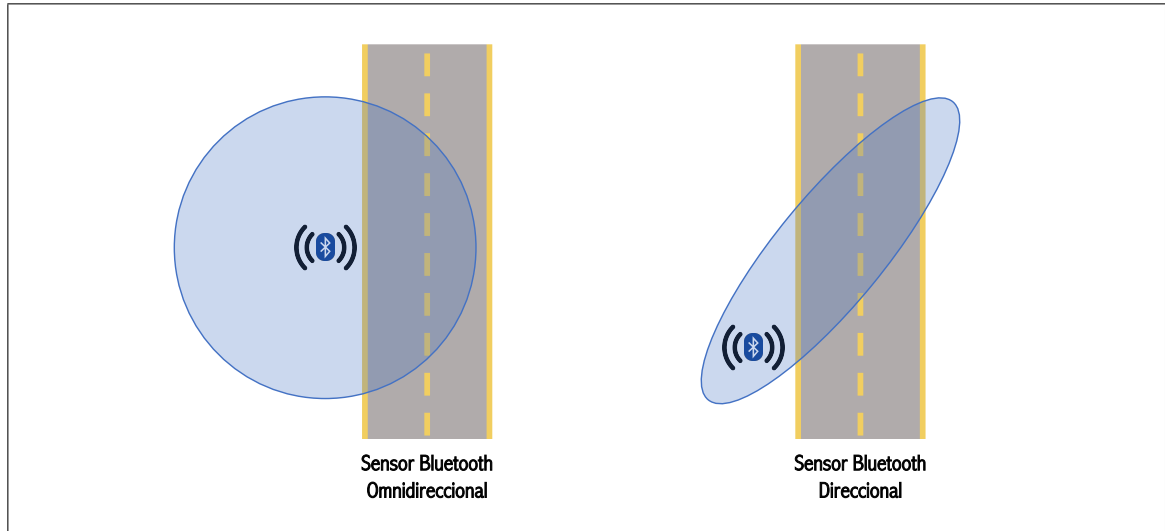


Figura 2.1. Geometrías de zonas de detección para sensores Bluetooth direccionales y omnidireccionales

2.1.4. Limitaciones propias de la tecnología

La principal limitación que presentan este tipo de señales para registrar dispositivos, es que los SB en determinadas condiciones no son capaces de detectar a todos los dispositivos que se encuentran o atraviesan la zona de detección. Abedi (2014) menciona que las detecciones fallidas producen errores significativos y su causa puede depender de factores intrínsecos, como la potencia o los tipos de dispositivos involucrados, o efectos externos, entre los que se destacan:

1. Objetos físicos: Algunos árboles, edificaciones o cualquier otro tipo de construcción cercana.

2. Radio frecuencias: Microondas o cualquier otro tipo de señales inalámbricas, tales como teléfonos inalámbricos.
3. Aparatos electrónicos: Artefactos como computadoras, iluminarias y cableados. Esto incluye a los demás dispositivos Bluetooth cercanos.
4. Factores ambientales: Condiciones climáticas, contaminación y condiciones de luz.
5. Saturación: Al existir una alta cantidad de dispositivos en la zona de detección de un mismo sensor, este puede saturarse y no registrar ciertos dispositivos.
6. Tiempo de conexión: Tiempo que el dispositivo permanece dentro del área de detección. Este se relaciona directamente con la velocidad de pasada del dispositivo por la zona de detección del SB.

2.2. Usos de tecnologías inalámbricas en transporte

Las redes inalámbricas de comunicación como el BT y el Wi-Fi han tomado un rol fundamental en la vida cotidiana, sumado al auge de los dispositivos inteligentes, este tipo de señales conviven la mayor parte del día con las personas. Este hecho ha llamado la atención de la comunidad científica para utilizar la información que este tipo de señales proveen. En general, permiten identificar la pasada de un mismo dispositivo por distintos sensores (localizados en distintos lugares). Dentro del área de transporte se pueden encontrar investigaciones que utilizan esta información (asumiendo que un dispositivo corresponde a una entidad o vehículo) para detectar la presencia de personas en oficinas o espacios públicos cerrados, predecir tiempos de viajes, reconstruir rutas en autopistas y entornos urbanos, construir matrices origen destino dinámicas en el tiempo, entre otras. En las siguientes secciones de este capítulo se presenta cómo se han utilizado este tipo de señales inalámbricas a lo largo del tiempo, qué inconvenientes se han presentado y cómo la comunidad científica ha desarrollado nuevas aplicaciones.

En 2001 se realiza una primera evaluación para utilizar las redes BT en la comunicación y detección de dispositivos en un entorno cerrado. La investigación, desarrollada por Kasten y Langheinrich (2001), muestra por qué el BT es la tecnología adecuada dentro de las redes inalámbricas emergentes para comunicar un conjunto de dispositivos en espacios cerrados. Los investigadores se basan en que el BT posee un bajo uso de energía, no requiere infraestructura, funciona bajo una banda libre y trabaja con *frequency hopping* para reducir las interferencias. El estudio no busca mostrar una herramienta para posicionar dispositivos (entidades), sino que muestra una posibilidad de utilizar el BT para comunicar varios dispositivos entre sí, sin la necesidad de estar directamente conectados por cables. De acuerdo Friesen y McLeod (2015), las primeras referencias del uso del BT con el fin de realizar seguimiento de entidades, no estaban relacionadas al tráfico vehicular en específico. Sin embargo, con el paso del tiempo se realizaron algunos desarrollos en este ámbito. De los trabajos pioneros, se destacan una aplicación de seguimiento de niños en un zoológico realizado por la empresa BlueTags (GeekZone, 2003) y el seguimiento de estudiantes en una universidad (Haase y Handy, 2004). El uso de redes inalámbricas en general toma un rol más significativo a mediados de los 2000s, pero no se menciona de forma explícita el BT (Wen, Pan, y Le, 2007). Antes de esa fecha, no existía el interés de utilizar estas tecnologías para el posicionamiento de entidades, dado que no eran usadas por el común de las personas, sino que estaban en escasos celulares y prácticamente su señal no se encontraba activa en el uso diario. El uso potencial del BT, específicamente en monitoreo de tráfico, comienza a aparecer en la literatura académica cerca de 2010 (Friesen y McLeod, 2015). Pero años antes, ya existían algunos trabajos, los que se mencionan en la Sección 2.2.1.

Cabe destacar que, para los objetivos de este trabajo, el Wi-Fi posee las mismas características que el BT. La principal diferencia es el uso que se le dan a las redes y no el cómo estas se detectan. El Wi-Fi se utiliza mayoritariamente para transmitir señales de internet, aunque igualmente es utilizada para conectar dispositivos. Fuera de esto, el Wi-Fi

también cuenta con una dirección MAC y sus señales también pueden ser detectadas por un sensor. Por lo tanto, las conclusiones a las que se pueden llegar por utilizar redes BT son replicables al uso de redes Wi-Fi. Incluso ambas tecnologías pueden ser utilizadas de forma simultánea teniendo en cuenta que es necesario filtrar la información, dado que se pueden duplicar los registros al detectar un mismo dispositivo por cada una de las redes (si es que este cuenta con ambas redes disponibles).

2.2.1. Posicionamiento de entidades en entornos cerrados

No fue hasta la masificación del BT en 2007 que comienza el interés por desarrollar herramientas de detección de entidades con esta tecnología. Una de las primera aplicaciones en temas de transporte y movilidad fue la identificación y/o posicionamiento de personas en entornos cerrados usando BT y/o Wi-Fi. La motivación nace a raíz del creciente uso de dispositivos móviles BT y Wi-Fi en entornos cotidianos. Para el caso de BT, la Figura 2.2 muestra que en 2005 más del 80 % de celulares ya contaban con esta tecnología. Además, Nielsen (2013) reporta que la penetración de las señales BT desde celulares en Estados Unidos era de un 64 % al año 2013.

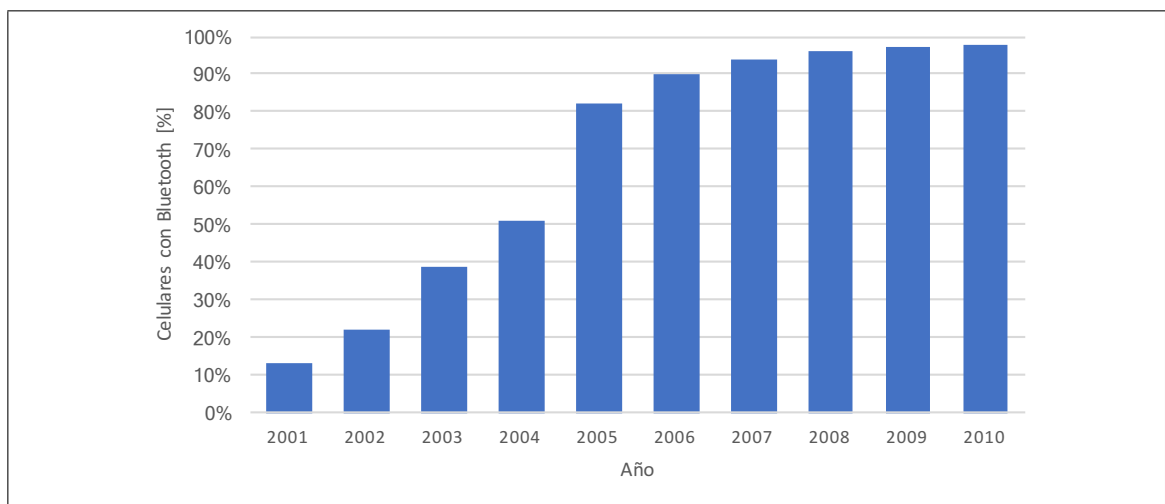


Figura 2.2. Porcentaje de celulares Bluetooth desde 2001 a 2010
Fuente: GSMArena (2010)

La principal idea detrás del posicionamiento de entidades se basa en utilizar un SB que sea capaz de medir la potencia de la señal de los dispositivos que se encuentran en la zona de detección de este. Gracias a este valor se puede estimar la distancia que existe entre el sensor y la entidad (dispositivo BT). Por lo tanto, si se tienen instalados varios sensores (con posiciones conocidas) se podrá estimar la posición de un dispositivo BT dentro de la zona donde se encuentran los sensores. Esto se puede llevar a cabo, gracias a que se puede inferir la distancia de un dispositivo a todos los SB y luego precisar la posición por medio de triangulación.

Bargh y de Groote (2008) buscan posicionar con BT a trabajadores en oficinas. Los autores analizan la tasa de respuesta de las señales BT entre sensores y celulares en un edificio de oficinas con el fin de saber la ubicación de los empleados. En cada computador del personal fue instalado un sensor BT capaz de detectar la tasa de respuesta de conexión de los celulares y así determinar la posición de cada persona dentro de las dependencias del edificio. La precisión alcanzada fue de 98 %, pero al disminuir parcialmente el número de sensores, la predicción bajó a un 75 %, lo que demuestra la importancia de contar con una alta densidad de SB. Además, los autores afirman que es necesario que exista una superposición de las zonas de detección de cada SB con el fin de realizar una mejor triangulación y obtener resultados más precisos. Por otro lado, el Wi-Fi al ser una red similar al BT, también ha sido utilizado con el mismo objetivo. Pérez, Aparicio, Bernardos, Casar, y An (2007) y Aparicio, Pérez, Bernardos, y Casar (2008) muestran metodologías capaces de estimar la posición de personas en entornos cerrados. En este caso utilizan las redes BT para identificar la zona de detección y con Wi-Fi posicionan a los individuos.

En determinadas circunstancias este tipo de redes pueden fallar, por lo que es importante contar con modelos capaces de suplir las pérdidas de información. Adesso, Bruno, y Restaino (2010) y Altini, Brunelli, Farella, y Benini (2010) toman conocimiento de este tipo de problemas y asisten los datos de las señales inalámbricas con modelos estadísticos, como filtros de Kalman (Kalman, 1960), y redes neuronales respectivamente. En el caso

de Altini *et al.* (2010), justifican el uso de estas herramientas dada las altas variaciones que presentan los índices de potencia de señal, lo que genera dificultades en la precisión de la posición. Gracias al uso de modelos estadísticos se logra aumentar la precisión sin aumentar los costos de procesamiento.

2.2.2. Predicción de tiempos de viaje

La necesidad de conectar los dispositivos móviles con el vehículo, ya sea para realizar llamadas, utilizar servicios de GPS o escuchar música, dieron el espacio para que las redes BT logran incorporarse dentro de los vehículos. La Figura 2.3 muestra la evolución de la cantidad de vehículos en Europa con presencia de señales BT, que en 2010 alcanzó una penetración cercana al 50 %. Este hecho logró motivar a la comunidad científica para estudiar fenómenos viales usando las redes BT y/o Wi-Fi.

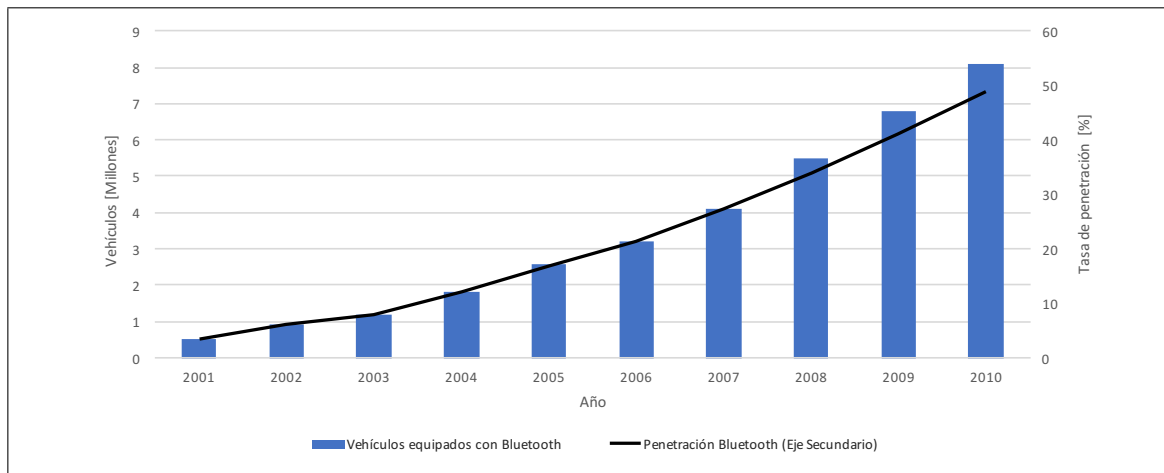


Figura 2.3. Vehículos con redes Bluetooth y tasa de penetración de la red en Europa desde 2001 a 2010.

Fuente: Frost (2005)

La obtención de tiempos de viaje en autopistas y caminos es una de las aplicaciones viales del BT. Para obtener el tiempo de viaje se instalan una serie de SB a lo largo de una

vía o autopista y se registran las direcciones MAC (descrita en la Sección 2.1.2) al paso de vehículos por los SB. De este modo, se conoce el tiempo de pasada en cada sensor y es posible determinar el tiempo de viaje entre cada par de SB para cada vehículo. En base a este concepto, una serie de investigaciones se han realizado para estudiar los tiempos de viaje. Por ejemplo, Haseman, Wasson, y Bullock (2010) presentan una metodología donde miden el tiempo de viaje con sensores BT en una autopista alemana a zonas de trabajo y realizan comparaciones del tiempo de viaje a lo largo de día.

Por otro lado, es importante conocer cuán confiable son los tiempos de viaje determinados con este tipo de señales. Aliari y Haghani (2012) muestran que existe entre un 98 % y 99 % de precisión al comparar los tiempos de viaje obtenidos de SB con los registros reales. Este estudio revela que los datos BT presentan buenos índices de predicción, pero a la fecha no se cuenta con altos índices de penetración de la señal. Así, uno de los problemas existentes, es que no todos los vehículos cuentan con algún dispositivo que emita las señales. En la Figura 2.3 vemos que la penetración ha ido en aumento, pero no es posible asegurar que la cantidad de datos sea suficiente o que pueda existir algún tipo de sesgo. En otro estudio, Araghi, Pedersen, Christensen, Krishnan, y Lahrman (2015) miden tiempos de viaje en una vía mediante BT y GPS a fin de medir el impacto que tiene la penetración y tasa de fallo del BT sobre la precisión en la predicción. Los autores logran demostrar que los datos BT replican muy bien los datos de tiempo de viaje, específicamente el mínimo y la mediana, que presentan mejor ajuste que el máximo y el promedio. Esto es explicado dado que los vehículos lentos tienden a ser más detectados que los vehículos rápidos.

Dentro de este mismo contexto han habido otras investigaciones más aplicadas, donde se utilizan las redes y registros BT para comparar el tiempo de viaje de buses en servicio y fuera de servicio con tiempos de viaje en automóvil (Kieu, Bhaskar, y Chung, 2012) o evaluar los niveles de congestión midiendo el tiempo que les toma a los vehículos viajar entre dos SB (Tsubota, Bhaskar, Chung, y Billot, 2011). Este tipo de metodologías no solo se han aplicado en vehículos. Dado el nivel de penetración que ha alcanzado el BT en

los teléfonos celulares, también se han estudiado los tiempos de viaje que experimentan ciclistas en pequeños corredores (Mei, Wang, y Chen, 2012).

2.2.3. Generación de matrices origen destino

Las detecciones BT no solo son capaces de entregar posiciones o tiempos de viaje, sino que también son capaces de saber desde dónde se inició y finalizó un viaje (en una red cubierta de sensores). Esta información da lugar a las matrices origen destino (OD) de viajes, que comúnmente se construyen con encuestas origen destino. Construir matrices OD con este tipo de tecnologías permite tener información actualizada regularmente y disminuir considerablemente los costos de generación de la información.

La idea principal detrás de la construcción de matrices OD con información BT consiste en registrar la primera dirección MAC de un vehículo al ingresar a una red equipada con sensores BT y luego ver dónde se registra la misma dirección MAC por última vez. Barceló, Montero, Marqués, y Carmona (2010), Barceló, Montero Mercadé, Bullejos, Serch, y Carmona (2012), Carpenter, Fowler, y Adler (2012) y Blogg, Semler, Hingorani, y Troutbeck (2010) desarrollan herramientas que permiten realizar el seguimiento de vehículos en autopistas por medio de la dirección MAC. Para conocer el par OD en autopistas basta con poner SB en las entradas y salidas de estas. Además, estas investigaciones determinan tiempos de viaje tal y como fueron descritos en la Sección 2.2.2. Barceló *et al.* (2012) logran aplicar la construcción de matrices OD con BT en redes urbanas usando filtros de Kalman (Kalman, 1960) para corregir los tiempos de viaje *a priori* (obtenidos de las detecciones BT) y dispositivos convencionales (como espiras magnéticas) para complementar y expandir la muestra a valores poblacionales.

2.2.4. Reconstrucción de rutas de vehículos

En la línea de la generación de matrices OD, se puede profundizar no solo accediendo a la información del punto de partida y final de un viaje en vehículo, sino que también es posible realizar un seguimiento completo de las detecciones BT en una red equipada con SB. De este modo, es posible reconstruir la ruta que ha seguido un vehículo entre su origen y destino. La reconstrucción de rutas es posible de llevar a cabo por medio de un registro de las detecciones que han realizado los SB en una red, para una dirección MAC específica. Las primeras pruebas de reconstrucción de rutas con señales inalámbricas se realizaron en redes de autopistas dada las facilidades que estas entregan, tales como, bajas interferencias de señales y la baja cantidad de rutas que se pueden realizar en ellas. Hainen *et al.* (2011) y Carpenter *et al.* (2012) desarrollan este tipo de análisis.

En una nueva escala de complejidad y como trabajos previos a esta investigación, algunas investigaciones aplican la reconstrucción de rutas de vehículos en redes urbanas. Musa y Eriksson (2012) desarrollan una tecnología para reconstruir las rutas de vehículos mediante señales Wi-Fi. La motivación de su trabajo busca conocer el comportamiento de vehículos en una ciudad captando las señales de teléfonos inteligentes. Para tener un orden de magnitud de la penetración de la señal Wi-Fi a la fecha, se detectaron 7.000 vehículos de un total de 37.000 en 12 horas (en una vía de un sentido), por lo que, aproximadamente 1 de cada 5 vehículos poseía un dispositivo con señales Wi-Fi habilitadas en los Estados Unidos.

Las grandes distancias que cubren las zonas de detección de redes Wi-Fi significaron un desafío para localizar los vehículos mientras estos circulan por la red. Dadas estas complejidades, Musa y Eriksson (2012) desarrollan un modelo Markoviano para intuir el comportamiento de los vehículos al enfrentar las intersecciones. Las detecciones son utilizadas como dato de entrada para conocer los flujos que circulan por distintas vías de la red. Luego se discretiza cada calle en celdas, con las que se construye la red Markoviana,

donde se definen probabilidades de transición para pasar de una celda a otra. Uno de los supuestos utilizados por estos investigadores es que todas las probabilidades de transición que emergen de cada celda en particular son iguales, es decir, si un vehículo circula por una vía, este tiene 0,5 de probabilidad de quedarse en la misma celda y 0,5 de avanzar a la siguiente (dado que solo existen esas dos opciones). En las intersecciones ocurre algo similar, es decir, si llega a una intersección donde puede seguir recto, virar a la izquierda o virar a la derecha (3 opciones), entonces se asume que este tiene $1/3$ de probabilidad en cada opción. A pesar de ello los autores afirman que la identificación de rutas muestra ser efectiva con el universo de vehículos testeados.

La investigación desarrollada por Musa y Eriksson (2012) marca un punto de partida en las intenciones de reconstruir rutas utilizando tecnologías inalámbricas, pero utiliza un supuesto poco realista a la hora de definir el comportamiento de los vehículos en una intersección, dado que no todas las opciones posibles son equiprobables.

Michau *et al.* (2014) desarrollan una nueva propuesta metodológica que busca resolver el mismo problema abordado por Musa y Eriksson (2012). En dicha investigación se presenta una metodología para reconstruir rutas con información BT en un entorno urbano. Michau *et al.* (2014) reconocen que uno de los grandes desafíos de trabajar con detecciones BT es que estas pueden fallar a la hora de reconocer todos los dispositivos cercanos. Este hecho genera vacíos de información en el seguimiento de un vehículo, lo que imposibilita una directa inferencia de la ruta que pudo escoger el vehículo. La Figura 2.4 muestra una situación donde se da este caso. El vehículo que recorrió la red de izquierda a derecha solo fue detectado por los SB que se destacan con un *check mark*, mientras que el SB marcado con una cruz (que fue por donde efectivamente circuló el vehículo) falló en detectar. En este caso no se puede hacer una inferencia directa sobre la ruta que usó el vehículo dado que los SB de las rutas disponibles no presentan registro (una falló en detectar y por la otra no circuló el vehículo).

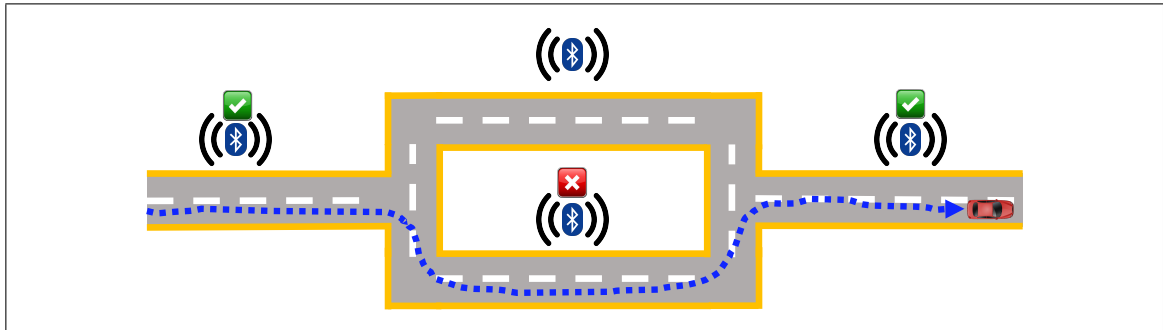


Figura 2.4. Ejemplo de una situación cuando no es posible conocer la ruta que realizó un vehículo producto de la falla de un sensor Bluetooth

Para solucionar este problema, Michau *et al.* (2014) desarrollan una metodología de reconstrucción de rutas en una ciudad que presenta una alta densidad de sensores. Esto implica que la ruta que conecta dos detecciones sucesivas (para un mismo vehículo) corresponde a la que pasa por la menor cantidad de sensores. Así, Michau *et al.* (2014) asignan la ruta que posee menos sensores (en una situación como la de la Figura 2.4). De ser igual la cantidad de sensores entre las distintas rutas, se asigna a cualquiera de ellas con la misma probabilidad.

A partir de esta investigación, Michau *et al.* (2017) presentan un segundo trabajo para reconstruir rutas en un entorno urbano con señales BT. El estudio detalla las problemáticas de la tecnología y mejoran el algoritmo presentado por Michau *et al.* (2014). La nueva metodología es aplicada a datos simulados y reales (GPS de taxis) de la ciudad de Brisbane, Australia. Esta ciudad cuenta con 580 sensores BT en una superficie aproximada de 900 kilómetros cuadrados. Los autores reconocen que las trayectorias reales de vehículos son difíciles de obtener e imponen desafíos tecnológicos y éticos. Con el fin de cuantificar el problema presentado en la Figura 2.4, Michau *et al.* (2017) realizan un estudio en una autopista para medir la tasa de fallo de los SB en la detección de vehículos. El experimento concluye que la cota inferior para la proporción de vehículos no detectados es de un 20 %.

La metodología de Michau *et al.* (2017) construye la ruta del vehículo a partir de la ruta mínima entre de detecciones sucesivas y las compara con dos metodologías simples. Una de ellas solo considera la primera y última detección de cada vehículo y otra considera la primera, la última y la detección intermedia de cada vehículo a lo largo de la ruta. Cada uno de los modelos es testado en una red modelada en Aimsun, donde se considera que la probabilidad de detección (fija en cada experimento) varía entre un 40 % y 80 % y el radio de la zona de detección varia entre 10 metros y 150 metros. Finalmente, obtienen resultados del orden del 84 % de predicción en las rutas.

2.2.5. Consideraciones del uso de datos Bluetooth en transporte

En las investigaciones de Bhaskar y Chung (2013), Michau *et al.* (2014) y Michau *et al.* (2017) se reconocen algunos aspectos importantes de considerar a la hora de trabajar con registros de señales inalámbricas. Algunas de estas son las siguientes:

- (i) Las señales BT pueden provenir de dispositivos que se encuentren viajando en automóvil, bus, en bicicleta o a pie. Esto obliga a filtrar previamente los datos con el fin de inferir el modo de viaje. Además, de un mismo vehículo o persona puede generarse múltiples señales, tantas como dispositivos BT posea. Por lo que también es importante que se filtren los datos para evitar el doble conteo de la información. Los autores mencionan que este alcance se puede corregir por medio de una estimación de las velocidades o los patrones de movilidad, pero su complejidad requiere que sea estudiado en detalle.
- (ii) Solo se cuenta con información dentro de las zonas cubiertas por los SB. Esto quiere decir, que no se puede saber nada (directamente de los datos) de lo que realizaron los vehículos fuera de las zonas de detección de sensores. De este modo, se busca construir un modelo que permita predecir el comportamiento fuera de las zonas de detección mediante el uso de estados de tráfico.

- (iii) La obtención de múltiples detecciones simultáneas radica en que pueden instalarse sensores cercanos y esto provoca que el vehículo pueda ser registrado por más de uno en el mismo instante de tiempo. Al ocurrir este fenómeno, se podrían considerar velocidades de circulación extremadamente elevadas o que un vehículo se encuentre en más de un lugar a la vez. Frente a esto Michau *et al.* (2014) desarrollaron un criterio para corregir el error (que se da mayoritariamente en redes densas de antenas).
- (iv) No todos los dispositivos BT son detectados al momento de atravesar la zona de detección de un sensor. Este problema ya fue detallado anteriormente.

El surgimiento de este tipo de tecnologías ha iniciado la discusión respecto a la invasión de privacidad que se puede producir al momento de trabajar con los datos que se generan. Frente a esto, los datos son trabajados de forma anónima y como mencionan Bhaskar y Chung (2013) no es posible realizar una conexión directa entre la dirección MAC y la identidad de una persona. Por otro lado, algunos desarrolladores de *software* y *hardware* especializado para captar este tipo de redes encriptan las direcciones MAC de forma diferente cada día con el fin disminuir la posibilidad de rastreo.

2.3. Resumen del capítulo

Este Capítulo ha presentado la tecnología BT en general, su funcionamiento y las fallas que presenta en la realidad. Los usos en ingeniería de transporte son variados: posicionamiento de entidades en espacios físicos, estimación de tiempos de viaje y generación de matrices origen destino dinámicas. Dentro del marco de esta investigación, se ha reportado que existen metodologías para la reconstrucción de rutas usando diferentes enfoques. Esta investigación aborda este último problema y levanta supuestos impuestos en investigaciones anteriores. Los aspectos más relevantes a considerar en los aportes de este trabajo, es que incorpora estados de tráfico (producto de las mismas detecciones BT) para suplir la

falta de información producida por las detecciones fallidas. Además, entrega las probabilidades de uso para cada una de las posibles rutas que pudo escoger cada vehículo. En el siguiente Capítulo se procede a presentar la metodología propuesta para la reconstrucción de rutas.

CAPÍTULO 3. METODOLOGÍA

¡En esta Sección se detalla la metodología desarrollada para inferir las elecciones de rutas realizadas por vehículos equipados con tecnología Bluetooth (BT) que circulan en una red parcialmente equipada con sensores Bluetooth (SB).

Es importante mencionar que este trabajo considera los siguientes aspectos respecto a los SB:

1. En la modelación se utilizan solo SB omnidireccionales (Figura 2.1). Aun así, la metodología permite el uso de SB direccionales si se conoce la sección de la red que cubre la zona de detección.
2. El modelo considera que los SB son instalados en las intersecciones de la red, pero la metodología no pierde generalidad al posicionarlos en cualquier parte de esta (arcos o intersecciones).
3. Se asume que no existe traslape de las zonas de detección de los SB instalados.
4. Se considera que la probabilidad de detección para un SB solo depende de la velocidad del vehículo. En la realidad esta probabilidad depende de varios factores, pero la metodología trabaja únicamente con la información que es posible obtener de las detecciones Bluetooth.
5. Se asume que los datos provenientes de los SB se encuentran previamente filtrados, por lo que solo cuenta con vehículos particulares e/o individuales. Este aspecto es importante, dado que, los SB detectan todo tipo de señales BT, ya sea *smartphones*, audífonos, radios de automóviles, entre otras. De este modo, la señal puede corresponder a distintos tipos de entidades, tales como: automóviles, peatones, bicicletas, buses de transporte público, etc. La metodología fue desarrollada en primera instancia para inferir rutas de vehículos motorizados, por lo que es necesario realizar un filtro de las detecciones, lo que queda fuera de los alcances de esta investigación. Por consiguiente, la base de datos consta solo de detecciones

de vehículos, en donde cada registro consiste en el ID del SB, la dirección MAC del vehículo y la fecha y hora en que se realizó la detección. La Tabla 3.1 muestra la composición de un registro de la base de datos. Esta indica que el sensor SB-1 detectó un vehículo a las 15:24 hrs.

La metodología trabaja con un conjunto de datos previamente almacenados. Es decir, esta no puede funcionar precisamente en tiempo real, sino que deben procesarse datos en un período en específico (*e.g.* 15 minutos) y luego la metodología será capaz de predecir rutas para ese período.

Cuadro 3.1. Ejemplo de un registro de la base de datos de detecciones Bluetooth

Sensor Bluetooth	Dirección MAC	Fecha y hora
SB-1	0A:B6:8B:24:E6:50	21-02-16 15:24:13

La metodología esta compuesta por tres grandes etapas estructuradas en la Figura 3.1. La primera corresponde a la **definición de la red**, donde se construye un grafo dirigido a partir de la red vial original y la distribución de SB dispuestos en ella. La segunda es la etapa de **precarga de información**, donde se almacenan los tiempos de permanencia en cada nodo y los tiempos de viaje en cada arco del grafo como distribuciones de probabilidad. La tercera y última etapa de **reconstrucción de rutas** se encarga de inferir las probabilidades de uso para las secciones de ruta entre SB por medio de inferencia Bayesiana, para finalmente calcular las probabilidades de uso de la ruta completa para cada vehículo.

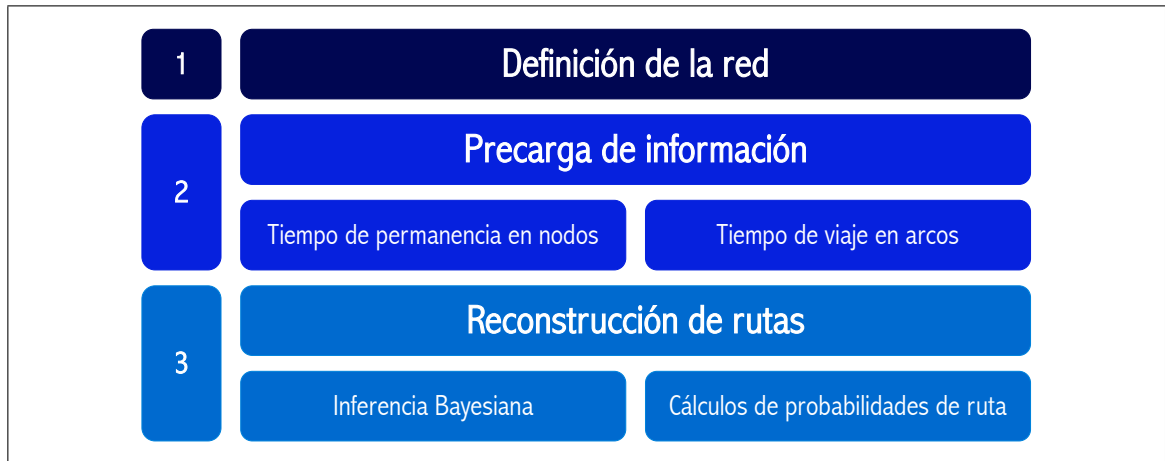


Figura 3.1. Etapas de la metodología

El resto de este Capítulo se estructura con la definición y explicación en detalle de cada una de las etapas y sub etapas de la metodología.

3.1. Etapa 1: Definición de la red

La primera etapa consiste en construir un grafo dirigido a partir de la configuración vial y la distribución de SB instalados en la zona donde se busca reconstruir las rutas de vehículos, que de ahora en adelante se conocerá como zona de estudio. Para construir correctamente el grafo, es necesario definir esta zona, donde el tamaño de esta depende exclusivamente de la superficie que cubren los SB, dado que es imposible tener información de áreas externas a la zona.

Una vez definida el área en la que se realizará el estudio, es importante conocer la vialidad presente y la ubicación de los sensores instalados en ella. A partir del conjunto de intersecciones que cuentan con un SB y de la configuración vial, se procede a construir el grafo necesario para la aplicación de la metodología. En el ejemplo de la Figura 3.2 se

observa una zona de estudio con una red tipo Manhattan de 3 por 4 cuadras, con sentidos de circulación intercalados y 11 SB (símbolos BT) instalados en las intersecciones.

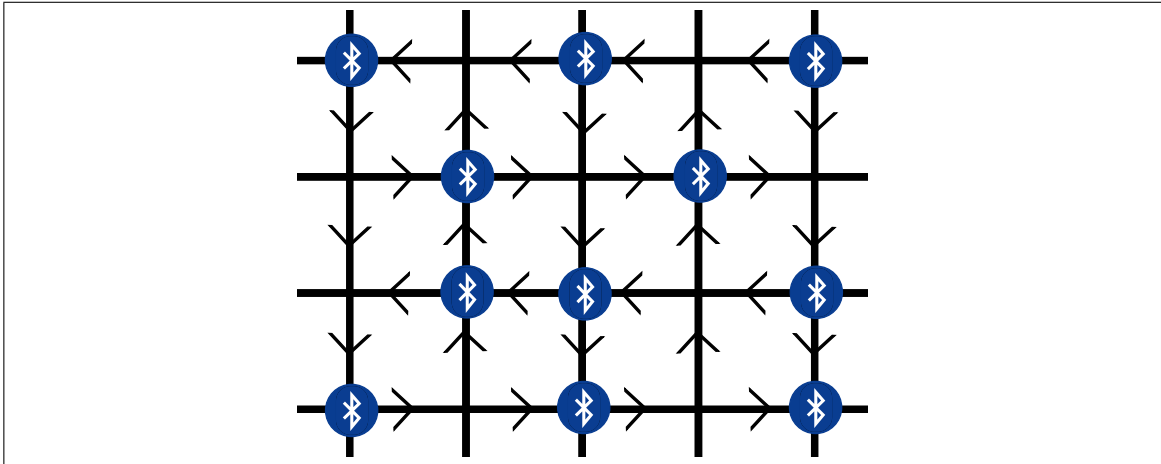


Figura 3.2. Ejemplo de una red vial y una posible configuración de sensores Bluetooth

El grafo a construir, definido como $G(N, L)$, donde N corresponde al conjunto de nodos de G y L al conjunto de arcos dirigidos de G , se construye por medio de los criterios que se definen a continuación:

- $N \rightarrow$ Conjunto de intersecciones de la red original que tienen instalado un SB.
- $L \rightarrow$ Conjunto de arcos, donde el arco entre el nodo a y el nodo b ($a, b \in N$) indica si es posible viajar de a a b sin pasar por una intersección equipada con SB.

Al considerar la Figura 3.2, el conjunto N está formado por los 11 nodos especificados y la Figura 3.3 muestra el grafo G construido a partir de la red vial anterior. Los arcos de L se obtienen al analizar cada par de nodos de N . Por ejemplo, el nodo superior izquierdo posee solo dos arcos que salen, ya que, solo se puede ir a dos nodos (SB) sin pasar por un tercer SB.

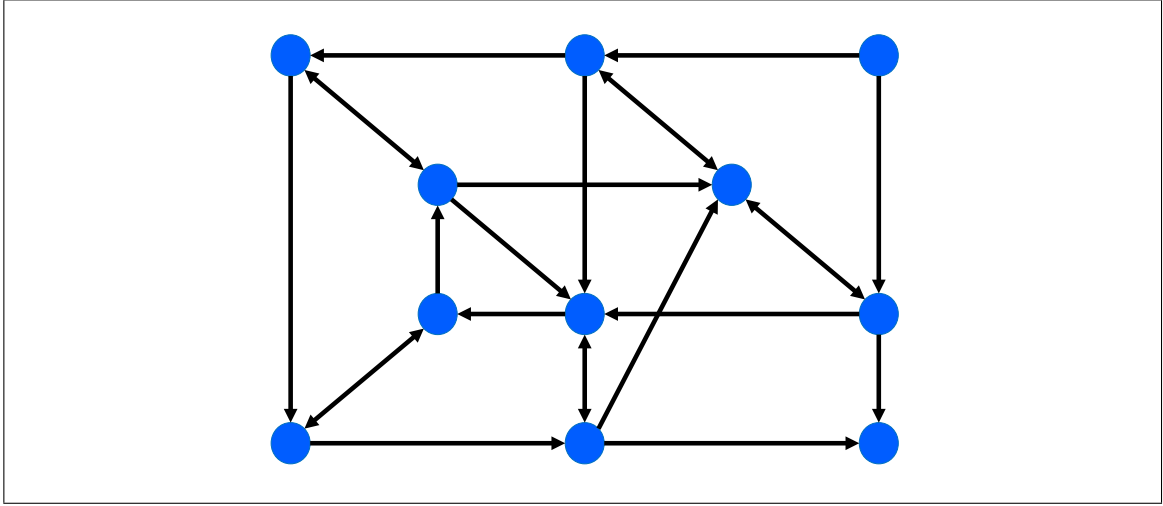


Figura 3.3. Grafo $G(N, L)$ construido a partir de la red vial de la Figura 3.2

Una vez construido el grafo, se procede a asignar un costo a cada arco, el que corresponde a la distancia mínima entre los nodos de la red original. Para esto se ejecuta una variación del algoritmo de rutas mínimas propuesto por Dijkstra (1959). De este modo cada arco del grafo representa la ruta de menor distancia (si es que existe) entre cada par de SB (sin pasar por otros SB). Al finalizar estos pasos, se cuenta con un grafo de arcos dirigidos que representa a la red original y considera la cantidad y distribución de SB instalados en las intersecciones.

Sin pérdida de generalidad se considera que los sensores son instalados en las intersecciones de la red, pero es posible que estos se encuentren en los arcos, mientras se cubra uno a la vez. Si llegase a darse este caso, basta con crear una intersección ficticia y considerarla como un nodo más en el grafo.

3.2. Etapa 2: Precarga de información

En la segunda etapa se procede a realizar la carga de información de tiempos de permanencia en cada nodo y tiempo de viaje en cada arco.

El objetivo de esta etapa es representar los diferentes estados de congestión que se presentan en la red. Para ello es necesario medir los tiempos de permanencia y viaje con la tecnología BT. Para lo anterior, distribuciones de tiempo de permanencia en nodos (DT_i) y distribuciones de tiempo de viaje en arcos (TT_j) son calibradas, donde $i \in N$ y $j \in L$.

Los fenómenos que puedan presentarse en la red, tales como congestión, accidentes o lluvia, pueden cambiar la forma de estas distribuciones por lo que es importante actualizarlas regularmente (*e.g.* cada 15 minutos como se utiliza en este trabajo). Es por esto que cada distribución será denotada como DT_i^h y TT_j^h , donde h es el período del día que representa cada distribución.

3.2.1. Tiempo de permanencia en nodos (DT_i^h)

Este tiempo corresponde al tiempo total que un vehículo está dentro de la zona de detección de un SB. Puede ser estimado como la diferencia temporal entre la primera y la última detección ocurrida dentro de esa zona para un mismo vehículo. Este tiempo incorpora parte de los efectos de semáforos, congestión o cualquier otro fenómeno que involucre detenerse o reducir la velocidad en las cercanías del nodo.

Cuando un vehículo atraviesa el área de detección de un SB se pueden presentar tres casos:

- (i) El vehículo no es detectado por el SB.
- (ii) El vehículo es detectado solo una vez.
- (iii) El vehículo es detectado más de una vez.

Los casos (i) y (ii) no puede ser utilizados para conocer el tiempo de permanencia de un vehículo en el nodo (intersección), ya que no es posible calcular una diferencia de tiempos (no existen al menos dos detecciones). Por esto, las distribuciones de tiempos de permanencia en nodos solo son calibradas con tiempos entregados por vehículos del caso (iii).

El hecho de usar solo vehículos del tercer caso puede producir un sesgo pues estos tienden a ser los vehículos más lentos. Esto se debe a que cuando los vehículos más rápidos atraviesan la zona de detección de un SB, permanecen menos tiempo dentro de la zona que los vehículos lentos, por lo que menos detecciones podrían llevarse a cabo sobre ellos (vehículos de casos (i) y (ii)).

Para vehículos del caso (iii) se calcula una aproximación del tiempo de permanencia (Ecuación 3.1) donde se realizan W detecciones en el SB i para el vehículo de MAC m .

$$dt(m, i) = t(m, i, W) - t(m, i, 1) \quad (3.1)$$

$dt(m, i)$ es la estimación del tiempo que el vehículo de MAC m permanece en la zona de detección del SB en el nodo i y $t(m, i, w)$ es el instante de la w -ésima detección del vehículo m en el nodo i , $w \in [1, W]$.

La Figura 3.4 muestra un ejemplo de cómo se obtiene el tiempo de permanencia en nodos para un vehículo en específico. El vehículo de MAC m atravesó la zona de detección del SB i y logró ser captado tres veces. La diferencia entre el tercer y el primer registro dan origen a la estimación de tiempos de permanencia para el nodo i .

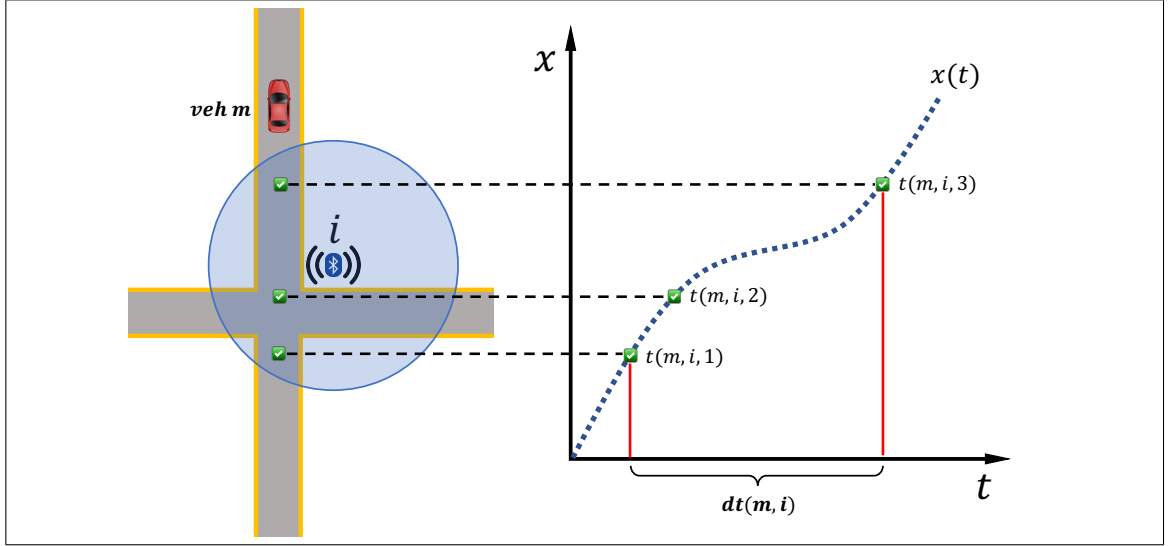


Figura 3.4. Tiempo de permanencia en nodo para un vehículo del caso (iii)

Dado que el tiempo de permanencia del vehículo puede incluir tiempos de parada u otros efectos que no son propios del comportamiento del tráfico, por ejemplo, pasar a dejar a alguien cerca de un sensor o detenerse por algún otro motivo. Por esto, se define τ_{perm} como el tiempo máximo que un vehículo podría estar dentro de la zona de detección de un SB. Es decir, si $dt(m, i) > \tau_{perm}$, $dt(m, i)$ no está representando la permanencia de un vehículo sino que además incluye efectos externos por lo que no es considerado en la muestra. El valor de τ_{perm} puede ser determinado con previo análisis de los datos o escogiendo un tiempo apropiado para representar al conjunto de intersecciones.

Dado que el objetivo de esta etapa es calibrar una distribución para los tiempos de permanencia (DT_i^h), es necesario agrupar los $dt(m, i)$ cada h unidades de tiempo. Para cada grupo de datos se construye un histograma y una posterior distribución. Esta puede ser continua o discreta en función de la capacidad computacional para estimar una correcta distribución y de la cantidad de datos recolectados. La metodología presentada trabaja con distribuciones discretas. Por lo tanto, DT_i^h es calibrada con el grupo de datos $dt(m, i)$ que ocurren en el periodo h en el nodo i . Cabe destacar que si la primera y última detección del

tiempo de permanencia ocurren en dos períodos h diferentes, el registro será almacenado en el grupo de datos de la primera detección.

Es importante mencionar que un vehículo puede pasar por un mismo SB más de una vez en el mismo viaje. Esto implica que es necesario reconocer cada una de las pasadas del vehículo por el sensor. Para ello se puede asumir que τ_{perm} es el tiempo máximo y así subdividir el tiempo total en pasadas individuales.

3.2.2. Tiempo de viaje en arcos (TT_j^h)

El tiempo de viaje en arcos corresponde al tiempo empleado por un vehículo para viajar entre dos SB unidos por un arco en el grafo G . Es decir, mientras exista una ruta directa entre dos SB (sin pasar por otro SB), el viaje será asignado al arco directo. Este es uno de los supuestos de la metodología, dado que el vehículo no necesariamente viaja por este arco. Este tiempo incorpora los efectos de todos los fenómenos que puedan ocurrir mientras un vehículo se encuentre viajando entre dos sensores, tales como cruces peatonales, accidentes viales, colas de vehículos, entre otros.

Cuando un vehículo pasa por dos nodos conectados directamente en el grafo G , se pueden dar tres casos de detecciones:

- (i) El vehículo no es detectado por ninguno de los SB.
- (ii) El vehículo es detectado por solo uno de los SB.
- (iii) El vehículo es detectado en ambos SB.

Los casos (i) y (ii) no pueden ser utilizados para conocer el tiempo de viaje en el arco. Si no hay detecciones es imposible saber si ocurrió el evento y si solo uno de los sensores realiza una detección no se puede calcular una diferencia de tiempo ni saber a qué arco asignar el tiempo. En consecuencia, las distribuciones de tiempo de viaje en arcos solo son calibradas con datos provenientes de vehículos del caso (iii). Nuevamente el utilizar

vehículos de este caso propicia a tener mayor cantidad de datos de los vehículos más lentos.

Para vehículos del caso (iii) se calcula una aproximación del tiempo de viaje en arcos de acuerdo a las detecciones en el SB aguas abajo (d) y en el SB aguas arriba (u) (Ecuación 3.2).

$$tt(m, j) = t(m, d, 1) - t(m, u, W) \quad (3.2)$$

$tt(m, j)$ es el tiempo de viaje del vehículo de MAC m en el arco j y W el número de detecciones del vehículo m en el nodo u (aguas arriba).

La Figura 3.5 muestra un ejemplo de cómo se obtiene el tiempo de viaje en arcos para un vehículo m del tercer caso. Análogamente al tiempo de permanencia en nodos, El tiempo de viaje en arcos puede incluir efectos que no son propios de los efectos de tráfico. Por lo que se define τ_{viaj} como el máximo tiempo posible que un vehículo puede tardar al viajar por un arco. Por lo que no existirán tiempo de viaje en un arco mayores a es que τ_{viaj} .

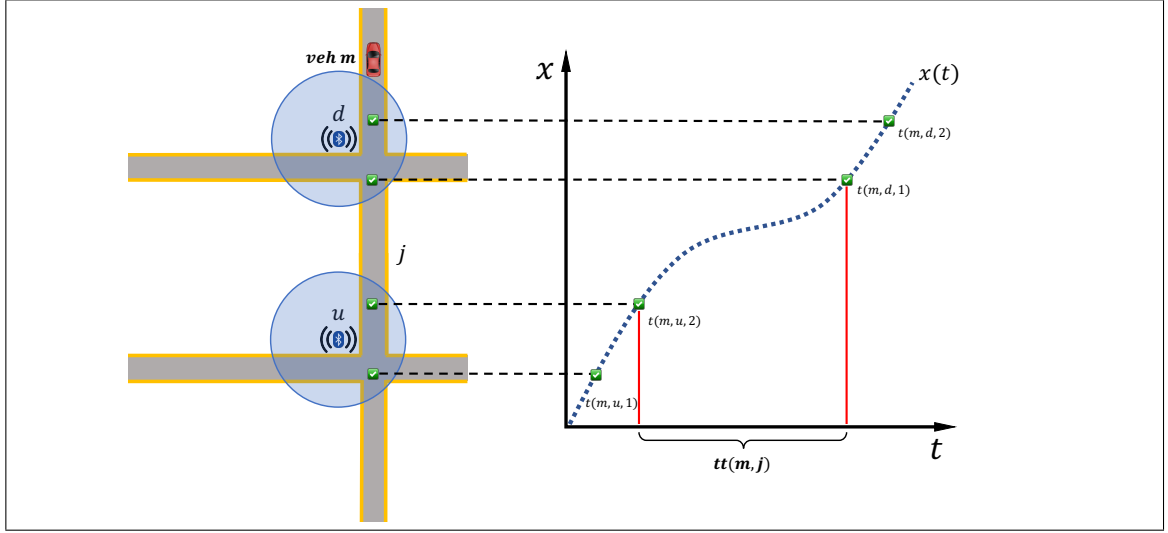


Figura 3.5. Tiempo de viaje en arcos para un vehículo del caso (iii)

Las expresiones obtenidas de la Ecuación 3.2 ($tt(m, j)$) son agrupadas en cada período h . De forma análoga se determinan las distribuciones de tiempo de viaje TT_j^h calibradas con el histograma de tiempos de viaje ocurridos en el arco j en el periodo h .

Un vehículo puede pasar más de una vez por un arco en el mismo viaje por lo que es necesario subdividir los tiempos obtenidos y que sean mayores a τ_{viaj} . Con el fin de precisar en los valores de τ_{perm} y τ_{viaj} pueden ser calibrados para cada nodo y arco respectivamente en función de las magnitudes de los tiempos observados en cada uno (τ_{perm}^i y τ_{viaj}^j).

A finalizar esta etapa, cada nodo y cada arco de la red tendrá asociada una distribución de tiempo de permanencia y tiempo de viaje respectivamente en cada periodo h del día. La Figura 3.6 esquematiza en una red ficticia el resultado de las etapa 2 de la metodología. En cada arco se tiene una distribución de tiempos de viaje por periodo y en cada nodo una distribución de tiempos de permanencia.

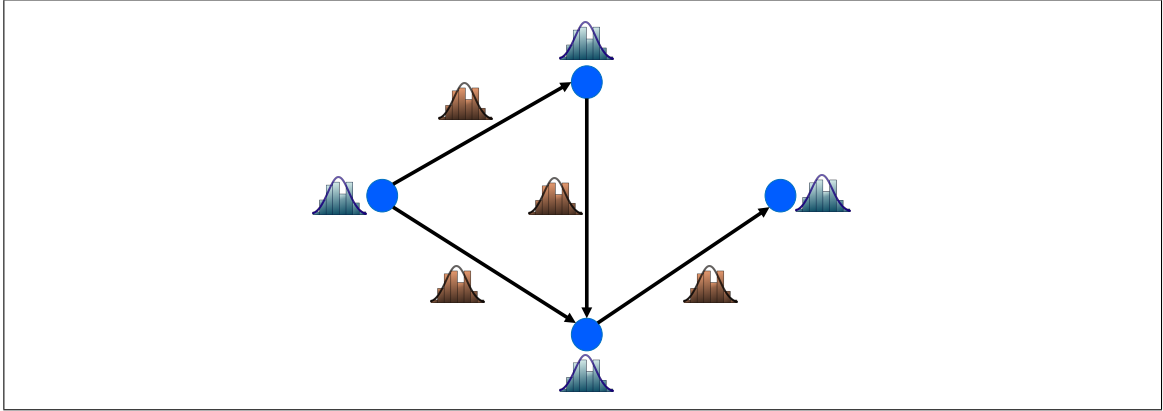


Figura 3.6. Distribuciones de tiempo de permanencia en nodos y tiempo de viaje en arcos para un red ejemplo al finalizar la etapa 2 de la metodología

3.3. Etapa 3: Reconstrucción de rutas

Una vez ajustadas las distribuciones de tiempos de permanencia en nodos (DT_i^h) y tiempos de viaje en arcos (TT_j^h) para cada nodo (i) y arco (j) en cada periodo del día (h) se procede a realizar la inferencia de rutas a los vehículos que no se puede inferir directamente su ruta. Es decir, si los registros BT no son suficientes para conocer el detalle de la ruta, entonces se procede a aplicar la inferencia Bayesiana. Esta etapa consta de dos fases. En la primera fase se aplica una inferencia Bayesiana para cada par de detecciones sucesivas de un mismo vehículo. Luego, en la segunda fase se obtienen las probabilidades de la ruta completa.

3.3.1. Inferencia Bayesiana

En esta fase se determinan las probabilidades de elección de una sección de ruta entre dos detecciones sucesivas en el tiempo para un mismo vehículo. Si los SB que realizan este par de detecciones se encuentran conectados por un arco directo, entonces el vehículo será asignado con probabilidad 1 a ese arco, tal y como se mencionó en la Sección 3.2.2.

Si no existe un arco que conecte directamente este par de nodos, es necesario analizar las rutas disponibles para ir desde un nodo al otro e inferir la elección de dichas rutas.

Sean $t_a = t(m, a, W)$ y $t_b = t(m, b, 1)$ instantes en que ocurren dos detecciones consecutivas del vehículo m ocurridas en ($t_a < t_b$). t_a es el instante de la última detección en el nodo a (aguas arriba) y t_b el instante de la primera detección en el nodo b (aguas abajo). Los nodos a y b no se encuentran directamente conectados en G (ver Figura 3.7). Sobre las rutas que conectan es que se realiza la inferencia Bayesiana.

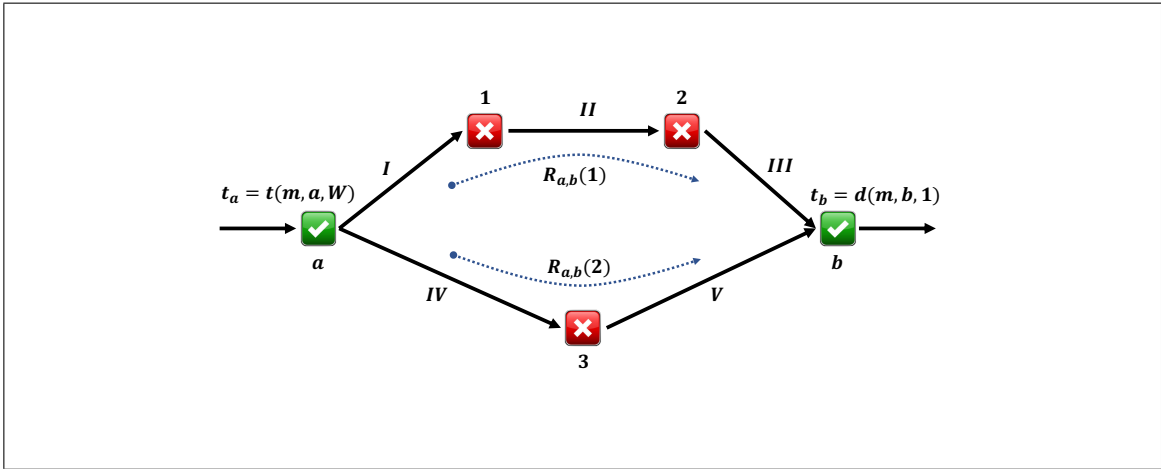


Figura 3.7. Detecciones de un vehículo en las que se puede realizar una inferencia de ruta

Definición del conjunto de alternativas

Es necesario definir un conjunto de rutas para cada par de detecciones sucesivas sobre las que se realiza la inferencia Bayesiana. En particular para el ejemplo, se define un mecanismo para construir el conjunto de rutas entre los nodos a y b (Figura 3.7). A cada una de estas rutas se le asigna una probabilidad de uso luego de aplicar la metodología. Para definir el conjunto de rutas es posible utilizar cualquier criterio coherente y que en principio entregue el mayor número de rutas posibles, dado que la metodología se encarga

de decidir cuáles rutas son las más probables de las que pudo escoger el vehículo que fue registrado por los sensores.

En este estudio se aplicará el algoritmo presentado por Yen (1971), que permite encontrar un total de k rutas mínimas entre dos nodos de una red. Para esto, se procede a encontrar la ruta mínima con algún algoritmo estándar como el presentado por Dijkstra (1959). A partir de ella se analizan desviaciones de la ruta mínima utilizando solo una parte de esta y completándola con una nueva aplicación de ruta mínima. Es decir, el algoritmo de Yen (1971) consiste en tomar una porción de la ruta mínima y completarla por algún camino alternativo. Este proceso se repite con distintas partes iniciales de la ruta mínima original y se almacenan todas las rutas encontradas, para luego escoger las k rutas más cortas.

En este estudio se utilizaron tres funciones de costo en el algoritmo de Yen (1971) para decidir qué rutas son las mínimas: (i) la distancia, que corresponde a la longitud almacenada en cada arco en la Etapa 1 de la metodología; (ii) el número de nodos, que corresponde a la cantidad de SB que posee cada ruta, siendo mejor aquella que tenga un menor número de nodos, que es el criterio utilizado por Michau *et al.* (2014) para seleccionar la ruta elegida por los vehículos; y (iii) el costo angular (Raveau, Muñoz, y De Grange, 2011) donde se define que las rutas mínimas son las más directas (menor desviación de la línea recta entre los puntos). Este último costo se define como la sumatoria de las distancias de cada arco ponderadas por los senos de los semi-ángulos desde cada punto al destino.

Para ejemplificar cada uno de los costos utilizados en la obtención del conjunto de rutas alternativas, la Figura 3.8 presenta una red con un nodo origen (a) y un nodo destino (b). Existen dos posibles rutas para ir desde a a b : la ruta superior (a -1-2- b) y la ruta inferior (a -3- b). La ruta superior posee un costo de distancia (i) igual a $d_1 + d_2 + d_3$, mientras que la ruta inferior $d_4 + d_5$. Respecto al número de nodos (ii), la ruta superior presenta un costo

igual a 2 (nodos 1 y 2) y la inferior un costo igual a 1 (solo el nodo 3). Finalmente, el costo angular (iii) de la ruta superior viene dado por la distancia de ir de a a 1 (d_1) ponderado por seno del semi-ángulo que forma con el camino directo ($\sin(\theta_1/2)$), más la distancia de ir de 1 a 2 (d_2) ponderado por $\sin(\theta_2/2)$, quedando luego el camino directo de 2 a b . De esto resulta $d_1 \cdot \sin(\theta_1/2) + d_2 \cdot \sin(\theta_2/2)$. De forma análoga para la ruta inferior, el costo angular es $d_4 \cdot \sin(\theta_3/2)$.

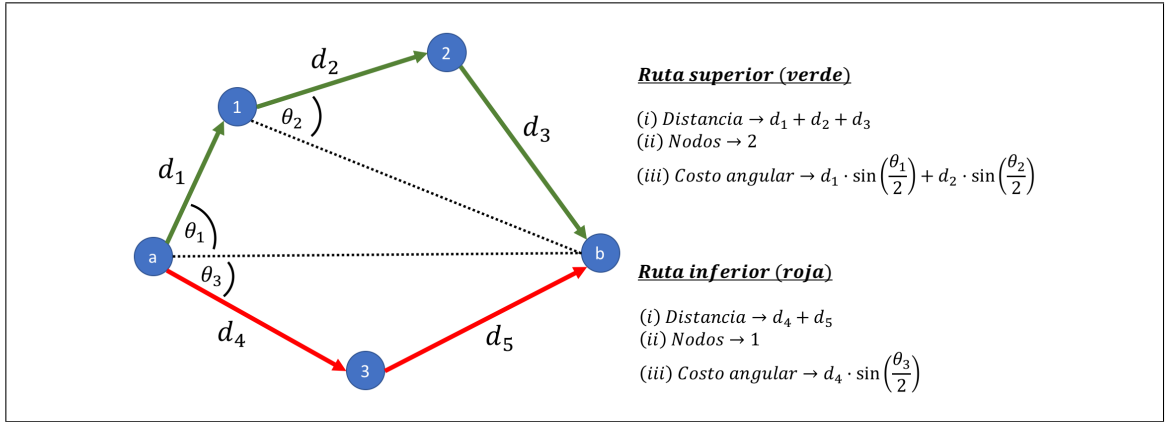


Figura 3.8. Definición de costos utilizados en el algoritmo de k rutas mínimas

El algoritmo de Yen (1971) se aplica un total de tres veces entre cada par de nodos de la ruta. Cada aplicación entregará a lo más k rutas mínimas y dado que puede existir repetición de rutas obtenidas con criterios distintos, se tendrán a lo más $3k$ rutas dentro de las que se aplicará directamente la inferencia Bayesiana. De este modo se obtendrá una probabilidad de uso para cada una de las rutas encontradas. Es importante destacar que el algoritmo de Yen (1971) solo entrega rutas sin ciclos. El valor de k puede ser definido en función del tamaño de la red, con el fin de no dejar rutas fuera de análisis. De este modo, la metodología será capaz de decidir qué rutas descartar.

Una vez definido el conjunto de rutas para cada par de nodos se procede a realizar la inferencia. Sea $R_{a,b}$ el conjunto de rutas alternativas (obtenidas con el algoritmo de Yen) entre los nodos a y b . La expresión $R_{a,b}(n)$ es utilizada para denotar la n -ésima ruta del

conjunto $R_{a,b}$. Cada ruta corresponde a una secuencia de nodos y arcos. Por ejemplo, en la Figura 3.7, la ruta $R_{a,b}(1)$ tiene 3 arcos y 2 nodos, mientras que la ruta $R_{a,b}(2)$ tiene 2 arcos y un nodo.

Inferencia del conjunto de rutas

La inferencia Bayesiana, como la describe McElreath (2015), es un tipo de inferencia estadística en donde se busca conocer la probabilidad de un suceso en base a evidencias y datos disponibles que se tienen sobre el suceso mismo u otro que se relacione con él. Este concepto deriva del teorema de Bayes, en que se actualiza una probabilidad en base a condicionantes de los que ya se tiene certeza. En la Ecuación 3.3 se enuncia la fórmula de Bayes, donde se busca conocer la probabilidad del evento A_i sujeto a que se conoce la ocurrencia del evento B .

$$Pr \{A_i | B\} = \frac{Pr \{B | A_i\} \cdot Pr \{A_i\}}{\sum_z Pr \{B | A_z\} \cdot Pr \{A_z\}} \quad (3.3)$$

Antes de aplicar esta propiedad en este estudio, es necesario caracterizar la probabilidad buscada (A_i) y la información de la que ya se tiene certeza de su ocurrencia (B). En base a la definición del problema, A_i corresponde al evento de elección de cada una de las rutas disponibles, listadas como $R_{a,b}(n)$. Por otro lado, B corresponde a dos tipos de información conocida:

1. No existen detecciones entre los nodos que realizaron las detecciones (entre los nodos a y b de la Figura 3.7)
2. El tiempo entre estas detecciones (considerando una aproximación del tiempo de viaje entre a y b) es conocido e igual al tiempo entre detecciones ($T_{obs} = t_b - t_a$ en la Figura 3.7).

Con esta información es posible definir $Pr\{R|Nd \wedge T_{obs}\}$ que corresponde a saber que probabilidad tiene la ruta R sabiendo que no se hicieron detecciones Nd y que el tiempo de viaje T_{obs} es conocido. En concreto por tanto es $Pr\{R = R_{a,b}(n)|Nd = q \wedge T_{obs} = t\}$, es decir, la probabilidad que un vehículo escoja la ruta n del conjunto $R_{a,b}$ condicionado a que ninguno de los q SB de la ruta n realizaron detecciones (Nd) para este vehículo y que el tiempo de viaje entre a y b (T_{obs}) es t . Con el fin de simplificar la notación, la probabilidad descrita será representada por $Pr\{R_n|Nd_q \wedge t\}$. Luego, aplicando la fórmula de Bayes (Ecuación 3.3) se obtiene la expresión de la Ecuación 3.4.

$$Pr\{R_n|Nd_q \wedge t\} = \frac{Pr\{Nd_q \wedge t|R_n\} \cdot Pr\{R_n\}}{\sum_z Pr\{Nd_{q(z)} \wedge t|R_z\} \cdot Pr\{R_z\}} \quad (3.4)$$

Donde z son las rutas del conjunto de alternativas ($R_{a,b}(z) \in R_{a,b}$) y $q(z)$ es el número de SB en la ruta z . Luego, al desarrollar la Ecuación 3.4 tanto en el numerador como en el denominador, resulta:

$$Pr\{R_n|Nd \wedge t\} = \frac{Pr\{Nd_q|t \wedge R_n\} Pr\{t|R_n\} Pr\{R_n\}}{\sum_z Pr\{Nd_{q(z)}|t \wedge R_z\} Pr\{t|R_z\} Pr\{R_z\}} \quad (3.5)$$

La suma en el denominador de la Ecuación 3.5 considera todos los elementos del conjunto $R_{a,b}$. Para evaluar la expresión que representa la probabilidad buscada es necesario entender y obtener los tres tipos de probabilidades que componen la Ecuación $R_{a,b}$. Estas son: $Pr\{Nd_q|t \wedge R_n\}$, $Pr\{t|R_n\}$ y $Pr\{R_n\}$. La primera de ellas corresponde a la probabilidad que no se realicen detecciones en el paso por la ruta n . La segunda de ellas es la probabilidad de que un vehículo tarde un cierto tiempo t al usar la ruta n . Por último, la tercera de ellas corresponde a la probabilidad *a priori* de usar la ruta n , antes de contar con cualquier tipo de información.

A continuación, se describe como determinar cada una de las probabilidades ya mencionadas. La Figura 3.7 es de gran utilidad para comprender cada una de las probabilidades que se presentan.

Probabilidad de no detección ($Pr \{Nd_q | t \wedge R_n\}$)

Cada ruta del conjunto $R_{a,b}$ posee un número fijo de SB (uno por cada nodo intermedio). Cuando un vehículo pasa por cada uno de los sensores puede ser detectado o no en función de variados factores, como mencionan Michau *et al.* (2014). Dentro de los más importantes, ya mencionados en la Sección 2.2.5, están la infraestructura cercana a los sensores, interferencias de otro tipo de señales, condiciones climáticas y tiempo total que un vehículo se encuentra en en la zona de detección de un sensor, entre otras. De estas, la información con la que se cuenta a partir de las detecciones BT es la velocidad de los vehículos a lo largo de la red, valor que incide directamente en el tiempo de permanencia de estos dentro del área de detección. Entre más rápido pase el vehículo por un sensor, menor será la probabilidad de detectarlo. Así, debiera ser posible construir una función que relacione la velocidad de un vehículo y su probabilidad de ser detectado al menos una vez al pasar por un sensor. Luego de experimentación se concluye que a velocidades superiores a los 72 km/h es improbable realizar una detección (Michau *et al.*, 2014). Es importante construir una relación entre la velocidad y la probabilidad de detección $p(v)$. Para ello se puede realizar una calibración en detalle si se cuenta con los medios para realizar una experimentación. Dicha experimentación está fuera de los alcances de este estudio.

Para conocer la velocidad promedio de un vehículo m en la ruta n , v_n^m se divide la distancia que existe entre los dos sensores por la ruta n ($R_{a,b}(n)$) por el tiempo entre las detecciones (T_{obs}). Si se desea conocer la velocidad para una ruta en el ejemplo de la Figura 3.7, se utiliza la Ecuación 3.6.

$$v_n^m = \frac{\text{distancia}(R_{a,b}(n))}{T_{obs}^m} \quad (3.6)$$

Con esta información es posible calcular la probabilidad de no registrar detecciones condicionada a que el vehículo viaja por una ruta específica y a que tarda cierto tiempo en recorrerla, es decir, viaja a cierta velocidad. Se considera que la ruta $R_{a,b}(n)$ tiene q SB, entonces la expresión que representa la probabilidad buscada corresponde a:

$$Pr \{Nd = q | T_{obs} = t \wedge R = R_{a,b}(n)\} = (1 - p(v_n^m))^q \quad (3.7)$$

Aplicado al ejemplo de la Figura 3.7 resulta que para la ruta 1 es $(1 - p(v_n^m))^2$, dado que tiene dos SB, y para la ruta 2 es $(1 - p(v_n^m))$, dado que tiene solo un SB. Para explicar la estructura de esta ecuación basta pensar que la detección de los vehículos por los sensores componen experimentos Bernoulli, donde $p(v_n)$ es la probabilidad de detectar y $(1 - p(v_n^m))$ es la probabilidad de no detectar. Como se trata de q detecciones fallidas, es que resulta $(1 - p(v_n^m))^q$. Esto asume que la probabilidad de ser detectado en un SB es independiente de que ocurre en otros SB.

Probabilidad del tiempo total de viaje ($Pr \{t | R_n\}$)

Para cada ruta n perteneciente a $R_{a,b}$, se ajusta una distribución que represente el tiempo total de viaje por dicha ruta. Para llevar a cabo este ajuste es necesario recurrir a las distribuciones obtenidas en la primera etapa (tiempo de permanencia en nodos y tiempo de viaje en arcos). Para entender cómo es posible obtener una nueva distribución a partir de las anteriores, notar que cada ruta está formada por una secuencia de nodos y arcos. Por ejemplo, en la Figura 3.7 se observan 2 rutas, la primera de ellas está formada por la secuencia: Arco I, Nodo 1, Arco II, Nodo 2 y Arco III; y la segunda ruta está compuesta por Arco IV, Nodo 3 y Arco V. Para cada periodo de tiempo (h), cada uno de los elementos mencionados (nodos y arcos) posee una distribución de tiempo. La distribución buscada

corresponde a la agregación de estas distribuciones, las que generan una nueva distribución de probabilidad que representa el tiempo total de viaje desde el nodo a al b por la ruta n para cada periodo h (TTT_n^h).

Existe una operación de funciones matemáticas que transforma dos funciones en una tercera función que representa la superposición de una de ellas con la invertida de la otra. En distribuciones de probabilidad permite la unión de distribuciones de probabilidad, con el fin de obtener una nueva distribución que representa la suma de las variables explicadas por las distribuciones iniciales. Esta operación lleva el nombre de convolución de funciones. En la Ecuación 3.8 se muestra la convolución para dos funciones, $f(x)$ y $g(x)$.

$$(f \otimes g)(x) = \int_{-\infty}^{\infty} f(c) g(x - w) dc \quad (3.8)$$

Donde $\otimes$ corresponde al operador de una convolución. Es importante mencionar que esta operación es conmutativa, por lo que no importa el orden en que se convolucionan las distribuciones. Además, las distribuciones con las que trabaja este estudio son discretas, por lo que es necesario particularizar la Ecuación 3.8 al caso discreto, de lo que resulta:

$$(f \otimes g)(x) = \sum_c f(h) g(x - c) \quad (3.9)$$

En el ejemplo de la Figura 3.7, para cada ruta es posible determinar una distribución que represente el tiempo total de viaje por ella. Este tiempo está compuesto por la suma de los tiempos de permanencia de todos los nodos y de viaje de todos arcos que componen la ruta. Para el caso de la ruta 1, se tiene que TTT_1^h es el resultado de convolucionar las distribuciones DT_1^h y DT_2^h (respecto a nodos) y TT_I^h , TT_{II}^h y TT_{III}^h (respecto a arcos). A modo general, para obtener la distribución del tiempo de viaje de cada ruta, basta con realizar la operatoria de la Ecuación 3.10.

$$TTT_n^h = TT_1^h \otimes TT_2^h \otimes \dots \otimes TT_{q+1}^h \otimes DT_1^h \otimes DT_2^h \otimes \dots \otimes DT_q^h \quad (3.10)$$

Donde q corresponde al número de nodos, por lo que existen $q + 1$ arcos.

Posterior a la construcción de esta distribución es necesario calcular la probabilidad buscada que corresponde a la probabilidad $Pr\{t = T_{obs} | R = R_{a,b}(n)\}$ que un vehículo demore un tiempo específico al viajar por una de las rutas disponibles. Dado el condicionante de esta probabilidad, ya se conoce la estructura de la ruta y por ende su distribución de tiempo de viaje.

Previo a determinar la probabilidad, es importante destacar que el valor de T_{obs} obtenido por la diferencia de los tiempos de detecciones en a y en b posee errores debido a la precisión de la tecnología Bluetooth. Para entender este error, la Figura 3.9 muestra una de las rutas disponibles entre los nodos a y b . Los tiempos t_a y t_b corresponden a los instantes de la última detección en el nodo a (aguas arriba, cuadro *tick* verde en la zona de detección del nodo a) y la primera detección en el nodo b (aguas abajo, *tick* verde en la zona de detección del nodo a), respectivamente. El error puede ser dividido en dos componentes. El primero de ellos corresponde a la diferencia que existe entre la detección en el tiempo t_a y el instante real en que el vehículo dejó la zona de detección del SB en a (borde de la zona de detección). De forma similar, el segundo corresponde a la diferencia entre el instante real en que el vehículo ingresó a la zona de detección del sensor en b (borde del área de de detección) y el tiempo en que se realizó la detección t_b .

Para estimar la magnitud de este error, se procede a contabilizar el número total de detecciones que se realizaron para el vehículo en cuestión en cada uno de los SB (a y b). Sean W_a y W_b el número total de registros realizados en a y b , respectivamente. Si se asume que las detecciones ocurren equidistantes en el tiempo, entonces la Ecuación 3.11

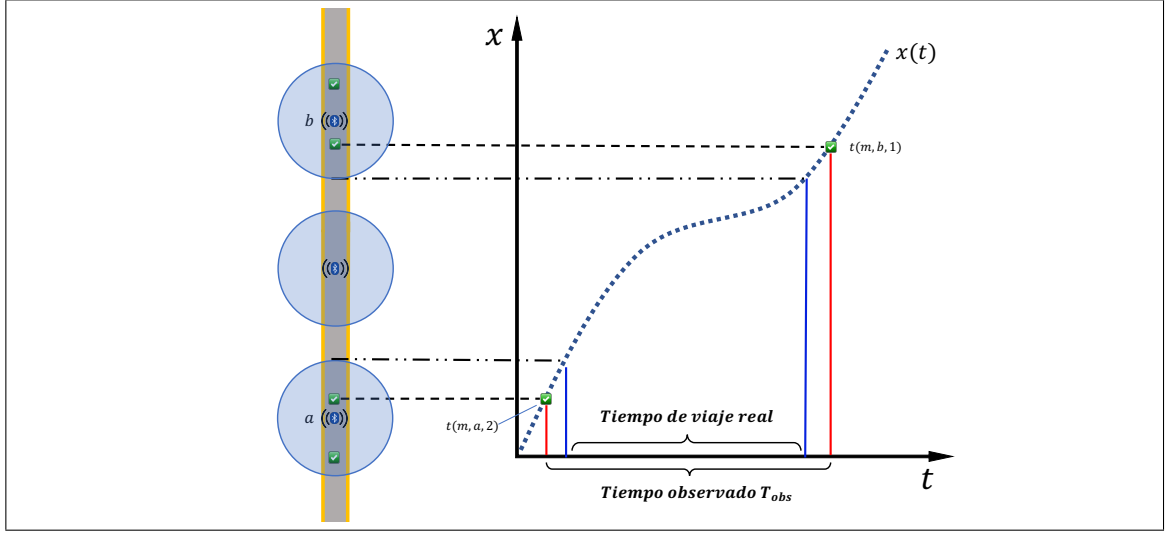


Figura 3.9. Error de medición de la variable T_{obs}

muestra la forma de calcular este error.

$$\varepsilon = \frac{E[DT_a^k]}{W_a + 1} + \frac{E[DT_b^k]}{W_b + 1} \quad (3.11)$$

El primer sumando corresponde al error de medición en el sensor a , donde se divide el valor esperado del tiempo de permanencia del vehículo en la zona de detección del sensor en el número total de detección más 1. Este valor representa cada uno de los tiempos entre detecciones que es utilizado para estimar el error. El ejercicio es análogo para el error presente en b (segundo sumando).

Con este procedimiento es posible saber que el valor del tiempo de viaje del vehículo pertenece al intervalo $[T_{obs} - \varepsilon, T_{obs}]$. Además, sabemos que la distribución detrás de este tiempo de viaje es TTT_n^h . Así, solo basta conocer con qué probabilidad este tiempo puede ser explicado por la distribución. Para ello, es necesario calcular la integral presentada en la Ecuación 3.12 cuando se trata de cualquier distribución para TTT_n^h .

$$Pr \{t = T_{obs} | R = R_{a,b}(n)\} = Pr \{T_{obs} - \varepsilon < t \leq T_{obs}\} = \int_{T_{obs}-\varepsilon}^{T_{obs}} TT_n^h(t) dt \quad (3.12)$$

Dado que en este trabajo TT_n^h es una distribución discreta, entonces la probabilidad se debe determinar a partir de la Ecuación 3.13

$$Pr \{t = T_{obs} | R = R_{a,b}(n)\} = Pr \{T_{obs} - \varepsilon < t \leq T_{obs}\} = \sum_{T_{obs}-\varepsilon}^{t=T_{obs}} TT_n^h(t) \quad (3.13)$$

Probabilidad *a priori* de uso de ruta ($Pr \{R_n\}$)

Este último término representa la probabilidad *a priori* de uso de cada ruta, es decir, muestra que proporción de flujo tiene cada ruta n por sobre las demás. Es posible realizar distintos supuestos para su obtención; por ejemplo, se puede asumir que todas las rutas poseen igual probabilidad *a priori* de uso. También se pueden asignar las probabilidades *a priori* en función de algún estudio previo acerca de la utilización de las diferentes rutas. Una vez que se cuenta con las probabilidades *a priori*, estas serán actualizadas por las dos probabilidades anteriores (probabilidad del tiempo total de viaje y probabilidad de no detección) para obtener una nueva probabilidad de elección de ruta basada en la información obtenida de los SB.

Si se cuenta con una estimación de flujos de cada ruta, es necesario determinar la relación de flujo de cada ruta f_n respecto al flujo total (Ecuación 3.14).

$$Pr \{R_r\} = \frac{f_r}{\sum_n f_n} \quad (3.14)$$

A continuación se presenta una propuesta metodológica para estimar este flujo f_n de cada ruta realizando conteos en la base de datos.

Cuando un vehículo circula por una ruta de N nodos, este puede ser detectado por solo uno de los sensores, dos de ellos, tres o hasta por los N disponibles. La probabilidad que sea detectado por un número específico de sensores se puede estimar por medio de la distribución Binomial, la que determina la probabilidad de obtener cierta cantidad de éxitos (detecciones) en una sucesión de eventos Bernoulli (paso de vehículos por la zona de detección de los SB). Dada esta información, es posible conocer la probabilidad de que un vehículo sea detectado por todos los sensores de una ruta. Esto, con el previo conocimiento del número de sensores de la ruta y la probabilidad de detección en cada uno de ellos.

Para determinar el valor del flujo de vehículos que pasa por una ruta n específica (f_n) se busca plantear una ecuación con los vehículos que son detectados en todos los sensores de la ruta. Con el fin de facilitar la comprensión de esta relación se utiliza el ejemplo de la Figura 3.7. Como se conocen las detecciones de los extremos de la ruta (en los nodos a y b de la Figura 3.7), también se conocen los tiempos de detección (t_a y t_b) y la longitud de la ruta. Con esta información se puede determinar la velocidad promedio de pasada en la ruta para un vehículo en específico.

Dentro de la base de datos, se puede realizar un conteo de todos los vehículos que hayan sido detectados en a , b y en todos los sensores de la ruta en cuestión, llámese DB_n al número total de vehículos contados para la ruta n que fueron detectados por todos los sensores de esa ruta. Luego, es posible determinar la velocidad de pasada de todos estos vehículos por la ruta (v_{prom}). Finalmente es posible plantear una relación que permita obtener el valor de f_n . Por un lado se tiene que DB_n (vehículos que son detectados por todos los sensores en la ruta n) y por otro se tiene que la proporción de vehículos que cumplen esta característica es $(p(v_{prom}))^q$, donde q es el número de sensores de la ruta n y $p(v_{prom})$ es la probabilidad de un vehículo para ser detectado por un sensor) y que el total de vehículos que circula por la ruta es f_n (valor buscado). Al plantear esta igualdad y despejando el valor de f_n resulta la Ecuación 3.15.

$$f_n = \frac{DB_n}{(p(v_{prom}))^q} \quad (3.15)$$

A partir de la Ecuación 3.15 se pueden determinar todos los flujos involucrados, los que por ejemplo en la Figura 3.7 corresponden a los flujos f_1 y f_2 . Luego debe aplicarse la Ecuación 3.14 para obtener las respectivas probabilidades *a priori* de uso de cada una de las rutas. Si excepcionalmente no se registran conteos para una ruta en específico ($DB_n = 0$), entonces el valor del flujo asociado a esa ruta será cero.

3.3.2. Cálculo de probabilidades de ruta

Ya presentada la metodología para determinar cada elemento de la Ecuación 3.5, es posible calcular las probabilidades de elección de ruta, para cada par de detecciones sucesivas en donde exista más de una ruta posible. Por lo tanto, para conocer las probabilidades de cada una de las ruta se debe seguir la siguiente secuencia de pasos:

1. Para cada vehículo, se reconoce la secuencia de nodos en donde este fue detectado.
2. Para cada par de detecciones consecutivas de la secuencia, que constituyen una sección de la ruta, se evalúa si existe más de una ruta para viajar desde el primer al segundo nodo.
 - 2.1. Si existe solo una ruta (*i.e.* un arco directo que conecte ambos nodos), entonces esta ruta tiene una probabilidad del 100 % de haber sido elegida.
 - 2.2. De lo contrario, se procede a aplicar la metodología propuesta y así obtener las probabilidades individuales de cada una de las rutas entre el par de nodos.
3. La probabilidad total de elección de una ruta corresponde al producto de las probabilidades individuales de cada Sección de ruta.

A modo de ejemplo, considere la Figura 3.10 en la que un vehículo circuló de izquierda a derecha, y es detectado en los sensores marcados con un *tick* verde y no detectado en los sensores con una cruz roja. Se procede a aplicar la metodología entre los dos primeros

nodos de detección (sección de ruta). Entre ellos existen tres rutas diferentes, a las que se les asignó probabilidades de elección de 20 %, 30 % y 50 % mediante la Ecuación 3.5. Luego se contabilizan las rutas disponibles entre el segundo y tercer nodo de detección, que al resultar solo una, se asigna una probabilidad del 100 %. Finalmente, entre los dos últimos nodos de detección se observan dos rutas, a las que se les asigna probabilidades de elección del 80 % y 20 %. De este ejemplo, se puede notar que existen seis rutas totales con diferentes probabilidades de uso. Por ejemplo, para la ruta destacada en la Figura 3.10 (línea punteada) la probabilidad de uso es 4 % ($20 \% \cdot 100 \% \cdot 20 \%$).

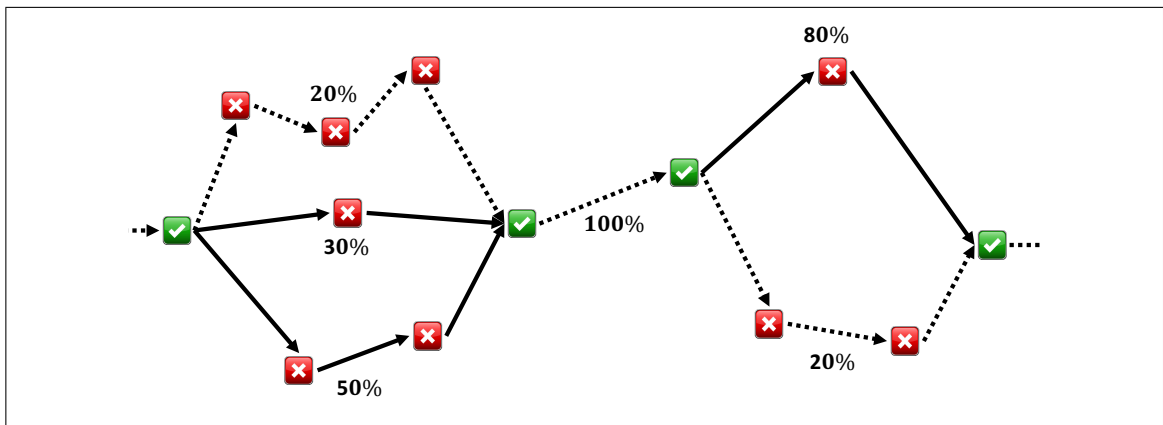


Figura 3.10. Ejemplo esquemático de la aplicación de la metodología sobre la ruta completa

CAPÍTULO 4. VALIDACIÓN

En esta sección se detalla cómo fue implementada la metodología con el fin de validarla. Además se describen las herramientas utilizadas para llevar a cabo esta validación. Para este último punto, se describe la experimentación, se presentan los escenarios a los que es sometido el trabajo realizado y se detallan los mecanismos e indicadores utilizados para medir la predicción de la metodología. El desarrollo y programación de la metodología es descrito en el Anexo B.

Con el fin de validar la capacidad predictiva de la metodología se procede a desarrollar una simulación en el *software* de microsimulación de tráfico Aimsun. Este programa permite conocer la posición de cada vehículo en todo instante de tiempo, de modo que se pueda conocer la ruta elegida por cada uno de ellos. De esta forma es posible comparar la ruta real con la predicción realizada por la metodología.

Dado que el *software* no cuenta con una herramienta para la recolección de datos a partir de SB, se desarrolló una API en *Python* que se encarga de simular un SB en todas las intersecciones de la red. En caso de que no todas las intersecciones cuenten con un SB, esta API permite seleccionar un subconjunto de las detecciones BT que solo pertenezcan a intersecciones con SB. De este modo, se pueden evaluar distintas configuraciones de sensores para una red dada.

4.1. Definición del experimento

A continuación se definen los aspectos generales del experimento al que será sometida la metodología. Para ello es necesario describir las características y comportamiento de los SB a utilizar, la red en donde se instalarán, la demanda de vehículos que va a circular

por la red, la matriz origen destino utilizada en la simulación y la definición de los diferentes escenarios que se van a testear. La experimentación utiliza 1,5 horas de horizonte de simulación con una discretización de $h = 15$ minutos.

4.1.1. Sensores Bluetooth

Los SB dispuestos en la red almacenan el identificador de todos los vehículos equipados con BT que se encuentran en un radio de 25 metros a su alrededor. En práctica, este valor puede ser superior a los 100 m, pero es posible regular en base a la potencia configurada en el SB. Los identificadores de vehículos se almacenan cada cierto tiempo a definir por el usuario.

Luego, a cada uno de estos vehículos se le asigna una probabilidad de detección en función de su velocidad instantánea, la que es determinada por la función $p(v)$ definida en la Sección 3.3.1 para obtener la probabilidad de no detección. Esta función permite conocer la probabilidad de ser detectado en función de su velocidad instantánea de circulación. Como se mencionó en la Sección 3.3.1, esta puede ser calibrada mediante la utilización de datos o supuestos de comportamiento. En concreto, los escenarios utilizan la función presente en la Figura 4.1 basada en las descripciones de Bhaskar y Chung (2013).

Esta relación establece que a partir de 80 km/h la probabilidad de detección es 0, mientras que estando en reposo esta será de un 80 %. Entre ambos valores la probabilidad decrece linealmente.

Finalmente se genera un experimento Bernoulli con dicha probabilidad para saber si existió detección o no. De ser así, se almacena la detección en la base de datos; en caso contrario, simplemente se descarta el registro. De este modo, se construye una base de datos de detecciones BT. Estas detecciones son las utilizadas en la etapa de precarga de información y en la posterior reconstrucción de rutas.

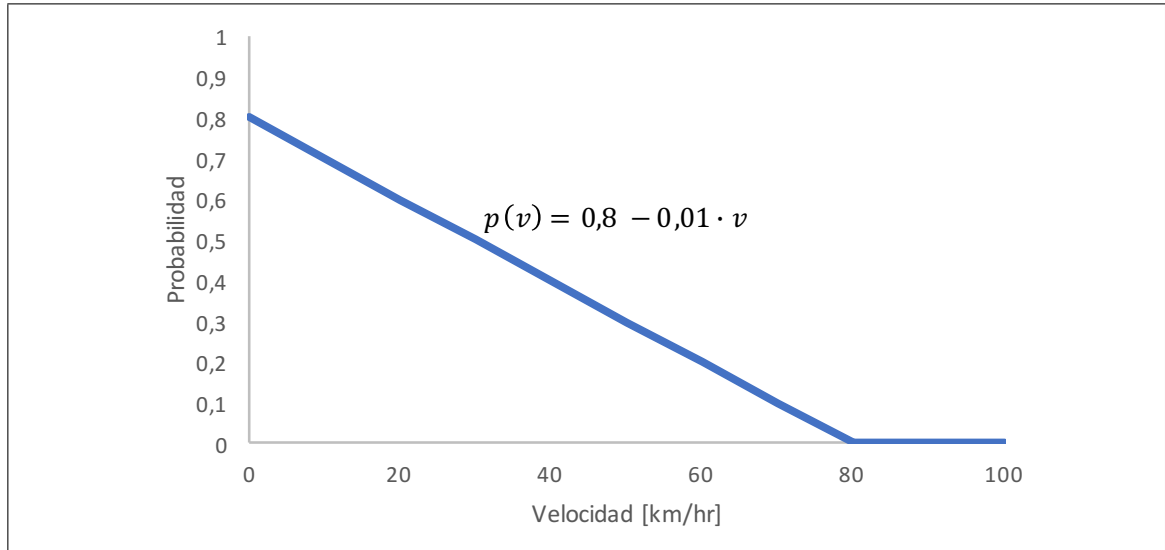


Figura 4.1. Probabilidad de detección en función de la velocidad de circulación

Con esta información se genera una base de datos que contiene las rutas realmente escogidas por cada vehículo a partir de la secuencia de detecciones sin probabilidad de fallo. Esta sirve para tener una base de datos de comparación respecto a la inferencia de ruta que entrega el *software* de la metodología con las detecciones de los SB. El cómo se obtiene esta información se encuentra explicado en el Anexo B.

4.1.2. Red

La red modelada se compone de una grilla tipo Manhattan de seis por seis intersecciones, como se muestra en la Figura 4.2.

La distancia entre dos intersecciones consecutivas es 100 metros. Los sentidos de cada eje vial se encuentran intercalados y se indican por las flechas de la Figura 4.2. Cada vía posee dos pistas de 2,75 metros de ancho cada una y cada intersección posee un semáforo. Cada cruce semafórico opera con un ciclo de 90 segundos y dos fases, donde cada una de ellas es de 45 segundos y los tiempos de amarillo y rojo-rojo no son considerados

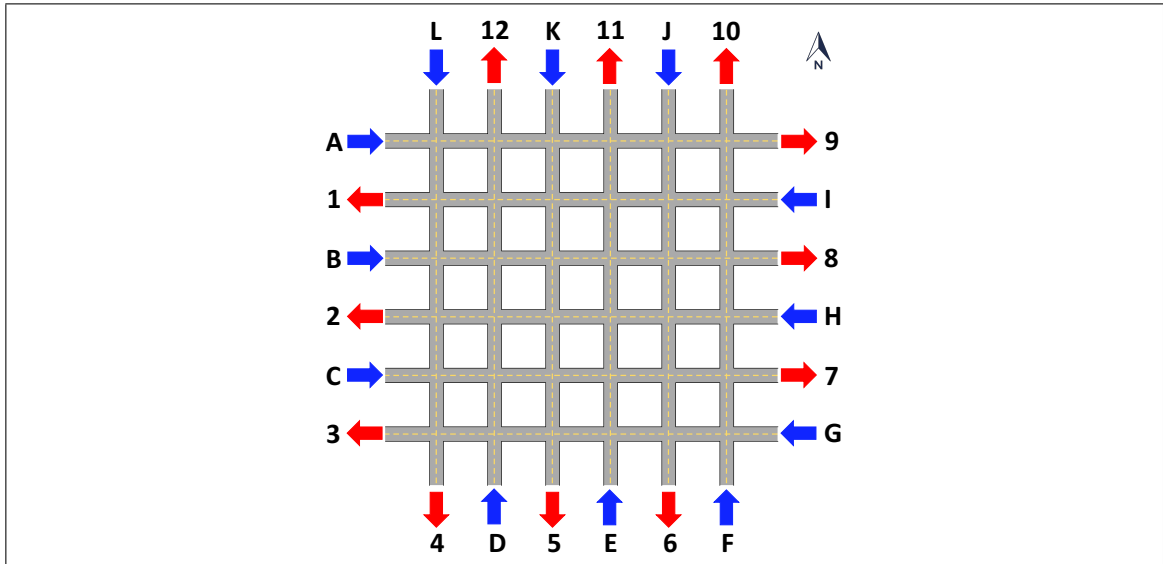


Figura 4.2. Red vial utilizada en la simulación

por simplicidad. Además, existe un desfase de 5 segundos en los semáforos de los ejes este-oeste (horizontales) en el sentido del movimiento de cada uno.

4.1.3. Demanda

La red recibe una demanda total de 16.250 vehículos en un periodo de 90 minutos. Los vehículos hacen ingreso a la red por las vías marcadas con letras y salen por las marcadas con números (ver Figura 4.2). Los niveles de demanda cambian cada 30 minutos con los valores indicados en la Figura 4.3.

Se observa un periodo inicial de carga de la red, luego un periodo de alta demanda y finalmente un periodo de muy baja demanda. La entrada de vehículos por cada vía de entrada (marcadas con letras) es aleatoria y se rige por una distribución de Poisson. Además, la demanda es heterogénea en la red, es decir, cada vía de acceso presenta diferentes tasas de entrada, las que también fueron asignadas de forma aleatoria. Del total de vehículos, un 40 % de ellos (6.500 vehículos) se encuentran equipados con tecnología BT y son los

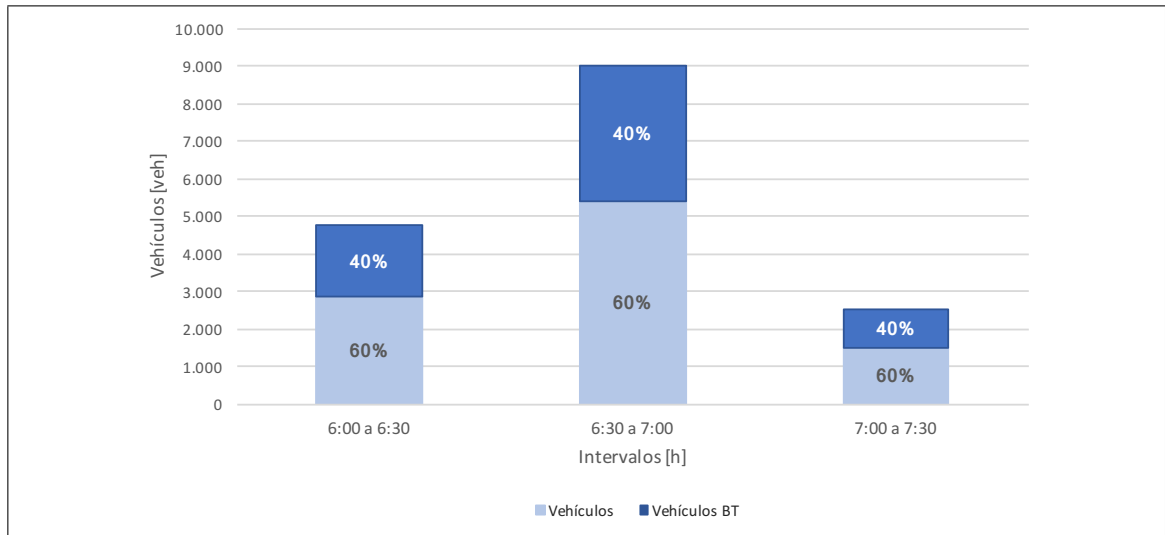


Figura 4.3. Distribución de la demanda sobre la red vial

únicos que pueden ser detectados por los SB. Los demás vehículos solo circulan por la red, sin ser detectados.

La composición de la demanda puede ser clasificada respecto a como fue detectado cada vehículo. La Figura 4.4 muestra la clasificación de cada uno.

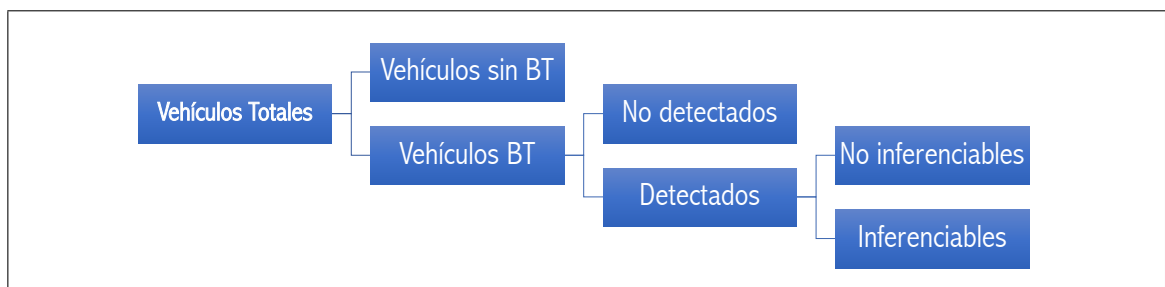


Figura 4.4. Clasificación de la demanda respecto a las detecciones realizadas por los SB

- Vehículos totales: Corresponde al conjunto de todos los vehículos que ingresan a la red, incluso si no emiten algún tipo de señal BT.

- Vehículos sin BT: Corresponden a un subconjunto de los vehículos totales (60 %) que no puede ser detectados dado que no poseen ni emiten ningún tipo de señal BT.
- Vehículos BT: Corresponden a un subconjunto de los vehículos totales (40 %) que cuentan con señales BT y pueden ser detectados por los SB de la vialidad.
- Vehículos detectados: Corresponden a un subconjunto de los vehículos BT que son detectados al menos una vez por alguno de los SB de la red.
- Vehículos inferenciables: Corresponden a un subconjunto de los vehículos detectados que tienen dos o más detecciones y existe la posibilidad de aplicar al menos una inferencia Bayesiana para inferir su ruta (*i.e* un vehículo que tiene más de una opción de ruta entre dos detecciones sucesivas de alguno de sus tramos).

Para conocer la distribución de los viajes entre las distintas vías de entrada y salida, en la Figura 4.5 se detalla la matriz origen destino de los vehículos que circulan por la red. Como se mencionó, los valores en esta matriz fueron asignados de forma aleatoria con tal de tener una red de viajes heterogéneamente distribuidos.

O\D	1	2	3	4	5	6	7	8	9	10	11	12	Total
A	164	220	172	164	164	172	172	172	164	184	164	172	2084
B	184	82	172	164	164	164	172	180	184	164	82	123	1835
C	164	180	172	172	180	164	164	172	164	172	172	186	2062
D	184	82	172	164	180	180	184	172	186	164	180	164	2013
E	180	164	184	164	184	164	164	184	172	172	180	180	2093
F	184	164	172	164	164	164	184	143	164	180	172	180	2036
G	164	186	201	201	186	166	164	164	186	172	184	172	2146
H	201	184	172	186	164	184	102	186	172	164	184	164	2063
I	172	164	164	61	172	186	82	61	172	186	164	82	1667
J	186	186	164	164	184	164	164	172	186	180	164	164	2079
K	201	186	164	186	164	184	184	123	82	164	172	180	1991
L	172	164	172	184	186	172	164	172	186	164	184	164	2085
Total	2157	1962	2081	1974	2093	2064	1901	1903	2019	2066	2002	1931	24153

Figura 4.5. Matriz OD de flujos en vehículos por hora para la red vial

La elección de ruta de los vehículos dentro de la red se rige por un modelo Logit Multinomial. La Ecuación 4.1 muestra la probabilidad de elección de la ruta k dentro del universo de rutas disponibles entre el origen y destino del vehículo (P_k). La variable V_i es el negativo del costo de la ruta i en horas y θ el factor de escala. Este último fue ajustado a 4,2 con el fin de obtener una distribución de vehículos razonable entre dos rutas de tiempos de viaje diferentes. Para encontrar este valor se procedió a definir que, si existe una diferencia de 2 minutos ó más entre dos rutas diferentes, entonces, el 80 % de los vehículos se irá por la ruta más rápida. Si la diferencia es de 30 minutos entonces el 98 % se irá por la más rápida. La probabilidad representada en la Ecuación 4.1 se actualiza en cada segundo de la simulación para vehículo nuevos en función de la congestión que se observe en la red.

$$P_k = \frac{1}{1 + \sum_{l \neq k} e^{(V_l - V_k)\theta}} \quad (4.1)$$

4.1.4. Escenarios a evaluar

Los escenarios a evaluar están determinados por tres características:

- (i) El “número y distribución de SB” se analiza con cuatro configuraciones diferentes: una heterogénea de 18 SB, una uniforme de 18 SB, una de 27 SB y una de 36 SB (el máximo posible). El detalle de cada distribución de sensores será descrito más adelante en esta misma Sección. La simbología utilizada para cada una de las configuraciones mencionadas es: 18h, 18u, 27 y 36, respectivamente.
- (ii) El “tiempo de interdetección” del SB se define como el tiempo que existe entre dos instancias de detección, es decir, cada cuanto tiempo el sensor busca dispositivos dentro de su zona de detección. En la práctica el tiempo de interdetección depende del dispositivo, pero su promedio ronda los 5 segundos (Franssens, 2010). Dado

esto, en este trabajo se experimenta con 1, 5 y 8 segundos.

- (iii) El “modelo *a priori*” busca determinar la probabilidad de elección de ruta *a priori* definida en la Sección 3.3.1. En este caso, se experimenta con dos modelos diferentes. El primero y más simple es el modelo equiprobable (se denota con la letra *E*) que asigna la misma probabilidad a cada una de las rutas previo a la aplicación de la inferencia Bayesiana. El segundo es el modelo de conteos (se denota con la letra *C*) que contabiliza vehículos en cada ruta para obtener una aproximación de la probabilidad de uso (modelo definido en la Sección 3.3.1).

Dadas estas características, se evalúan un total de 24 escenarios diferentes. La Tabla 4.1 muestra los escenarios que son posibles de generar y asigna un identificador para cada uno. Este se construye en base a la simbología señalada en cada características. 1, 5 y 8 para el tiempo de interdetección; *E* y *C* para el modelo *a priori*; 18*u*, 18*h*, 27 y 36 para el número y distribución de los SB. El identificador del escenario agrupa estos valores y los separa por “:”. Por ejemplo en un escenario de 27 SB, con 5 segundos de interdetección y un modelo *a priori* equiprobable su identificador es “27:5:E”.

Cuadro 4.1. Resumen de escenarios en función de los segundos de interdetección, modelo de probabilidad *a priori* y cantidad de sensores Bluetooth

Nº SB\Interdetección (s)	Modelo de probabilidad <i>a priori</i>					
	Conteos			Equiprobable		
	1	5	8	1	5	8
18 <i>h</i>	18h:1:C	18h:5:C	18h:8:C	18h:1:E	18h:5:E	18 h:8:E
18 <i>u</i>	18u:1:C	18u:5:C	18u:8:C	18u:1:E	18u:5:E	18u:8:E
27	27:1:C	27:5:C	27:8:C	27:1:E	27:5:E	27:8:E
36	36:1:C	36:5:C	36:8:C	36:1:E	36:5:E	36:8:E

La Tabla 4.1 muestra un total de cuatro distribuciones diferentes para posicionar los SB en la red. Todas estas configuraciones consideran SB en el borde de la red, ya que de lo contrario se estaría reduciendo la zona de estudio y no la densidad de SB (*i.e.* los bordes de la red son definidos por los SB).

La primera de ellas, $18h$ corresponde a una configuración de 18 SB dispuestos de forma aleatoria y heterogénea dentro de la red. En la Figura 4.6 es posible notar los SB instalados en las intersecciones de la red y se observa una mayor densidad de estos en la zona superior derecha, en contraste con la inferior izquierda. Este tipo de configuración permite medir la calidad de la predicción de la metodología cuando se tienen sectores de la red con baja visibilidad de dispositivos BT.

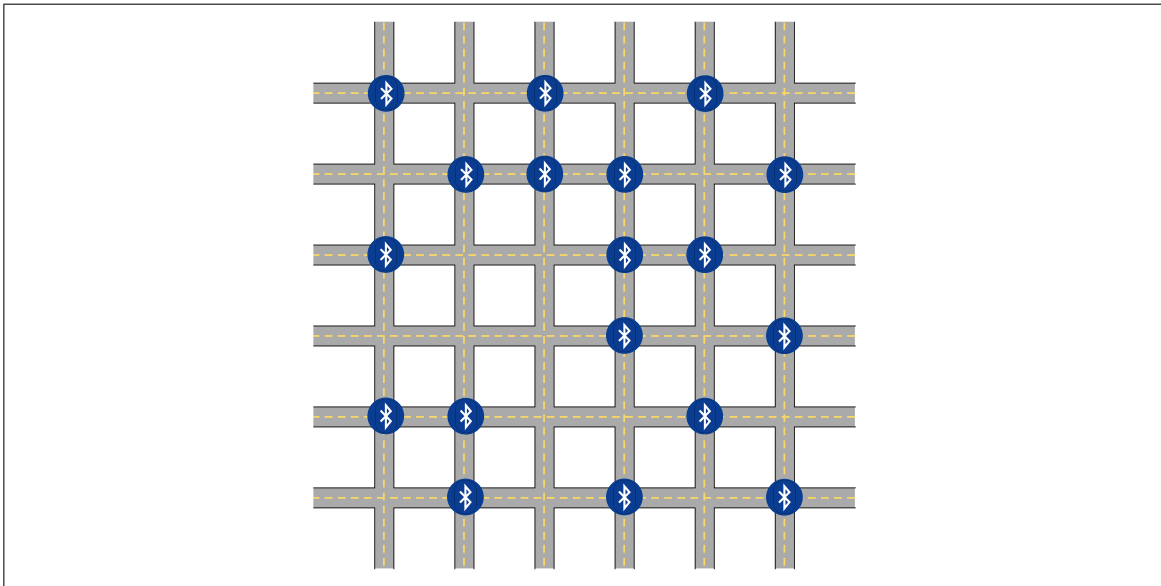


Figura 4.6. Distribución de SB en configuración $18h$

La distribución $18u$, corresponde a una distribución uniforme de 18 sensores instalados en las intersecciones de la red. Esta distribución obedece a un damero (*i.e.* distribución intercalada como tablero de ajedrez) que se puede apreciar en la Figura 4.7.

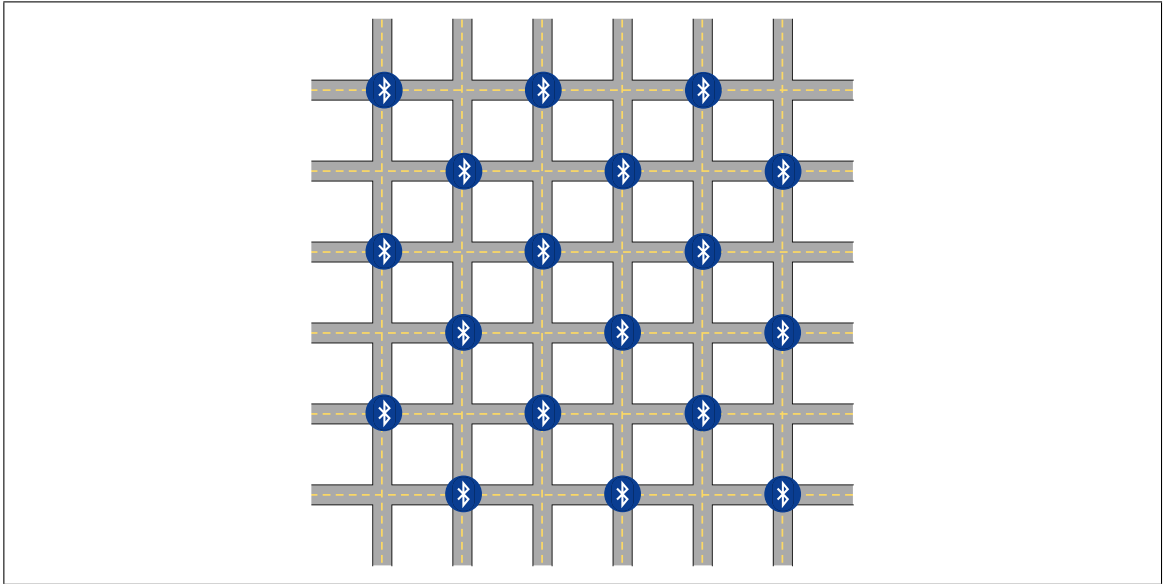


Figura 4.7. Distribución de SB en configuración $18u$

La distribución 27 también considera una distribución aleatoria (como en $18h$) pero con un total de 27 SB. Dada la alta densidad de SB, es difícil observar heterogeneidad en la configuración. La Figura 4.8 muestra la distribución de sensores utilizada por la configuración 27.

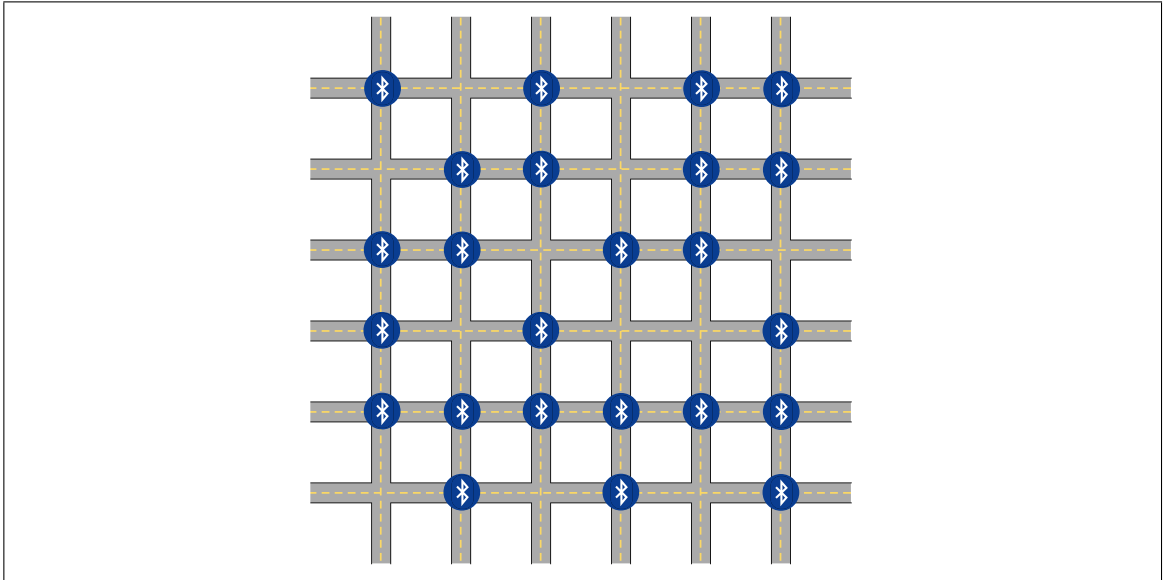


Figura 4.8. Distribución de SB en configuración 27

La última configuración considerada corresponde a la distribución 36, donde todas las intersecciones cuentan con un SB. La Figura 4.9 muestra la configuración única de los 36 sensores.

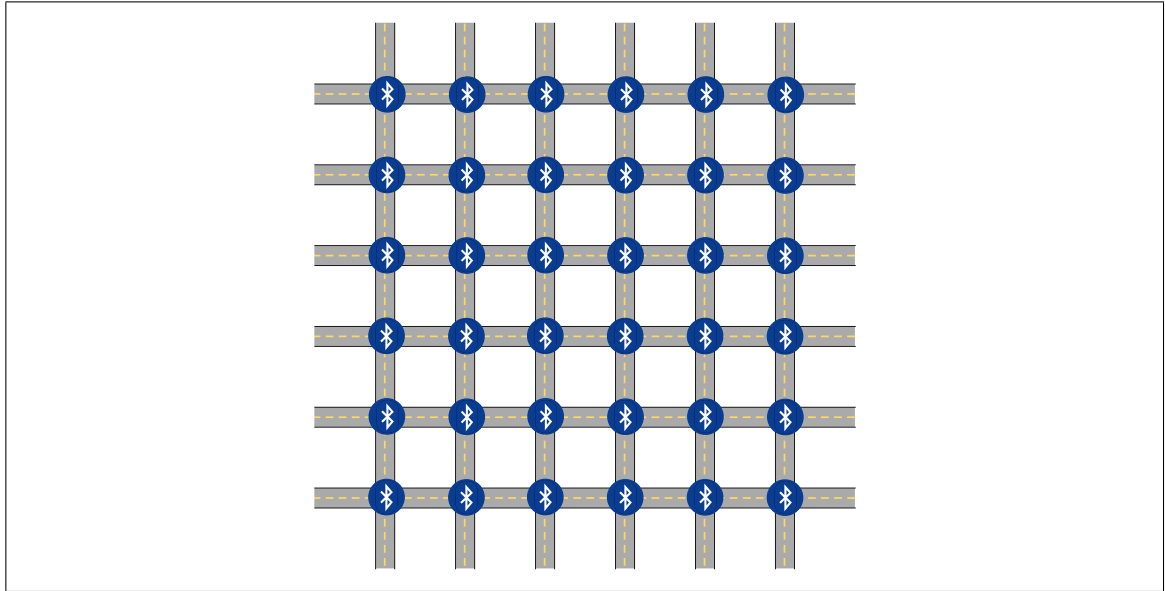


Figura 4.9. Distribución de SB en configuración 36

Además de los escenarios descritos, se pretende comparar la metodología con uno de los modelos de reconstrucción de rutas presentados por Michau *et al.* (2014) y Michau *et al.* (2017). Los que serán descritos en el Capítulo 5.

4.2. Medidas de desempeño

Con el fin de analizar el rendimiento de la metodología propuesta en cada uno de los escenarios, se propone una serie de medidas de desempeño. Estas permiten comparar los distintos escenarios y evaluar cómo inciden los cambios sobre los resultados. En esta sección se describen las medidas a utilizar.

4.2.1. Análisis de ajuste predictivo

Para evaluar la calidad de la predicción de la metodología, en cada escenario se calcula el porcentaje correctamente predicho (*First Preference Recovery*, FPR) (Ortúzar y

Willumsen, 2011). Este indicador es una medida agregada que contabiliza la proporción de éxitos que se producen luego de la asignación del modelo. Se define éxito si la alternativa elegida presenta la mayor probabilidad predicha de elección.

Como se explicó en el Capítulo 3, el modelo realiza una inferencia Bayesiana entre cada par de detecciones sucesivas, y luego estas se concatenan lo largo del tiempo para conocer las probabilidades de la ruta completa. Por esto se determinan dos indicadores FPR, uno para contabilizar el total de éxitos de inferencias Bayesianas entre dos detecciones sucesivas (se llamarán “secciones”) y otro para contabilizar el total de éxitos en la predicción de la ruta completa del vehículo (se llamarán “viajes”).

Con el fin de comparar el modelo respecto a una asignación aleatoria de rutas, se determina el indicador de proporción de recuperación de preferencias al azar (*Chance Recovery*, CR). Este indicador contabiliza los aciertos de la ruta correcta frente a una elección aleatoria dentro de las alternativas (*i.e.* al elegir al azar una ruta del conjunto de alternativas, se considera como acierto si la ruta elegida es la ruta que utilizó el vehículo). A modo de ejemplo, si se tienen 3 rutas posibles, se escoge una ruta aleatoria, si esta coincide con la elegida por el vehículo es un acierto, de lo contrario es un fracaso. Se espera que el indicador CR sea menor que el FPR, ya que el modelo propuesto no es superior a una elección al azar.

Finalmente, se calcula un tercer indicador que corresponde al promedio de las probabilidades asignadas a la ruta elegida por cada vehículo. A diferencia del FPR, que mide la cantidad de éxitos, este indicador mide, en cierta forma, la calidad de dichos éxitos. Esto, porque no es lo mismo afirmar que un vehículo utilizó cierta ruta con un 30 % de probabilidad máxima versus un 90 %. Dado que este indicador mide la calidad de la predicción, se reconocerá con la sigla PQ (*Prediction Quality*).

Estos tres indicadores pueden ser utilizados tanto para secciones como para viajes completos y se determinan para cada escenario modelado.

4.2.2. Análisis de flujos obtenidos

Para analizar la calidad del modelo desde un punto de vista agregado, se analizan los flujos a nivel de arcos de la red. Dado que la metodología es capaz de entregar probabilidades de elección entre varias rutas para cada vehículo, se asignará esa probabilidad a cada arco que compone la ruta. De este modo, cada vehículo aportará flujo a todas las rutas en las que se le asignó probabilidad.

Para comparar los modelos mediante este indicador, se construyen gráficos que relacionan los flujos reales de vehículos equipados con BT con los inferidos por la metodología. Además, se determinarán indicadores de medidas de error. Estos son el Error Cuadrático Medio (MSE) y el Error Absoluto Ponderado Porcentual Medio (WMAPE).

El MSE mide el promedio del cuadrado de las diferencias entre flujos reales e inferidos. En él se da mayor importancia a los errores más grandes, pero solo permite comparaciones cuando el total del flujo a comparar es igual. Esto último se explica dado que si dos escenarios a comparar tienen distintos flujos totales, las diferencias de estos con el valor real en cada uno, serán proporcionales a la magnitud de estas diferencias, por lo que la comparación sería poco informativa.

El WMAPE mide el promedio de las diferencias porcentuales absolutas entre los flujos reales e inferidos, ponderadas por el flujo real que se busca predecir. Así, el WMAPE corresponde al MAPE (Error Porcentual Absoluto Medio) con una modificación que penaliza más a los errores que se cometen en arcos con alto nivel de flujo.

Además de estos dos indicadores numéricos, se presentan los errores porcentuales mínimos, máximos y su desviación estándar. Las expresiones matemáticas de los errores descritos se presentan en el Anexo A.

4.2.3. Análisis de predicción dependiendo de la velocidad

Entre cada par de detecciones sucesivas se conocen los tiempos de pasada por cada sensor, por lo que se puede determinar el tiempo de viaje entre ellos. Además, como se conoce la ruta utilizada por cada vehículo, es posible determinar una velocidad media de circulación entre ambas detecciones. Como entre cada par de detecciones sucesivas se realiza una inferencia Bayesiana (si es que existe más de una alternativa) se puede conocer el número de éxitos que se producen a distintas velocidades de circulación. Este análisis consta de construir un histograma del porcentaje de éxitos que se producen en distintos umbrales de velocidad (FPR en diferentes rangos de velocidad). Con este análisis se espera obtener conclusiones respecto a como afecta la velocidad de circulación de los vehículos en la predicción de su ruta.

CAPÍTULO 5. RESULTADOS

En este Capítulo se presentan los resultados obtenidos de la aplicación de la metodología en los diferentes escenarios simulados que fueron descritos en el Capítulo 4. Los resultados son evaluados con los indicadores descritos en el capítulo anterior, además de ser comparados entre sí. Finalmente se realiza una comparación adicional con uno de los modelos de reconstrucción de rutas presentados por Michau *et al.* (2014) y Michau *et al.* (2017).

5.1. Comparaciones entre escenarios

En esta sección se comparan desde una perspectiva general, diferentes características observadas en cada uno de los 24 escenarios. El objetivo es identificar el impacto que tiene sobre la predicción el número, distribución tiempo de interdetección de SB y las distintas formas de obtener una probabilidad a priori de uso de rutas. Para ello se evalúa la calidad de la predicción realizada en cada modelo a nivel de secciones, viajes y estimación de flujos.

5.1.1. Estimación de flujos por escenario

Este análisis busca evaluar la estimación de flujos asignados a los arcos de la red. En este caso, se utiliza el indicador de error WMAPE. Este representa la diferencias que existen entre los flujos de vehículos con BT y los estimados por la metodología en base a las detecciones. La Figura 5.1 muestra el indicador WMAPE para la estimación de arcos en los 24 escenarios. El color de cada burbuja diferencia el modelo utilizado y el tamaño corresponde a la magnitud del error WMAPE.

En la Figura 5.1 es claro notar que al aumentar el número de SB en la red (de izquierda a derecha en el gráfico) disminuye el error de predicción. Además, en vista general se

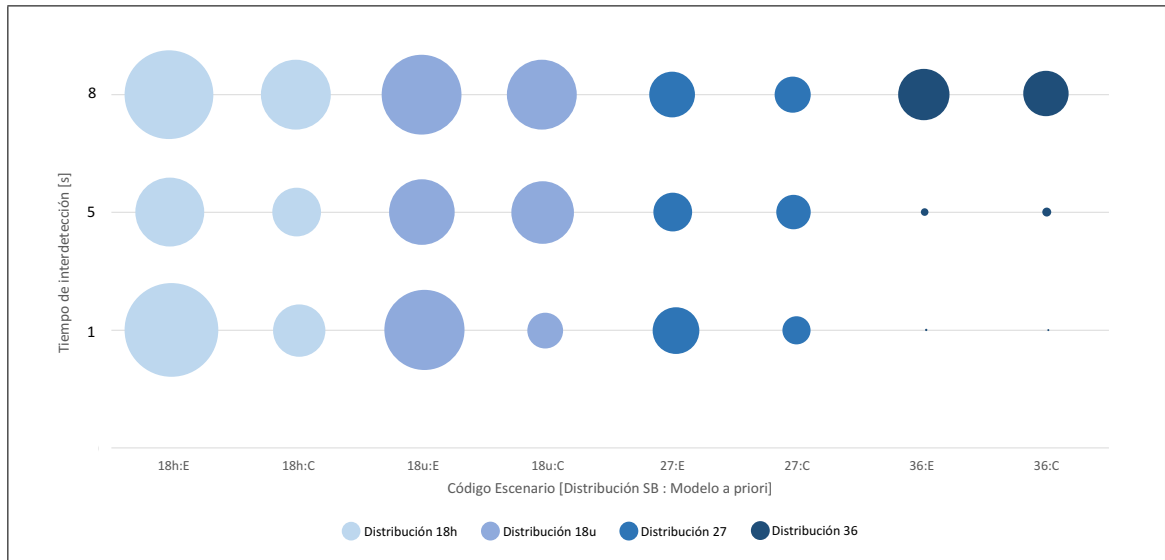


Figura 5.1. Indicador de error WMAPE determinado para todos los escenarios

observa que el modelo de conteos presenta un menor error que el modelo equiprobable. Respecto al tiempo de interdetección, no se observa una tendencia muy clara, pero entre más SB se encuentren instalados, más significativa se hace la disminución del tiempo de interdetección. En general, los resultados indican que a mayor cantidad de datos se tendrá un mayor calidad en la predicción. En este caso, a menor tiempo de interdetección, mayor número de SB se tienen mejores predicciones, al registrar una mayor cantidad de datos.

En específico, el modelo equiprobable muestra cierta inestabilidad, dado que no hay un patrón claro de cómo afecta el tiempo de interdetección en la medida del error. Además, el modelo equiprobable presenta una minimización del error a los 5 segundos de interdetección, lo que se puede explicar porque este tiempo es cercano al tiempo que varios vehículos recorren la red, disminuyendo el número de detecciones y por ende disminuyendo las alternativas de estos. De igual modo es necesario evaluar más tiempos de interdetección, en la vecindad de los 5 segundos para dar una afirmación más robusta.

Al comparar la magnitud del error WMAPE entre modelos de asignación equiprobable y conteos de la base de datos, es posible notar que entre menos sensores tenga la red, más importante se hace el uso de un modelo más sofisticado que una asignación equiprobable.

5.1.2. Impactos del uso de los modelo *a priori*

Se realiza una comparación entre modelos de igual demanda con el indicador error cuadrático medio (MSE) en la estimación de flujos con el fin de evaluar en detalle diferencias en el uso de modelos *a priori*. Este error permite comparar de forma justa escenarios de igual demanda y ver el efecto aislado del modelo *a priori* utilizado. La Figura 5.2 muestra el MSE para cada uno de los escenarios contruidos. En el eje horizontal se especifican los escenarios y en el eje vertical primario se indica el valor que toma el indicador. Cada barra corresponde a uno de los dos modelos *a priori* testeados en esta investigación. Cabe destacar que el MSE solo es comparable en cada par de columnas (escenarios con igual demanda) y no es válida una comparación entre dos grupos de columnas si es que no se toma en cuenta la demanda (vehículos inferenciados en Figura 5.5) de cada escenario. Por esta razón, se incluye en el gráfico una línea con la diferencia porcentual del modelo Equiprobable respecto del modelo de Conteo, que permite conocer la diferencia de esta variable entre modelos.

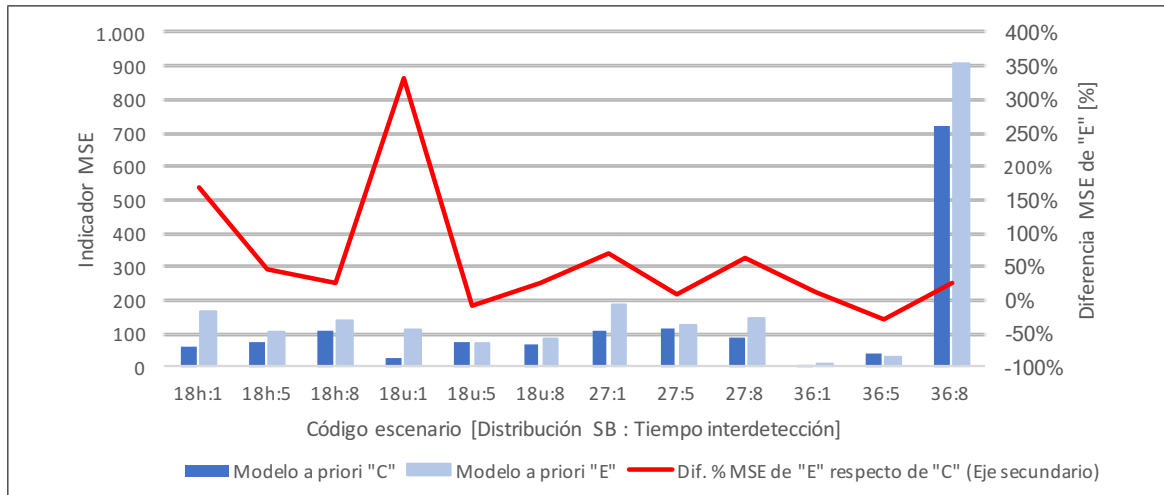


Figura 5.2. Error cuadrático medio para comparar modelos *a priori* entre escenarios de igual demanda

Consistentemente con el primer análisis, al observar cada par de columnas es posible notar que en la mayoría de los casos el MSE es mayor cuando se utiliza el modelo equiprobable. Este hecho justifica el desarrollo de metodologías más sofisticadas que permitan determinar una probabilidad *a priori* más efectiva. Solo en tres casos se obtuvo el resultado opuesto (donde el gráfico de líneas está por debajo del 0 %). El escenario que se vio más favorecido por el modelo de conteos fue el 18u de 1 segundo de interdetección, donde el modelo equiprobable presenta cerca de un 350 % más error. Por otro lado, el escenario 36 se vio perjudicado dado que el equiprobable tiene aproximadamente 35 % menos de error que el modelo de conteos.

5.1.3. Capacidad de predicción en la reconstrucción de rutas

La capacidad predictiva del modelo se puede medir a nivel de la predicción realizada luego de cada inferencia (secciones) o a luego de predecir la ruta completa (viajes). Es por esto que en esta sección se evalúa la capacidad de la metodología en secciones y viajes con la ayuda del indicador FPR. La Figura 5.3 presenta, para los 24 escenarios, el FPR en

secciones y la Figura 5.4 FPR para viajes. Cada grupo de barras muestra el valor de FPR para los modelos *a priori* evaluados en cada escenario.

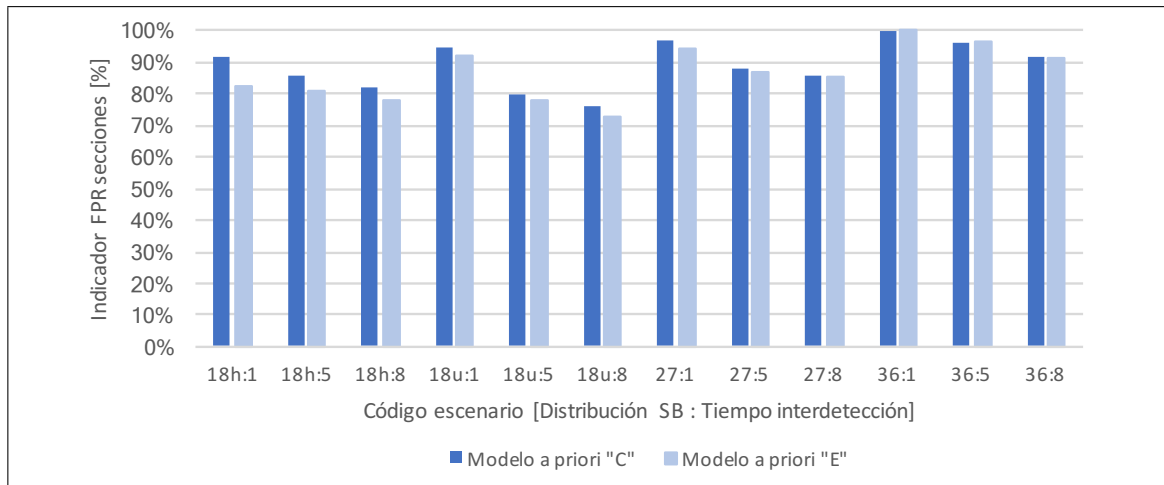


Figura 5.3. Valores del indicador FPR analizado por secciones

En las Figuras 5.3 y 5.4 se observa que cuando existe una alta densidad de SB, las diferencias de FPR son menores a cuando existe una baja densidad. Esto quiere decir que a medida que baja la densidad de SB de la red, y por lo tanto aumenta el número de rutas posibles entre SB, el modelo equiprobable pierde efectividad. De esto se desprende que cuando existe un mayor número de opciones de ruta, el método de conteos es capaz de predecir mejor las secciones utilizadas y viajes.

Respecto a la predicción de rutas completas, la Figura 5.4 muestra un comportamiento bastante similar al mencionado en secciones. Adicionalmente el FPR disminuye para ambos modelos al aumentar el tiempo de interdetección. Este hecho radica en que a mayor tiempo de interdetección, más distantes (en el espacio) son las detecciones, por lo que se generan más rutas posibles lo que se dificulta el problema. Pese a lo mencionado, no se observan grandes diferencias en el valor de FPR frente a los distintos modelos *a priori* utilizados.

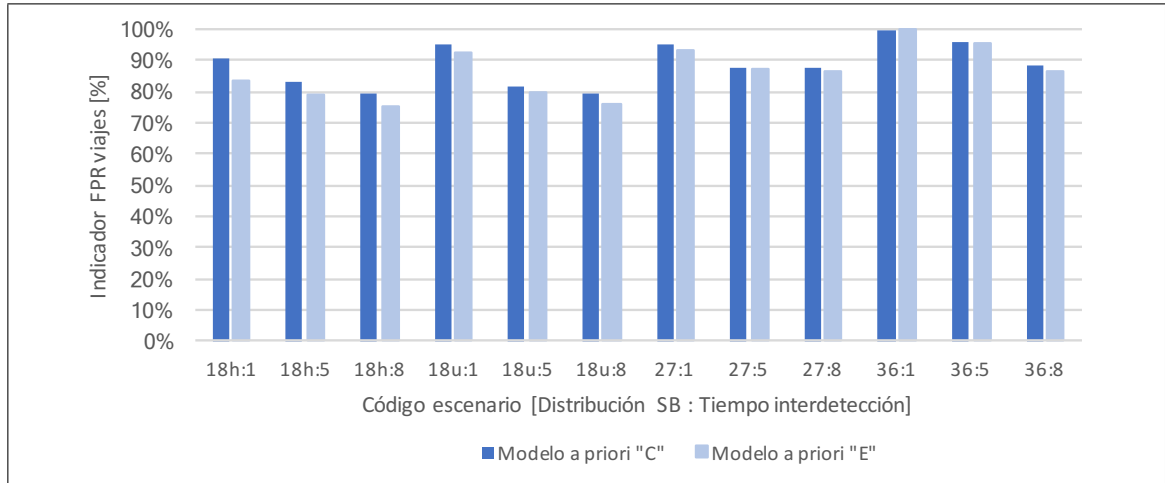


Figura 5.4. Valores del indicador FPR analizado por viajes

5.1.4. Cantidad de detecciones producidas en cada escenario

La cantidad de vehículos que son detectados solo depende de dos factores: número y distribución de SB y el tiempo de interdetección. En la Figura 5.5, la línea horizontal en 6.500 vehículos corresponde al universo de vehículos equipados con BT, equivalente al 40 % del total de vehículos que ingresan a la red. De estos, la barra de la izquierda de cada grupo corresponde la cantidad de vehículos que son detectados (del total de vehículos BT) al menos una vez en cada escenario, la barra de la derecha de cada grupo la cantidad de vehículos aptos para ser inferenciados (*i.e.* poseen al menos un par detecciones consecutivas con más de una opción de ruta). Dado que estos últimos son una porción de los vehículos detectados, el gráfico de líneas muestra esta proporción. En el eje horizontal se identifican los código de los escenarios sin especificar el modelo *a priori* utilizado, dado que este no afecta la cantidad de detecciones realizadas.

La Figura 5.5 verifica que a medida que aumenta el tiempo de interdetección para cada configuración de SB, disminuye el total de vehículos que son detectados al menos una vez. Por otro lado, se observa que esta cantidad de detección mantiene el orden de magnitud entre configuración y cantidad de SB. Esto deja claro que la variable que afecta al número

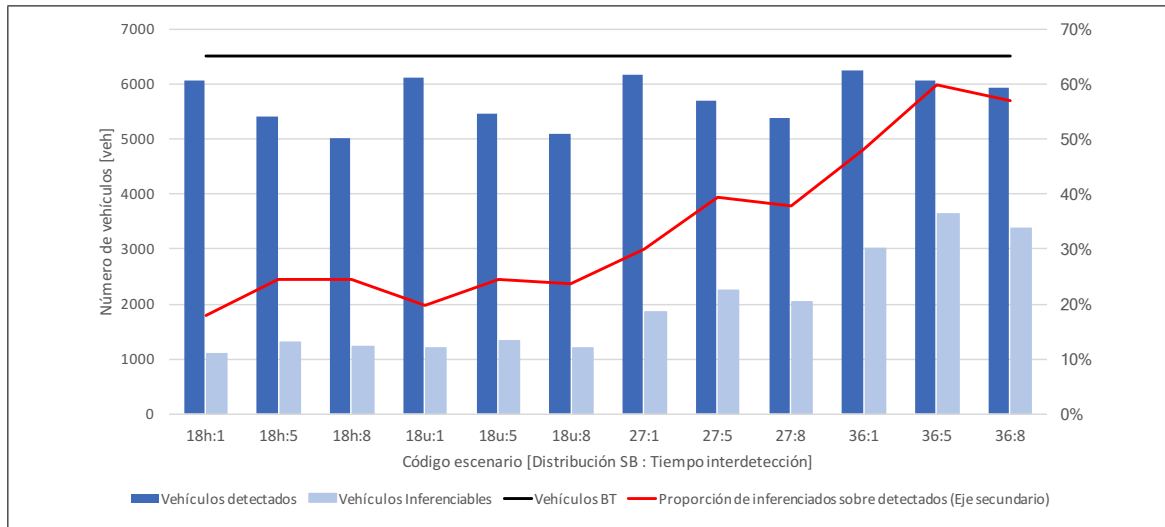


Figura 5.5. Total de detecciones producidas por escenario en función de las características de cada uno

de detecciones totales viene dada por el tiempo de interdetección del SB y no en gran medida por la cantidad ni configuración de SB utilizados, que produce un leve aumento en las detecciones a medida que se instalan más SB. Este comportamiento es lógico y consistente dado que al ocurrir mayor número de detecciones para un intervalo de tiempo fijo, entonces, habrán más detecciones.

La cantidad de vehículos en los que es posible realizar una inferencia Bayesiana (*i.e.* barras de la derecha de cada grupo), posee un comportamiento que en primera instancia parece poco intuitivo, ya que, presenta un *peak* a los 5 segundos de interdetección y no en alguno de los extremos (1 u 8 segundos). La hipótesis para explicar este comportamiento es que a medida que el tiempo de interdetección aumenta tiende a generarse una mayor brecha de distancia y tiempo entre detecciones sucesivas de cada vehículo, lo que explicaría el aumento de vehículos inferenciados al pasar de 1 a 5 segundos de interdetección. Pero, por otro lado, al aumentar el tiempo de interdetección también aumenta la probabilidad que se realicen menos detecciones en un paso del vehículo por la red (sumado a la probabilidad de detección). Además, este hecho provoca un aumento de la cantidad de

vehículos que son detectados solo una vez. Por lo tanto, la relación que existe entre el tamaño de la red y el tiempo de interdetección juega un rol fundamental, estipulando que el tiempo de interdetección debiera estar muy por debajo del tiempo mínimo de viaje en alguna de las posibles rutas de la red. Esto quiere decir que si la red fuera más grande, con 8 segundos de interdetección se esperaría una mayor cantidad de vehículos inferenciables. El radio de detección del SB también juega un rol importante, dado que al ser más grande, los vehículos estarán más tiempo en la zona de detección del SB.

5.2. Resultados de escenarios evaluados

En esta Sección solo se presentan los resultados de cinco de los 24 escenarios de la Tabla 4.1, el resto se encuentra disponible en el Anexo C. Los escenarios seleccionados para ser presentados en esta sección entregan resultados concluyentes respecto a cambios en las características utilizadas para la construcción de escenarios (número y distribución de antenas, tiempo de interdetección y modelo *a priori*). En estos resultados se busca observar en detalle cambios en solo una de estas variables y ver si son consistentes con los resultados obtenidos de forma agregada.

Se toma como base el escenario más favorable en términos de inferencia de ruta, es decir, el que posee 36 SB, un segundo de interdetección y el modelo de conteos para obtener la probabilidad *a priori* (36:1:C).

Luego se sensibiliza el tiempo de interdetección y se cambia este valor de 1 segundo a 8 segundos (máximo valor dado a esta variable en esta investigación). De este modo se analiza un escenario (36:8:C) que evalúa los efectos del tiempo de interdetección.

A partir del escenario base, se sensibiliza el número de SB (y su distribución) y se cambia la configuración 36 por las 18 h y 18 u , que poseen el menor número de SB evaluados en esta investigación. En estos escenarios (18 h :1:C y 18 u :1:C) solo cambia la configuración

y número de SB respecto al escenario base (36:1:C), por lo que se espera obtener conclusiones respecto de los efectos de esta característica en la inferencia de rutas. Además, pueden analizarse los efectos de una distribución homogénea y heterogénea con el mismo número de SB.

Finalmente, para evaluar los efectos de la probabilidad *a priori* de uso de rutas sobre la inferencia de ruta completa, se modifica respecto al escenario base, el modelo *a priori* de conteos al equiprobable (36:1:E).

A continuación se detallan los resultados en cada uno de los modelos descritos.

5.2.1. Escenario 36:1:C

Los resultados del análisis predictivo se presenta en la Tabla 5.1. De los 6.500 vehículos BT, 6.235 son vehículos detectados (*i.e.* detectados al menos una vez). Dadas las detecciones generadas se realizaron 3.736 inferencias a nivel de sección y 3.200 a nivel de viaje, lo que genera un promedio de 1,24 inferencias por viaje.

En ambos casos (secciones y viajes), el indicador FPR supera el 99 %, lo que significa que se predijo correctamente casi la totalidad de las rutas de vehículos inferenciables. Este valor es muy superior al valor CR que ronda el 30 % en ambos análisis (secciones y viajes). Respecto al índice PQ, que mide la calidad de los aciertos reconocidos por el indicador FPR, casi en la totalidad de las inferencias la metodología asignó probabilidades cercanas a 1 en la ruta que efectivamente utilizó cada vehículo.

Cuadro 5.1. Resultados análisis predictivo Escenario 36:1:C

Indicador	$N_{inferencias}$	FPR	CR	PQ
Secciones	3.736	99,7 %	32,6 %	99,7 %
Viajes	3.002	99,7 %	26,9 %	99,7 %

Dado que el indicador FPR y PQ son extremadamente altos, es de esperar que la asignación de flujos a la red tenga un buen ajuste a la realidad. Esto se comprueba en la Figura 5.6, donde se observan los puntos que relacionan ambos flujos (inferido y real) en la recta $X = Y$ con muy baja dispersión. Esta relación revela que las estimaciones que se pueden realizar de los flujos de vehículos equipados con BT sobre los arcos son prácticamente iguales a la realidad (en un escenario de estas características).

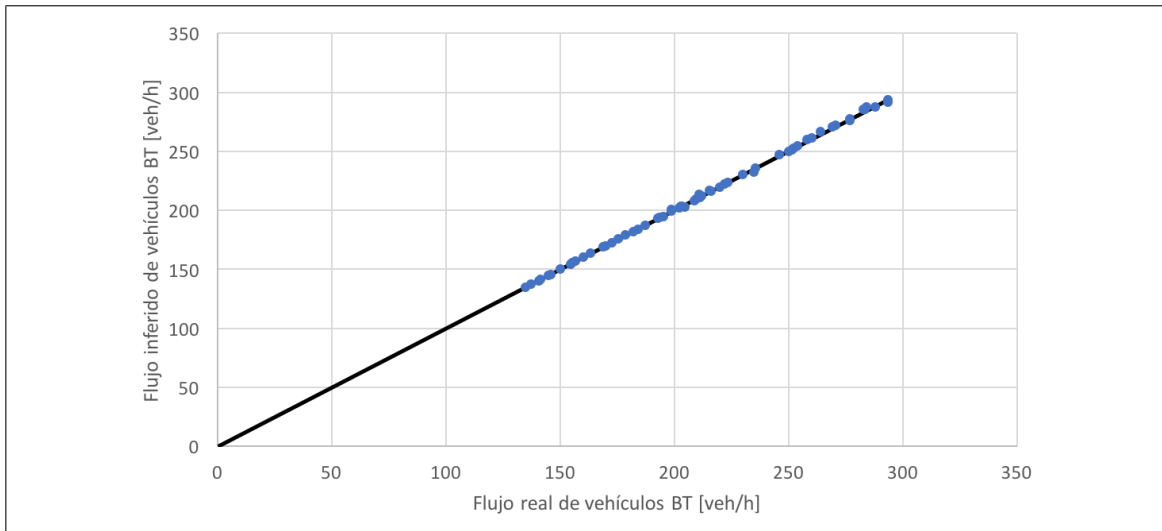


Figura 5.6. Comparación de flujos reales e inferidos para Escenario 36:1:C

Con el fin de cuantificar el ajuste que los puntos muestran en la Figura 5.6, se determinan los indicadores MSE y WMAPE, que se resumen en la Tabla 5.2. Los errores en general son prácticamente nulos, siendo el máximo tan solo 0,03 %. Esto reafirma la calidad de la predicción en este escenario.

Cuadro 5.2. Indicadores de error en flujos inferidos de viajes Escenario 36:1:C

Indicador	MSE	WMAPE	MIN	MAX	DESV_STD
Valor	2,9	0,01 %	0 %	0,03 %	0,01 %

Como último análisis del escenario, se procede a analizar la proporción de éxitos en función de la velocidad media del vehículo al aplicar la metodología. Las barras en la Figura 5.7 corresponden al total de inferencias de predicciones de secciones realizadas en cada intervalo de velocidad. El gráfico de línea muestra la proporción de éxitos para cada intervalo de velocidades. Como los indicadores de predicción son elevados para este escenario, es difícil notar diferencias de predicción a distintas velocidades.

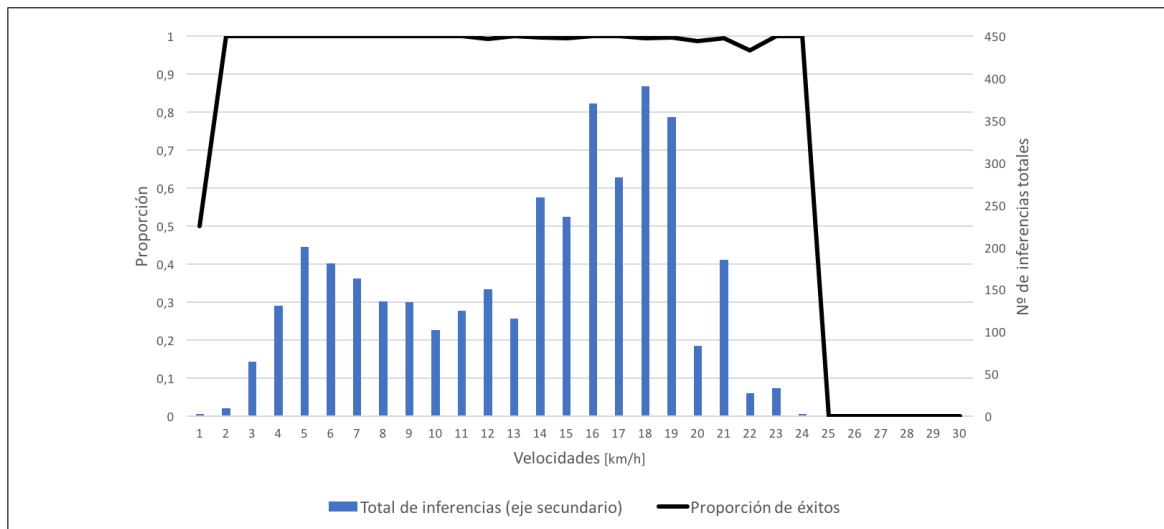


Figura 5.7. Histograma de proporción de aciertos en secciones por velocidades de vehículos para Escenario 36:1:C

Si bien la metodología mostró ser efectiva en este escenario (el más favorable), es importante evaluar cómo cambia la calidad de los resultados al variar el número de sensores o tiempo interdetección, dado que estas variables depende exclusivamente del entorno físico en donde se aplique la metodología. Por otro lado, es importante evaluar cuán importante es el modelo de asignación de probabilidad *a priori* en la calidad de la predicción.

5.2.2. Escenario 36:8:C

Este escenario considera una modificación del tiempo de interdetección del SB respecto al primer escenario presentado reduciéndose de 1 segundo a 8 segundos. En 36:1:C

se detectaron 6.235 vehículos, cifra que disminuye a 5.925 en este escenario. Esta baja es explicada por el tiempo de interdetección, ya que al aumentar, es menos probable detectar vehículos porque el SB registra menos veces. Debido a esto, el número de viajes inferenciados crece de 3.002 a 3.381. Al existir mayor tiempo entre detecciones, más espaciadas son y más rutas es posible construir entre nodos más distantes. Incluso la cantidad de inferencias que se realizan en promedio por viaje aumentan de 1, 24 a 1, 41.

A continuación se presentan los resultados del análisis predictivo del modelo con 8 segundos de interdetección en el SB. La Tabla 5.3 resume los indicadores de predicción.

Cuadro 5.3. Resultados análisis predictivo Escenario 36:8:C

Indicador	$N_{inferencias}$	FPR	CR	PQ
Secciones	4.767	91, 9 %	28, 4 %	92, 4 %
Viajes	3.381	88, 2 %	19, 5 %	89, 0 %

Los indicadores de FPR para secciones y viajes disminuyeron en 10 puntos porcentuales aproximadamente respecto al escenario 36:1:C. El indicador CR de secciones varía en menor cantidad que el de viajes. Aún así, el indicador FPR se encuentra muy por encima del valor de CR, lo que justifica aplicar una metodología para buscar una mejor probabilidad *a priori*. La calidad de la predicción también cayó 10 puntos porcentuales aproximadamente respecto al primer escenario.

El análisis de los flujos muestra claramente una mayor dispersión de los puntos en el gráficos de la Figura 5.8 respecto al del escenario 36:1:C. Además, se muestra una tendencia a sobrestimar los flujos reales. Esto quiere decir que cuando la metodología comete un error en la inferencia de la ruta, asigna el viaje a una ruta más larga que la real. De este modo, habrá más flujo distribuido por la red, lo que genera la sobrestimación observada.

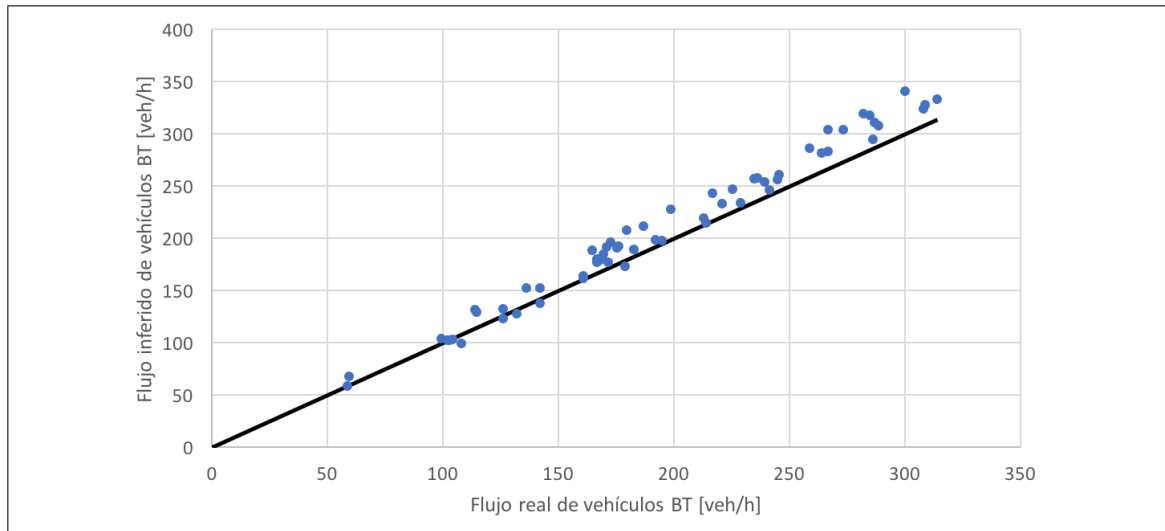


Figura 5.8. Comparación de flujos reales e inferidos para Escenario 36:8:C

La Tabla 5.4 resume los indicadores de error determinados para la estimación de flujos en arcos. El indicador WMAPE es superior al obtenido en el escenario 36:1:C (0,12 % versus 0,01 %). Si bien el valor del error cuadrático medio es mayor que en el experimento anterior, este valor no es comparable dado que es aplicado sobre demandas diferentes. Aun así, se observa una diferencia de 720 puntos en este error y la diferencia de demanda entre cada experimento es solo de 310 (6.235 y 5.925 en cada experimento). El máximo error reportado es 0,35 %, 0,32 puntos porcentuales mayor que el base.

Cuadro 5.4. Indicadores de error en flujos inferidos de viajes Escenario 36:8:C

Indicador	MSE	WMAPE	MIN	MAX	DESV_STD
Valor	723,38	0,12 %	0 %	0,35 %	0,09 %

La Figura 5.9 muestra que las velocidades de circulación media de los viajes inferidos no superan los 15 km/h, valor inferior al registrado en el primer escenario (Figura 5.7). Esto se explica porque las rutas entre detecciones sucesivas son más largas y por ende tienen

mayores probabilidades de detención, ya sea por los semáforos instalados o producto de la congestión. Al observar la línea negra de la proporción de éxitos, es posible notar como esta es inferior con velocidades superiores a los 11km/h . La baja en la proporción de éxitos se puede explicar dado que los vehículos más rápidos tienen una menor probabilidad de ser detectados, por ende las detecciones sucesivas tienden a estar más distantes.

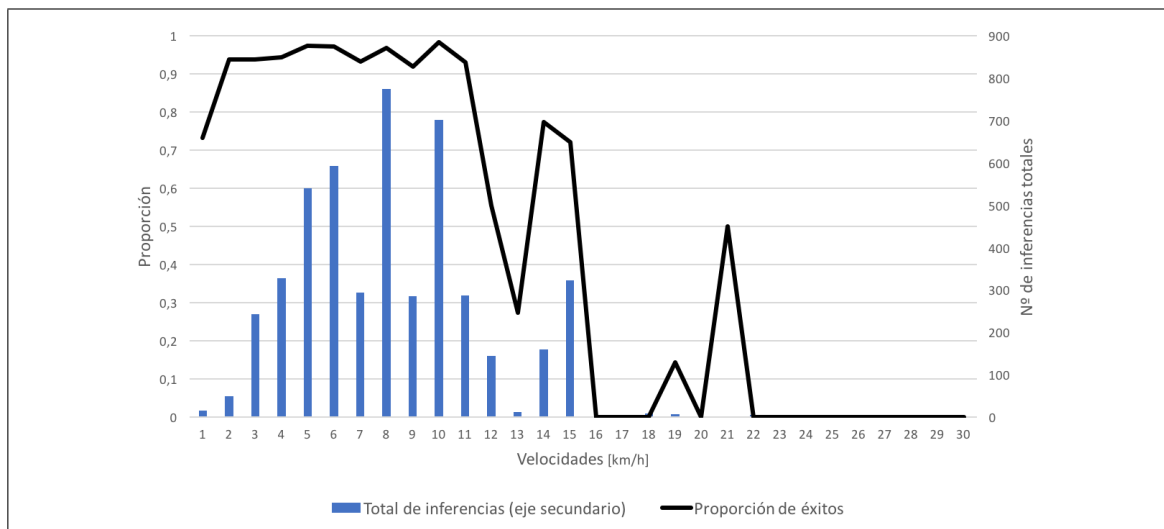


Figura 5.9. Histograma de proporción de aciertos en secciones por velocidades de vehículos para Escenario 36:8:C

En general, el aumentar el tiempo de interdetección del SB produce un menor número de detecciones, por lo que estas se encuentran más espaciadas en el tiempo y en el espacio. Esto genera que se realicen más inferencias por viaje, y sea más difícil inferir la ruta correcta, ya que aumenta el número de alternativas o la longitud de cada una de las rutas a evaluar.

5.2.3. Escenario 18h:1:C

En este tercer experimento se vuelve a utilizar un tiempo de interdetección de 1 segundo, pero se disminuye el número de sensores a la mitad (18 en total). Además, estas son localizadas de forma aleatoria en la red como se muestra en la Figura 4.6.

En este escenario el universo de viajes inferenciables cae a 1.093, muy por debajo de los registros obtenidos en los experimentos anteriores. La cantidad promedio de inferencias por viaje cae a 1,04.

La Tabla 5.5 resume los indicadores de predicción. De la Tabla 5.5 se puede inferir que el efecto de eliminar la mitad de los sensores es similar al de aumentar el tiempo de interdetección de 1 a 8 segundos (en términos del indicador FPR y PQ). El CR ronda el 25 % de predicción en secciones y viajes, lo que revela nuevamente una alta diferencia respecto al FPR (justificando la utilidad de contar con una metodología).

Cuadro 5.5. Resultados análisis predictivo Escenario 18h:1:C

Indicador	$N_{inferencias}$	FPR	CR	PQ
Secciones	1.141	91,4 %	25,7 %	90,3 %
Viajes	1.093	90,5 %	24,9 %	89,2 %

La estimación de flujos de la Figura 5.10 se comporta levemente mejor que el escenario 36:8:C (segundo escenario presentado). Además, en este escenario (18h:1:C) no se observa el fenómeno de sobre estimación de flujos. Una posible hipótesis para explicar lo anterior es que la sobrestimación de flujos la produce la distancia entre detecciones sucesivas (número de nodos sin detectar entre el par de detecciones realizadas) y no por la cantidad de SB.

La Tabla 5.6 muestra los indicadores de error en la estimación de flujos. El WMAPE es 0,02 puntos porcentuales más grande que en el escenario 36:8:C, pese a que en términos relativos esos errores pesan más. En el escenario 36:8:C, casi todos los flujos reales eran mayores que 100 veh/h. Ahora casi todos son menores que 100 veh/h. Pese a que el error cuadrático medio es bastante menor, este valor no es comparable con el entregado en el segundo escenario analizado, porque el nivel de demanda detectada es mucho menor que

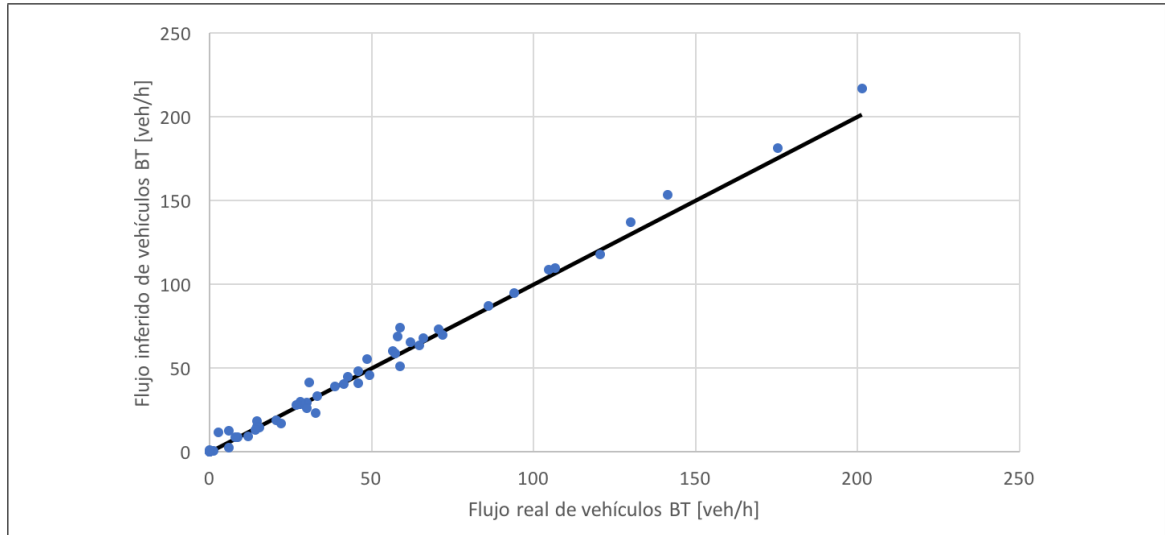


Figura 5.10. Comparación de flujos reales e inferidos para Escenario 18h:1:C

en el otro caso. Por esto, se espera un error de magnitud absoluta menor. De esto se puede inferir que no existen grandes diferencias entre los escenarios 36:8:C y 18h:1:C.

Cuadro 5.6. Indicadores de error en flujos inferidos de viajes Escenario 18h:1:C

Indicador	MSE	WMAPE	MIN	MAX	DESV_STD
Valor	62,79	0,14 %	0 %	0,62 %	0,16 %

El análisis por velocidad presente en la Figura 5.11 revela un alza en la proporción de predicción a medida que aumenta la velocidad de circulación de los viajes inferidos en el intervalo 0 km/h y 8 km/h. Este efecto puede ser explicado por el hecho que los vehículos más lentos son los experimentan mayor congestión y por ende mayor variabilidad en el tiempo de viaje, de este modo, vehículos con velocidad baja son inferidos con datos con mayor variabilidad, lo que afecta la calidad de la predicción. En una red de 18 SB, el viajar entre dos sensores puede significar que en el red real se realizó un trayecto más largo (dado que hay intersecciones sin instalación BT), lo que significa que existe un tiempo de viaje

con mayor variabilidad que un tiempo de viaje de un tramo corto. Esto impacta en la calidad de la predicción, ya que, con estos tiempos el modelo opta a qué ruta asignar el viaje.

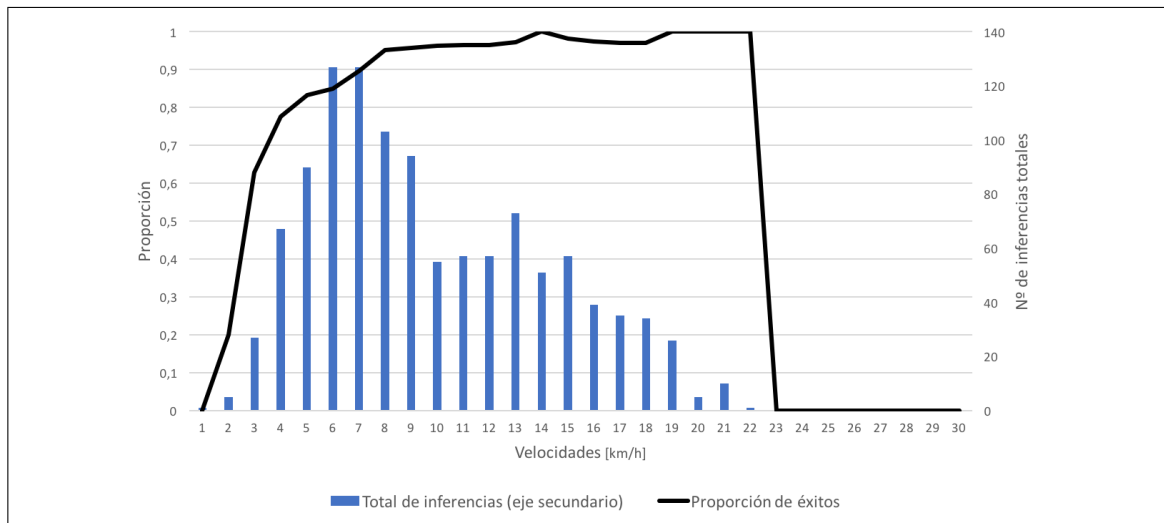


Figura 5.11. Histograma de proporción de aciertos de secciones por velocidades de vehículos para Escenario 18h:1:C

Al comparar el escenario base (36:1:C) con este escenario se desprende que la cantidad de SB y su distribución sí afecta en la calidad de la predicción. Es importante analizar qué variable afecta en mayor medida la calidad de la predicción. Por un lado puede ser la reducción a la mitad en el número de sensores o su distribución aleatoria. Para analizar esta interrogante, el cuarto escenario considera una distribución de 18 sensores más uniforme.

5.2.4. Escenario 18u:1:C

Este experimento replica las mismas características que el escenario 18h:1:C de la Sección 5.2.3 con la diferencia que los 18 sensores se encuentran distribuidos uniformemente a lo largo de red. La Tabla 5.7 muestra los indicadores de predicción para este escenario.

El universo de vehículos es levemente mayor (100 viajes inferidos más aproximadamente) y el indicador de FPR es casi 3 puntos porcentuales mejor que en la distribución aleatoria de SB. Por otro lado, el indicador CR baja en más de 10 puntos porcentuales respecto a la distribución aleatoria. El índice PQ sube 3 puntos respecto a lo registrado en el experimento anterior (18h:1:C)

Cuadro 5.7. Resultados análisis predictivo Escenario 18u:1:C

Indicador	$N_{inferencias}$	FPR	CR	PQ
Secciones	1.240	94,5 %	13,2 %	93,3 %
Viajes	1.211	94,7 %	13,0 %	93,5 %

En este caso se utilizan los indicadores de error descritos en la Tabla 5.8 para comparar las estimaciones de flujo dada la gran similitud en la predicción. Pero se puede ver que el flujo real de BT es bastante menor.

Al comparar el escenario con el 36:1:C, el FPR cae 8 puntos porcentuales. Es decir, al eliminar la mitad de los sensores, se infiere correctamente un menor porcentaje de rutas. Esto puede estar relacionado con la cantidad de datos disponibles para calibrar distribuciones de tiempo. Es decir, al tener mayor cantidad de SB, más registros se cuentan, por lo que existe una mayor cantidad de datos para calibrar las distribuciones. Esto incide también en la calidad de la predicción, donde el índice PQ cae 9 puntos porcentuales.

Cuadro 5.8. Indicadores de error en flujos inferidos de viajes Escenario 18u:1:C

Indicador	MSE	WMAPE	MIN	MAX	DESV_STD
Valor	25,62	0,10 %	0 %	0,41 %	0,09 %

Respecto al análisis de predicción por velocidades, los valores de la Figura 5.12 se observan similares entre los dos últimos escenarios. Es posible notar que existe una mejor predicción para los vehículos lentos que en el escenario 18h, lo que se puede explicar por el mismo fenómeno: al encontrarse los SB uniformemente espaciados, ya no se forman zonas ciegas de la red (con baja densidad de SB), lo que reduce las distancias entre detecciones, la variabilidad y por ende aumenta el porcentaje de predicción.

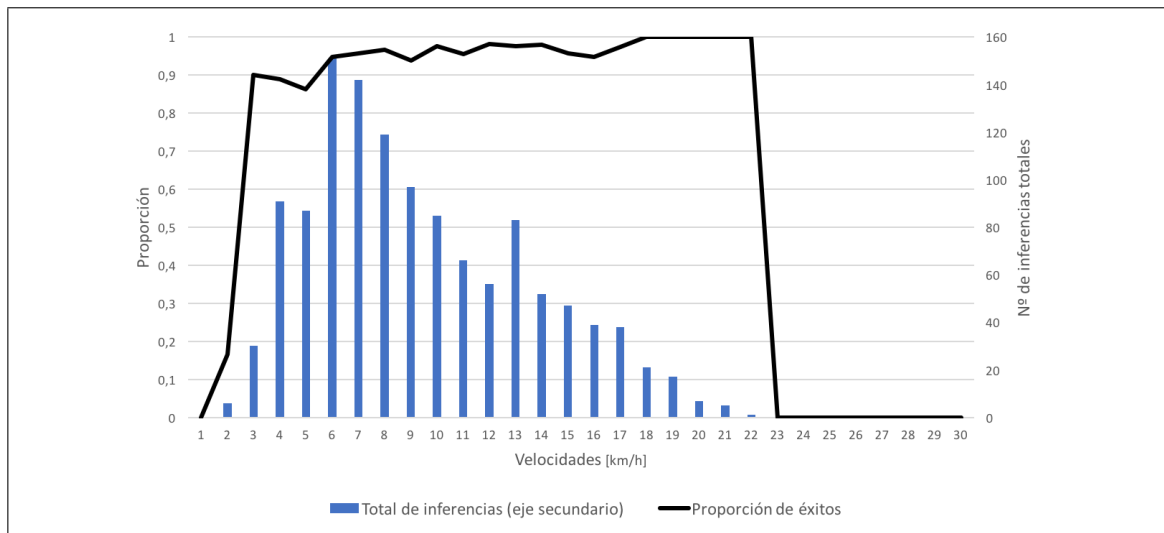


Figura 5.12. Histograma de proporción de aciertos de secciones por velocidades de vehículos para Escenario 18u:1:C

Este escenario confirma que no solo importa el número de SB, sino que también su distribución. En este caso se utilizó una distribución uniforme de damero, pero quizás, si se conoce la demanda de una red, se pueda utilizar dicha información para determinar una distribución óptima de los SB que maximiza la predicción.

5.2.5. Escenario 36:1:E

En contraste al primer escenario 36:1:C (Sección 5.2.1), se busca evaluar el impacto que tiene el modelo de asignación de probabilidad *a priori* en la probabilidad Bayesiana dada por la Ecuación 3.5. Para ello, se considera que el vehículo puede haber elegido cada

ruta (entre detecciones sucesivas) con igual probabilidad. Primero se presenta el análisis de predicción, la Tabla 5.9 resume los indicadores ya presentados en los modelos anteriores.

No existen grandes diferencias entre cada modelo (36 conteo versus 36 equiprobable) respecto a los tres indicadores. Los FPR de secciones y viajes tuvieron un aumento de 0,1 puntos porcentuales respecto al primer escenario presentado. El indicador de calidad PQ también tienen variaciones del mismo orden. Para estos escenarios, el modelo de asignación *a priori* no marca una diferencia significativa a la hora de comparar la predicción de cada uno.

Cuadro 5.9. Resultados análisis predictivo Escenario 36:1:E

Indicador	$N_{inferencias}$	FPR	CR	PQ
Secciones	3.736	99,9 %	32,6 %	99,5 %
Viajes	3.002	99,9 %	26,9 %	99,4 %

La Figura 5.13 compara los flujos estimados con los reales. La gráfica no permite realizar comparaciones a simple vista, dado que estas se ven similares. Para ello, es mejor analizar los indicadores de error que se resumen en la Tabla 5.10.

La Tabla 5.10 muestra los indicadores de error para este escenario. Esta comparación sí permite utilizar el error cuadrático medio como variable de comparación, en la que vemos una diferencia de 0,37 puntos en el indicador y que favorece al modelo con conteos. Aún así, las diferencias que se evidencian de esta comparación señala que son modelos con prácticamente igual nivel de predicción probablemente por que ambos cuentan con 36 SB. El resto de indicadores no presenta diferencia respecto al modelo inicial.

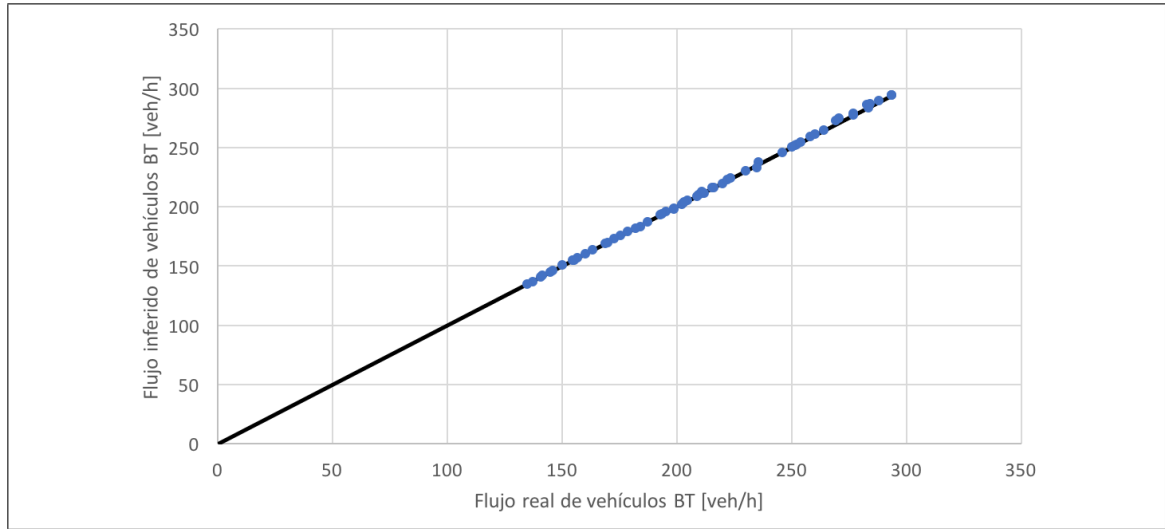


Figura 5.13. Comparación de flujos reales e inferidos para Escenario 36:1:E

Cuadro 5.10. Indicadores de error en flujos inferidos Escenario 36:1:E

Indicador	MSE	WMAPE	MIN	MAX	DESV_STD
Valor	3,27	0,01 %	0 %	0,03 %	0,01 %

La Figura 5.14 muestra el análisis de predicción por velocidad y nuevamente no se observan grandes diferencias con el modelo al que se compara.

Esta comparación señala que no existen diferencias sustanciales en la calidad de la predicción de cada modelo en las características señaladas (cobertura total de sensores y tiempo de interdetección de 1 segundo). Pero aun así, no se puede afirmar que siempre se comportaran de forma similar. Para ello, en la siguiente sección se realizan un análisis comparativo mucho más general que permite evaluar el impacto de las distintas variables que rigen los escenarios, sobre los resultados de la metodología.

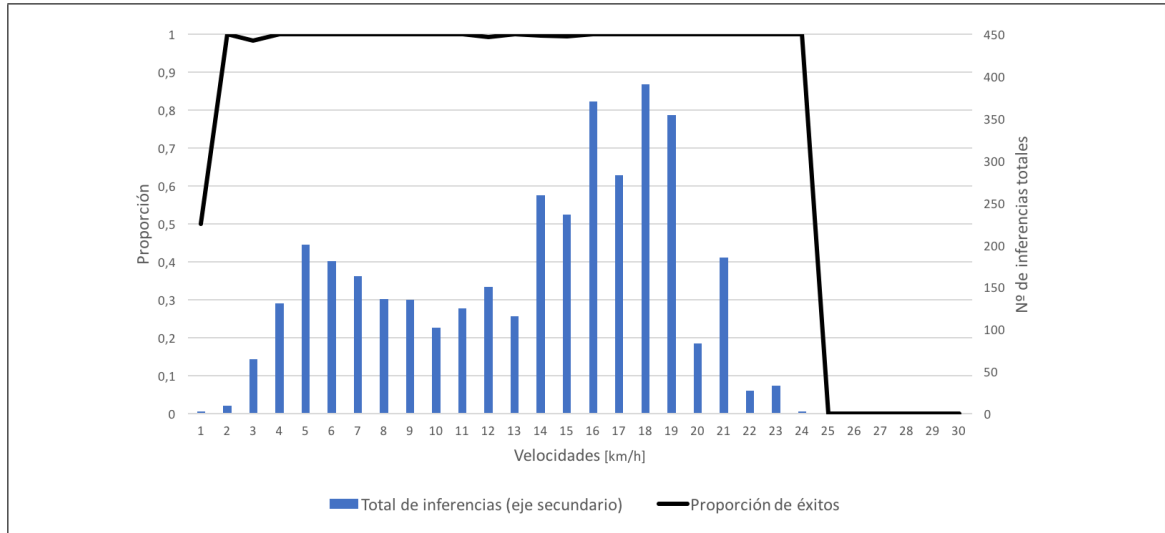


Figura 5.14. Histograma de proporción de aciertos de secciones por velocidades de vehículos para Escenario 36:1:E

5.3. Comparación con metodología existente en literatura

En esta última sección se procede a comparar la metodología propuesta (utilizando el modelo de conteos), con alguna alternativa reportada en la literatura. Dentro de los documentos revisados, la investigación de Michau *et al.* (2014) cumple con las mayores similitudes con el presente trabajo. Además el trabajo de Michau *et al.* (2014) es el que cuenta con mayor documentación y descripción del proceso, como fue descrita en la Sección 2.2.

La comparación entre ambas metodologías se realiza en términos de FPR y error WMAPE. Ambas metodologías son capaces de inferir los vehículos que sean detectados a lo menos dos veces (para tener al menos un par OD). Las metodologías entregan la misma ruta cuando existe asignación directa, es decir, cuando existe una única ruta entre un par de detecciones sucesivas. Es por esto que estas comparaciones se realizan específicamente con vehículos inferenciables.

El modelo propuesto por Michau *et al.* (2014) identificado con una “L” se compara con el modelo de conteos (identificado con una C) desarrollado en este trabajo. Las

Figuras 5.15 y 5.16 muestran el indicador FPR para la inferencia de secciones y viajes, respectivamente, y para los distintos escenarios.

El indicador FPR de secciones de la Figura 5.15 es bastante similar a la hora de comparar la predicción de la metodología propuesta con la metodología de Michau *et al.* (2014).

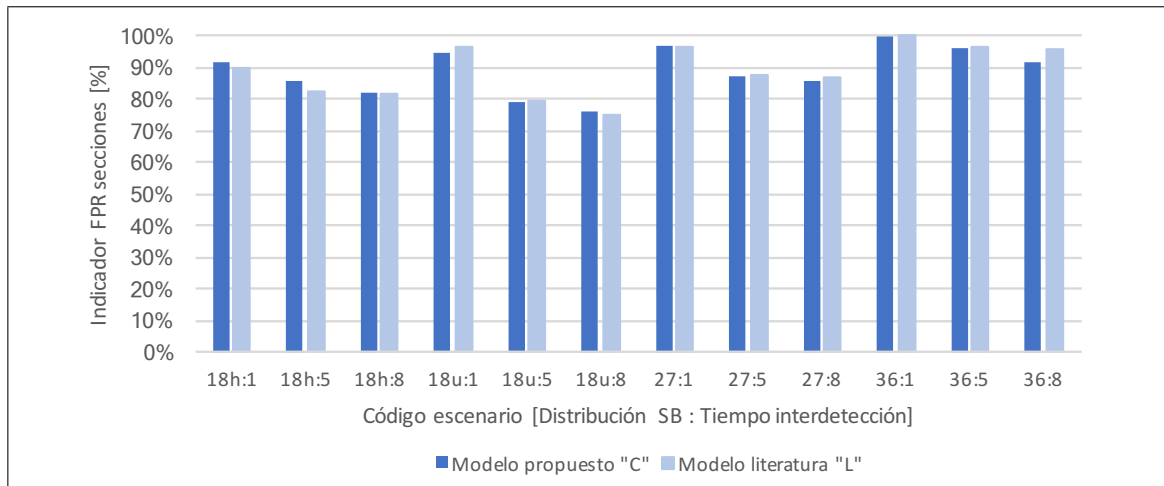


Figura 5.15. Valores de FPR de secciones entre el escenario de conteos en la BD y la literatura escenarios evaluados

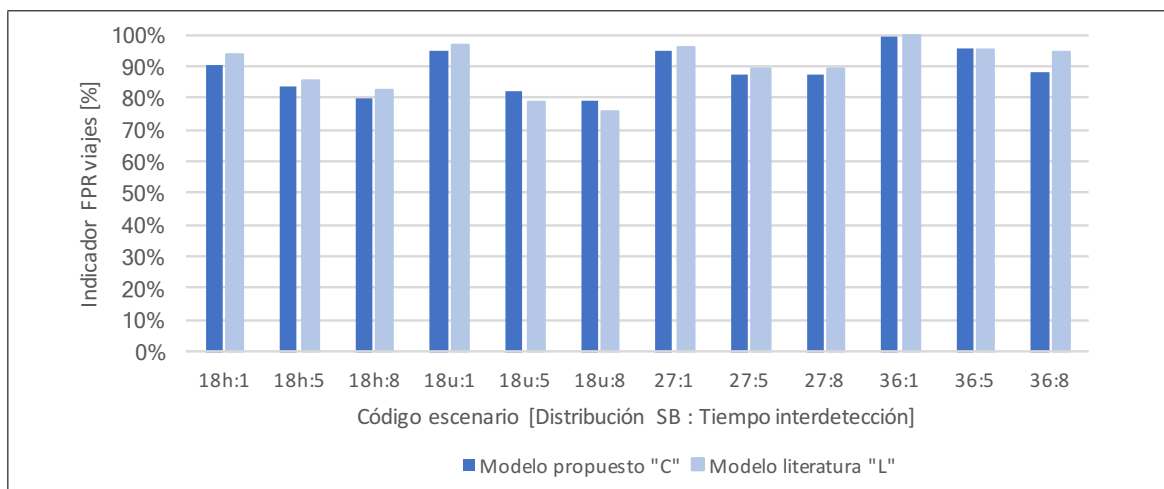


Figura 5.16. Valores de FPR de viajes entre el escenario de conteos en la BD y la literatura escenarios evaluados

Al observar el indicador FPR de viajes no se encuentran grandes diferencias en este indicador. Luego de analizar cada gráfico de FPR se puede afirmar que no existen grandes aportes de este trabajo en términos de FPR. Pero es importante analizar la calidad de estas predicciones y saber cuánto error se comete al fallar en el acierto de una ruta.

La metodología de Michau *et al.* (2014) solo entrega la ruta inferida por la que pudo pasar el vehículo mientras que esta metodología es capaz de entregar el conjunto de probabilidades de las rutas posibles. Gracias esto se puede estimar el flujo presente en cada uno de los arcos y medir el error. En este caso se utiliza el indicador WMAPE para evaluar el error cometido al asignar flujos a los arcos. La Figura 5.17 muestra la información recolectada de este indicador.

El indicador de error WMAPE es mayor en la metodología de Michau *et al.* (2014) para los escenarios de 18 y 27 sensores. Cuando la interdetección es baja y/o la densidad de sensores es máxima, entonces el modelo propuesto por Michau *et al.* (2014) muestra mejores desempeños.

A medida que la densidad de sensores dispuestos en la red disminuye, los errores de la metodología propuesta por Michau *et al.* (2014) crecen más rápido que los de la metodología propuesta en esta investigación. Esta diferencia es aún mayor si el tiempo de interdetección aumenta. Así, si el método de Michau *et al.* (2014) tiene muy buena información, es capaz de predecir las rutas casi sin errores, pero pierde efectividad rápidamente a medida que pierde información.

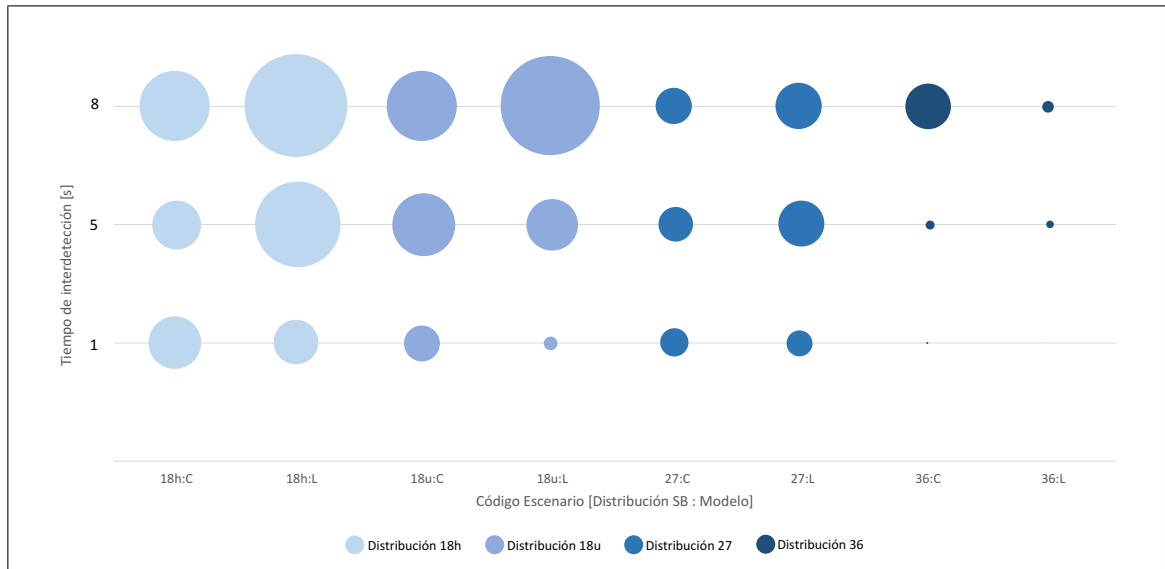


Figura 5.17. Error WMAPE entre los escenarios de conteos en BD y la literatura (vista 2)

CAPÍTULO 6. CONCLUSIONES

En esta tesis se presentó una nueva metodología que permite inferir la elección de rutas en una red vial urbana. Esto se lleva a cabo al aplicar inferencia estadística Bayesiana sobre un conjunto de detecciones Bluetooth. Conocer en detalle la ruta escogida por cada vehículo permite estimar flujos vehiculares sobre la red, estudiar cambios en los patrones de viaje frente a modificaciones viales, entre otras.

La metodología es capaz de inferir las probabilidades de todas las posibles rutas incluso cuando no se realizan detecciones en un tramo específico. Este aspecto diferencia esta investigación de otras metodologías existentes en la literatura.

Para evaluar la metodología y nivel de predicción se utilizó una serie de indicadores que fueron comparados en una microsimulación desarrollada en Aimsun. El indicador FPR fue utilizado para conocer el porcentaje de aciertos que se obtuvo, donde un acierto se define como asignar la mayor probabilidad a la ruta elegida por el vehículo. Junto con este indicador se utilizaron los índices CR y PQ que miden el nivel de predicción de un modelo aleatorio equiprobable y la calidad de las los aciertos contabilizados por el FPR, respectivamente. Además, se realizó un estudio de los flujos asignados a cada arco, en función de las probabilidades determinadas. Estos fueron comparados con los reales obtenidos por la simulación. Finalmente se determinó el nivel de predicción que existe a rangos de diferentes velocidades.

Se desarrollaron 24 escenarios diferentes donde se modificó el número y distribución de sensores dispuestos en la red, el tiempo de interdetección del sensor y el modelo utilizado para la construcción de probabilidades *a priori*.

Los resultados indican que el mejor índice de predicción FPR se obtuvo en el modelo de mayor densidad de sensores (una por cada intersección), el menor tiempo de interdetección de 1 segundo y el modelo desarrollado para calcular la probabilidad de elección *a*

priori mediante conteos en la base de datos. El valor de FPR fue 99,70 % rutas correctamente predichas de un total de 3,002 analizadas. El índice PQ alcanzó un valor similar de 99,67 %, por lo que se puede inferir que las predicciones se realizan con una asignación de probabilidad muy cercana a 1 en la mayoría de los casos. Al comparar el índice FPR con CR (26,94 %) se desprende que el uso de un modelo es necesario para la predicción de rutas, ya que de lo contrario se obtendrían bajos niveles de acierto.

A medida que el tiempo de interdetecciones aumenta, se observa una disminución en la calidad de la predicción. Por ejemplo, al aumentar de 1 segundo a 8 segundos el valor de FPR disminuye a 88,17 %. Por otro lado el efecto de disminuir la cantidad de sensores en la red, también produce un efecto negativo en la predicción. El valor de FPR que se reduce a 90,48 % reduciendo de 36 a 18 sensores heterogéneos. En general el tiempo de interdetección afecta negativamente la calidad del modelo dado que al existir menos detecciones, las distribuciones de tiempo de viaje en arcos y tiempos de permanencia en nodos tendrán menos datos para ser ajustadas. Además se tienen rutas a inferir más largas.

Respecto a la cantidad de sensores, se compararon dos situaciones de iguales características pero con distintos número de sensores. Los resultados indican que a mayor número de sensores, entonces se observa una mejor predicción por parte de la metodología. Primero, porque se tiene un mayor número de detecciones y, segundo, por que estas se encuentran más cercanas en términos de distancia y tiempo. Esto provoca que los tramos de ruta en los que se desea inferir son más pequeños y poseen menor variabilidad, aumentando la probabilidad de tener éxito en la predicción.

En dos escenarios con 18 sensores cada uno (mitad del máximo posible) se aplicó la metodología. Uno de ellos posee una distribución uniforme de sensores en el espacio y el otro cubre las intersecciones por zonas. Los resultados indican que instalar los sensores uniformemente distribuidas en la red ayuda a obtener resultados de mejor calidad que disponerlos heterogéneamente (90,48 % versus 94,72 %).

En la Sección 3, se analizó la metodología en dos escenarios que se diferencian en el modelo utilizado para asignar la probabilidad *a priori*. Uno de ellos se basa en los conteos que se pueden realizar en la base de datos y el otro es una simple asignación equiprobable. Los resultados indican que este modelo cobra relevancia cuando se tiene una menor cantidad de datos, ya sea por un tiempo de interdetección más alto o por menor número de sensores. Al tener el máximo número de sensores y el mínimo tiempo de interdetección se presentaron diferencias de un punto porcentual a favor del modelo equiprobable. En caso contrario (con el mínimo número de sensores y el máximo tiempo de interdetección) se registró una diferencia de 4 puntos porcentuales a favor del modelo de conteos.

Respecto de la velocidad media a la que circula el vehículo entre dos detecciones sucesivas, no impacta la calidad de la predicción. Cuando el número de sensores es bajo los vehículos más rápidos tienden a ser más difíciles de detectar. Aún así, esta velocidad incluye tiempos de parada, congestión y otros fenómenos. Es por esto que no se puede realizar una afirmación contundente respecto de su impacto en la calidad de la inferencia de elección de rutas.

Al comparar los flujos reales con los estimados, se observan las mismas tendencias que experimenta el FPR. Es decir, a medida que aumenta el tiempo de interdetección o disminuye el número de sensores, aumentan los errores de flujos estimados sobre los reales.

Dentro de la literatura se cuenta con un limitado número de trabajos que presenten metodologías para la inferencia de rutas con tecnologías inalámbricas. En particular, la investigación de Michau *et al.* (2014) es una de las pioneras. La diferencia con la presentada en este trabajo radica en cómo se trata a un vehículo cuando los sensores fallan a la hora de detectar vehículos y se generan puntos ciegos dentro de la información recopilada por los sensores. Por un lado Michau *et al.* (2014) asumen que todos los vehículos viajan por la ruta con menor número de SB entre detecciones sucesivas. La metodología propuesta

en esta investigación, desarrolla un modelo para inferir rutas en estas situaciones en base a la red, tiempos registrados por otros vehículos y probabilidades de detección de sensores. Ambas metodologías presentan altos niveles de predicción (FPR por sobre el 90 %) pero es posible apreciar que a medida que se pierde información este trabajo muestra mejores desempeños. A menor número de sensores y mayor tiempo de interdetección, la metodología propuesta se comporta mejor que lo propuesto por Michau *et al.* (2014). Además, este es el primer trabajo en entregar un conjunto de probabilidades para todas las posibles rutas y no solo indicar cual es la ruta inferida.

La metodología es capaz de inferir las rutas de vehículos solo con información Bluetooth y entregando buenos niveles de desempeño (sobre 90 % en los escenarios más realistas). El *software* desarrollado es capaz de entregar el conjunto de rutas para cada vehículo, además de entregar un resumen con indicadores globales.

Este trabajo es capaz abrir nuevas líneas de investigación dada la reciente explosión tecnológica y su incorporación en los sistemas inteligentes de transporte. Es importante desarrollar modelos que sean capaces de optimizar la localización de sensores con el fin de obtener mejores resultados. Otro aspecto interesante de analizar es el filtro previo que es necesario realizar a los registros de las señales y que se permita diferenciar cuando el registro provino de un peatón, vehículo, ciclista o incluso transporte público. De este último modo, es interesante analizar cómo es posible realizar un seguimiento de un conjunto de señales Bluetooth, por ejemplo, que identifiquen un bus. De este modo, se podría contar con un modelo de control de flota de transporte público. Por último, es sumamente importante experimentar físicamente con un sensor Bluetooth a fin de ajustar correctamente la curva de comportamiento de un sensor frente a las pasadas de vehículos y frente a diferentes condiciones viales y/o atmosféricas.

REFERENCIAS

- Abedi, N. (2014). *Monitoring spatiotemporal dynamics of human movement based on mac address data* (Tesis Doctoral no publicada). Queensland University of Technology.
- Addesso, P., Bruno, L., & Restaino, R. (2010). Adaptive localization techniques in wi-fi environments. En *Wireless pervasive computing (iswpc), 2010 5th ieee international symposium on* (pp. 289–294).
- Aliari, Y., & Haghani, A. (2012). Bluetooth sensor data and ground truth testing of reported travel times. *Transportation Research Record: Journal of the Transportation Research Board*(2308), 167–172.
- Altini, M., Brunelli, D., Farella, E., & Benini, L. (2010). Bluetooth indoor localization with multiple neural networks. En *Wireless pervasive computing (iswpc), 2010 5th ieee international symposium on* (pp. 295–300).
- Aparicio, S., Pérez, J., Bernardos, A. M., & Casar, J. R. (2008). A fusion method based on bluetooth and wlan technologies for indoor location. En *Multisensor fusion and integration for intelligent systems, 2008. mfi 2008. ieee international conference on* (pp. 487–491).
- Araghi, B. N., Pedersen, K. S., Christensen, L. T., Krishnan, R., & Lahrmann, H. (2015). Accuracy of travel time estimation using bluetooth technology: Case study limfjord tunnel aalborg. *International Journal of Intelligent Transportation Systems Research*, 13(3), 166–191.

Barceló, J., Montero, L., Marqués, L., & Carmona, C. (2010). Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. *Transportation Research Record: Journal of the Transportation Research Board*(2175), 19–27.

Barceló, J., Montero Mercadé, L., Bullejos, M., Serch, O., & Carmona, C. (2012). A kalman filter approach for the estimation of time dependent od matrices exploiting bluetooth traffic data collection. En *Trb 91st annual meeting compendium of papers dvd* (pp. 1–16).

Bargh, M. S., & de Groote, R. (2008). Indoor localization based on response rate of bluetooth inquiries. En *Proceedings of the first acm international workshop on mobile entity localization and tracking in gps-less environments* (pp. 49–54).

Bhaskar, A., & Chung, E. (2013). Fundamental understanding on the use of bluetooth scanner as a complementary transport data. *Transportation Research Part C: Emerging Technologies*, 37, 42–72.

Blogg, M., Semler, C., Hingorani, M., & Troutbeck, R. (2010). Travel time and origin-destination data collection using bluetooth mac address readers. En *Australasian transport research forum* (Vol. 36).

Bluetooth. (2017). *How it works — what is bluetooth technology*. Descargado 21 de octubre de 2017, de <https://www.bluetooth.com/what-is-bluetooth-technology/how-it-works>

Carpenter, C., Fowler, M., & Adler, T. (2012). Generating route-specific origin-destination tables using bluetooth technology. *Transportation Research Record: Journal of the Transportation Research Board*(2308), 96–102.

- Chen, L.-J., & Hung, H.-H. (2011). A two-state markov-based wireless error model for bluetooth networks. *Wireless Personal Communications*, 58(4), 657–668.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1), 269–271.
- Franssens, A. (2010). *Impact of multiple inquires on the bluetooth discovery process: And its application to localization* (Tesis de Master no publicada). University of Twente.
- Friesen, M. R., & McLeod, R. D. (2015). Bluetooth in intelligent transportation systems: a survey. *International Journal of Intelligent Transportation Systems Research*, 13(3), 143–153.
- Frost. (2005). *Wireless communication technologies – hi-fi with wi-fi and bluetooth*. Descargado 22 de mayo de 2018, de <http://www.frost.com/prod/servlet/market-insight-print.pag?docid=51683460>
- GeekZone. (2003). *Bluetags to install world's first bluetooth tracking system*. Descargado 14 de mayo de 2018, de <https://www.geekzone.co.nz/content.asp?contentid=1070>
- GSMArena. (2010). *Mobile phones evolution*. Descargado 15 de mayo de 2018, de https://www.gsmarena.com/mobile_phones_evolution_features_review-501p3.php
- Haase, M., & Handy, M. (2004). Bluetrack—imperceptible tracking of bluetooth devices. En *Ubicomp poster proceedings*.

Hainen, A., Wasson, J., Hubbard, S., Remias, S., Farnsworth, G., & Bullock, D. (2011). Estimating route choice and travel time reliability with field observations of bluetooth probe vehicles. *Transportation Research Record: Journal of the Transportation Research Board*(2256), 43–50.

Haseman, R., Wasson, J., & Bullock, D. (2010). Real-time measurement of travel time delay in work zones and evaluation metrics using bluetooth probe tracking. *Transportation Research Record: Journal of the Transportation Research Board*(2169), 40–53.

IEEE, C. S. (2002). *Wireless medium access control (mac) and physical layer (phy) specifications for wireless personal area networks (wpans)*. *ieee standard 802.15.1-2002*.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1), 35–45.

Kasten, O., & Langheinrich, M. (2001). First experiences with bluetooth in the smart-its distributed sensor network. En *Workshop on ubiquitous computing and communications, pact* (Vol. 1).

Kieu, L., Bhaskar, A., & Chung, E. (2012). Bus and car travel time on urban networks: integrating bluetooth and bus vehicle identification data. En *Arrb conference, 25th, 2012, perth, western australia, australia*.

Lindsey, R., & Verhoef, E. (2001). Traffic congestion and congestion pricing. En *Handbook of transport systems and traffic control* (pp. 77–105). Emerald Group Publishing Limited.

McElreath, R. (2015). *Statistical rethinking. texts in statistical science*. CRC Press.

Mei, Z., Wang, D., & Chen, J. (2012). Investigation with bluetooth sensors of bicycle travel time estimation on a short corridor. *International Journal of Distributed Sensor Networks*, 8(1), 303521.

Michau, G., Nantes, A., Bhaskar, A., Chung, E., Abry, P., & Borgnat, P. (2017). Bluetooth data in an urban context: Retrieving vehicle trajectories. *IEEE Transactions on Intelligent Transportation Systems*.

Michau, G., Nantes, A., Chung, E., Abry, P., & Borgnat, P. (2014). Retrieving dynamic origin-destination matrices from bluetooth data.

Murphy, P., Welsh, E., & Frantz, J. P. (2002). Using bluetooth for short-term ad hoc connections between moving vehicles: a feasibility study. En *Vehicular technology conference, 2002. vtc spring 2002. ieee 55th* (Vol. 1, pp. 414–418).

Musa, A., & Eriksson, J. (2012). Tracking unmodified smartphones using wi-fi monitors. En *Proceedings of the 10th acm conference on embedded network sensor systems* (pp. 281–294).

Nielsen. (2013). *Smartphone switch: Three-fourths of recent acquirers chose smartphones*. Descargado 14 de mayo de 2018, de <http://www.nielsen.com/us/en/insights/news/2013/smartphone-switch--three-fourths-of-recent-acquirers-chose-smart.html>

Ortúzar, J. d. D., & Willumsen, L. G. (2011). *Modelling transport*. John Wiley & Sons.

Pérez, J., Aparicio, S., Bernardos, A., Casar, J., & An, J. (2007). An indoor location system based on bluetooth and wlan. En *Proc. workshop on user-centric technologies and applications* (pp. 77–86).

Raveau, S., Muñoz, J. C., & De Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A: Policy and Practice*, 45(2), 138–147.

Tsubota, T., Bhaskar, A., Chung, E., & Billot, R. (2011). Arterial traffic congestion analysis using bluetooth duration data. En *Australasian transport research forum (atrf)*, 34th, 2011, adelaide, south australia, australia (Vol. 34).

Wen, Y., Pan, J., & Le, J. (2007). Survey on application of wireless sensor networks for traffic monitoring. En *International conference on transportation engineering 2007* (pp. 2079–2084).

Yen, J. Y. (1971). Finding the k shortest loopless paths in a network. *management Science*, 17(11), 712–716.

ANEXOS

A. ERRORES DE ESTIMACIÓN

A.1. Error cuadrático medio (MSE)

El error cuadrático medio mide el promedio del cuadrado de los errores. Es decir, mide la diferencia de los flujos reales con los estimados. Este error se encarga de penalizar más los errores más grandes, pero solo permite comparaciones entre demandas iguales. Es decir, en situaciones donde se trabaje con una alta demanda, los valores reales y estimados serán más grandes y por ende sus errores también. Hecho que imposibilita utilizarlo frente a modelos con demandas más pequeñas.

La expresión matemática de este error corresponde a la presentada en la Ecuación A.1

$$MSE = \frac{\sum_{i=1}^n (\hat{f}_i - f_i)^2}{n} \quad (A.1)$$

Donde $\hat{f}_i$ corresponde al flujo estimado en el arco i , f_i al flujo real en el arco i y n al número de estimaciones realizadas.

A.2. Error porcentual ponderado absoluto medio (WMAPE)

El error porcentual ponderado absoluto medio deriva del error porcentual absoluto medio (MAPE). El problema del indicador MAPE es que penaliza de igual forma los pequeños errores y los grandes. Para ello se construye este nuevo indicador WMAPE, que consta de ponderar el valor de MAPE por una proporción del flujo respecto del total. De este modo, se corrige esta diferencia de penalización, donde los errores grandes son más penalizados que los pequeños. Además, a diferencia del indicador MSE, este puede ser utilizado para comparar flujos sin importar la demanda, ya que son determinados de forma porcentual.

La expresión matemática de este error corresponde a la presentada en la Ecuación A.2

$$\frac{\sum_{i=1}^n \frac{f_i}{\sum_{j=1}^n f_j} \cdot \frac{|\hat{f}_i - f|}{f_i}}{n} \quad (\text{A.2})$$

Donde $\hat{f}_i$ corresponde al flujo estimado en el arco i , f_i al flujo real en el arco i y n al número de estimaciones realizadas.

B. DESARROLLO Y PROGRAMACIÓN DE LA METODOLOGÍA

La metodología fue desarrollada en el lenguaje de programación *c#* del entorno de *Visual Studio*. La herramienta es capaz de procesar una base de datos de detecciones BT (ver Tabla 3.1), determinar un conjunto de rutas para cada vehículo y calcular las respectivas probabilidades de elección.

Para que el *software* sea capaz de predecir las rutas a partir de la base de datos de detecciones BT, es importante proporcionar una codificación de la red de la zona de estudio. En este trabajo se opta por desarrollar un archivo que tiene la información de los nodos y otro para los arcos. Por esto, los tres archivos necesarios para la ejecución del *software* son: base de datos de detecciones BT, codificación de nodos y codificación de arcos. Estos archivos deben estar en formato “.csv” y seguir la estructura que se detalla en las secciones.

B.1. Archivo de base de datos de detecciones Bluetooth

Este archivo de texto contiene toda la información que proporcionan los SB de la red de estudio. Cada fila representa una detección, en donde se registra el ID del sensor que realizó la detección, la MAC del vehículo detectado y la hora en que se realizó el registro. La Figura B.1 muestra un ejemplo de la estructura que debe tener el archivo utilizado por el *software* para leer una base de datos.

1,	99:64:5A:42:2C:B9,	06:00:15
2,	4E:90:0F:6F:D7:6A,	06:00:20
1,	99:64:5A:42:2C:B9,	06:00:20
3,	4E:90:0F:6F:D7:6A,	06:00:25
4,	1B:2B:7E:9F:20:98,	06:00:25
1,	99:64:5A:42:2C:B9,	06:00:25
1,	50:77:24:79:24:A8,	06:00:25

Figura B.1. Ejemplo de un extracto del archivo de base de datos
detecciones Bluetooth

B.2. Archivo de codificación de nodos

Este archivo detalla la cantidad de nodos (intersecciones) presentes en la red original y sus características. Dentro de estas encontramos: la coordenada de latitud, la coordenada de longitud y la presencia de SB. Este último es una variable binaria, donde un 1 representa que ese nodo posee un SB, mientras que 0 significa que no posee instalación alguna. La Figura B.2 muestra la estructura del archivo que almacena esta información. La primera fila indica la cantidad de nodos presentes en la red y a partir de la segunda fila se define cada nodo: ID del nodo, coordenada latitud, longitud y presencia o no de sensor.

6
1, -33.163, -70.528, 1
2, -33.524, -70.710, 0
3, -33.804, -70.759, 1
4, -33.773, -70.406, 0
5, -33.134, -70.588, 1
6, -33.414, -70.637, 1

Figura B.2. Ejemplo de un extracto del archivo de codificación de nodos de la red

B.3. Archivo de codificación de arcos

Este archivo detalla la cantidad de arcos reales (calles entre intersecciones) presentes en la red original y sus características. En la Figura B.3 se puede ver un ejemplo del archivo de arcos. La primera fila indica la cantidad de arcos presentes en la red, y a partir de la segunda fila se indica la información de cada arco. Esta corresponde a: ID del arco, ID del nodo de origen del arco e ID del nodo de destino del arco.

6
1, 1, 2
2, 3, 4
3, 5, 2
4, 2, 1
5, 6, 1
6, 2, 4

Figura B.3. Ejemplo de un extracto del archivo de codificación de arcos de la red

C. RESULTADOS GENERALES DE ESCENARIOS

La Tabla C.1 resumen los resultados de cada escenario evaluado. La columna escenario representa el código del escenario. Este utiliza una configuración de SB, el tiempo intersecciones y el modelo *a priori*. E significa equiprobable y C conteos de base de datos. Se presentan los valores de FPR para secciones y arcos, MSE y WMAPE de la estimación de flujos.

Cuadro C.1. Resumen resultados generales de escenarios

Escenario	$FPR_{seccion}$	FPR_{viaje}	MSE	WMAPE
18h:1:E	82,4 %	83,7 %	167,90	0,26 %
18h:1:C	91,4 %	90,5 %	62,79	0,14 %
18h:5:E	80,7 %	79,2 %	105,69	0,19 %
18h:5:C	85,8 %	83,4 %	72,17	0,13 %
18h:8:E	77,7 %	75,2 %	135,41	0,24 %
18h:8:C	82,1 %	79,6 %	109,08	0,19 %
18u:1:E	92,0 %	92,2 %	110,51	0,22 %
18u:1:C	94,5 %	94,7 %	25,62	0,10 %
18u:5:E	77,6 %	79,9 %	68,50	0,18 %
18u:5:C	79,4 %	81,8 %	74,05	0,17 %
18u:8:E	72,5 %	75,9 %	81,39	0,22 %
18u:8:C	76,1 %	79,1 %	64,97	0,19 %
27:1:E	94,1 %	93,0 %	187,60	0,13 %
27:1:C	96,5 %	95,0 %	111,17	0,08 %
27:5:E	86,7 %	87,4 %	127,09	0,11 %
27:5:C	87,5 %	87,7 %	115,38	0,09 %
27:8:E	85,5 %	86,4 %	142,28	0,13 %
27:8:C	86,0 %	87,3 %	87,32	0,10 %
36:1:E	99,9 %	99,9 %	3,27	0,01 %
36:1:C	99,7 %	99,7 %	2,90	0,01 %
36:5:E	96,5 %	95,4 %	30,95	0,02 %
36:5:C	96,3 %	95,5 %	42,84	0,02 %
36:8:E	91,3 %	86,7 %	907,18	0,14 %
36:8:C	91,9 %	88,2 %	723,38	0,12 %