PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

ESCUELA DE INGENIERÍA

SENSORES MÓVILES:
LA RELACIÓN ENTRE LA DISPONIBILIDAD DE DATOS Y LA ESTIMACIÓN DE ESTADOS DE TRÁFICO

CHRISTOPHER PAUL BUCKNELL RIDERELLI

Tesis para optar al grado de

Magíster en Ciencias de la Ingeniería

Profesor Supervisor:

JUAN CARLOS HERRERA MALDONADO

Santiago de Chile, Enero 2014

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

ESCUELA DE INGENIERÍA

SENSORES MÓVILES:
LA RELACIÓN ENTRE LA DISPONIBILIDAD DE DATOS Y LA ESTIMACIÓN DE ESTADOS DE TRÁFICO

CHRISTOPHER PAUL BUCKNELL RIDERELLI

Tesis presentada a la Comisión integrada por los profesores:

JUAN CARLOS HERRERA MALDONADO

JUAN ENRIQUE COEYMANS AVARIA

RODRIGO FERNÁNDEZ AGUILERA

ÁLVARO SOTO ARRIAZA

Para completar las exigencias del grado de

Magíster en Ciencias de la Ingeniería

Santiago de Chile, Enero 2014

A mi familia.

AGRADECIMIENTOS

En primer lugar, quiero agradecer el apoyo de mi profesor supervisor, Juan Carlos Herrera. Su apoyo, tanto en lo personal como en lo profesional, fue clave para el exitoso desarrollo de esta tesis. Sin duda fue un desafío exigente, que no hubiese sido lo mismo sin su presencia y dedicación.

También me gustaría agradecer a la comisión evaluadora, y en especial al profesor Juan Enrique Coeymans. Sus consejos a lo largo de mi estadía en la universidad han sido muy valiosos y certeros.

A todos los funcionarios, administrativos, profesores, y compañeros del Departamento de Ingeniería de Transporte y Logística, ya que de alguna u otra manera me apoyaron en la realización de mi magíster. Especialmente aquellos que me ayudaron en el análisis de la metodología y resultados: Gabriel, Cristian, Matías, Sebastián y Marco. Me llevo grandes amistades de todos los rincones del mundo.

Un agradecimiento muy especial a Isidora, quien me apoyó constantemente en el desarrollo de la tesis, especialmente en las preparaciones de las presentaciones del Congreso Chileno y la defensa.

Quiero agradecer a mi familia, pilar fundamental de mi vida. Me han entregado todo el apoyo necesario para que pueda desarrollarme en lo que me gusta.

Finalmente, agradecer y reconocer el apoyo de la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) a través del Fondo de Desarrollo Científico y Tecnológico (FONDECYT #11110206).

ÍNDICE GENERAL

Pág.

ÍNDICE DE TABLAS

Pág.

ÍNDICE DE FIGURAS

Pág.

RESUMEN

La creciente incorporación de sistemas de navegación geo-referenciados a teléfonos celulares entrega nuevas opciones para monitorear el tráfico. Estos dispositivos pueden registrar y enviar gran cantidad de datos de tráfico, los cuales pueden ser utilizados para la estimación de estados de tráfico en las distintas vías. Sin embargo, es necesario entender la relación entre la cantidad y calidad de los datos recolectados y la estimación de estados de tráfico. En esta tesis se analiza esta relación a través del análisis sistemático de dos bases de datos de trayectorias vehiculares de secciones de autopista. La primera contiene trayectorias reales, mientras que la segunda trayectorias obtenidas de microsimulación. Sobre ellas se crearon escenarios hipotéticos en cuanto a cantidad y composición de datos a enviar por los vehículos, dado que un mismo número total de datos puede obtenerse de distintas formas. Esto se realizó a través de la definición de valores de variables de importancia relacionadas con el envío de datos: la penetración de sensores móviles en el flujo vehicular, la frecuencia con que cada uno de ellos envía un dato, y la anonimidad del dato enviado. Cada combinación de estos valores definió un escenario en particular. Para cada escenario, la estimación de estados de tráfico se realizó a través de dos métodos simples y uno más avanzado al incorporar teoría de tráfico vehicular.

En cuanto a los resultados obtenidos, destacan los siguientes: i) la formulación propuesta para cuantificar la cantidad de datos a recibir entrega errores relativos menores al 6%, ii) para los métodos evaluados, escenarios en los cuales se preserva la anonimidad de los datos no presentan desventajas en la estimación de estados de tráfico comparados con aquellos que no preservan la privacidad, iii) se estimó una expresión del error en la estimación en función de las características de cada escenario, y iv) se determinaron los escenarios en los cuales un aumento en la penetración a fin de aumentar el número total de datos, y mejorar así la estimación de estados de tráfico, es más conveniente que un aumento en la frecuencia de muestreo (y vice-versa).

Estos resultados permiten entender de mejor manera la relación que existe entre los datos recolectados de sensores móviles y la estimación de estados de tráfico, entregando en definitiva políticas de uso de estos sensores relacionadas a la privacidad de usuarios y calidad de la estimación.

Palabras Clave: Sensores móviles, Monitoreo de tráfico, Recolección de datos

ABSTRACT

The rapid-growth of smartphones with embedded navigation systems such as GPS modules provides new ways of monitoring traffic. These devices can register and send a great amount of data related to traffic, that can be used for traffic state estimation. However, we need to understand the relation between the amount of data received and its composition, and how traffic states are estimated. This thesis examines this relationship through an in-depth analysis of two datasets of vehicle trajectories in freeways. The first dataset consists of vehicle trajectories over a real freeway, while the second dataset is obtained through microsimulation. Hypothetical scenarios of data sent by equipped vehicles were created, based on three fundamental variables related to data collection: penetration rate of sensors in traffic flow, their data sampling frequency and the anonymity of data sent. Different values of these variables were used, and each unique combination defined a specific scenario. Traffic states were estimated through two simple methods, and a more advanced one that incorporates traffic flow theory.

Results obtained suggest the following: i) the proposed formulation to quantify data to be collected yielded estimation errors below 6% for every scenario in each dataset, ii) scenarios where anonymity is preserved do not show disadvantages related to traffic state estimation performance compared with those scenarios that do not preserve it, iii) an expression to infer the traffic state estimation error based on each scenario's characteristics was proposed, and iv) scenarios in which an increase in penetration rate is more effective to reduce estimation error than an increase in sampling frequency were identified (or vice-versa).

These results allow us to understand more precisely the relation that exists between data collected from mobile devices and traffic state estimation, delivering usage policies related to user's privacy and quality of the estimation.

Keywords: Mobile sensors, Traffic monitoring, Data collection

1. INTRODUCCIÓN

La llegada del internet móvil y la incorporación de GPS a los teléfonos celulares entrega nuevas oportunidades para realizar monitoreo de tráfico, utilizando como pilar datos enviados como posición y velocidad. Esta tecnología posee beneficios importantes, tales como nulo costo del dispositivo (es asumido por el usuario) y pequeños costos de mantención para los sistemas de información (Sanwal & Walrand, 1995), aspecto clave para países en desarrollo. Adicionalmente, proveen una gran cobertura tanto temporal como espacial.

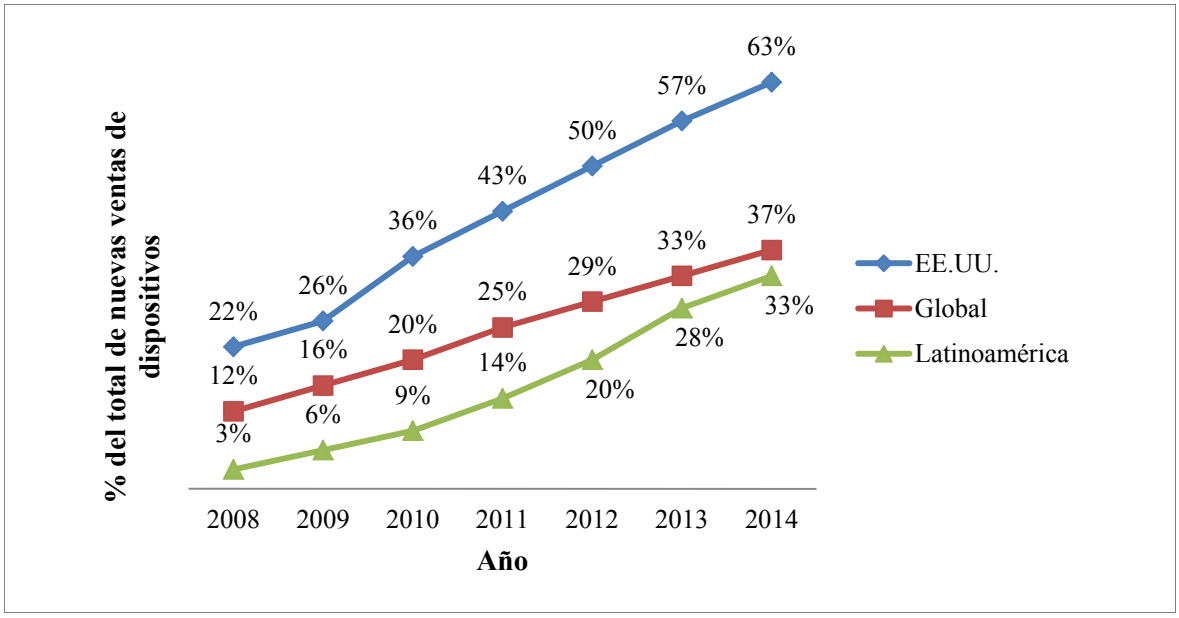

Figura 1-1: Penetración de teléfonos inteligentes en ventas

Fuente: GMSA (2011)

La penetración de este tipo de dispositivos en la población ha ido en exponencial aumento (Figura 1-1). Esta alta tenencia de teléfonos inteligentes con GPS abre una puerta para que sean utilizados como sensores en la estimación de estados de tráfico en una ciudad. En el caso de la aplicación Waze, existen 1,9 millones de usuarios en Chile (Ministerio de Transportes y Telecomunicaciones, 2013).

A pesar de los claros beneficios de esta tecnología, existen limitaciones asociadas a la cantidad de datos que se puede obtener de ellos (Herrera et al., 2010). Primero, no toda la población tiene este sensor, luego sólo parte de la realidad se puede observar. Sigue el consumo de energía del aparato, que aumenta al tener la función GPS encendida. Asociado a lo anterior, está cuán frecuentemente se envían datos. Pueden existir restricciones de ancho de banda, lo que limita la cantidad de datos que el sensor podrá enviar (los datos pueden significar un costo para el usuario, o bien pueden estar dentro de un límite mensual). Finalmente está el concepto de privacidad, de alta importancia en la nueva era informática, en especial si se desea contar con una alta penetración de sensores en la población (Rose, 2006).

Debido a las limitaciones mencionadas anteriormente, los datos a recibir en un sistema como este es acotada, y es principalmente función de (i) la penetración del dispositivo en la población (i.e. cuántos vehículos envían información), (ii) la tasa y forma de muestreo del dispositivo (i.e. cómo y cuán frecuentemente se envían datos) y (iii) los datos que se envían (i.e. qué contiene y si viola la privacidad). Estos datos son usualmente utilizados para estimar estados de tráfico.

Así, la calidad de la estimación dependerá tanto los datos recolectados (cantidad, tipo y forma de recolección), como la forma en que se estiman los estados de tráfico. Bajo el supuesto que la cantidad, tipo y forma de recolección de datos a partir de sensores móviles influye en la estimación de estados de tráfico, es de interés entender a cabalidad la relación entre los datos recolectados y la calidad de las estimaciones realizadas.

1.1 Objetivos

El objetivo general de esta investigación es examinar la relación entre la cantidad y composición de datos recolectados a partir de sensores móviles y la forma de estimar estados de tráfico.

Los objetivos específicos son los siguientes:

- Definir una medida que permita cuantificar la cantidad de datos con que se cuenta.
- Cuantificar el efecto de la anonimidad de sensores en la calidad de la estimación de estados de tráfico.
- Analizar la precisión obtenida con diferentes métodos de estimación de estados de tráfico, en relación a la cantidad de datos con que se trabaja, y bajo qué condiciones de operación y tipo de vía.
- Obtener una métrica que permita inferir un error en la estimación estados de tráfico, en base a variables conocidas.
- Determinar el efecto de la composición de datos en la estimación de estados de tráfico, dado que una misma cantidad de datos puede obtenerse de diferentes maneras.

1.2 Alcances

Los objetivos propuestos solamente se analizarán en un entorno de autopistas (flujo ininterrumpido). Asimismo, se dejará de lado temas técnicos de los sensores móviles que enviarán información y se asumirá que la información enviada por estos sensores está libre de error. También se hará el análisis usando una sola forma de muestreo de información.

1.3 Contenidos

Aparte de esta introducción, la tesis contiene otros cuatro capítulos. El capítulo segundo muestra el marco teórico, y se trata de una revisión del estado del arte de los diferentes tópicos atingentes a la investigación. El capítulo tercero expone la forma en que se alcanzarán los objetivos propuestos en esta investigación. En el cuarto capítulo, se muestran los resultados de aplicar la metodología a dos bases de datos distintas. El capítulo final presenta las principales conclusiones e implicancias del trabajo, así como las recomendaciones para futuras investigaciones.

2. MARCO TEÓRICO

El presente capítulo tiene como objetivo realizar una revisión del estado del arte en torno a los diferentes tópicos atingentes a la investigación.

2.1 Teoría de tráfico

Un estado de tráfico se define como como una combinación de las variables flujo, densidad y velocidad ($q\ \left[\frac{veh}{hr}\right]$, $k\ \left[\frac{veh}{km}\right]$ y $v\ \left[\frac{km}{hr}\right]$ respectivamente) atribuibles a cierta sección de vía en un instante determinado. La relación entre dos pares de estas variables lleva como nombre relación bivariada (Cassidy, 2003), y en general, se define de manera explícita la relación entre flujo y densidad (conocida como función flujo o diagrama fundamental). A partir de la relación fundamental ($q = k \cdot v$) es posible obtener el valor de la velocidad. Diferentes relaciones bivariadas se han planteado en la literatura, desde el trabajo seminal de Greenshields (1935), pasando por la relación de Smulders (1990) y hasta la simple, pero ampliamente utilizada, relación triangular (Newell, 1993), también conocida como Daganzo-Newell (Figura 2-1). Una revisión bibliográfica de varias de estas relaciones propuestas en la literatura es presentada por Wang et al. (2013).

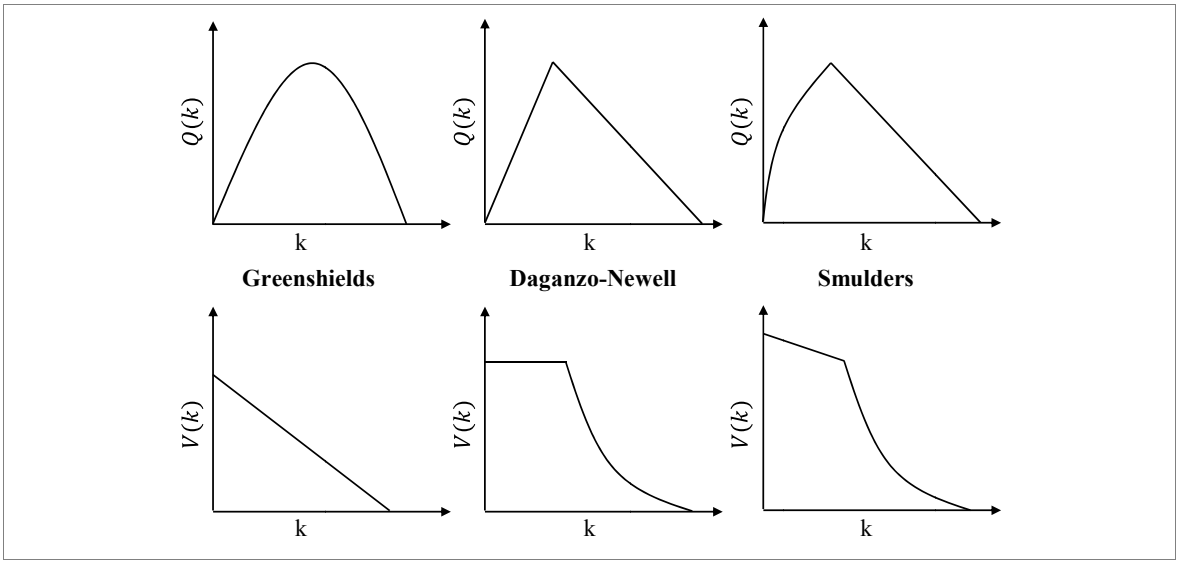

Figura 2-1: Diagrama fundamental y funciones de velocidad para diferentes modelos

Fuente: Elaboración propia a partir de Work et al. (2010)

2.2 **Midiendo la realidad**

Si para una sección de vía de largo $\Delta x = L$ que se analiza durante un periodo de tiempo $\Delta t = T$ existe la información detallada de las trayectorias que cruzan por ella (Figura 2-2): ¿cómo se calcula el estado de tráfico de esta sección? En general, las mediciones de tráfico se realizan durante un intervalo de tiempo en una posición fija (medida temporal) o bien para una sección de vía en un instante de tiempo fijo (medida especial). Edie (1963) define el flujo, la densidad y la velocidad de forma independiente al proceso de medición. Esto evita supuestos de la forma en que fueron obtenidos los datos (ya sea de manera temporal o espacial).

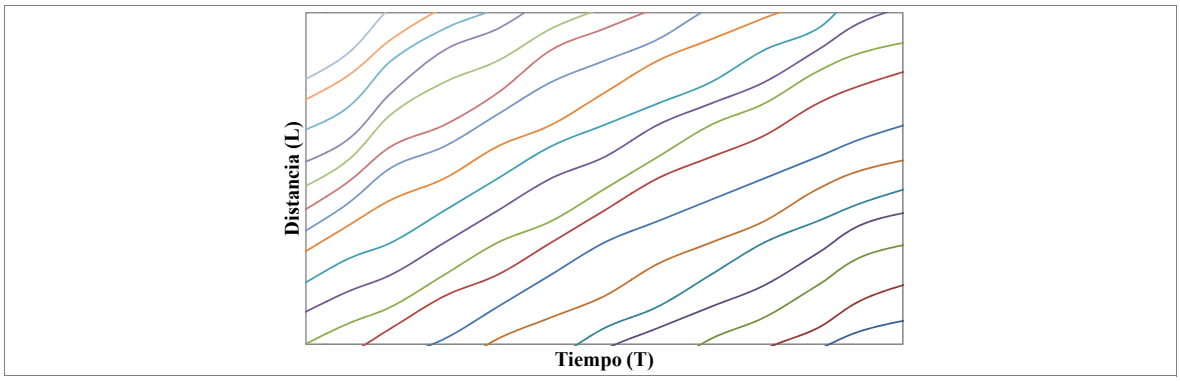

Figura 2-2: Trayectorias en una celda

Fuente: Elaboración propia

Si se considera un área A (de tamaño $L \cdot T$) dentro del dominio espacio-tiempo, es posible obtener las definiciones generalizadas para las variables macroscópicas como flujo, densidad y velocidad esta área. La variable flujo en esta región, $q(A)$, se obtiene como la distancia agregada de todos los vehículos dentro del área ($d(A)$), dividida por el área del dominio en inspección ($|A|$). La densidad ($k(A)$) se obtiene como el tiempo agregado de los vehículos dentro del área ($t(A)$), dividido por el área del dominio en inspección. La velocidad ($v(A)$) se puede obtener a partir de la relación fundamental. El cálculo de cada variable se puede apreciar en las ecuaciones (2-1), (2-2) y (2-3) respectivamente:

$$q(A) = \frac{d(A)}{|A|} \tag{2-1}$$

$$k(A) = \frac{t(A)}{|A|} \tag{2-2}$$

$$v(A) = \frac{q(A)}{k(A)} = \frac{d(A)}{t(A)} \tag{2-3}$$

2.3 **Obteniendo información de la realidad**

Para conocer el estado de tráfico de una vía, es necesario recoger datos de ella. Es por esto que existen diversos tipos de sensores que permiten conocer parte de la realidad con el objetivo final de reconstruirla. Esta información es valiosa pues les permite tomar una

más informada decisión respecto a su viaje. También es valiosa para las agencias encargadas de la operación de las vías, ya que permite identificar la presencia de congestión o incidentes. Existen principalmente dos familias de sensores para conocer parte de la realidad: los estáticos y móviles. También son distinguidos como Eulerianos y Lagrangianos respectivamente (Herrera & Bayen, 2010). Estos se analizarán a continuación.

2.3.1 Sensores estáticos

Se entiende por sensores estáticos aquellos que proveen mediciones a través de un área de control fija. Existen diversos sensores estáticos, tales como espiras inductivas, sensores magnéticos, cámaras de procesamiento de imágenes, radares de microondas, sensores infrarrojos, sensores láser, entre otros. Estos sensores son principalmente utilizados para conocer flujos, velocidades y densidades promedio. El más utilizado es la espira inductiva, desplegada ampliamente en distintos sistemas de transporte en el mundo.

La limitación principal de estos sensores es la necesidad de inversión en infraestructura física para obtener datos. En el caso de las espiras, la operación de éstas se ven afectadas notoriamente con el deterioro del pavimento y la instalación incorrecta. Además, la precisión de los datos obtenidos pueden no ser suficiente para detectar de manera precisa incidentes (Klein et al., 2006a, 2006b).

2.3.2 Sensores móviles

Estos sensores son aquellos que se mueven con el flujo, los cuales son geo-referenciados mediante satélites (en el caso de dispositivos GPS[1] portables o antenas de celulares) o bien infraestructura física local (en el caso de *tags* RFID[2] o re-identificación de la

[1] Del inglés *Global Positioning System.*

[2] Del inglés *Radio Frequency Identification*.

dirección MAC[3] de un dispositivo Bluetooth). A diferencia de los sensores estáticos, los sensores móviles no sólo entregan una cobertura temporal de los datos, sino también una espacial. Esta característica hace sumamente interesante su inclusión en la estimación de estados de tráfico. El constante aumento de sensores como los teléfonos inteligentes (que incluyen GPS o GLONASS[4]) promueve el uso de éstos como los principales sensores a usar.

Estos sensores en general entregan datos relacionados con la posición (longitud, latitud), velocidad (instantánea o de cierta medida temporal) y además de una etiqueta que permita la re-identificación del vehículo con cierta frecuencia.

Existen tres ejes principales interesantes de analizar al utilizar teléfonos inteligentes como sensores móviles: el muestreo de datos, la penetración de sensores en los vehículos circulantes, y la privacidad. A continuación, se presenta una caracterización de cada uno de estos ejes:

a. **Muestreo de datos**

El muestreo se refiere a la forma (esquema) y frecuencia con que el dispositivo enviará datos. Estos se relacionan con el uso de batería, la capacidad de hardware del sensor, restricciones de capacidad de datos (dentro de un límite mensual), costo directo (según el consumo de datos) y la preservación de la privacidad.

Los sensores pueden muestrear datos de tres maneras distintas (Herrera, 2009; Nanthawichit et al., 2003):

- Muestreo temporal: los dispositivos envían datos cada cierto intervalo de tiempo fijo, independiente de la posición. Un estudio acerca del consumo de energía del aparto y la precisión de los datos con un muestreo temporal se puede apreciar en Frank et al. (2012).

[3] Dirección física única para cada dispositivo, del inglés *Media Access Control*.

[4] Del ruso *Globalnaya Navigatsionnaya Sputnikovaya Sistema* o *Global Navigation Satellite System.*

- Muestreo espacial: los dispositivos envían datos cada vez que se cruza algún lugar espacial en específico (generalmente algún punto de interés). Estos lugares son conocidos como Líneas Virtuales de Viaje (*Virtual Trip Lines* o VTL). Ver Hoh et al. (2008).
- Muestreo accionado por evento: los dispositivos envían datos ante ciertas situaciones específicas y predefinidas, como frenado (cambio en la velocidad, detectado con un acelerómetro) o ante sonidos de una bocina (utilizando un micrófono). Estos ejemplos son presentados en Mohan et al. (2008).

Un muestreo temporal tiene la ventaja de entregar información a lo largo de la trayectoria del vehículo, independiente de su posición, permitiendo obtener información que tal vez no estuvo disponible en el caso del muestreo espacial. Por otro lado, el muestreo espacial asegura el envío de datos en un punto, lo que puede ayudar a estimar de mejor manera variables en torno a esa posición. El muestreo accionado por evento entrega datos ante situaciones de interés, como cuando un auto frena. Esta información puede ser más valiosa que una que fue enviada sin un evento en específico. Al mismo tiempo, puede que no existan eventos de interés, donde no habrá envío de datos. Estos últimos dos esquemas de muestreo son de alta demanda energética para el dispositivo, pues debe estar midiendo variables constantemente mediante sensores de ubicación u otro tipo.

b. **Penetración de sensores**

La tasa de penetración de sensores se refiere a la proporción de vehículos que cuenta con un sensor móvil y envía datos. Esta tasa principalmente depende de las características socioeconómicas de las personas, pero cada vez en menor grado por la tendencia a la baja del costo de los dispositivos. También depende del grado de retribución de información que existe al enviar información: una mayor retribución hace más atractivo el sistema.

c. Privacidad

El concepto de privacidad es de suma importancia en esta nueva era informática, especialmente en el uso de sensores móviles que permiten inferir la ubicación de personas. Esta información puede ser utilizada para dañar a una persona de manera económica o física (Krumm, 2007), como también es posible inferir condiciones médicas, afiliaciones políticas, infracciones de tráfico o bien algún grado de involucración en accidentes (Hoh et al., 2012).

Existen diversos intentos para preservar la privacidad, como la pseudoanonimidad estática (Pfitzmann & Shostack, 2000; Sun et al., 2013). Esto último se refiere a la posibilidad de asociar puntos de información con etiquetas, donde una misma etiqueta indica que la información fue enviada por el mismo sensor (y que es únicamente un número de identificación, que no permite identificar el usuario de manera directa). También existen otros métodos para preservarla, como añadir ruido, redondear posiciones, entre otros (Krumm, 2007; Sun et al., 2013). En el caso de sistemas de estimación de tráfico, Hoh et al. (2012) propusieron la utilización de Líneas Virtuales de Viaje para preservar la privacidad, donde los vehículos envían información únicamente en posiciones específicas.

Una de las principales maneras de preservar la privacidad es mediante sistemas que permitan separar la comunicación de los datos, con el análisis de éstos (Hoh et al., 2006). En el sistema propuesto, existen dos servidores: el de comunicación y el de tráfico. El servidor de comunicación sabe quién envió el dato, pero no la información asociada a éste (posición y velocidad). Al mismo tiempo, el servidor de tráfico conoce la información asociada, pero no la identificación. También se han propuesto sistemas más sofisticados (Hoh et al., 2012; Jeske, 2013).

Estudios han mostrado que a partir de datos pseudoanónimos, es posible inferir la ubicación de hogares de las personas (Hoh et al., 2006; Krumm, 2007). Incluso es posible identificar el dueño de casa con búsquedas invertidas a partir de la ubicación, utilizando buscadores de internet (Krumm, 2007). Hoh et al. (2006) analizan que es

posible reducir la tasa de identificación de hogares si el intervalo de muestreo de los sensores se agranda, lo que es útil como medida para preservar la privacidad. Cabe mencionar que trayectorias pueden ser inferidas con datos completamente anónimos si la frecuencia de muestreo es suficientemente alta (Hoh et al., 2006).

Jeske (2013) analiza el concepto de privacidad para dos aplicaciones de teléfonos inteligentes que estiman estados de tráfico: Google Live Traffic (a través de Google Maps de la plataforma Android) y Waze (popular aplicación multiplataforma). El concepto de privacidad utilizado radica en la posibilidad de seguir o rastrear a usuarios (que es equivalente a la definición de pseudoanonimidad). El autor demuestra que la privacidad no está asegurada, pues ambas aplicaciones permiten reconstruir trayectorias a partir de los datos enviados por los usuarios.

2.4 **Estimación de estados de tráfico**

A medida que no toda la información está disponible, es necesaria la implementación de modelos matemáticos que permitan entender las relaciones entre las variables fundamentales del tráfico y su evolución en el tiempo. Estos modelos permitirán estimar el estado de tráfico en lugares e instantes de tiempo en donde no hay datos.

Para estimar un estado de tráfico, existen principalmente dos enfoques para realizarlo (Herrera, 2009). El primer enfoque (inductivo) se basa en estadística sobre datos actuales (y posiblemente históricos), sin usar un modelo de tráfico vehicular. El segundo (deductivo) se basa en conceptos físicos del flujo vehicular (basado en analogías a propiedades fluidos, gases y la conservación de masa). También se puede considerar un enfoque intermedio entre ambos (Papageorgiou, 1998).

Independiente del enfoque, la estimación de estados de tráfico suele discretizar tanto el espacio (celdas) como el tiempo (intervalos). Así, cada cierto intervalo de tiempo h se desea conocer el estado de tráfico de cada celda i. El estado de tráfico se caracteriza usualmente por la velocidad o densidad. Así $\hat{v}_i^h$ indica la velocidad estimada de la celda i en el instante de tiempo h. En general, ésta se subdivide en I intervalos iguales de

distancia Δx (indexado por i) y H intervalos de tiempo Δt (indexado por h). Por lo tanto, se obtiene un total de $I \cdot H$ celdas-intervalos. Lesort et al. (2003) analizan las diferentes escalas a las cuales se puede analizar el tráfico (tanto en el tiempo como en el espacio) y sus implicancias. Papageorgiou (1998) analiza los requerimientos computacionales para distintos tamaños de discretización, y las ganancias en la precisión de la modelación. La discretización de una vía también puede estar sujeta a otro tipo de restricciones relacionadas con modelos de estimación, como cuando un vehículo no puede atravesar más de una celda en un intervalo de tiempo (Courant et al., 1967). Una vez discretizada la vía, es de interés poder obtener información relevante de ella.

Considerando la presencia de sensores móviles, y para un intervalo dado, en ciertas celdas no se tendrán observaciones, mientras que en otras sí. En las celdas con observaciones, la velocidades observada[5] (que se denotará como u_i^h) se computa generalmente como el promedio aritmético de todas las velocidades enviadas por los sensores en esa celda (Herrera et al., 2010). Esto es correcto sólo cuando los intervalos de tiempo son suficientemente pequeños (en otro caso, la velocidad obtenida no es la velocidad media espacial de los vehículos). Para las celdas sin observación, se necesita de métodos que permitan imputar datos, los cuales pueden ser derivados desde un enfoque estadístico o uno físico.

2.4.1 Enfoque estadístico

Un proceso estadístico permite estimar estados en base a información pasada o colindante, sin utilizar procesos físicos que faciliten el conocimiento del proceso deseado. Por ejemplo, el uso de cadenas de Markov de tiempo discreto en autopistas (Yeon et al., 2008), regresiones lineales (Zhang & Rice, 2003), regresiones logísticas y modelos autoregresivos para arterias (Herring et al., 2010), redes neuronales (Liu et al., 2006), e incluso filtros de paso bajo que permiten imitar comportamientos físicos del tráfico (Treiber & Helbing, 2002). Krause et al. (2008) utilizan un enfoque probabilístico

[5] En realidad, es una estimación de la velocidad observada, pues depende de cómo se compute.

para velocidades en un tramo. En general, los modelos presentados necesitan de datos de aprendizaje para calibración, por lo que su uso no es inmediato. Además, su complejidad puede ser inherente.

Un esquema básico utilizado por Work et al. (2008) como *benchmark* para estimar un campo de velocidades ($\hat{v}_i^h$ $\forall i$ y $\forall h$), en base a datos de sensores móviles, es el siguiente:

$$\hat{v}_i^h = \begin{cases} u_i^h & \text{si } u_i^h \neq \emptyset \\ \hat{v}_i^{h-1} & \text{si } u_i^h = \emptyset \end{cases} \quad (2\text{-}4)$$

Donde u_i^h es la velocidad observada en la celda i e intervalo h. Es decir, si existe velocidad observada en una celda, la velocidad estimada será ese mismo valor. Por otro lado, si no existe una la velocidad observada, el valor de la estimación será el valor estimado de esa celda en el intervalo temporal anterior. Este método lo llamaremos PSH.

Otro esquema simple es uno utilizado con espiras inductoras (Chen et al., 2003), que es fácilmente aplicable en el caso de sensores móviles. Al igual que el esquema anterior, en caso de que exista una estimación de la velocidad observada en una celda, la estimación de la velocidad es el mismo valor. En caso que no exista estimación en la velocidad observada, se procede a realizar una interpolación lineal para el intervalo de tiempo h entre las celdas más cercanas aguas abajo y aguas arriba de la celda:

$$\hat{v}_i^h = \begin{cases} u_i^h & \text{si } u_i^h \neq \emptyset \\ \text{interpolación} & \text{si } u_i^h = \emptyset \end{cases} \quad (2\text{-}5)$$

Siempre existirá estimación en caso de considerar existencia de condiciones de borde. A este método lo llamaremos PSI.

2.4.2 Enfoque físico

La evolución en el espacio (x) y tiempo (t) de las macro variables fundamentales (q, k, v) recae en la teoría de tráfico continua, la cual se basa en la conservación de vehículos, imitando la conservación de masa de un fluido:

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q\big(k(x,t)\big)}{\partial x} = 0 \quad (2\text{-}6)$$

Esta relación fue propuesta de manera independiente por Lighthill & Whitham (1955) y Richards (1956), conocida en la literatura como LWR. Se define como velocidad de la onda de choque a la velocidad con que fluye la información entre dos estados de tráfico, digamos 1 y 2, y se calcula matemáticamente como:

$$\mu_s = \frac{\Delta q}{\Delta k} = \frac{q_1 - q_2}{k_1 - k_2} \tag{2-7}$$

El modelo LWR posee simplificaciones inherentes que limitan su uso (Blandin et al., 2013; Daganzo, 1995a; Papageorgiou, 1998) al tratar el flujo vehicular como un fluido, tales como:

- No considera diferencias entre conductores, entre vehículos y sus velocidades, no capturando la dispersión vehicular y adelantamiento.
- El modelo asume cambios instantáneos de velocidad (aceleraciones infinitas), lo que además crea discontinuidades en la densidad.
- No captura inestabilidades que se puedan observar ("*stop & go"*).
- No reproduce patrones de histéresis.
- No reproduce el fenómeno de caída de capacidad.

Modelos más complejos (conocidos como de orden superior) nacen a raíz del trabajo seminal de Payne (1971), agregando una ecuación extra que modela la aceleración de vehículos en función de situaciones locales. Daganzo (1995a) realiza una crítica ante las investigaciones a la fecha, indicando que los modelos propuestos poseen fallas estructurales, ya que pueden predecir flujos y velocidades negativas. El autor indica que estudios han demostrado que, a pesar de la complejidad de los modelos de mayor orden, no se obtienen mejoras en relación al modelo LWR. Nuevos modelos han emergido, tales como los propuestos por Zhang (2002), Aw & Rascle (2000) y Colombo (2002) con el objetivo de sanar estas fallas estructurales. Una revisión bibliográfica de los modelos hasta la fecha es propuesta en Blandin et al. (2013), además de presentar un nuevo modelo.

Dado que el flujo vehicular es un proceso físico, es de utilidad la aplicación de propiedades inherentes a este tipo de procesos. Una de las propiedades más utilizadas es la conservación de vehículos, dada por la ecuación (2-6), la cual puede ser discretizada aplicando el esquema de Godunov (Godunov, 1959; Lebacque, 1996). En el ámbito de ingeniería de transporte, esta discretización es conocida como el modelo de transmisión de celdas (*Cell Tranmission Model* o CTM) (Daganzo, 1994, 1995b). Dada condiciones de borde débiles (Strub & Bayen, 2006), es posible estimar y predecir la evolución del tráfico en celdas intermedias. Si una vía es discretizada en H intervalos de tiempo (de largo Δt) e I celdas espaciales (de largo Δx), el modelo puede estimar para cada intervalo de tiempo $h+1$ y celda i la densidad (k_i^{h+1}) según:

$$k_i^{h+1} = k_i^h - \frac{\Delta t}{\Delta x}\left(g\left(k_i^h, k_{i+1}^h\right) - g\left(k_{i-1}^h, k_i^h\right)\right) \tag{2-8}$$

Donde $g\left(k_i^h, k_{i+1}^h\right)$ es el flujo Godunov, y corresponde a la cantidad de vehículos que fluyen de la celda i a la $i+1$ en el intervalo h. Para un intervalo de tiempo h este flujo se define como:

$$g(k_1, k_2) = \begin{cases} Q(k_2) & \text{si } k_c \le k_2 \le k_1 \\ Q(k_c) & \text{si } k_2 \le k_c \le k_1 \\ Q(k_1) & \text{si } k_2 \le k_1 \le k_c \\ \min\left(Q(k_1), Q(k_2)\right) & \text{si } k_1 \le k_2 \end{cases} \tag{2-9}$$

Donde $Q(k)$ es la función flujo (diagrama fundamental) y k_c es la densidad crítica de la función (donde el flujo es máximo). Para asegurar estabilidad en la solución, es necesario garantizar la condición CFL (Courant-Friedrich-Levy), que indica que ningún vehículo avance más de una celda en un intervalo de tiempo (Courant et al., 1967):

$$\Delta t * v_f \le \Delta x \tag{2-10}$$

Donde v_f es la velocidad a flujo libre de los vehículos.

Varias extensiones y mejoras se han propuesto en base a este modelo, como el extendido para redes (Daganzo, 1995b), el con desfase (Daganzo, 1999), el de cambio de modos (Muñoz et al., 2006), el con desfase mejorado (Szeto, 2008) y el CTM basado en velocidades (Work et al., 2010). Este último trabajo propone una discretización basada

en velocidades, de manera de poder incorporar directamente datos de sensores móviles. En general, estos sensores entregan datos de velocidades, a diferencia de los estáticos, que en general entregan densidades. Para una función $V(k)$ invertible, es posible obtener el modelo de transmisión de celdas basado en velocidades (*Cell Transmission Model for velocity* o CTM-v) (Work et al., 2010):

$$v_i^{h+1} = V\left(V^{-1}(v_i^h) - \frac{\Delta t}{\Delta x}\left(\tilde{g}(v_i^h, v_{i+1}^h) - \tilde{g}(v_{i-1}^h, v_i^h)\right)\right) \tag{2-11}$$

Donde el flujo de velocidades (siendo una analogía al flujo de Godunov) es:

$$\tilde{g}(v_1, v_2) = \begin{cases} \tilde{Q}(v_2) & \text{si } v_1 \le v_2 \le v_c \\ \tilde{Q}(v_c) & \text{si } v_1 \le v_c \le v_2 \\ \tilde{Q}(v_1) & \text{si } v_c \le v_1 \le v_2 \\ \min\left(\tilde{Q}(v_1), \tilde{Q}(v_2)\right) & \text{si } v_2 \le v_1 \end{cases} \tag{2-12}$$

Con:

$$\tilde{Q}(v) = V^{-1}(v) * v \tag{2-13}$$

El CTM-v necesita una función velocidad-densidad $V(k)$ invertible. Una función con estas características es la del tipo Smulders (1990) obtenida de Work et al. (2010), donde la relación es:

$$V(k) = \begin{cases} v_{max}\left(1 - \frac{k}{k_J}\right) & \text{si } k \le k_c \\ -w\left(1 - \frac{k_J}{k}\right) & \text{e.o.c} \end{cases} \tag{2-14}$$

Donde v_{max} es la velocidad máxima de la vía, k_J es la densidad máxima admisible, k_c es la densidad crítica y $-w$ es la pendiente de la rama congestionada (Figura 2-3). Para conservar la continuidad del flujo en k_c, se debe satisfacer la relación $\frac{k_c}{k_J} = \frac{w}{v_{max}}$.

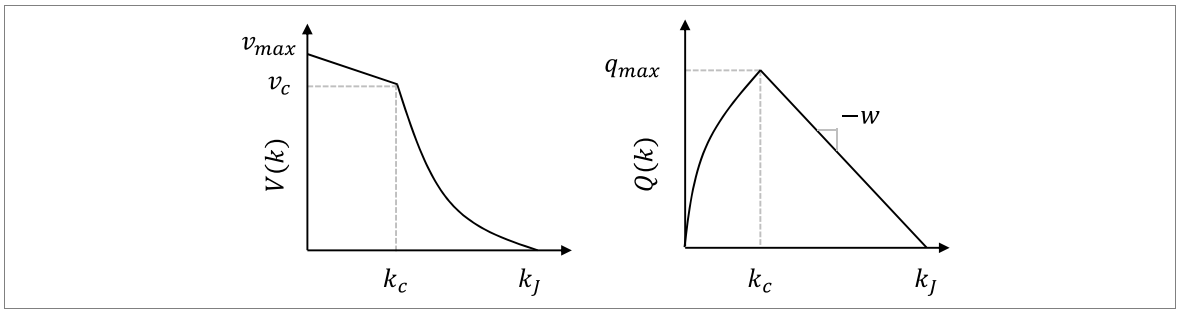

Figura 2-3: Parámetros del diagrama velocidad-densidad y flujo-densidad tipo Smulders

Fuente: Elaboración propia

Los modelos físicos anteriores han sido usados principalmente para estimar densidades o velocidades dadas ciertas condiciones de borde. Generalmente éstas son obtenidas a partir de sensores estáticos, por su muestreo espacial. Típicamente, la información entre las condiciones de borde es incorporada o asimilada al sistema mediante filtros, el cual corrige y actualiza la estimación del modelo a partir de diferentes fuentes de información.

El clásico algoritmo de procesamiento de datos recursivo es el filtro de Kalman (Kalman, 1960). En este filtro el estado de un proceso discreto en el tiempo es gobernado por una ecuación diferencial, lineal y estocástica en presencia de observaciones con ruido blanco (sin autocorrelación), distribución gaussiana y media cero (Gelb, 1974; Maybeck, 1979). Este filtro es óptimo en el sentido de minimizar la covarianza del error estimado. Existen variadas extensiones al filtro de Kalman para el caso de procesos no lineales, como el caso del filtro de Kalman linealizado y el extendido (*Extended Kalman Filter* o EKF) basados en expansiones locales de la función utilizando de series de Taylor (Gelb, 1974).

Algunos estudios utilizan sensores estáticos para asimilar datos. Gazis & Knapp (1971) y Szeto & Gazis (1972) utilizan el filtros de Kalman (lineal en el primero y extendido en el segundo) para monitorear el número de vehículos en un tramo de vía. Wang & Papageorgiou (2005) utilizan el mismo filtro sobre un modelo macroscópico de segundo orden con diferentes configuraciones de espiras. Luego aplican este modelo con datos

reales (Wang et al., 2008) utilizando espiras, radares y videocámaras. Mihaylova et al. (2006) utilizan un modelo macroscópico de segundo orden y aplican un *Unscented Kalman Filter* y lo comparan con un filtro de partículas (*Particle Filter* o PF) desarrollado anteriormente (Mihaylova & Boel, 2004). Sun et al. (2004) utilizan un *Mixture Kalman Filter* (MKF) para tratar la identificación de modos asociada al modelo de tráfico de cambio de modos (*Switching-Mode Model* o SMM).

Otros estudios utilizan sensores móviles para la asimilación de datos. Chu et al. (2005) utilizan un filtro de Kalman adaptativo (*Adaptive Kalman Filter* o AKF) para estimar errores que varían con el tiempo (el filtro de Kalman convencional los asume invariantes) utilizando un microsimulador. Nanthawichit et al. (2003) también utilizan microsimulación al aplicar un filtro de Kalman. Work et al. (2008) utilizan un *Ensemble Kalman Filter* (EnKF) ante la no-linealidad del CTM-v y la no-diferenciabilidad de la función de flujo de éste. Se utilizan datos reales del experimento *Mobile Century*, donde se recolectó datos a partir de celulares GPS en la autopista I880 de California. Herrera & Bayen (2010) utilizan dos métodos. El primero basado en el filtro de Kalman con cambio de modos donde se identifica condiciones de flujo libre o congestión y se obtiene un sistema de ecuaciones para aplicar el filtro. El segundo es la relajación Newtoniana, utilizado en el área de mecánica de fluidos. Se utilizan los datos del experimento *Mobile Century* y de NGSIM (datos altamente desagregados de una autopista real).

Herrera (2009) propone la utilización de una heurística basada en filtros (*Filter-based heuristic* o FBH) la cual permite incorporar la información de observaciones, sin la necesidad de identificar el modo existente, ni conocer la matriz de covarianzas del proceso y observación. Una vez obtenida la estimación de k_i^{h+1} de la ecuación (2-8), se procede a actualizar el valor estimado partir de la siguiente ecuación:

$$\hat{k}_i^{+,h} = \hat{k}_i^h + L_i^h \cdot (k_i^{o,h} - \hat{k}_i^h) \qquad (2\text{-}15)$$

Donde $\hat{k}_i^h$ es la estimación *a-priori* de la densidad, $k_i^{o,h}$ la densidad observada[6], y $\hat{k}_i^{+,h}$ la estimación *a-posteriori* de la densidad de la celda i en intervalo de tiempo h, y L_i^h una variable entre cero y uno. La ecuación (2-15) se puede entender como una combinación lineal entre la densidad *a-priori* y la densidad observada de las celdas. La variable L_i^h imita la ganancia de un filtro, corrigiendo la estimación basada en su valor. Su valor se determina basándose en la confianza relativa existente entre la estimación *a-priori* y el valor observado. Digamos que la relación entre la densidad real y estimada es la siguiente:

$$\hat{k}_i^h = k_i^h + \xi_i^h \tag{2-16}$$

$$k_i^{o,h} = k_i^h + \phi_i^h \tag{2-17}$$

Donde k_i^h es la densidad real, ξ_i^h el error asociado a la densidad estimada por el modelo, y ϕ_i^h a la densidad de la observación. Si las varianzas de los errores ξ_i^h y ϕ_i^h son respectivamente $\sigma_{\xi_i^h}$ y $\sigma_{\phi_i^h}$, en algunas circunstancias es posible demostrar que el valor óptimo de L_i^h es (Herrera, 2009):

$$L_i^{h^*} = \frac{\sigma_{\xi_i^h}}{\sigma_{\xi_i^h} + \sigma_{\phi_i^h}} = \frac{1}{1 + \frac{\sigma_{\phi_i^h}}{\sigma_{\xi_i^h}}} \tag{2-18}$$

Si se espera que el modelo reproduzca de mejores estimaciones en relación a las observaciones obtenidas, se tendrá un valor de $L_i^{h^*}$ cercano a cero. Asimismo, si se valoran más las observaciones, $L_i^{h^*}$ tendrá un valor cercano a uno. Mejoras en el modelo permiten que $\frac{\sigma_{\phi_i^h}}{\sigma_{\xi_i^h}}$ disminuya.

Asumiendo un L_i^h invariante en el tiempo y en el espacio ($L_i^h = L$), Herrera (2009) encontró valores óptimos según la cantidad de observaciones. Para un alto número de observaciones, es conveniente un valor de entre 0,4 y 0,6; para un bajo número de observaciones, sobre 0,8.

---

[6] Se puede obtener a partir del diagrama fundamental utilizando $\hat{u}_i^h$, la velocidad observada.

3. METODOLOGÍA

En este capítulo se presenta la metodología propuesta para alcanzar los objetivos de esta investigación. La metodología supone conocidas las trayectorias de todos los vehículos de una sección de autopista de largo determinada, durante un periodo de tiempo definido. La fuente de estas trayectorias puede ser de datos obtenidos de la realidad, o bien, mediante microsimulación. Para el análisis, se trabajará con una discretización determinada de la vía, en I celdas espaciales y H intervalos de tiempo.

Primero se presenta cómo, a partir de estas trayectorias, se generan distintos escenarios hipotéticos de información entregada por sensores móviles. Estos escenarios son utilizados como base de análisis para responder a los objetivos propuestos. Luego, se presentan secciones asociadas a los objetivos, proponiendo metodologías para alcanzarlos.

3.1 Datos de sensores móviles y generación de escenarios

De la totalidad de los vehículos presentes en el conjunto de trayectorias, los vehículos equipados con sensores representarán una proporción P del total de vehículos. Esta proporción se conoce como tasa de penetración.

Cada uno de los vehículos equipados presentará un muestreo temporal de datos, donde cada vehículo equipado envía datos a intervalos regulares de tiempo de largo τ. A este periodo τ le llamaremos intervalo de muestreo. Por lo tanto, $z = 1/\tau$ es la tasa o frecuencia de muestreo (cantidad de datos que cada vehículo equipado envía por unidad de tiempo).

Los datos que enviará el vehículo equipado g, en el instante t, serán su posición (x_t^g) y velocidad promedio (u_t^g) de los últimos S segundos, para aminorar los cambios inestables de velocidad (Herrera & Bayen, 2010). En otras palabras, la velocidad que el vehículo equipado g enviará en el instante t será:

$$u_t^g = \frac{x_t^g - x_{t-S}^g}{S} \quad (3\text{-}1)$$

Además, la información podrá contar con una etiqueta que permita re-identificar el vehículo (escenario pseudoanónimo).

La información anterior la envían sólo aquellos vehículos equipados. Una vez definidos la tasa de penetración P y la tasa de muestreo z será posible conocer la posición y el tiempo en el que se genera y envía un dato. La Figura 3-1 muestra el instante y ubicación en que se podrían generar datos. Con líneas punteadas se muestra la trayectoria real del vehículo.

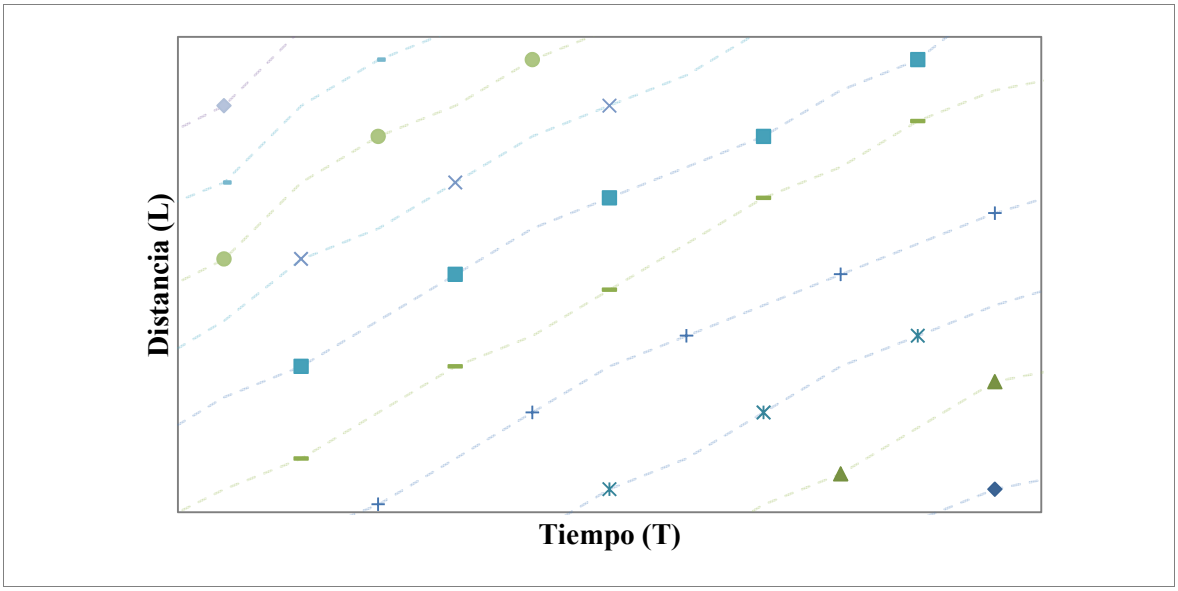

Figura 3-1: Datos enviados por vehículos equipados con etiquetas (pseudoanonimidad)

Los distintos símbolos representan a un vehículo en específico (etiquetas).

Fuente: Elaboración propia

A partir de esta información, es posible reconstruir trayectorias uniendo puntos en caso que la información contenga una etiqueta que identifique individualmente al vehículo (Figura 3-1), como el caso de un escenario pseudoanónimo. Por razones de preservación de la privacidad, puede ser posible que la información no contenga estas etiquetas, donde identificar trayectorias de manera directa no es posible (Figura 3-2).

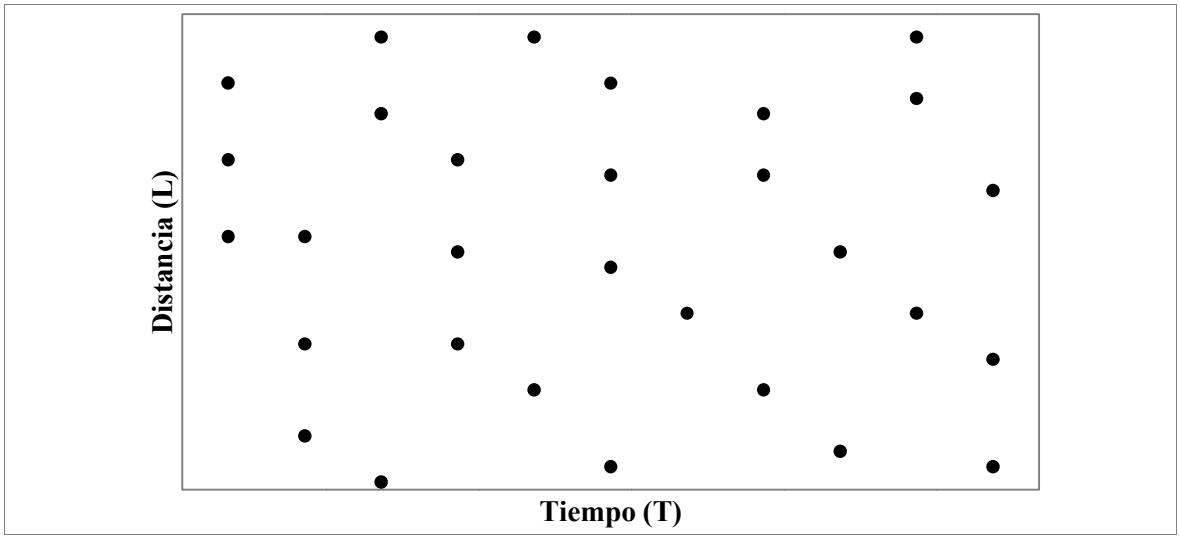

Figura 3-2:Datos enviados por vehículos equipados sin etiquetas (anonimidad)

Fuente: Elaboración propia

Así, la cantidad de datos disponibles (i.e. la cantidad de puntos y si tienen etiquetas) depende directamente de P y z, mientras que el tipo de dato depende del tipo de anonimidad del escenario. Por lo tanto, cada combinación de estas variables define un escenario en particular. Notar que la única diferencia entre un escenario anónimo y otro que no lo es, es la existencia de una etiqueta para cada dato que permita identificar el vehículo que lo envió (el resto de la información enviada es la misma). Una vez que se genera el dato, éste se asocia a una celda i e intervalo de tiempo h en específico.

3.2 Cuantificación de datos

Para cada escenario generado, o combinación de P y z, se desea obtener una medida que permita determinar la cantidad total de observaciones con la que se trabaja. Esto es, por ejemplo, la cantidad de puntos que posee la Figura 3-2. Considerando un tramo de vía de largo L durante un periodo de tiempo T, una medida aproximada de la cantidad de observaciones (Obs) esperada para un escenario (P,z) se puede calcular como:

$$\mathrm{Obs} = \bar{t_v} \cdot z \cdot P \cdot N_{veh} \tag{3-2}$$

Donde N_{veh} es el número total de vehículos y $\bar{t_v}$ es el tiempo de viaje promedio de los vehículos equipados. Así, $\bar{t_v} \cdot z$ es la cantidad de datos que en promedio envía un vehículo equipado, y $P \cdot N_{veh}$ es la cantidad de vehículos equipados para los distintos escenarios generados. Asimismo, y considerando una vía con N_L pistas, es posible obtener una medida de la densidad espacio-temporal de observaciones (DObs) en la vía:

$$\mathrm{DObs} = \frac{\mathrm{Obs}}{L \cdot T \cdot N_L} = \frac{\bar{t_v} \cdot P \cdot N_{veh}}{\tau \cdot L \cdot T \cdot N_L} \qquad \left[\frac{\mathrm{obs}}{\mathrm{km-h-pista}}\right] \tag{3-3}$$

Dada la aleatoriedad de los escenarios, estas medidas pueden no ser exactas. Por ejemplo, la presencia de salidas y entradas en la autopista genera una mayor variabilidad en el tiempo de viaje. Por lo tanto, se desea cuantificar estas medidas de cantidad de observaciones y verificar su ajuste. Se realizarán replicaciones para estimar un número esperado de observaciones por escenario, y además capturar la variabilidad de este número.

3.3 **Efecto de anonimidad de sensores**

Es de interés cuantificar el efecto que tiene la anonimidad de sensores en la calidad de la estimación de estados de tráfico. Para esto, se analizará si es posible obtener, a partir de un escenario pseudoanónimo (Figura 3-1), una mejor estimación de la velocidad que un escenario anónimo (Figura 3-2). Dado que en ambos casos la metodología de estimación de estados de tráfico es la misma, en definitiva interesa determinar cuál de los dos entrega mejor input al modelo. En esta investigación se asumirá que no es posible garantizar la anonimidad con datos pseudoanónimos, pero sí en el caso de que no sea posible para el sistema asociar a cada información, una etiqueta identificadora.

En aquellas celdas con observaciones, se compararán dos métodos de estimación de velocidades en celdas con observaciones (u_i^h): promedios simples (aritméticos) de todas las observaciones en la celda i en el intervalo h y definiciones generalizadas (ecuación (2-3)) a partir de las mismas observaciones. El primer método permite obtener estimaciones de velocidad sin necesidad de violar la privacidad, mientras que el

segundo, al necesitar trayectorias para estimar la velocidad, necesita identificar individualmente vehículos.

Estas estimaciones se compararán con la realidad, de manera cuantificar el efecto de preservar la privacidad en la estimación, al tener menos información. Esto dará indicios de cuán valiosa es la anonimidad para la estimación de estados de tráfico, o visto de otra forma, si preservando la privacidad es aún posible obtener buenas estimaciones.

3.4 Comparación y requerimiento de heurística

De las $I \cdot H$ celdas-intervalos, sólo una proporción tendrá observaciones de velocidad a partir de los sensores móviles (u_i^h). Para el resto, es necesario estimarla. Es decir, para cada celda i e intervalo de tiempo h se desea estimar su velocidad ($\hat{v}_i^h$) para generar un campo de velocidades. La estimación de este campo de velocidades se realizará mediante tres métodos, que se denominarán de la siguiente manera:

- Promedios simples históricos: ecuación (2-4).
- Promedios simples interpolados: ecuación (2-5).
- FBH-v

FBH-v es la combinación del esquema CTM-v (ecuación (2-11)) con el filtro FBH (ecuación (2-15)). Ahora la incorporación de observaciones se realiza de la siguiente manera:

$$\hat{v}_i^{+,h+1} = \hat{v}_i^{h+1} + L^* \cdot (u_i^{h+1} - \hat{v}_i^{h+1}) \tag{3-4}$$

En lugar de utilizar una ganancia de filtro L fija o dependiente del tiempo o espacio, se utilizará un L^* que depende del escenario en que se encuentre. Dentro de un rango de valores posibles de la ganancia, el método elegirá aquel que minimice el error en la estimación (dado que se conoce la realidad). Por lo tanto, la elección de L^* está directamente relacionada con la penetración (P) y frecuencia de muestreo (z) utilizada.

A partir de los tres métodos de estimación de estados de tráfico (promedios simples históricos, promedios simples interpolados y la heurística FBH-v) es posible analizar,

para distintos valores de P y z, qué método entrega mejores resultados. También es posible vislumbrar posibles efectos de saturación de información: en qué circunstancias más datos no entregan mucha información adicional.

Como métrica de error, se utilizará el error relativo medio (ERM). Se compara la velocidad real ($\bar{\bar{v}}_i^h$) y la estimada ($\hat{v}_{i,j}^{h,r}$) de la celda i en el intervalo h, para cada escenario j y réplica r, de manera de construir un indicador para cada escenario j a partir de R réplicas. Se construye el error relativo medio (ERM_j) a partir del error relativo como:

$$ER_{j,r} = \frac{1}{I \cdot H} \sum_{h=1}^{H} \sum_{i=1}^{I} \frac{\left| \hat{v}_{i,j}^{h,r} - \bar{\bar{v}}_i^h \right|}{\bar{\bar{v}}_i^h} \tag{3-5}$$

$$ERM_j = \frac{\sum_{r=1}^{R} ER_{j,r}}{R} \tag{3-6}$$

Es también posible crear intervalos de confianza para el error, para cada escenario j (a partir de R réplicas) y para un nivel de significancia α:

$$ERM_j \pm t_{(R-1,1-\frac{\alpha}{2})} \sqrt{\frac{\sigma_j^2}{R}} \tag{3-7}$$

Donde σ_j^2 es:

$$\sigma_j^2 = \frac{\sum_{r=1}^{R} (ER_{j,r} - ERM_j)^2}{R - 1} \tag{3-8}$$

También se puede definir la precisión de la estimación como:

$$pe_j = \frac{t_{(R-1,1-\frac{\alpha}{2})} \sqrt{\frac{\sigma_j^2}{R}}}{ERM_j} \tag{3-9}$$

3.5 **Métrica para inferir errores**

Se desea obtener una función f(·) que permita, a partir de variables conocidas, obtener un error esperado ε que se comete al estimar estados de tráfico. Las variables conocidas que se utilizarán para predecir este error serán la penetración (P) y la frecuencia de muestreo

(z) de sensores en vehículos equipados. Por lo tanto, $\varepsilon = f(P, z)$. Se considerará como métrica de error el error relativo medio (ecuación (3-6)), es decir $\varepsilon_{ERM} = f(P, z)$.

Para obtener las formas funcionales $\varepsilon = f(P, z)$, en primer lugar, es de conveniencia graficar en un diagrama P v/s z distintos escenarios y su valor de error. A partir de lo observado en este gráfico, se proponen formas funcionales para la función $\varepsilon = f(P, z)$. Es esperable que el error disminuya a medida que aumenta *P* y/o *z*, es decir, *P* y *z* son factores substitutos. Esta información puede ser útil para proponer formas funcionales. Mediante técnicas econométricas, y a partir de los datos generados, se puede ajustar una forma funcional de manera correcta.

3.6 El efecto de la composición de información

Una vez caracterizada la función $\varepsilon = f(P, z)$, se propone evaluar el efecto de la penetración *P* y la tasa de muestreo *z* en el error ε mediante sus elasticidades, $E_{\varepsilon,P}$ y $E_{\varepsilon,z}$, respectivamente.

$$E_{\varepsilon,P} = \frac{\Delta\varepsilon}{\Delta P} \cdot \frac{P}{\varepsilon} = \frac{\partial\varepsilon}{\partial P} \cdot \frac{P}{\varepsilon} = \frac{\partial \ln(\varepsilon)}{\partial \ln(P)} \tag{3-10}$$

$$E_{\varepsilon,z} = \frac{\Delta\varepsilon}{\Delta z} \cdot \frac{z}{\varepsilon} = \frac{\partial\varepsilon}{\partial z} \cdot \frac{z}{\varepsilon} = \frac{\partial \ln(\varepsilon)}{\partial \ln(z)} \tag{3-11}$$

Estas elasticidades indican el cambio porcentual en el error al variar *P* o *z* en un 1%. Dado que un cambio de un X % en P o z provoca el mismo aumento en el número de observaciones[7], es correcto comparar estas elasticidades.

Luego de obtener las elasticidades, es posible compararlas para diferentes valores de P y z. Por ejemplo, se plantea graficar las diferencias entre éstas o su razón. A partir de esto, es posible determinar combinaciones de P y τ en las cuales es de conveniencia aumentar

[7] Recordar que $Obs = (\overline{t_v} \cdot N_{veh}) \cdot P \cdot z$. Un aumento en la penetración en X %: $Obs = (\overline{t_v} \cdot N_{veh}) \cdot P \cdot (1 + X\%) \cdot z$. Un aumento en la frecuencia en X %: $Obs = (\overline{t_v} \cdot N_{veh}) \cdot P \cdot z \cdot (1 + X\%)$. Por lo tanto, Obs aumenta en X% independiente de dónde provenga el aumento.

la penetración en lugar de la frecuencia, y viceversa. También puede suceder que es indiferente de dónde provenga el aumento.

3.7 **Implementación computacional**

A grandes rasgos, la implementación computacional de la metodología consta de tres etapas principales: (i) obtención de trayectorias, (ii) programación del simulador y (iii) generación de resultados (Figura 3-3).

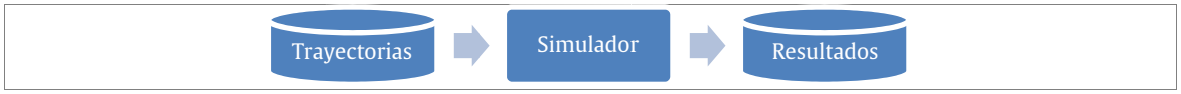

Figura 3-3: Esquema implementación computacional

Fuente: Elaboración propia

La primera etapa es la codificación de las trayectorias, que son archivadas en una base de datos MySQL. Luego, con un simulador implementado en Java (utilizando Eclipse) conectado a la base de datos, se generan nuevas tablas de información con los distintos escenarios ficticios a analizar (subconjunto de información de las trayectorias totales). A éstas se aplican las distintas metodologías de estimación de estados de tráfico, obteniendo los resultados de las estimaciones. Estos posteriormente se analizan mediante la utilización de Microsoft Excel, MATLAB y STATA.

4. RESULTADOS Y ANÁLISIS

En este capítulo se presentan los principales resultados obtenidos al aplicar la metodología del capítulo anterior. Primero, se muestran los resultados obtenidos a partir de una base de datos de trayectorias reales (*Next Generation Simulation* o NGSIM), las cuales fueron recolectadas con videocámaras de alta velocidad en una porción corta de autopista. Esto permite dar a luz resultados basados en comportamientos reales de los vehículos. Luego, se utiliza microsimulación para generar trayectorias vehiculares en un tramo de autopista más largo que el anterior. Esto dado que no existe una base de datos que contenga trayectorias reales en una porción larga de autopista. Permitirá analizar la implicancia de tramos más largos en la metodología. Finalmente, se establece una discusión entre los resultados de ambas bases de datos.

4.1 NGSIM

Primero se ofrece una descripción de la base de datos usada, para luego dar lugar a los resultados y análisis al aplicar la metodología.

4.1.1 Descripción de la base de datos

Se utilizó la base de datos de trayectorias NGSIM (FHWA, 2007). Esta base de datos fue obtenida a partir de información recolectada a través ocho cámaras, en un tramo de 640 m en la autopista US 101 de Los Ángeles, CA (Figura 4-1). Posee trayectorias de 6.101 vehículos que cruzaron la sección de 5 pistas durante 45 minutos del periodo punta-mañana (de 7:50 a.m. a 8:35 a.m.), cuando la congestión va aumentando. Posee una rampa de entrada y salida, incluyendo una pista de transición entre ambas rampas.

Figura 4-1: Autopista US 101 en Los Ángeles, California

Fuente: Google Maps

Para cada vehículo, se tiene el dato de su posición (incluyendo la pista en que circula), velocidad instantánea y aceleración instantánea cada 0,1 segundos.

Esta sección de autopista fue discretizada en $I = 12$ celdas espaciales de $\Delta x = 50$ m (Figura 4-2), obteniendo un largo total de 600 m. A partir de los datos, y dada una discretización temporal, es posible obtener el campo de velocidades real al aplicar definiciones generalizadas (ver sección 2.2). La Figura 4-3 muestra el campo de velocidades real para una discretización temporal de $\Delta t = 1{,}5$ s.

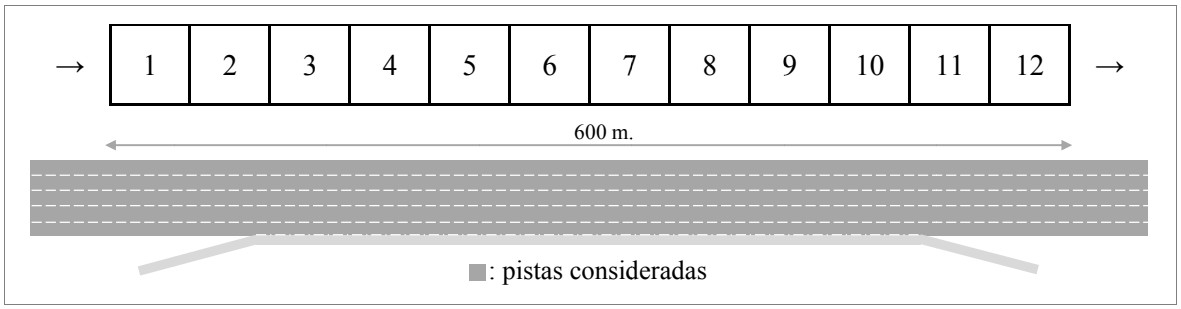

Figura 4-2: Discretización de la autopista US 101

Fuente: Elaboración propia

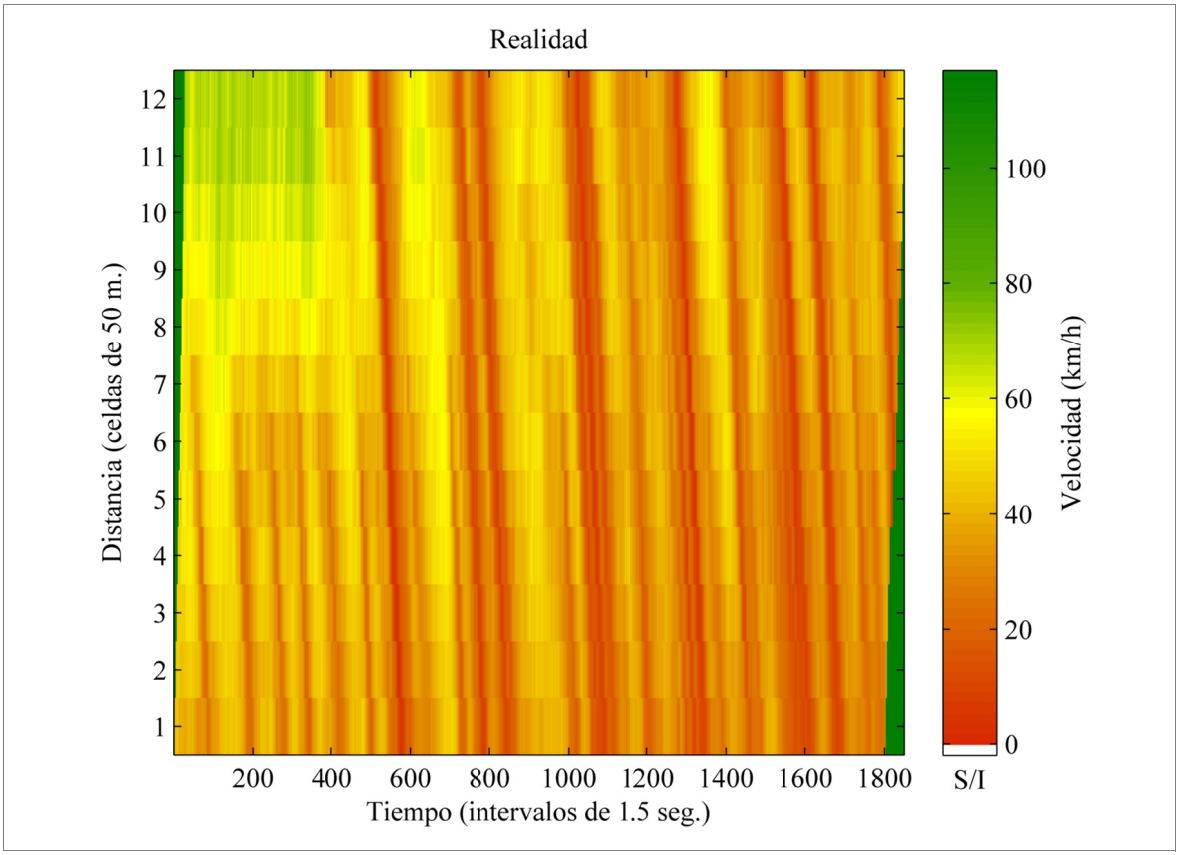

Figura 4-3: Campo de velocidades real para $\Delta t = 1{,}5\ s$ [NGSIM]

Fuente: Elaboración propia

4.1.2 Información de sensores móviles y generación de escenarios

Cada vehículo equipado enviará, cada τ segundos, su posición y velocidad promedio de los últimos $S = 3$ segundos, además de una etiqueta que permita la re-identificación del vehículo. La etiqueta tiene como objetivo evaluar el efecto de la anonimidad en la estimación de estados de tráfico.

Del total de vehículos en la base de datos (6.101), se eligió un subconjunto de vehículos que estarán equipados con sensores según los siguientes valores (en %): $\vec{P} = \{1, 2, 3, 4, 5, 10, 15, 20, 25, 30\}$. Esta elección se realizó en base a valores utilizados en la literatura y posibles tasas de penetración reales, más una proyección al futuro.

Asimismo, se consideraron los siguientes intervalos de tiempo de envío de datos para los vehículos equipados (en segundos): $\vec{\tau} = \{1, 3, 5, 10, 20, 30, 60, 120\}$. El inverso de estos valores define la tasa o frecuencia de muestreo z. El tiempo de viaje en la sección es de aproximadamente dos minutos, por lo que el último valor de intervalo de tiempo entrega aproximadamente un dato por vehículo equipado con sensor. El menor intervalo usado corresponde a la frecuencia máxima típica de estos sensores (Frank et al., 2012).

Dado los valores anteriores, en total existen 80 escenarios, donde cada escenario es una combinación de P y τ. Finalmente, para cada escenario se realizaron $R = 15$ réplicas. Es decir, para un P y τ dado, se eligieron de manera aleatoria un porcentaje P de vehículos, como también el instante en que empiezan a enviar datos (a pesar de que todos los vehículos equipados envían datos cada τ segundos, no todos lo enviarán en el mismo instante necesariamente).

4.1.3 Cuantificación de información

Para cada escenario generado, es posible obtener el número de observaciones promedio, considerando las 15 réplicas. Esta información se presenta en la Tabla 4-1.

Tabla 4-1: Número de observaciones promedio por escenario [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	3387	1129	676	338	169	112	56	28
2	6705	2234	1341	667	334	221	112	56
3	10145	3382	2026	1012	507	337	167	84
4	13275	4424	2654	1327	664	439	218	110
5	16992	5663	3399	1697	851	564	284	144
10	33911	11302	6783	3388	1697	1128	563	281
15	50699	16898	10134	5062	2534	1683	839	420
20	67898	22632	13574	6786	3397	2258	1127	565
25	84834	28278	16954	8468	4240	2822	1414	706
30	101818	33944	20349	10162	5095	3383	1692	841

Fuente: Elaboración propia

Esta tabla fue obtenida a partir de las observaciones entre el minuto 5 y el minuto 45 de la simulación, para las 12 celdas en inspección.

Para estimar el número de observaciones de manera a-priori de acuerdo a la ecuación (3-2), se debe estimar un número de vehículos en la autopista (N_{veh}) y un tiempo de viaje de vehículos equipados ($\bar{t_v}$). La Tabla 4-2 muestra el error relativo de la estimación en caso de conocer los valores reales de N_{veh} y $\bar{t_v}$ (iguales para todos los escenarios).

Tabla 4-2: Error relativo de la estimación de número de observaciones (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	0,7	0,7	0,9	0,9	0,8	1,3	1,0	2,0
2	1,8	1,8	1,7	2,3	2,2	2,9	1,5	1,3
3	0,9	0,9	1,0	1,1	1,0	1,2	1,9	1,8
4	2,8	2,8	2,8	2,8	2,7	3,6	4,2	3,5
5	0,4	0,4	0,4	0,5	0,2	0,8	0,1	1,2
10	0,6	0,6	0,6	0,7	0,5	0,8	1,0	1,3
15	0,9	0,9	1,0	1,1	1,0	1,4	1,7	1,4
20	0,5	0,5	0,5	0,5	0,4	0,7	0,9	0,6
25	0,5	0,5	0,6	0,7	0,6	0,7	0,5	0,7
30	0,5	0,5	0,6	0,7	0,4	0,8	0,8	1,4

Fuente: Elaboración propia

El error tiene como cota máxima un 4%, mientras que en muchos escenarios la estimación posee muy bajo error. La variabilidad del número de observaciones por escenario se puede observar a través de la precisión, definida anteriormente según la ecuación (3-9), pero utilizando el número de observaciones en vez del error. En este caso la precisión se define como la razón entre el intervalo de confianza del número de observaciones obtenido a partir de las réplicas ($t_{(R-1,1-\frac{\alpha}{2})}\sqrt{\frac{\sigma_j^2}{R}}$) y el número promedio de observaciones. La Tabla 4-3 muestra la precisión para del número de observaciones por escenario. Para $P \geq 10\%$ el número de observaciones posee muy poca variabilidad, tomando valores cercanos al 1%.

Tabla 4-3: Precisión del número promedio de observaciones según escenario (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	2,1	2,0	2,1	2,1	2,3	2,7	3,1	5,2
2	1,8	1,9	1,8	1,9	1,9	2,0	1,6	4,8
3	2,9	2,9	2,9	3,0	3,0	3,0	2,9	5,8
4	2,1	2,1	2,1	2,2	2,2	2,2	2,7	4,1
5	1,7	1,8	1,7	1,7	1,8	2,1	2,9	3,5
10	0,6	0,6	0,6	0,7	0,7	0,7	1,1	2,0
15	1,1	1,1	1,1	1,1	1,1	0,9	1,3	2,7
20	0,9	0,9	0,9	0,9	0,8	0,9	1,1	1,9
25	0,6	0,6	0,6	0,6	0,6	0,6	0,8	1,2
30	0,5	0,5	0,5	0,5	0,6	0,6	0,7	1,4

Fuente: Elaboración propia

También es posible obtener una medida de la densidad espacio-temporal de las observaciones (Tabla 4-4), dividiendo la cantidad de información por la distancia, tiempo inspeccionado y número de pistas (0,6 km, 40 minutos y 5 respectivamente).

Tabla 4-4: Densidad de observaciones por kilómetro-minuto-pista [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	28	9	6	3	1	1	0	0
2	56	19	11	6	3	2	1	0
3	85	28	17	8	4	3	1	1
4	111	37	22	11	6	4	2	1
5	142	47	28	14	7	5	2	1
10	283	94	57	28	14	9	5	2
15	423	141	85	42	21	14	7	4
20	567	189	113	57	28	19	9	5
25	708	236	142	71	35	24	12	6
30	850	283	170	85	43	28	14	7

Fuente: Elaboración propia

En resumen, para poder obtener el número de observaciones de una vía, es necesario conocer la frecuencia de envío de vehículos, la penetración, el número de vehículos y el tiempo de viaje promedio de los vehículos equipados de la sección. El error relativo que se cometa en la estimación de estas variables afecta de manera directa al número estimado de observaciones de una vía, a partir de la relación existente entre estas variables. La variabilidad del número de observaciones estimada depende del escenario en que se esté, y está relacionada con la estructura de la matriz origen-destino de entradas y salidas de la autopista. Para este conjunto de trayectorias, la variabilidad de la estimación es pequeña, y siempre menor a un 5% y en la gran mayoría de las veces menor a 3%.

4.1.4 Efecto de anonimidad de sensores

En esta sección se compararán dos métodos de estimación de velocidad de una celda a partir de observaciones de sensores móviles. El primero de ellos, promedios simples, funciona tanto para escenarios anónimos como pseudoanónimos. El segundo utiliza las definiciones generalizadas, y por lo tanto necesita reconstruir la trayectoria del vehículo

para obtener una estimación. Primero se analizará cada uno los métodos por separado, para posteriormente compararlos. La discretización, en este caso, considera $H = 12$ celdas espaciales, que se actualizan cada $\Delta t = 30$ s.

a. **Promedios simples**

La Tabla 4-5 presenta el error relativo medio cuando se utiliza promedios simples. Como se esperaba a medida que aumenta la cantidad de observaciones producto de un aumento en penetración o disminución del intervalo de muestreo el error relativo disminuye. Es decir, a medida que existen más datos disponibles, no sólo se estiman más celdas-intervalo, sino también se estiman de mejor manera. Cabe recordar que el error se computa exclusivamente para las celdas con observaciones (no se imputa datos). Cada escenario tiene un número distinto de celdas-intervalo sin estimación. De hecho, la Tabla 4-6 muestra el porcentaje de celdas sin estimación.

Tabla 4-5: Error relativo medio para promedios simples (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**
1	16	17	18	20	23	23	23	23
2	15	15	16	19	22	22	23	23
3	14	14	15	18	21	22	23	23
4	13	13	14	17	20	22	22	23
5	12	12	13	16	19	21	22	22
10	9	9	9	12	16	18	20	20
15	7	7	8	10	13	16	18	19
20	6	6	7	8	12	14	17	18
25	6	6	6	7	10	12	16	17
30	6	6	6	7	9	12	16	17

Fuente: Elaboración propia

Para penetraciones sobre 20% e intervalos de muestreo bajo 10 segundos, el error relativo se encuentra cercano a 6%. En estos escenarios, se logra estimar prácticamente el total de celdas-intervalo. Para penetraciones bajas e intervalos de muestreo altos, el

error alcanza niveles cercanos al 23%. En estos casos, la mayoría de las celdas-intervalo no tienen datos.

Tabla 4-6: Porcentaje de celdas sin estimación para promedios simples [NGSIM]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**
1	49	51	57	73	85	90	95	97
2	26	29	35	54	72	80	90	95
3	14	15	20	39	61	72	85	92
4	8	9	13	29	52	65	81	90
5	4	5	8	22	44	58	76	87
10	2	2	3	7	21	34	59	78
15	2	2	2	3	11	21	47	71
20	2	2	2	2	6	13	37	65
25	2	2	2	2	4	9	30	61
30	2	2	2	2	3	6	25	57

Fuente: Elaboración propia

La precisión de la estimación se presenta en la Tabla 4-7.

Tabla 4-7: Precisión de la estimación para promedios simples (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	4	4	4	4	6	5	8	15
2	2	2	2	2	3	3	5	11
3	3	3	3	3	2	3	4	5
4	3	3	2	2	3	4	5	6
5	2	2	1	**1**	2	2	3	3
10	2	3	2	2	2	2	2	4
15	2	2	2	2	2	2	2	3
20	1	1	1	**1**	2	2	1	3
25	2	2	2	**1**	1	2	2	2
30	1	1	1	**1**	1	2	1	2

Fuente: Elaboración propia

Para niveles altos de penetración, se alcanzan niveles altos de precisión en la estimación, independiente del intervalo de muestreo. Para bajos niveles de penetración, un intervalo de muestreo pequeño entrega mejor precisión, pero el mismo comportamiento no se replica claramente para intervalos mayores.

b. **Definiciones generalizadas**

En el caso de definiciones generalizadas, a medida que aumenta la penetración o disminuye el intervalo de muestreo, el error relativo disminuye, excepto para $\tau = 60$, donde el error llega a niveles del 40%, sin un patrón claro (Tabla 4-8).

Tabla 4-8: Error relativo medio para definiciones generalizadas (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**
1	16	16	16	17	23	29	38	-
2	15	15	15	16	23	29	37	-
3	13	13	13	15	22	29	40	-
4	13	12	13	15	22	28	40	-
5	11	11	11	14	21	29	40	-
10	8	8	8	11	19	27	39	-
15	6	6	6	10	18	26	39	-
20	5	5	5	9	17	25	38	-
25	4	4	5	8	17	25	38	-
30	4	4	4	8	16	24	38	-

Fuente: Elaboración propia

El porcentaje de celdas sin estimación (Tabla 4-9) es igual que en promedios simples para $\tau \leq 20$, y aumenta para valores mayores, debido a que el método necesita al menos dos datos de un mismo vehículo para reconstruir su trayectoria y así estimar velocidades de celdas-intervalo. De hecho, para $\tau = 120$ no logra estimar para ninguna celda-intervalo, lo que es de esperar si el tiempo de viaje promedio de la sección es del mismo orden, y por lo tanto cada vehículo envió un sólo dato. En el caso de $\tau = 60$, se tiene en promedio dos observaciones por vehículo equipado, lo que nuevamente no permite reconstruir la trayectoria completa.

Tabla 4-9: Porcentaje de celdas sin estimación para definiciones generalizadas [NGSIM]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**
1	49	51	57	73	85	91	98	100
2	26	29	35	54	73	82	96	100
3	14	15	20	39	62	74	95	100
4	8	9	13	30	53	68	93	100
5	4	5	8	22	45	62	92	100
10	2	2	3	7	22	39	85	100
15	2	2	2	3	11	26	80	100
20	2	2	2	2	6	18	76	100
25	2	2	2	2	4	13	72	100
30	2	2	2	2	3	10	70	100

Fuente: Elaboración propia

Finalmente, la razón entre el intervalo de confianza y el error relativo tiende a disminuir a medida que P aumenta para valores $\tau \geq 10$ (Tabla 4-10). Cabe mencionar que un error alto puede producir entregar una alta precisión (el error es más grande que el intervalo), con un bajo valor del indicador.

Tabla 4-10: Precisión de la estimación para definiciones generalizadas (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**
1	4	4	4	4	4	5	10	-
2	2	2	2	2	3	3	5	-
3	3	3	3	2	2	3	5	-
4	3	3	2	2	3	3	3	-
5	2	2	2	**1**	2	2	4	-
10	3	3	2	**1**	1	1	2	-
15	3	3	2	**1**	1	1	1	-
20	2	2	2	**1**	1	1	1	-
25	3	3	2	**1**	1	1	1	-
30	2	2	2	**1**	1	1	1	-

Fuente: Elaboración propia

c. Comparación: el efecto de la anonimidad en la estimación de estados de tráfico

Una vez analizados ambos métodos, es de interés compararlos. Para esto, se analizó la diferencia entre los errores estimados por definiciones generalizadas menos promedios simples (Tabla 4-11). También se realizó un test de medias[8] para confirmar estadísticamente si existen diferencias. Se utilizó un nivel de significancia del 95%.

[8] Para dos medias $\bar{x}_1$ y $\bar{x}_2$ con varianzas σ_1^2 y σ_2^2 el estadístico t es: $t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

Tabla 4-11: Diferencia entre errores relativos medios estimados (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	0*	-1*	-2	-3	1*	6	15	-
2	0*	-1	-2	-2	1*	7	14	-
3	0*	-1	-1	-2	1	7	17	-
4	0*	-1	-1	-2	2	7	18	-
5	0	-1	-1	-2	2	8	18	-
10	-1	-1	-1	0	3	9	19	-
15	-1	-1	-1	0	5	10	20	-
20	-2	-2	-1	1	6	11	21	-
25	-2	-2	-1	1	7	12	22	-
30	-2	-2	-1	1	7	12	22	-

(Definiciones generalizadas menos promedios simples. * indica igualdad estadística)

Fuente: Elaboración propia

En la Tabla 4-1, valores positivos indican que la estimación de definiciones generalizadas produjo mayores errores. El método de definiciones generalizadas pudo estimar levemente mejor o de igual manera para pequeños intervalos de muestreo ($\tau \leq 10$). Para mayores intervalos, promedios simples estimó notablemente mejor, lo que es de esperar ya que, en el caso de definiciones generalizadas, cada trayectoria es reconstruida con menos puntos. De hecho, para $\tau = 60$ no se asegura la reconstrucción completa de la trayectoria (dado que el tiempo de viaje es cercano a dos minutos bajo congestión y menos en flujo libre). Además, para $\tau \leq 30$ segundos, los dos métodos observan velocidades prácticamente en la misma cantidad de celdas. Sin embargo, para $\tau \geq 60$ promedios simples estima velocidades en más celdas.

Por otro lado, la pérdida de precisión en la estimación al preservar la privacidad es baja. Así, promedios simples permite estimar de manera correcta las velocidades.

4.1.5 Comparación y requerimiento de heurística

La estimación del campo de velocidades se realizará mediante los tres métodos expuestos en la sección 3.4. El input que reciben estos métodos proviene del promedio simple de las observaciones de velocidades en cada celda-intervalo. En el caso del

método FBH-v, se considerarán los siguientes valores de $L = \{0{,}1;\ 0{,}3;\ 0{,}5;\ 0{,}7;\ 0{,}9\}$ para computar L^* de cada escenario, como se explica en la sección 3.4. Se utiliza una discretización temporal (Δt) de 1,5 s y discretización espacial (Δx) de 50 m para todos los métodos, lo cual cumple con la condición dada por la ecuación (2-10) de estabilidad.

Asimismo, es necesario establecer parámetros de la relación $V(k)$ del tipo Smulders dada por la ecuación (2-14), los cuales fueron calibrados a partir de datos de Herrera (2009):

Tabla 4-12: Parámetros para relación Smulders [NGSIM]

Parámetro	Valor
q_{max}	$2040 \frac{\text{veh}}{\text{hr} \cdot \text{pista}}$
v_{max}	$116{,}97 \frac{\text{km}}{\text{hr}}$
v_c	$98{,}14 \frac{\text{km}}{\text{hr}}$
k_c	$20{,}51 \frac{\text{veh}}{\text{km} \cdot \text{pista}}$
k_j	$127{,}38 \frac{\text{veh}}{\text{km} \cdot \text{pista}}$
w	$18{,}83 \frac{\text{km}}{\text{hr}}$

Fuente: Elaboración propia a partir de Herrera (2009)

Con los métodos definidos, es posible obtener un campo de velocidades estimado para cada escenario y réplica. La Figura 4-4 muestra el campo de velocidades real y estimado para un escenario donde $P = 5\%$ y $\tau = 10 \frac{\text{seg}}{\text{obs} * \text{veh}}$.

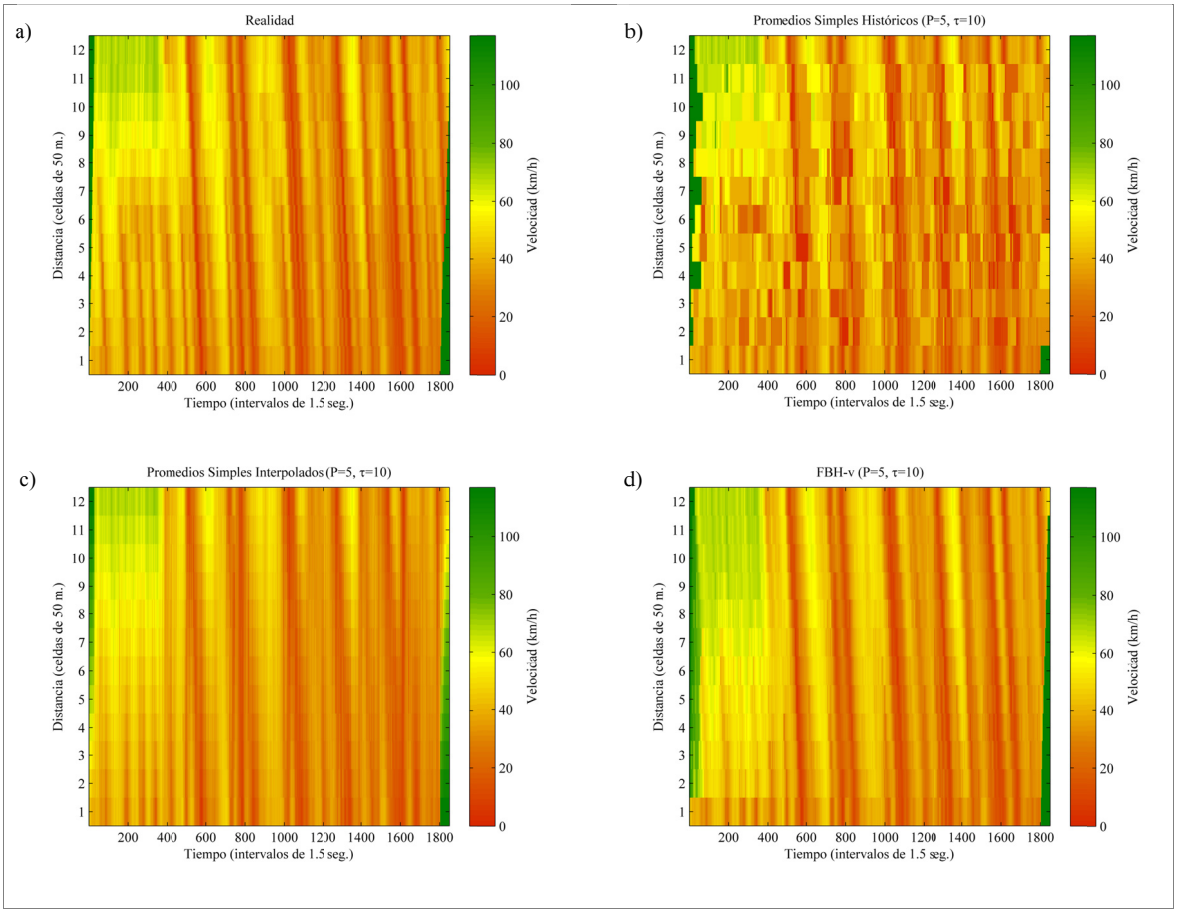

Figura 4-4: Campo de velocidades real y estimado para un escenario con $P = 5\%$ y $\tau = 10\ s$ [NGSIM]

a) Real; b) Promedios simples históricos; c) Promedios simples interpolados y d) FBH-v.

Estimación realizada a partir de 1.697 observaciones.

Fuente: Elaboración propia

Para este escenario, visualmente FBH-v parece estimar correctamente el campo. Por otro lado, promedios simples interpolados tiene un parecido a la realidad, pero la velocidad con que se propaga la información en la cola es muy alta (líneas casi verticales). Promedios simples históricos al no utilizar las condiciones de borde recae sólo en observaciones móviles internas. En este caso, no es inmediata la identificación del campo de velocidades real.

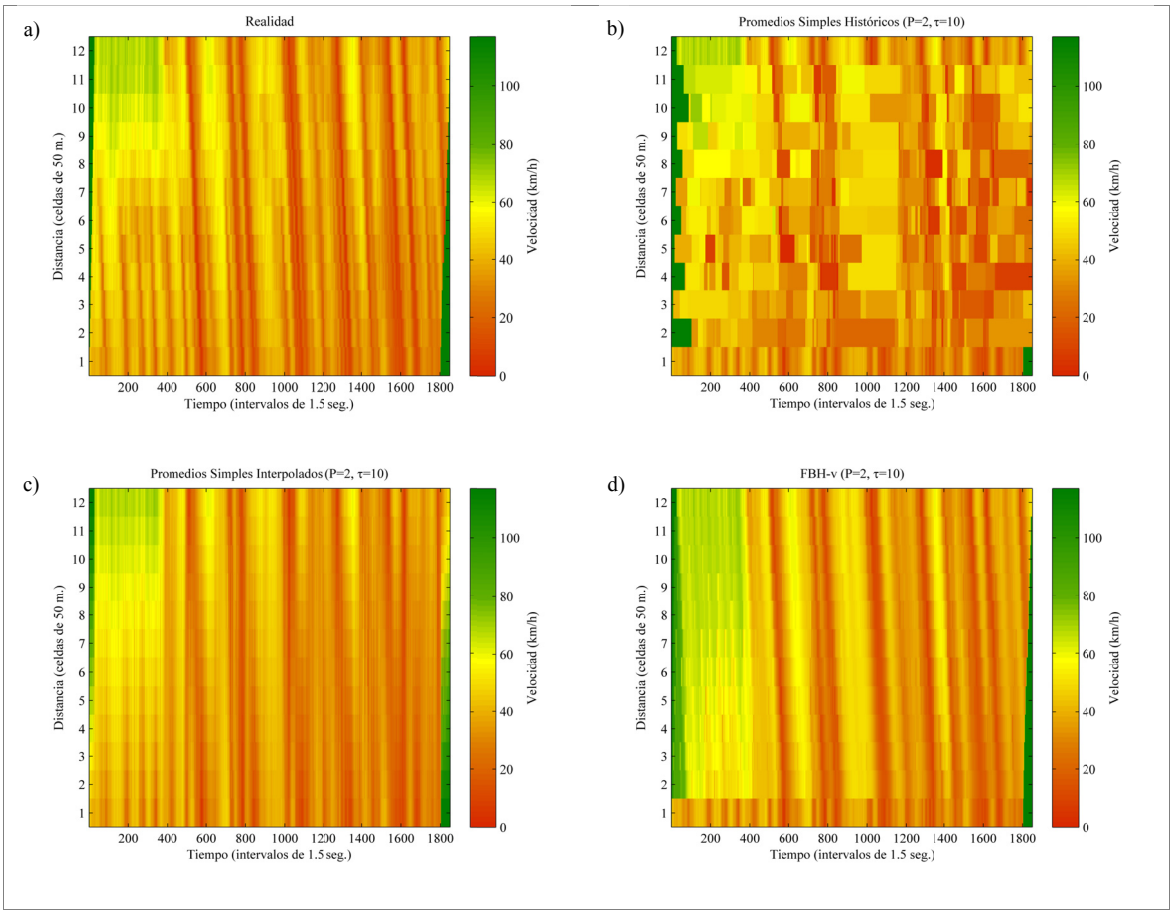

Figura 4-5: Campo de velocidades real y estimado para un escenario con $P = 2\%$ y $\tau = 10\ s$ [NGSIM]

a) Real; b) Promedios simples históricos; c) Promedios simples interpolados y d) FBH-v.

Estimación realizada a partir de 667 observaciones.

Fuente: Elaboración propia

Para un escenario donde $\mathrm{P} = 2\%$ y $\tau = 10\frac{\mathrm{seg}}{\mathrm{obs}*\mathrm{veh}}$ (Figura 4-5), se ve notoriamente afectado el método promedios simples históricos, donde la actualización de celdas es menos frecuente ya que hay menos datos. Se puede notar celdas donde no existe actualización, cerca del intervalo de tiempo 1000. En el caso de promedios simples interpolados, la velocidad de propagación de la cola es aún más alta. FBH-v es el método que se ve menos afectado con el cambio de cantidad de información.

Se desea evaluar cuantitativamente el error en la estimación, para realizar comparaciones entre los métodos de estimación. Se utilizará sólo la métrica de error relativo medio,

dada la similitud tendencial encontrada con el error relativo absoluto. Las tablas 4-13 a 4-15 presentan los errores relativos medios de los diferentes métodos de estimación.

Tabla 4-13: Error relativo medio de Promedios Simples Históricos (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**
1	38	39	41	46	54	60	84	127
2	30	32	33	38	47	50	62	83
3	26	27	29	35	42	46	53	69
4	24	25	27	32	40	43	51	64
5	21	23	25	29	36	41	49	55
10	16	19	20	23	29	33	40	49
15	14	17	19	21	25	28	37	44
20	12	16	18	20	23	26	34	43
25	11	15	17	19	22	24	32	40
30	10	14	16	18	21	23	30	39

Fuente: Elaboración propia

Promedios simples históricos (Tabla 4-13) posee altos errores en caso de tener muy poca información. Esto sucede al no utilizar información de las condiciones de borde y basarse en información pasada de gran antigüedad. A medida que se incorpora información, es posible disminuir errores, llegando a niveles de hasta 10%.

Tabla 4-14: Error relativo medio de Promedios Simples Interpolados (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	24	27	28	29	29	29	29	29
2	22	24	26	28	28	29	29	29
3	20	23	25	27	28	28	29	29
4	19	22	24	26	28	28	29	29
5	18	21	23	25	27	28	28	29
10	15	18	20	23	25	26	28	29
15	14	17	18	21	24	25	27	28
20	12	15	17	20	23	24	26	28
25	11	15	16	19	22	23	25	27
30	10	14	16	18	21	22	25	27

Fuente: Elaboración propia

Promedios simples interpolados (Tabla 4-14) posee una cota superior de error menor al caso anterior gracias a la utilización de las condiciones de borde. En un principio añadir información no disminuye el error, a pesar de tener 8 veces más información (por ejemplo, pasando de $P = 1\%$ a $P = 4\%$ y de $\tau = 120 \frac{seg}{obs*veh}$ a $\tau = 60 \frac{seg}{obs*veh}$). A medida que se incluye datos se llega a niveles de error del 10%, de hecho, muy similares a la Tabla 4-13.

Tabla 4-15: Error relativo medio de FBH-v (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	18	18	19	20	22	23	24	25
2	16	16	16	18	20	21	23	24
3	14	14	15	16	18	20	22	24
4	13	14	14	15	17	19	22	23
5	13	13	13	14	17	18	20	22
10	11	11	11	13	14	15	18	20
15	10	10	10	11	13	14	16	19
20	9	9	10	11	12	13	16	18
25	9	9	9	10	12	13	15	17
30	8	9	9	10	11	12	14	17

Fuente: Elaboración propia

FBH-v (Tabla 4-15) posee una cota superior del 25%, mientras que con la mayor cantidad de datos logra errores del 8%.

Tabla 4-16: Errores de Promedios Simples Históricos menos Promedios Simples Interpolados (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	2	12	13	18	25	31	55	97
2	9	7	7	11	18	22	33	54
3	6	4	5	8	15	17	25	40
4	5	3	3	6	12	15	22	35
5	3	2	2	4	9	13	20	26
10	1	1	0	1	4	6	13	20
15	1	0	0	0	1	3	10	16
20	0	0	0	0	1	2	8	15
25	0	0	0	0	0	1	6	13
30	0	0	0	0	0	1	5	12

Fuente: Elaboración propia

Al comparar los métodos de promedios simples (Tabla 4-16) es posible notar un área donde ambos son prácticamente iguales ($P \geq 10\% \,\&\, \tau \leq 10 \frac{seg}{obs*veh}$). En otro caso, la diferencia entre métodos es sustancial, por lo que siempre será conveniente utilizar

promedios simples interpolados. Si se compara este último método con FBH-v se obtiene la Tabla 4-17.

Tabla 4-17: Errores de Promedios Simples Interpolados menos FBH-v (%) [NGSIM]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**
1	6	8	9	8	7	6	5	5
2	6	8	10	10	8	7	6	5
3	5	8	10	10	9	9	7	6
4	5	8	10	11	10	9	7	6
5	5	8	10	11	11	10	8	7
10	4	7	8	10	11	11	10	8
15	4	7	8	10	11	11	11	9
20	3	6	8	9	10	11	11	10
25	2	6	7	8	10	10	10	10
30	2	5	7	8	9	10	11	10

Fuente: Elaboración propia

FBH-v rinde siempre mejor que promedios simples interpolados. En escenarios de pocos o muchos datos la diferencia no es tan notable como en escenarios con una cantidad de datos intermedios.

En resumen, las condiciones de borde permiten una reducción sustancial del error estimado en el caso de escenarios con poca información. Con un máximo de información, promedios simples históricos tienen igual rendimientos que promedios simples interpolados. Al comparar este último con FBH-v, las diferencias son mayores con escenarios de media cantidad de datos, y siempre es conveniente la utilización de FBH-v.

4.1.6 Métrica para inferir errores

El primer paso para identificar una relación funcional para inferir errores a partir de variables conocidas es la utilización de herramientas gráficas. Se realizaron gráficos de

contorno (Figura 4-6) explicando el error relativo medio[9] de los diferentes métodos de estimación en función de la penetración y la frecuencia de muestreo.

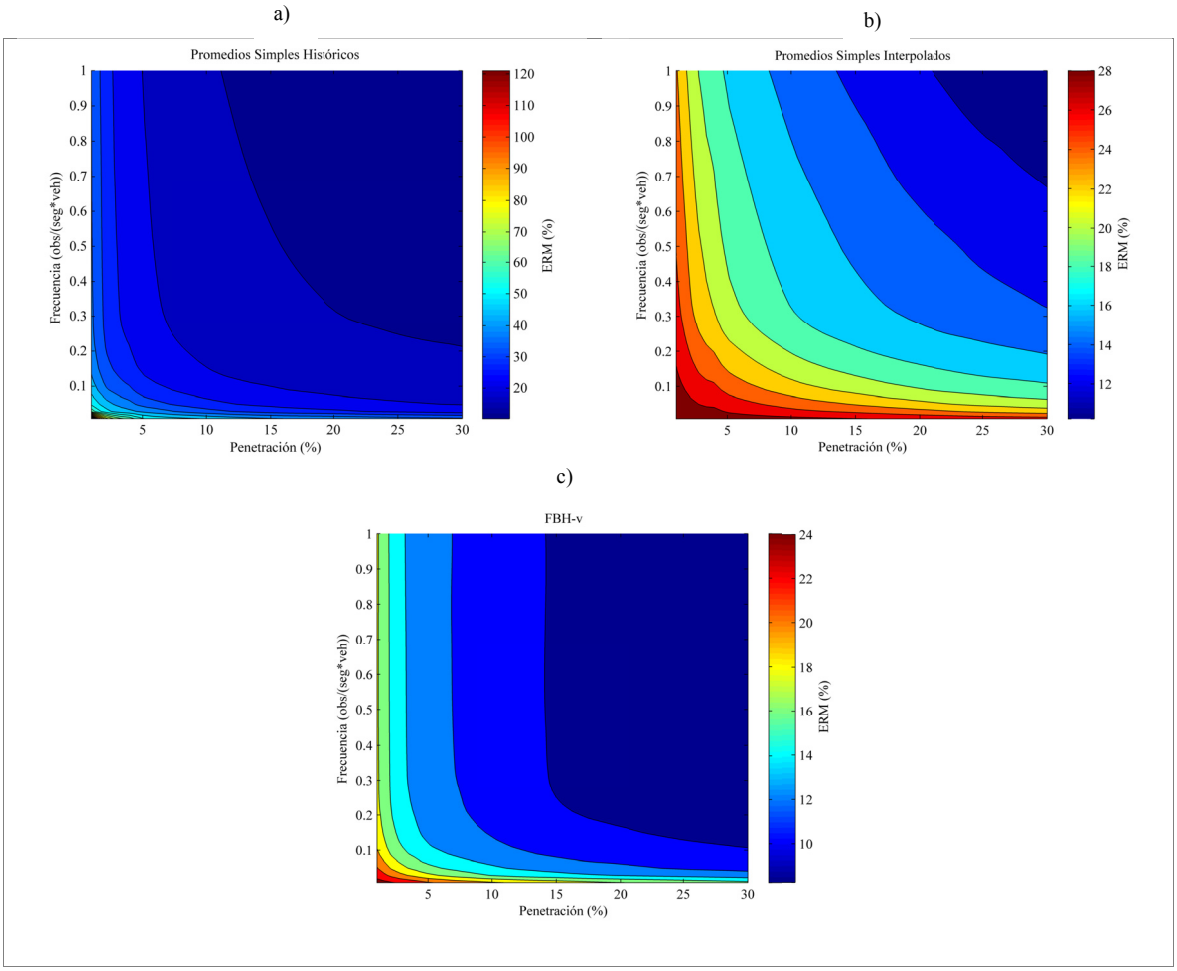

Figura 4-6: Gráficos de contorno del error relativo medio para diferentes métodos de estimación [NGSIM]

a) Promedios simples históricos; b) Promedios simples interpolados y c) FBH-v [diferentes escalas]

Fuente: Elaboración propia

Para los métodos presentados, se aprecian curvas de nivel o isocuantas totalmente convexas a lo largo del rango de frecuencias y penetraciones utilizadas. Esto tiene sentido, puesto que mayor cantidad de datos debe entregar menor error. Los métodos

[9] Graficar error absoluto medio tiene un comportamiento similar.

promedios simples históricos e interpolados poseen curvas estrictamente convexas, mientras que FBH-v posee una saturación de información en caso de frecuencias muy altas. Esto quiere decir que no se logra aprovechar datos extra. Un muestreo cada tres segundos es equivalente a uno cada un segundo.

Inicialmente se asumió que la relación entre el error y las variables en cuestión se podía explicar mediante una función Cobb-Douglas[10]. Este presentó graves problemas de heterocedasticidad al poseer el modelo una especificación incorrecta (Hayashi, 2000), ya que sus elasticidades son constantes, y una elasticidad de sustitución[11] igual a uno (Nicholson & Snyder, 2008). Una función CES[12] tampoco serviría, al presentar elasticidades de sustitución siempre constantes, siendo que los datos pueden indicar otra cosa (Nicholson & Snyder, 2008). Además su forma original presenta un grado de retornos a escala[13] constante (Hayashi, 2000). El problema de especificación fue resuelto aplicando una función Translog (Hayashi, 2000; Lau et al., 1973) de la siguiente forma:

$$\overrightarrow{\ln(\varepsilon)} = \hat{\alpha}_0 + \hat{\alpha}_1 \cdot \overrightarrow{\ln(P)} + \hat{\alpha}_2 \cdot \overrightarrow{\ln(z)} + \hat{\alpha}_3 \overrightarrow{\ln(P) \cdot \ln(z)} + \hat{\alpha}_4 \cdot \overrightarrow{\ln^2(P)} + \hat{\alpha}_5 \cdot \overrightarrow{\ln^2(z)} \qquad (4\text{-}1)$$

La función Translog es el caso general de una función Cobb-Douglas y de una aproximación lineal de un caso de la función CES (Kmenta, 1967). Esta forma funcional no impone restricciones en las elasticidades de sustitución y retornos a escala. Además, es de fácil manejo algebraico. También es posible justificar el uso de esta función al

---

[10] Cobb-Douglas: $\varepsilon = k \cdot P^{\alpha} \cdot z^{\beta}$

[11] Medida que permite identificar la "facilidad" de sustituir inputs (P y z) para mantener el nivel de error constante (a lo largo de una isocuanta). Mide el cambio proporcional de z/P relativo al cambio proporcional de la tasa técnica de sustitución (TTS) de P por z ($TTS = -\frac{dz}{dP}\Big|_{\varepsilon=\varepsilon_o} = \frac{\frac{\partial \varepsilon}{\partial P}}{\frac{\partial \varepsilon}{\partial z}}$) a lo largo de una isocuanta ε_o. Matemáticamente se puede definir como $\sigma = \frac{\partial \ln(z/P)}{\partial \ln(TTS)}$.

[12] *Constant elasticity of substitution*: $\varepsilon = \left(P^{\delta} + z^{\delta}\right)^{\frac{1}{\delta}}$

[13] Retornos a escala constantes: $\varepsilon(a \cdot P, a \cdot z) = a \cdot \varepsilon(P, z)$

analizar los términos de las elasticidades de P y z con respecto al error ε, $E_{\varepsilon,P}$ y $E_{\varepsilon,z}$ respectivamente:

$$E_{\varepsilon,P} = \frac{\partial \varepsilon}{\partial P} \cdot \frac{P}{\varepsilon} = \frac{\partial \ln(\varepsilon)}{\partial \ln(P)} = \hat{\alpha}_1 + \hat{\alpha}_3 \cdot \ln(z) + \hat{\alpha}_4 \cdot 2 \cdot \ln(P) \quad (4\text{-}2)$$

$$E_{\varepsilon,z} = \frac{\partial \varepsilon}{\partial z} \cdot \frac{z}{\varepsilon} = \frac{\partial \ln(\varepsilon)}{\partial \ln(z)} = \hat{\alpha}_2 + \hat{\alpha}_3 \cdot \ln(P) + \hat{\alpha}_5 \cdot 2 \cdot \ln(z) \quad (4\text{-}3)$$

Las elasticidades dependen de ambos factores P y z, lo que tiene sentido. Permite capturar efectos como la saturación de información (cuando ya hay demasiados datos, quizás no es necesario tener más, y puede depender tanto de P como z). Esto a través de la función logaritmo, al tener una derivada positiva y una segunda derivada negativa en toda su región. Además, captura el hecho que un aumento de *P* (o bien, *z*) afecta de distinta forma al error dependiendo del valor de *z* (o bien, *P*), indicando que cambios de P pueden incidir de distinta manera a cambios de z en el error.

Para obtener los estimadores $\hat{\alpha}_i$ de la ecuación (4-1) se realizaron regresiones lineales utilizando el error estándar robusto HC3 (Long & Ervin, 2000). Los parámetros estimados se encuentran en la Tabla 4-18.

Tabla 4-18: Parámetros estimados para diferentes métodos de estimación [NGSIM]

Parámetro	Variable	PS Históricos	PS Interpolados	FBH-v
$\hat{\alpha}_1$	$\ln(P)$	$-0{,}267$ $(-15{,}27)$	$-0{,}408$ $(-88{,}75)$	$-0{,}306$ $(-36{,}52)$
$\hat{\alpha}_2$	$\ln(z)$	$-0{,}128$ $(-14{,}35)$	$-0{,}345$ $(-134{,}13)$	$-0{,}0835$ $(-19{,}41)$
$\hat{\alpha}_3$	$\ln(P)\cdot\ln(z)$	$-0{,}0122$ $(-6{,}04)$	$-0{,}0507$ $(-94{,}44)$	$-0{,}0203$ $(-18{,}41)$
$\hat{\alpha}_4$	$\ln^2(P)$	$-0{,}0116$ $(-4{,}21)$	$-0{,}0287$ $(-37{,}13)$	$-0{,}0106$ $(-7{,}02)$
$\hat{\alpha}_5$	$\ln^2(z)$	$0{,}0290$ $(25{,}08)$	$-0{,}0180$ $(-50{,}69)$	$0{,}0211$ $(34{,}70)$
$\hat{\alpha}_0$	Constante	$-2{,}492$ $(-93{,}64)$	$-2{,}711$ $(-401{,}08)$	$-2{,}896$ $(-235{,}07)$
	N	1200	1065	1185

Estadísticos *t* entre paréntesis

Fuente: Elaboración propia

Todas las variables estimadas son significativas para un nivel de confianza del 95%. Se realizaron test-F (Ortúzar & Willumsen, 2011) para verificar que la función Translog posee un ajuste estadísticamente superior a una Cobb-Douglas (es decir, que los parámetros asociados a $\ln(P)\cdot\ln(z)$, $\ln^2(P)$, $\ln^2(z)$ son distintos a cero de manera conjunta). Los resultados indican que siempre es preferible la función Translog. Además, cabe notar que no existen restricciones directas de signos de parámetros. El efecto conjunto de ellos se expresa a través de las elasticidades, que deben ser no-positivas en el dominio analizado (se espera que más datos no empeoren las estimaciones). Se retiraron aquellos escenarios donde la elasticidad era positiva, dada que la forma funcional no permitía un ajuste correcto (en vez de positiva, debe haber sido cero, dado que los datos nunca mostraron este comportamiento). Se graficaron superficies de error para distintos valores de penetración y tasa de muestreo para cada método de estimación, de manera de asignar a cada escenario un error estimado. Estas se muestran en la Figura 4-7.

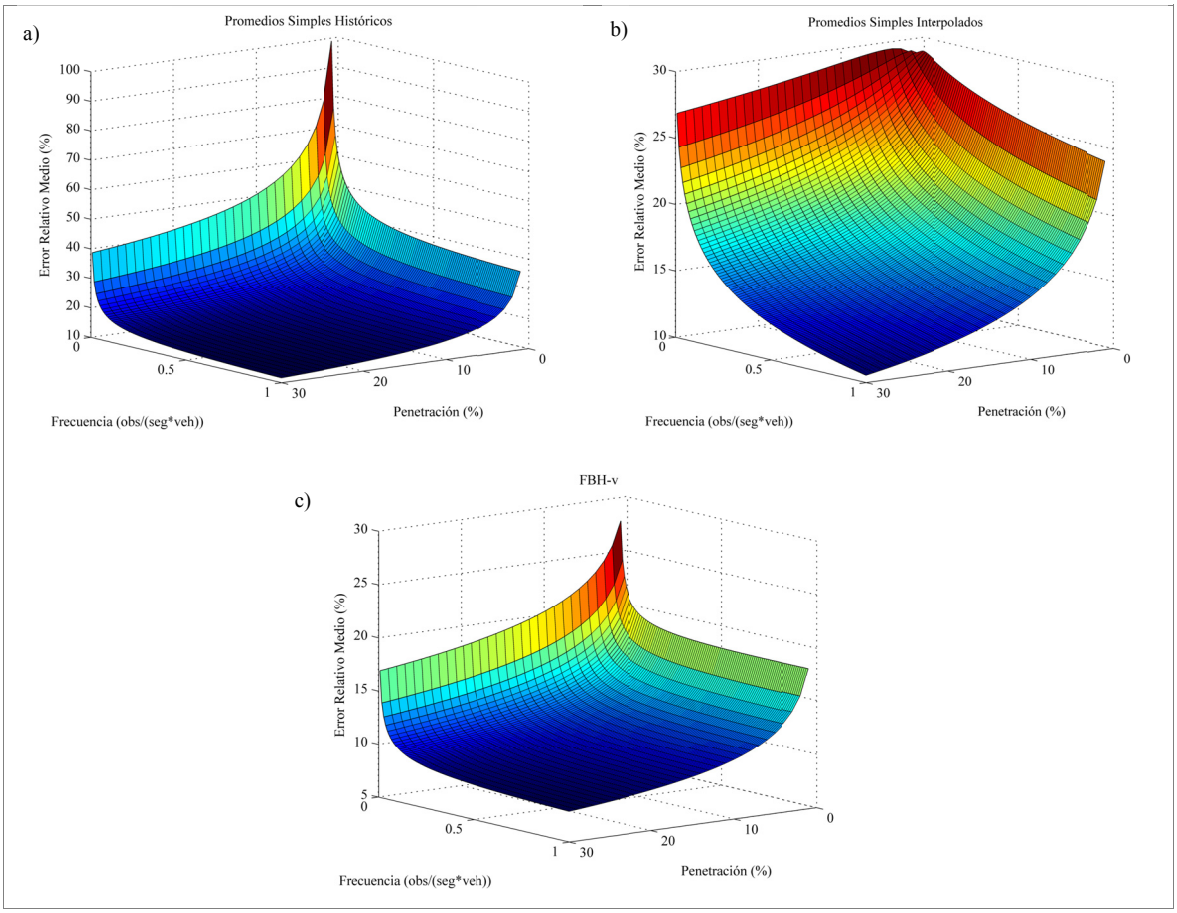

Figura 4-7: Superficies de error estimadas [NGSIM]

a) Promedios simples históricos; b) Promedios simples interpolados y c) FBH-v [diferentes escalas]

Fuente: Elaboración propia

4.1.7 El efecto de la composición de información

Una vez identificada la relación funcional que permite explicar el error alcanzado a partir de variables conocidas, es posible analizar el efecto de la composición de la información. Esto es, evaluar el efecto de cambios de la penetración (P) y/o tasa de muestreo (z) en el error estimado. En específico, se desea encontrar diferentes regiones $[P, z]$ que indiquen qué variable es conveniente aumentar para disminuir de mayor manera en error. En algunas situaciones puede ser de conveniencia aumentar la penetración en vez de la tasa de muestreo, y en otras puede suceder lo contrario. Entre

estas regiones se espera un sector donde sea indiferente el origen de la disminución del error.

Se propone analizar la razón escalada $\Delta = \ln\left(\frac{E_{\varepsilon,P}}{E_{\varepsilon,z}}\right)$, donde $\Delta > 0$ indica que es conveniente que un aumento de las observaciones provenga de un aumento de la penetración de sensores, $\Delta < 0$ que este aumento de las observaciones provenga de un aumento de la frecuencia y $\Delta = 0$ indica que es indiferente de dónde provenga el aumento. La Figura 4-8 muestra, para el método promedio simples históricos, valores de Δ para diferentes combinaciones de P y z. Como era de esperar existe un área de indiferencia ($\Delta = 0$), donde es equivalente aumentar P o z para disminuir el error.

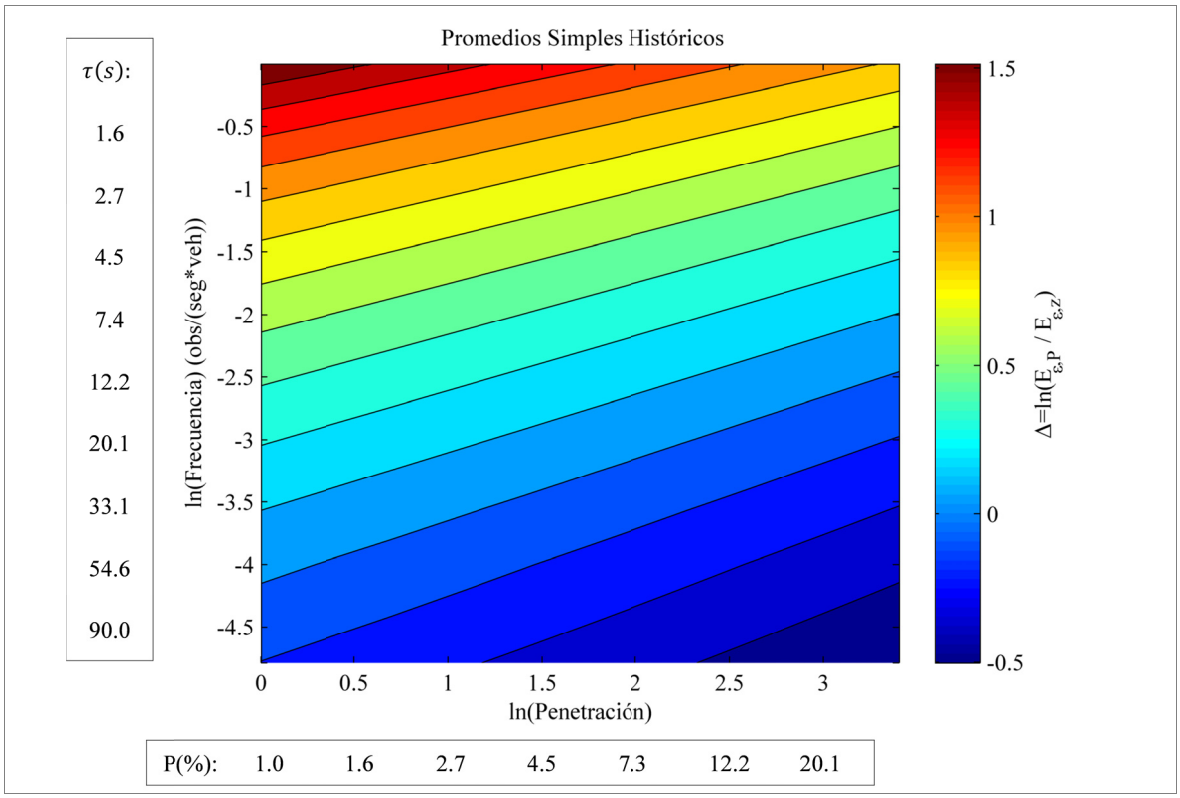

Figura 4-8: Razón escalada Δ para promedios simples históricos [NGSIM]

Fuente: Elaboración propia

Para cada método de estimación, se desea encontrar una expresión analítica para el lugar geométrico donde $\Delta = 0$. Esto es, una función $z(P)$ que permite encontrar, para cada penetración, una frecuencia de muestreo específica, tal que en ese punto $\Delta = 0$ (es decir, es indiferente de dónde proviene un aumento de las observaciones). Esta función se encuentra igualando las ecuaciones de elasticidades (4-2) y (4-3), obteniendo:

$$z(P)|_{E_{\varepsilon,P}=E_{\varepsilon,z}} = e^{\frac{\hat{\alpha}_1-\hat{\alpha}_2+(2\cdot\hat{\alpha}_4-\hat{\alpha}_3)\cdot\ln(P)}{2\cdot\hat{\alpha}_5-\hat{\alpha}_3}} \quad (4\text{-}4)$$

Donde $\hat{\alpha}_i$ son los parámetros estimados anteriormente para cada método de estimación de estados de tráfico. Utilizando la transformación $\tau = \frac{1}{z}$ es posible encontrar curvas de indiferencia para distintos valores de intervalo de muestreo y penetración (Figura 4-9).

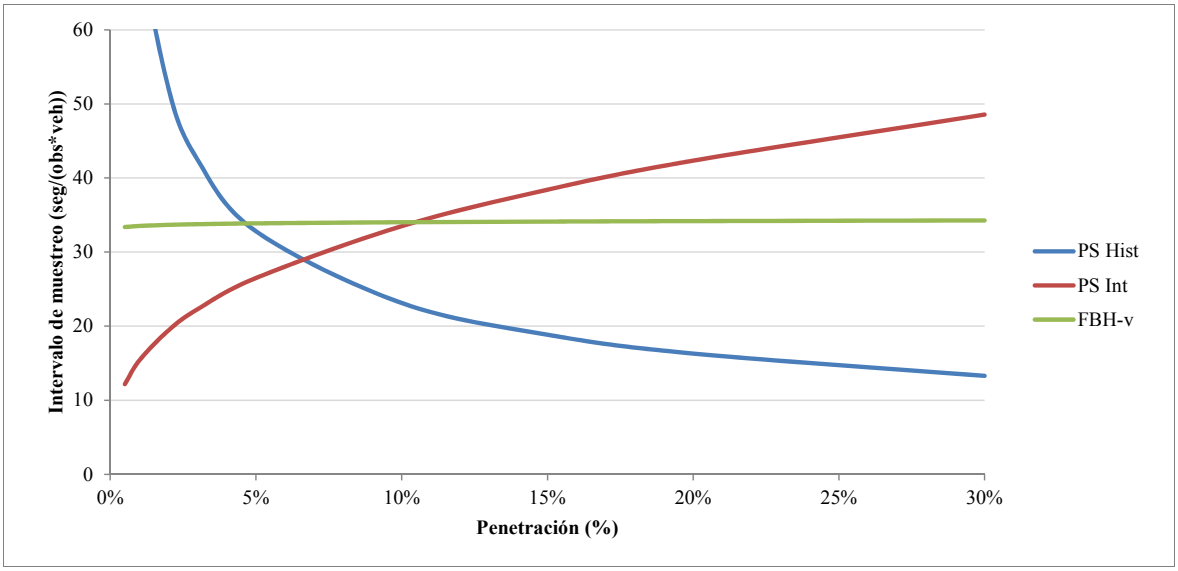

Figura 4-9: Curvas de indiferencia de aumento de observaciones [NGSIM]

Fuente: Elaboración propia

En el caso de promedios simples históricos, a medida que aumenta la penetración, el intervalo de muestreo asociado a la curva de indiferencia, baja desde 60 segundos hasta un valor cercano a 15 segundos. Con promedios simples interpolados el intervalo de muestreo asociado aumenta desde un valor cercano a 12 hasta 50 a medida que aumenta la penetración. Para FBH-v, la curva de indiferencia se encuentra cercana a los 34 segundos, independiente de la penetración. Sobre estas curvas, un aumento del número de observaciones, a través de un aumento de la penetración, disminuye más el error que este mismo aumento de observaciones mediante un aumento de la frecuencia de muestreo. Dado que los métodos de estimación son distintos, es esperable que las curvas sean diferentes.

Es interesante el problema de que dado un número de observaciones fijo, se desea encontrar la combinación óptima de P y z para minimizar el error. Por ejemplo, si no es posible procesar más datos que cierta cantidad. Esto es importante ya que un mismo número de observaciones genera diferentes magnitudes de error. La Figura 4-10 muestra isocuantas de error y observaciones, donde una misma cantidad de observaciones (15.687) presenta errores desde un 16 % a un 22 % dependiendo de los valores de P y τ.

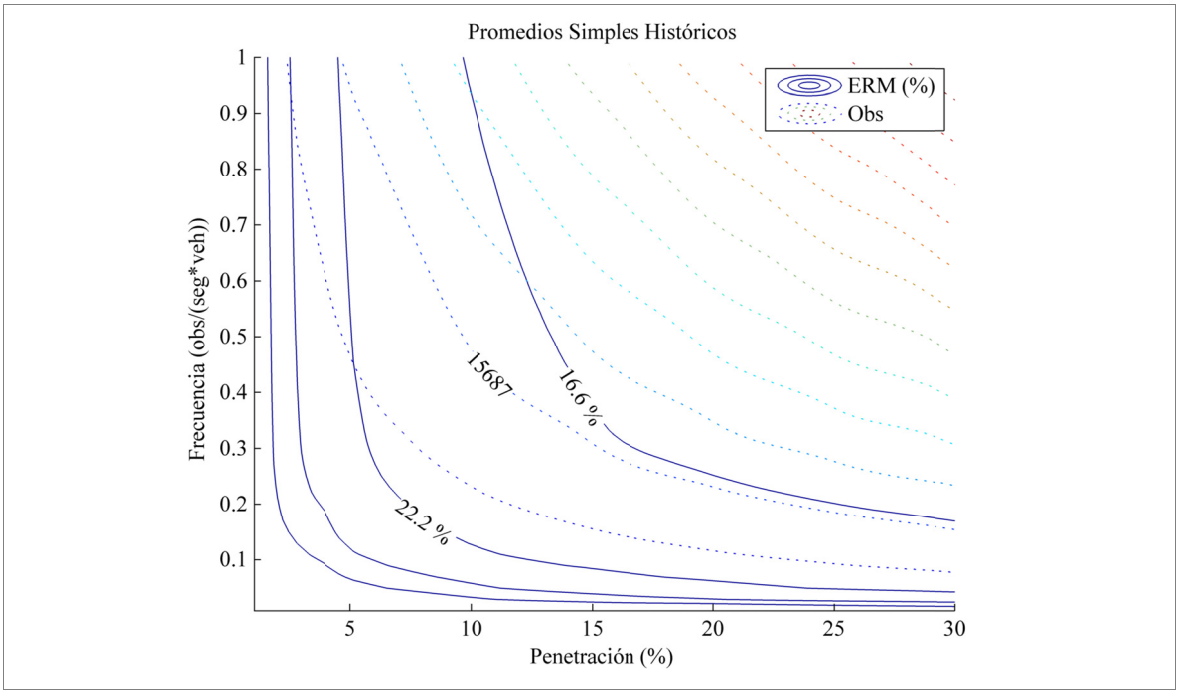

Figura 4-10: Diferente magnitud de error para igual número de observaciones [NGSIM]

Fuente: Elaboración propia

Es posible demostrar que las curvas de indiferencia de aumento de observaciones (presentes en la Figura 4-9) coinciden con la solución óptima al problema de minimización del error en la estimación dado un número de observaciones fijo (designado por N_o):

$$\boldsymbol{Min}\ \varepsilon(P, z)$$

$$\underline{\boldsymbol{s.a.}}\quad Obs(P, z) = N_0$$

$$0 \leq P \leq 1$$

$$0 \leq z \leq 1$$

Para solucionar este problema, es de conveniencia aplicar la función logaritmo natural a la función objetivo y a la restricción asociada al número de observaciones (recordando que la solución óptima se mantiene):

$$\boldsymbol{Min}\ \hat{\alpha}_0 + \hat{\alpha}_1 \cdot \overrightarrow{\ln(P)} + \hat{\alpha}_2 \cdot \overrightarrow{\ln(z)} + \hat{\alpha}_3 \overrightarrow{\ln(P) \cdot \ln(z)} + \hat{\alpha}_4 \cdot \overrightarrow{\ln^2(P)} + \hat{\alpha}_5 \cdot \overrightarrow{\ln^2(z)}$$

$$\underline{\boldsymbol{s.a.}} \quad \ln(\bar{t}_v \cdot N_{veh}) + \ln(P) + \ln(z) = \ln(N_0)$$

$$0 \leq P \leq 1$$

$$0 \leq z \leq 1$$

Basta aplicar las condiciones de Karush-Khun-Tucker (KKT) para obtener la solución óptima[14] (dada la convexidad de la función objetivo y restricción) al problema de minimización en función del número de observaciones:

$$P(N_o) = e^{\frac{-\hat{\alpha}_2+\hat{\alpha}_1-2\cdot\hat{\alpha}_5\cdot\ln(N_0)+2\cdot\hat{\alpha}_5\cdot\ln(\bar{t}_v\cdot N_{veh})+\hat{\alpha}_3\cdot\ln(N_0)-\hat{\alpha}_3\cdot\ln(\bar{t}_v\cdot N_{veh})}{2\cdot(\hat{\alpha}_3-\hat{\alpha}_5-\hat{\alpha}_4)}} \tag{4-5}$$

$$z(N_o) = e^{\frac{\hat{\alpha}_3\cdot\ln(N_0)-2\cdot\ln(N_0)\cdot\hat{\alpha}_4-\hat{\alpha}_3\cdot\ln(\bar{t}_v\cdot N_{veh})+2\cdot\ln(\bar{t}_v\cdot N_{veh})\cdot\hat{\alpha}_4+\hat{\alpha}_2-\hat{\alpha}_1}{2\cdot(\hat{\alpha}_3-\hat{\alpha}_5-\hat{\alpha}_4)}} \tag{4-6}$$

Al despejar N_o de la ecuación (4-5) se obtiene:

$$N_0(P) = e^{\frac{\hat{\alpha}_2-\hat{\alpha}_1+2\cdot\ln(P)\cdot\hat{\alpha}_3-2\cdot\hat{\alpha}_5\cdot\ln(\bar{t}_v\cdot N_{veh})-2\cdot\ln(P)\cdot\hat{\alpha}_4+\hat{\alpha}_3\cdot\ln(\bar{t}_v\cdot N_{veh})-2\cdot\ln(P)\cdot\hat{\alpha}_5}{-2\cdot\hat{\alpha}_5+\hat{\alpha}_3}} \tag{4-7}$$

Reemplazando la ecuación (4-7) en (4-6) se obtiene:

$$z(P)|_{E_{\varepsilon,P}=E_{\varepsilon,z}} = e^{\frac{\hat{\alpha}_1-\hat{\alpha}_2+(2\cdot\hat{\alpha}_4-\hat{\alpha}_3)\cdot\ln(P)}{2\cdot\hat{\alpha}_5-\hat{\alpha}_3}} \tag{4-8}$$

Demostrando la igualdad con la ecuación (4-4). Esto quiere decir que es posible construir las curvas de indiferencia eligiendo, para cada número de observaciones, la combinación óptima (P,τ) que provee el menor error en la estimación. La curva es el conjunto de soluciones óptimas para cada número de observaciones.

---

[14] Asumiendo que las restricciones $1 \geq P \geq 0$ y $1 \geq z \geq 0$ están inactivas.

Para las combinaciones (P,τ) que no recaen en las curvas de indiferencia, es interesante analizar cuán diferente es el impacto marginal de un aumento de P versus un aumento de τ. Para esto se utiliza Δ. Un valor de Δ mayor a uno indica que un aumento de P es $E_{\varepsilon,P}/E_{\varepsilon,z}$ veces más conveniente que un aumento de τ. Por otro lado, un valor menor a uno indica que un aumento de τ es $E_{\varepsilon,z}/E_{\varepsilon,P}$ veces más conveniente que un aumento de P. A continuación se muestra, para cada método de estimación, el cociente propuesto para los distintos escenarios analizados.

Tabla 4-19: Cociente de elasticidades para promedios simples históricos [NGSIM]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**
1	5,2	2,7	2,1	1,7	1,5	1,2	1,0	0,9
2	4,5	2,4	1,9	1,5	1,4	1,1	1,0	0,8
3	4,1	2,2	1,8	1,5	1,3	1,1	0,9	0,8
4	3,9	2,2	1,8	1,4	1,3	1,0	0,9	0,8
5	3,7	2,1	1,7	1,4	1,2	1,0	0,9	0,8
10	3,2	1,9	1,6	1,3	1,1	0,9	0,8	0,7
15	3,0	1,8	1,5	1,2	1,1	0,9	0,8	0,7
20	2,8	1,7	1,4	1,1	1,0	0,9	0,7	0,6
25	2,7	1,6	1,4	1,1	1,0	0,8	0,7	0,6
30	2,6	1,6	1,3	1,1	1,0	0,8	0,7	0,6

Fuente: Elaboración propia

Para promedios simples históricos (Tabla 4-19), existen escenarios donde es hasta cinco veces más conveniente un aumento de la penetración que uno de intervalo de muestreo. Por otro lado, un aumento del intervalo de muestreo puede ser hasta 1,6 veces más conveniente. Notar que los sectores donde el cociente toma el valor 1,0 crean la curva de indiferencia anteriormente calculada.

Tabla 4-20: Cociente de elasticidades para promedios simples interpolados [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	1,4	1,3	1,3	1,2	-	-	-	-
2	1,3	1,3	1,2	1,2	1,1	0,7	-	-
3	1,3	1,2	1,2	1,1	1,1	0,9	0,3	-
4	1,3	1,2	1,2	1,1	1,1	1,0	0,6	-
5	1,2	1,2	1,2	1,1	1,1	1,0	0,7	-
10	1,2	1,2	1,2	1,1	1,1	1,0	0,9	0,7
15	1,2	1,2	1,2	1,1	1,1	1,0	0,9	0,8
20	1,2	1,2	1,2	1,1	1,1	1,0	1,0	0,8
25	1,2	1,2	1,2	1,1	1,1	1,0	1,0	0,9
30	1,2	1,2	1,1	1,1	1,1	1,0	1,0	0,9

Fuente: Elaboración propia

Para promedios simples interpolados, el cociente presenta un dominio más acotado (Tabla 4-21). En otras palabras, un aumento de P o τ es genera cambios similares en el error. Su valor varía entre 1,4 y 0,3. Los escenarios de baja información presentaban elasticidades con valor cero, por lo que su interpretación no era válida. Esto sucede cuando más datos no brindan mejoras en la estimación.

Tabla 4-21: Cociente de elasticidades para FBH-v [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120
1	-	5,4	3,1	1,9	1,5	1,1	0,8	0,6
2	-	4,1	2,7	1,8	1,4	1,0	0,8	0,6
3	21,7	3,7	2,5	1,7	1,4	1,0	0,8	0,6
4	14,3	3,4	2,4	1,7	1,4	1,0	0,8	0,6
5	11,4	3,3	2,3	1,6	1,4	1,0	0,8	0,6
10	7,3	2,9	2,2	1,6	1,3	1,0	0,8	0,7
15	6,1	2,7	2,1	1,5	1,3	1,0	0,8	0,7
20	5,5	2,6	2,0	1,5	1,3	1,0	0,8	0,7
25	5,1	2,5	2,0	1,5	1,3	1,0	0,8	0,7
30	4,8	2,5	2,0	1,5	1,3	1,0	0,8	0,7

Fuente: Elaboración propia

En el caso de FBH-v (Tabla 4-21), para intervalos de muestreo muy altos, un aumento de penetración es varias veces más conveniente. Para el escenario $P = 1\%$ y $\tau = 120$ s, un aumento de intervalo de muestreo es hasta 1,7 veces más conveniente (1/0,6≈1,7). Se puede apreciar que la curva de indiferencia está presente en $\tau = 30$ s. Para los escenarios con (-), la elasticidad $E_{\varepsilon,z}$ fue cero, por lo que el cociente no tiene interpretación válida.

En la Figura 4-11 se muestra el cociente Δ para los diferentes métodos de estimación. Cabe destacar que la escala de Δ para esta figura va entre 0,5 y 1,5, por lo que no se aprecian todos los valores del indicador.

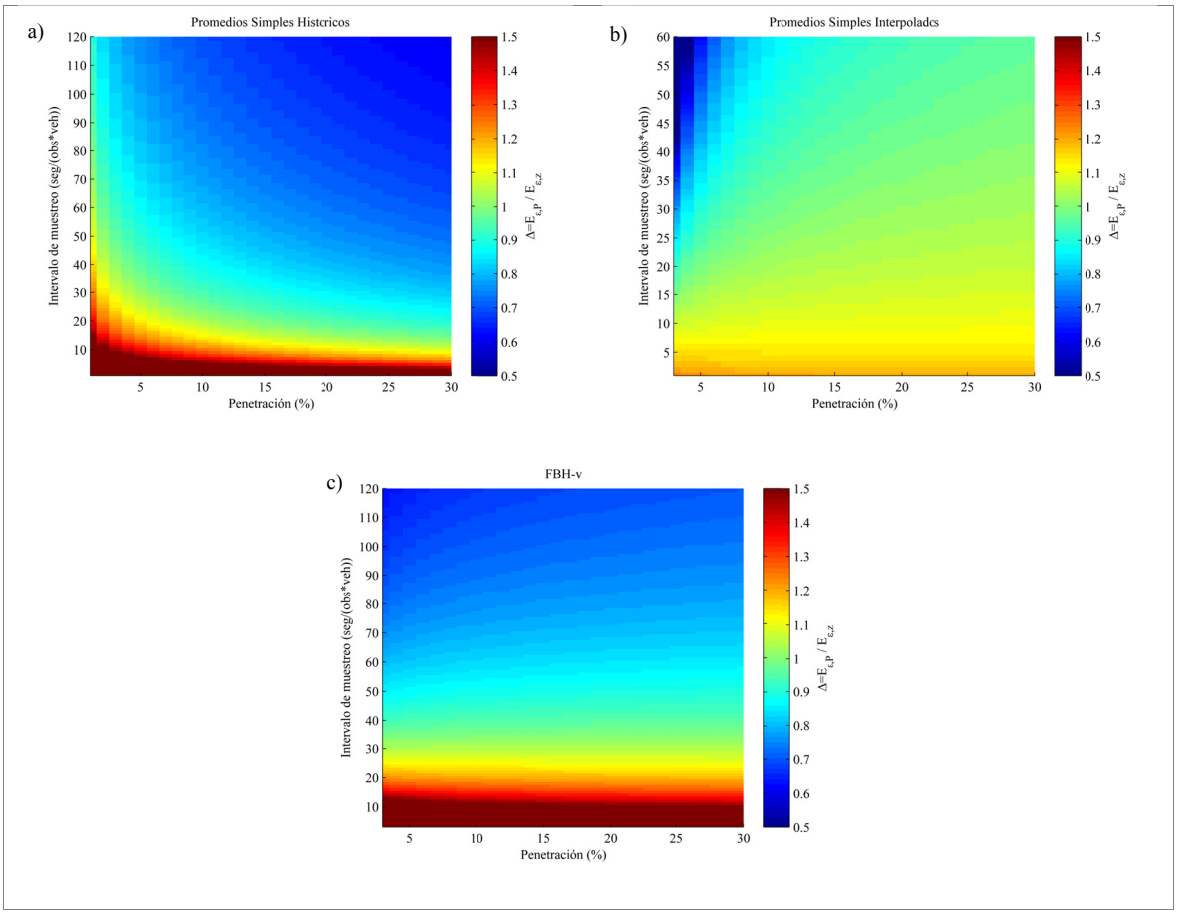

Figura 4-11: Cociente Δ para los diferentes métodos de estimación [NGSIM]

a) Promedios simples históricos; b) Promedios simples interpolados y c) FBH-v.

Valores mayores o menores que el límite de la escala se presentan del color del valor límite.

Fuente: Elaboración propia

En resumen, dado que un mismo número de observaciones puede tener errores diferentes a través de la composición, se analizó este fenómeno. Se identificó diferentes escenarios de penetración y tasa de muestreo. En algunos es conveniente aumentar la penetración, a fin de disminuir el error en la estimación. En otros, en cambio, es más conveniente ajustar la frecuencia de muestreo. La conveniencia de cada aumento se identificó a través del cociente $E_{\varepsilon,P}/E_{\varepsilon,Z}$. Su valor depende del escenario y método de muestreo analizado. Existen situaciones donde un cambio en la penetración o en el intervalo de muestreo produce el mismo efecto en el error (donde el cociente anterior es igual a uno). Estas situaciones constituyen una curva de indiferencia, donde es indiferente de dónde

provenga el aumento de observaciones. Esta curva de indiferencia coincide con el conjunto de soluciones del problema de minimización del error dado un número de observaciones fijo.

4.2 AIMSUN

El capítulo pasado utilizó como base de análisis una base de datos de trayectorias de vehículos para una porción de 600 m. de autopista. La longitud de la sección analizada no permite un análisis completo de los fenómenos estudiados, por lo que es necesario utilizar herramientas de microsimulación para generar nuevas trayectorias. Esta sección sigue la misma estructura que la sección 4.1 NGSIM.

4.2.1 Descripción de la base de datos

La base de datos fue obtenida a partir de microsimulación, mediante el programa AIMSUN. Se diseñó una autopista ficticia, con una longitud de 5 km (Figura 4-12). Esta posee tres entradas (E1, E2 y E3) y tres salidas (S1, S2 y S3), incluyendo pistas de aceleración y desaceleración respectivamente.

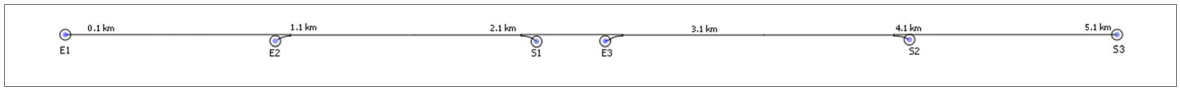

Figura 4-12: Autopista generada en AIMSUN

Fuente: Elaboración propia

La demanda se ingresó a través de la matriz O-D que se muestra en la Tabla 4-22. Se generó las trayectorias de 5.791 vehículos, durante un tiempo de simulación de una hora. Cada trayectoria contiene datos de posición, velocidad y aceleración cada un segundo de cada uno de ellos.

Tabla 4-22: Matriz Origen-Destino para AIMSUN

		Destino		
	Flujo (veh/hr)	S1	S2	S3
Origen	E1	500	500	4000
	E2	20	100	100
	E3	-	20	500

Fuente: Elaboración propia

Adicionalmente, se simuló un incidente que bloquea una porción de 100 m de la primera pista cercana a la salida S2, el cual empieza en $t = 10\ \text{min}$ y se extiende hasta $t = 30\ \text{min}$, para luego desaparecer. El incidente se muestra en la Figura 4-13, donde el rectángulo rojo simula el incidente.

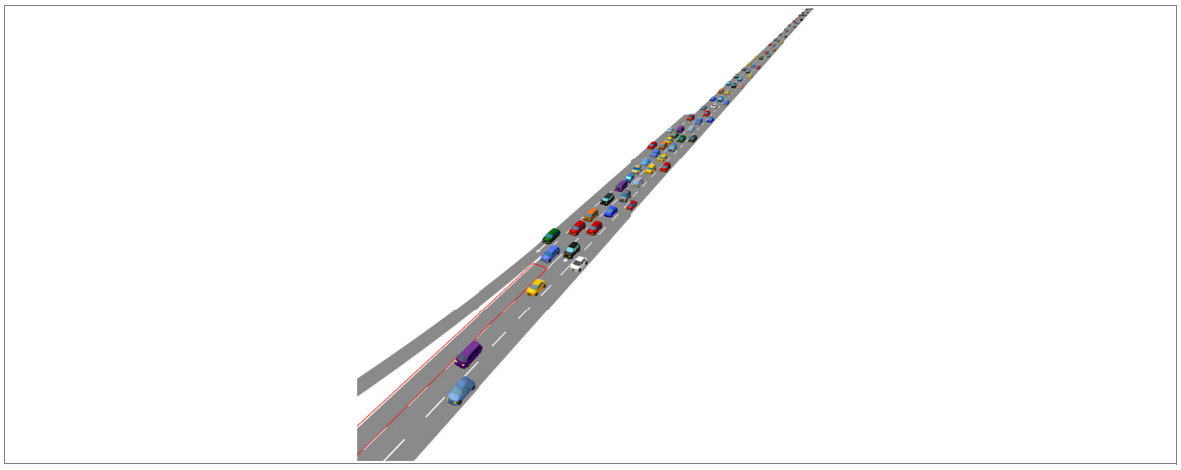

Figura 4-13: Incidente simulado en AIMSUN

Fuente: Elaboración propia

La sección de autopista fue discretizada en $I = 25$ celdas espaciales de longitud $\Delta x = 200\ \text{m}$. Se utilizó una discretización temporal que considera $H = 720$ intervalos de $\Delta t = 5\ \text{s}$ para cada celda. Esto genera un total de 18.000 celdas-intervalo.

La Figura 4-14 presenta el campo de velocidades real, donde se puede apreciar el inicio del incidente en la celda 20. Además, se puede apreciar congestión intermitente en la celda 14, dado el alto flujo proveniente de la entrada E3.

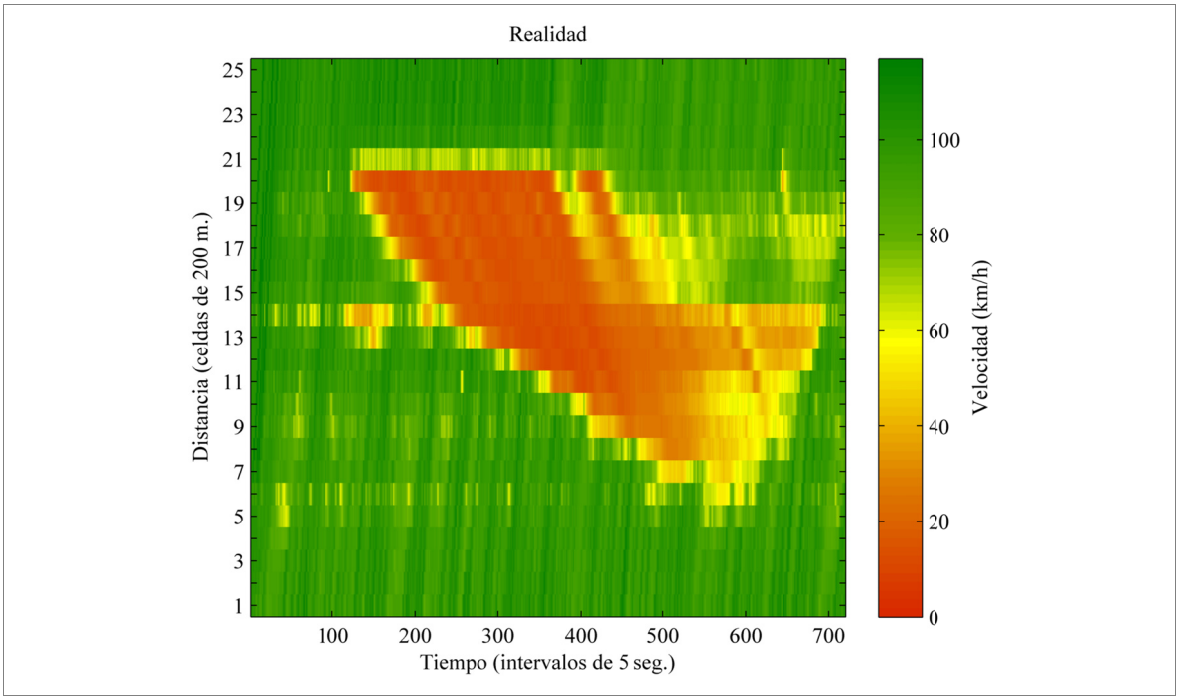

Figura 4-14: Campo de velocidades real [AIMSUN]

Fuente: Elaboración propia

4.2.2 Información de sensores móviles y generación de escenarios

Cada vehículo equipado enviará su posición y velocidad promedio del mismo de los últimos $S = 3$ segundos, además de una etiqueta que permita la re-identificación del vehículo. Los valores de penetración a usar son $\vec{P}(\%) = \{1, 2, 3, 4, 5, 10, 15, 20, 25, 30\}$. El intervalo de muestreo $\vec{\tau}(s) = \{1, 3, 5, 10, 20, 30, 60, 120, 240, 360, 480\}$. El último valor del intervalo de tiempo entrega aproximadamente una observación por vehículo equipado. En total se tienen 110 escenarios, con $R = 15$ réplicas para cada uno.

4.2.3 Cuantificación de información

La Tabla 4-23 presenta el número de observaciones promedio por escenario para el caso de AISMUN.

Tabla 4-23: Número de observaciones promedio por escenario [AIMSUN]

		τ (seg/(obs*veh))										
		1	**3**	**5**	**10**	**20**	**30**	**60**	**120**	**240**	**360**	**480**
P (%)	**1**	15268	5089	3053	1527	763	508	254	126	64	41	31
	2	30092	10032	6018	3008	1503	1004	501	251	124	83	61
	3	45671	15227	9136	4569	2284	1522	761	380	191	122	92
	4	61229	20409	12247	6124	3063	2039	1020	509	258	168	126
	5	76853	25616	15371	7685	3842	2561	1280	639	319	213	155
	10	151912	50638	30377	15190	7596	5061	2532	1262	634	413	311
	15	229682	76557	45934	22967	11483	7655	3831	1909	960	624	469
	20	304727	101574	60939	30469	15227	10153	5081	2538	1274	828	625
	25	381114	127036	76214	38107	19050	12696	6347	3168	1592	1042	780
	30	456695	152234	91327	45663	22826	15212	7612	3796	1906	1241	935

Fuente: Elaboración propia

Esta fue obtenida desde el minuto 5, hasta el fin de la hora. Considera una longitud de 5 km, desde los 100 m a 5.100 m. También es posible obtener la densidad de observaciones por kilómetro-minuto-pista (Tabla 4-24).

Tabla 4-24: Densidad de observaciones por kilómetro-minuto-pista [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	19	6	4	2	1	1	0	0	0	0	0
2	37	12	7	4	2	1	1	0	0	0	0
3	56	19	11	6	3	2	1	0	0	0	0
4	74	25	15	7	4	2	1	1	0	0	0
5	93	31	19	9	5	3	2	1	0	0	0
10	185	62	37	18	9	6	3	2	1	1	0
15	279	93	56	28	14	9	5	2	1	1	1
20	370	123	74	37	19	12	6	3	2	1	1
25	463	154	93	46	23	15	8	4	2	1	1
30	555	185	111	56	28	18	9	5	2	2	1

Fuente: Elaboración propia

La estimación de estas observaciones de acuerdo a la ecuación (3-2) entrega errores relativos inferiores al 6% (Tabla K-1 en Anexos). Por otro lado, la precisión del número promedio de observaciones según escenario tiene una variabilidad máxima del 7% (Tabla K-2 en Anexos).

4.2.4 Efecto de anonimidad de sensores

Al igual que para NGSIM, se compararán dos métodos de estimación de velocidades para celdas con observaciones de sensores móviles: promedios simples y definiciones generalizadas. Se utilizó una discretización que considera $H = 25$ celdas espaciales, que se actualizan cada $\Delta t = 5\ s$. Primero se analizarán ambos métodos por separado, para luego presentar una comparación.

a. **Promedios simples**

La Tabla 4-25 muestra el error relativo medio para el método promedios simples.

Tabla 4-25: Error relativo medio para promedios simples (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	16	16	17	18	18	19	19	19	20	19	19
2	14	14	15	17	18	18	18	18	20	19	18
3	13	13	14	16	18	18	18	19	19	19	17
4	12	12	13	16	17	18	18	19	19	18	18
5	11	11	12	15	17	18	18	19	20	19	18
10	8	9	9	12	15	16	18	19	19	19	18
15	7	7	8	10	13	15	17	18	19	19	18
20	6	6	6	9	12	14	16	18	19	19	18
25	5	5	6	8	11	13	16	18	19	18	18
30	4	5	5	7	10	12	15	17	19	18	18

Fuente: Elaboración propia

Para la mayoría de los escenarios, una mayor cantidad de datos permite disminuir el error, a pesar de estar estimando una mayor cantidad de celdas (Tabla 4-26). No siempre sucede lo anterior, como al pasar de un intervalo de muestreo de 480 a 360 o bien 240 segundos. Allí la estimación general empeora, al estar estimando cierta cantidad extra de celdas, o bien, la estimación del error no es robusta al obtener el indicador a través de pocas celdas.

Tabla 4-26: Porcentaje de celdas sin estimación para promedios simples [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	79,6	81,8	84,1	91,5	95,6	97,0	98,5	99,2	99,6	99,8	99,8
2	64,2	67,8	71,9	84,1	91,5	94,2	97,0	98,5	99,3	99,5	99,6
3	52,1	56,7	61,6	77,2	87,6	91,4	95,6	97,8	98,9	99,3	99,4
4	43,3	48,3	53,8	71,5	83,9	88,7	94,1	97,0	98,5	99,0	99,2
5	35,5	40,7	46,9	66,2	80,4	86,3	92,8	96,3	98,1	98,7	99,0
10	15,8	20,2	26,2	46,9	66,2	75,4	86,5	92,9	96,4	97,6	98,2
15	7,4	10,4	15,2	34,3	55,4	66,4	80,8	89,8	94,8	96,5	97,3
20	4,2	6,0	9,4	25,8	46,9	58,9	75,8	86,9	93,2	95,5	96,5
25	2,8	3,9	6,2	19,3	40,1	52,7	71,4	84,2	91,8	94,5	95,8
30	2,2	2,9	4,4	14,8	34,4	47,3	67,4	81,7	90,4	93,6	95,1

Fuente: Elaboración propia

El porcentaje de celdas sin estimación está directamente relacionado con la penetración y tasa de muestreo. Como es de esperar, a mayor cantidad de datos, más celdas se estiman. Para los escenarios analizados, este porcentaje varía desde un 99,8% a un 2,2%. La precisión de la estimación es cercana al 1% para escenarios con alta cantidad de datos, bajando a un 12,8% para el escenario de más baja cantidad (Tabla K-3 en Anexos).

b. **Definiciones generalizadas**

La Tabla 4-27 presenta los errores relativos medios para los distintos escenarios analizados para definiciones generalizadas.

Tabla 4-27: Error relativo medio para definiciones generalizadas (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	16	16	16	17	19	23	33	48	61	55	57
2	14	14	14	16	19	22	33	48	61	57	55
3	13	13	13	15	18	22	33	46	63	61	57
4	12	12	12	15	18	22	33	47	60	59	54
5	11	11	11	14	18	22	33	47	62	60	59
10	8	8	9	12	16	21	33	47	62	59	60
15	6	7	7	10	14	20	32	47	63	59	60
20	5	6	6	9	13	18	31	47	62	59	64
25	5	5	5	8	13	18	31	46	62	60	62
30	4	4	5	7	12	17	30	46	63	59	60

Fuente: Elaboración propia

En general, una mayor cantidad de datos permite estimar de mejor manera las velocidades de las celdas. En el caso de situaciones de pocos datos, más datos permite estimar más celdas, pero sin disminuir el error promedio. La Tabla 4-28 muestra el porcentaje de celdas sin estimación. Muy pocas celdas pueden estimarse en el caso de escenarios de baja cantidad de datos.

Tabla 4-28: Porcentaje de celdas sin estimación para definiciones generalizadas [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	79,6	81,8	84,1	91,5	95,6	97,1	98,6	99,4	99,8	100,0	100,0
2	64,3	67,9	71,9	84,2	91,6	94,3	97,2	98,7	99,7	99,9	100,0
3	52,2	56,7	61,6	77,3	87,7	91,6	95,8	98,1	99,5	99,9	100,0
4	43,3	48,3	53,9	71,6	84,0	88,9	94,4	97,5	99,3	99,9	100,0
5	35,6	40,7	46,9	66,2	80,6	86,5	93,1	96,8	99,1	99,8	100,0
10	15,8	20,2	26,3	47,0	66,6	75,9	87,2	94,0	98,3	99,7	100,0
15	7,4	10,4	15,2	34,4	55,9	67,0	81,7	91,3	97,5	99,5	99,9
20	4,2	6,0	9,4	25,8	47,4	59,6	77,0	88,8	96,8	99,4	99,9
25	2,8	3,9	6,2	19,4	40,6	53,6	72,8	86,5	96,1	99,2	99,9
30	2,2	2,9	4,4	14,8	35,1	48,2	68,9	84,3	95,4	99,1	99,9

Fuente: Elaboración propia

La precisión de la estimación es cercana al 1% para escenarios con alta cantidad de datos, mientras que la precisión baja drásticamente en el escenario de más baja cantidad, cercano al 100% (Tabla K-4 en Anexos).

c. Comparación: el efecto de la anonimidad en la estimación de estados de tráfico

La Tabla 4-29 muestra la diferencia entre errores estimados por ambos métodos, donde un valor negativo indica que definiciones generalizadas estimó mejor, mientras que uno positivo lo contrario.

Tabla 4-29: Diferencia entre errores relativos medios estimados (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	0*	0*	-1	-1	1	4	14	29	41	35	37
2	0*	0	-1	-1	1	4	15	30	42	38	37
3	0*	0	-1	-1	1	4	14	27	44	42	40
4	0*	0	-1	-1	1	4	15	28	40	41	37
5	0*	0	-1	-1	1	4	14	28	43	40	41
10	0*	0	-1	-1	1	4	15	29	42	41	42
15	0	0	-1	-1	1	5	15	29	43	41	42
20	0	0	-1	0	1	4	15	29	43	41	46
25	0	0	-1	0	1	5	15	29	43	41	44
30	0	0	0	0	1	5	15	29	44	41	42

(Definiciones generalizadas menos promedios simples. * indica igualdad estadística)

Fuente: Elaboración propia

Para intervalos de muestreo de uno y tres segundos, prácticamente no hubo diferencias entre las estimaciones. Para intervalos de muestreos de cinco y diez segundos, definiciones generalizadas logró estimar levemente superior. Para valores mayores, promedios simples logra estimar mejor, y de manera significativa. Para intervalos grandes, definiciones generalizadas no logra reconstruir la totalidad de la trayectoria (o bien, de manera imprecisa) por lo que aumenta el error en la estimación.

4.2.5 Comparación y requerimiento de heurística

Al igual que con la anterior base de datos, la estimación del campo de velocidades se realizará mediante los métodos expuestos en la sección 3.6. Para FBH-v, se utilizarán los siguientes valores de $L = \{0{,}1;\ 0{,}3;\ 0{,}5;\ 0{,}7;\ 0{,}9\}$ para la elección de L^*. Además, se calibró la relación $V(k)$ obteniendo los parámetros mostrados en la Tabla 4-30. Esta calibración se realizó implementando diversas espiras virtuales a lo largo de la autopista simulada en AIMSUN.

Tabla 4-30: Parámetros para relación Smulders [AIMSUN]

Parámetro	Valor
q_{max}	$2.113{,}92\ \frac{\text{veh}}{\text{hr}\cdot\text{pista}}$
v_{max}	$102{,}08\ \frac{\text{km}}{\text{hr}}$
v_c	$88{,}08\ \frac{\text{km}}{\text{hr}}$
k_c	$24{,}0\ \frac{\text{veh}}{\text{km}\cdot\text{pista}}$
k_j	$175{,}0\ \frac{\text{veh}}{\text{km}\cdot\text{pista}}$
w	$14{,}0\ \frac{\text{km}}{\text{hr}}$

Fuente: Elaboración propia

Con lo anterior, es posible realizar la estimación de estados de tráfico utilizando los distintos métodos anteriormente expuestos. La Figura 4-15 muestra el campo de velocidades real y estimado para un escenario donde $P = 5\%$ y $\tau = 10\ \frac{\text{seg}}{\text{obs}*\text{veh}}$.

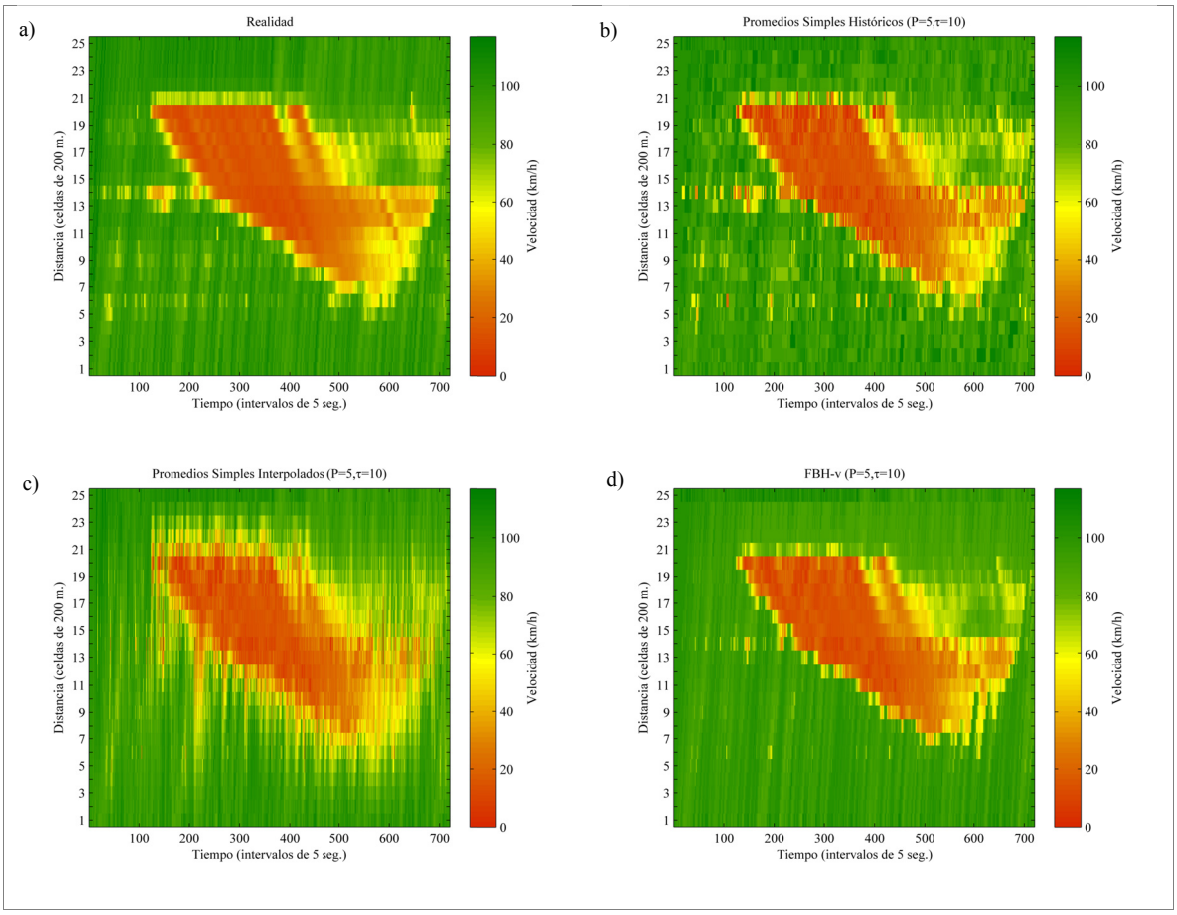

Figura 4-15: Campo de velocidades real y estimado para un escenario con P=5% y τ=10 s [AIMSUN]

a) Real; b) Promedios simples históricos; c) Promedios simples interpolados y d) FBH-v.

Estimación realizada a partir de 7.685 observaciones.

Fuente: Elaboración propia

A partir de inspección visual, los métodos parecen replicar correctamente el incidente. Promedios simples interpolados presenta velocidades más bajas en torno al accidente que la realidad. Promedios simples históricos replica la realidad, pero con una baja resolución.

La Figura 4-16 muestra las estimaciones al disminuir la penetración a un 2%, manteniendo el intervalo de muestreo.

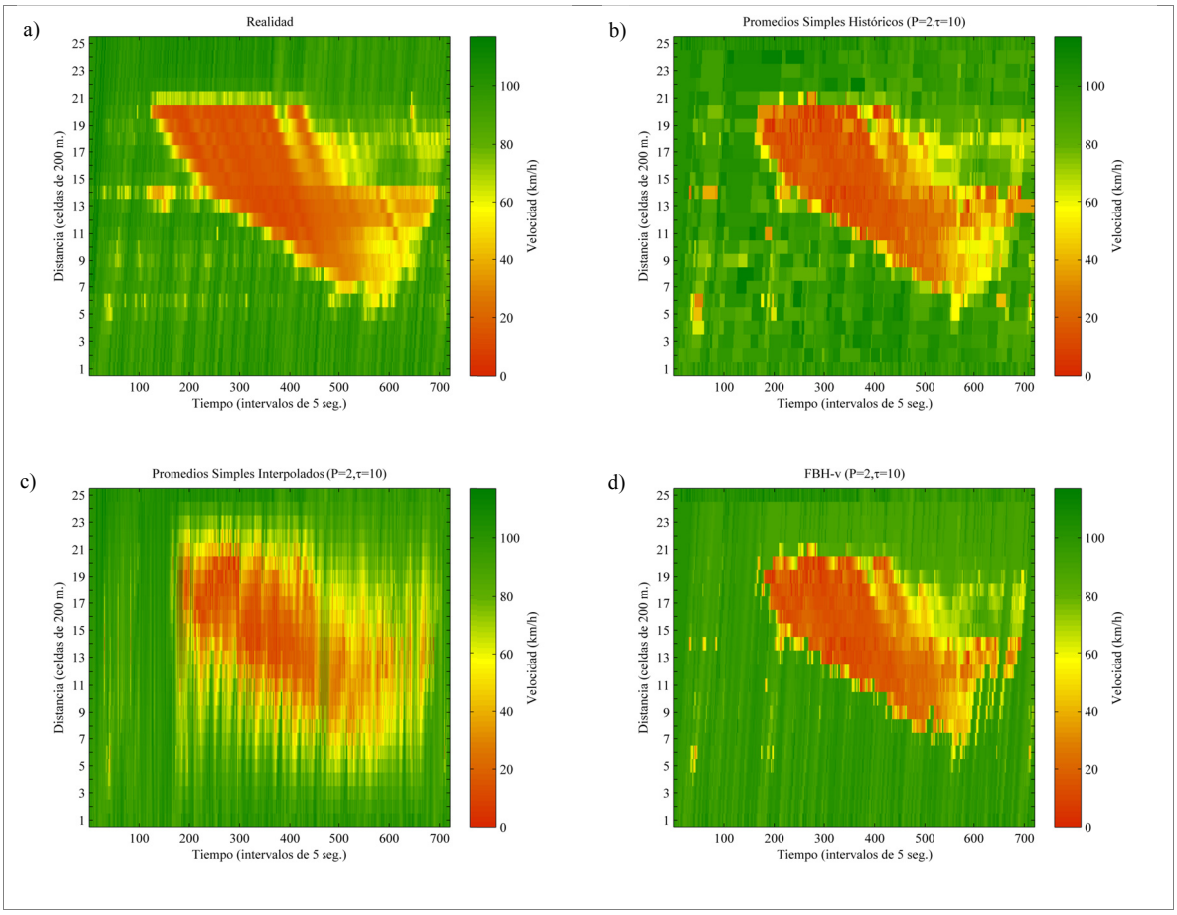

Figura 4-16: Campo de velocidades real y estimado para un escenario con P=2% y τ=10 s [AIMSUN]

a) Real; b) Promedios simples históricos; c) Promedios simples interpolados y d) FBH-v.

Estimación realizada a partir de 3.008 observaciones.

Fuente: Elaboración propia

Se aprecia una estimación difusa por parte de promedios simples interpolados. Los dos métodos restantes replican correctamente la forma del incidente, a pesar de una baja resolución en el caso de promedios simples históricos.

La Figura 4-17 muestra las estimaciones para un escenario donde la penetración alcanza un 1% y el intervalo de muestreo 120 segundos.

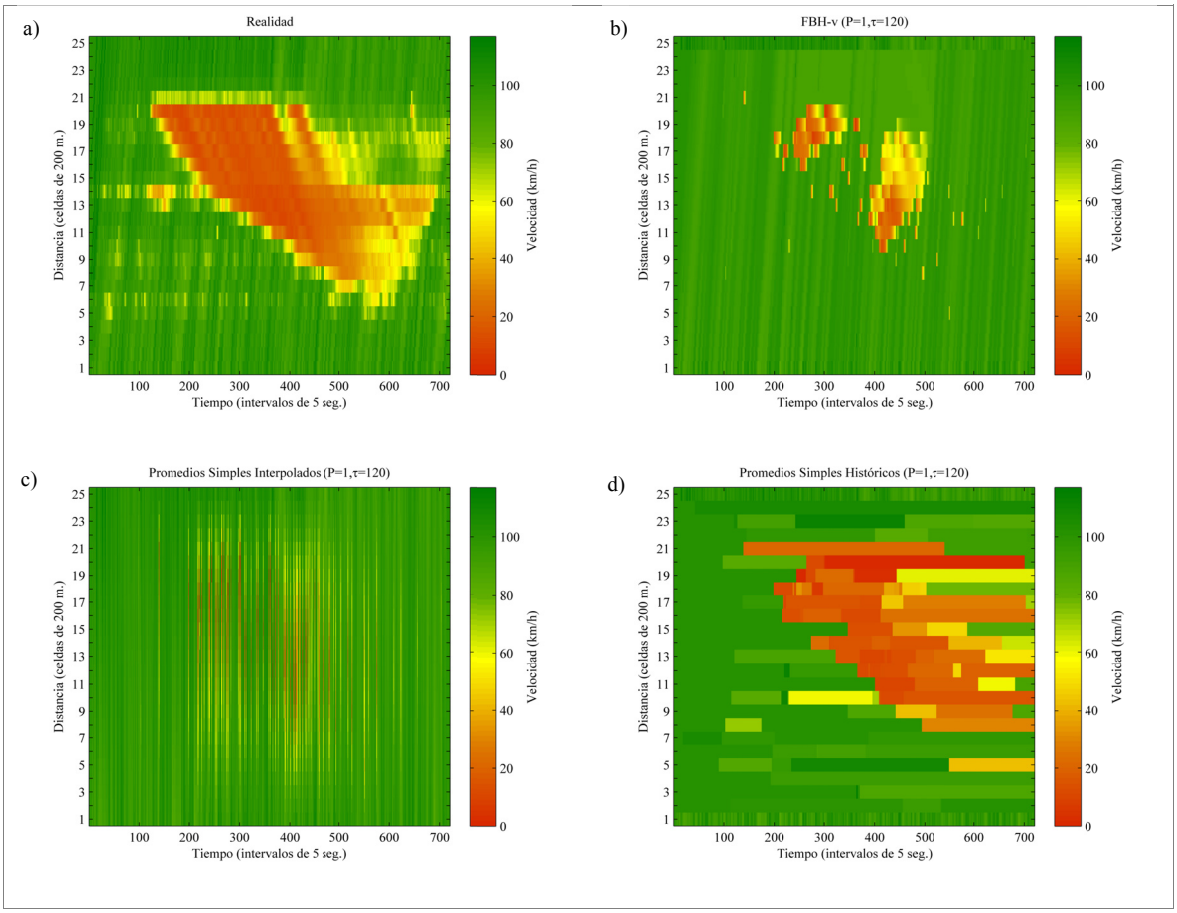

Figura 4-17: Campo de velocidades real y estimado para un escenario con P=1% y τ=120 s [AIMSUN]

a) Real; b) Promedios simples históricos; c) Promedios simples interpolados y d) FBH-v.

Estimación realizada a partir de 126 observaciones.

Fuente: Elaboración propia

Para este escenario, promedios simples interpolados no logra estimar correctamente, al utilizar de manera espacial y no temporal los datos recibidos. Por otro lado, FBH-v disipa rápidamente la congestión, pues el modelo no logra interpretar el cuello de botella que existe luego del incidente, con los pocos datos que recibe. Promedios simples históricos siempre presenta el incidente a pesar de su simpleza.

Los errores relativos en la estimación de cada método y escenario se presentan en el Anexo (Tabla K-5, Tabla K-6 y Tabla K-7). A continuación se procederá a compararlos.

La Tabla 4-31 muestra la diferencia de errores entre promedios simples históricos e interpolados. Sectores negativos indican que promedios simples históricos estimó mejor.

Tabla 4-31: Errores de Promedios Simples Históricos menos Promedios Simples Interpolados (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	**1**	**3**	**5**	**10**	**20**	**30**	**60**	**120**	**240**	**360**	**480**
1	-10	-10	-10	-23	-39	-48	-57	-53	-40	-28	-21
2	-5	-5	-6	-12	-26	-34	-49	-57	-52	-41	-30
3	-3	-3	-3	-8	-17	-25	-41	-54	-53	-45	-35
4	-2	-2	-2	-6	-12	-19	-34	-49	-52	-42	-36
5	-1	-1	-2	-5	-10	-15	-29	-44	-50	-44	-36
10	0	0	0	-2	-5	-8	-15	-29	-40	-38	-34
15	0	0	0	-1	-3	-5	-10	-20	-31	-30	-29
20	0	0	0	-1	-2	-3	-7	-15	-23	-23	-23
25	0	0	0	0	-1	-2	-6	-12	-19	-19	-18
30	0	0	0	0	-1	-2	-4	-9	-16	-13	-15

Fuente: Elaboración propia

A lo largo de todos los escenarios, se pudo verificar que promedios simples históricos estimó igual o mejor que promedios simples interpolados. Las diferencias absolutas alcanzan hasta una diferencia de un 57%. También existe un sector donde no existe diferencia en la calidad de la estimación ($P \geq 10\%$ & $\tau \leq 5$).

La diferencia entre los errores de promedios simples históricos y FBH-v se presenta en la Tabla 4-32. Un valor negativo indica que promedios simples históricos estimó mejor.

Tabla 4-32: Errores de Promedios Simples Históricos menos FBH-v (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	-9	-9	-9	-10	-12	-16	-27	-40	-35	-23	-17
2	-2	-2	-2	-2	-3	-5	-10	-24	-39	-35	-26
3	1	1	1	1	1	0	-3	-14	-32	-35	-30
4	2	2	2	2	2	1	-1	-9	-21	-26	-28
5	2	2	2	2	2	2	0	-5	-16	-25	-27
10	2	2	2	3	3	3	2	-1	-8	-15	-21
15	1	2	2	3	4	4	3	0	-6	-12	-19
20	1	1	1	3	4	4	3	1	-4	-10	-17
25	1	1	1	2	4	4	3	1	-3	-8	-15
30	0	1	1	2	3	4	4	1	-3	-8	-16

Fuente: Elaboración propia

Existen escenarios con valores negativos y otros con positivos. Para situaciones con una baja penetración ($P \leq 2\%$) o alto intervalo de muestreo ($\tau \geq 240$) promedios simples logra estimar mejor que FBH-v. Esto se debe a que las condiciones de borde impuestas en la simulación están siempre a flujo libre, y la baja cantidad de datos recibidos no permiten detectar el cuello de botella producido por el incidente. Así, la congestión inicialmente detectada mediante incorporación de datos de sensores móviles es rápidamente disipada. Al existir una mayor cantidad de datos, el cuello de botella es detectado, por lo que la congestión se propaga efectivamente, simulando correctamente la realidad.

También es posible notar que no necesariamente existe una relación entre los escenarios con valores negativos con la densidad de observaciones. Por ejemplo, el escenario $P = 1\%, \tau = 1\,s$ tiene una densidad de observaciones de 19, al igual que el escenario $P = 5\%, \tau = 5\%$, pero existe una diferencia de 11 puntos. Esto resalta la importancia de, no sólo el número de datos, sino también cómo se logra (composición).

Es posible incorporar de manera explícita en el modelo una restricción de capacidad en la celda en que sucede el incidente. Esto permite reducir en hasta un 20% el error en la estimación para el caso de FBH-v.

4.2.6 Métrica para inferir errores

Para explicar el error en función de la penetración y frecuencia de muestreo, y al igual que en el capítulo anterior, se utilizó una función Translog, descrita por la ecuación (4-1). La Tabla 4-33 muestra los parámetros $\hat{\alpha}_i$ estimados a través de una regresión lineal:

Tabla 4-33: Parámetros estimados para diferentes métodos de estimación [AIMSUN]

Parámetro	Variable	PS Históricos	PS Interpolados	FBH-v
$\hat{\alpha}_1$	$\ln(P)$	$-0{,}278$ $(-17{,}63)$	$-0{,}790$ $(-33{,}51)$	$-0{,}315$ $(-23{,}74)$
$\hat{\alpha}_2$	$\ln(z)$	$-0{,}134$ $(-14{,}09)$	$-0{,}355$ $(-33{,}38)$	$-0{,}0305$ $(-4{,}88)$
$\hat{\alpha}_3$	$\ln(P) \cdot \ln(z)$	$-0{,}0223$ $(-16{,}54)$	$-0{,}0487$ $(-25{,}75)$	$-0{,}0372$ $(-30{,}67)$
$\hat{\alpha}_4$	$\ln^2(P)$	$0{,}001317$ $(0{,}59)$	$-0{,}0356$ $(-9{,}00)$	$0{,}0213$ $(9{,}75)$
$\hat{\alpha}_5$	$\ln^2(z)$	$0{,}0197$ $(21{,}08)$	$0{,}0241$ $(22{,}43)$	$0{,}0488$ $(72{,}81)$
$\hat{\alpha}_0$	Constante	$-3{,}757$ $(-111{,}28)$	$-4{,}250$ $(-113{,}30)$	$-4{,}060$ $(-193{,}57)$
	N	1500	1650	1500

Estadísticos *t* entre paréntesis

Todos los parámetros son significativos, salvo $\ln^2(P)$ salvo para promedios simples históricos, el cual se mantuvo en el modelo por criterio del modelador. Se removieron aquellos escenarios que producían elasticidades no-positivas al evaluar las expresiones (4-2) y (4-3), terminando con 1500, 1650 y 1500 observaciones para los métodos expuestos. En la Figura 4-18 se presentan las superficies estimadas para cada método.

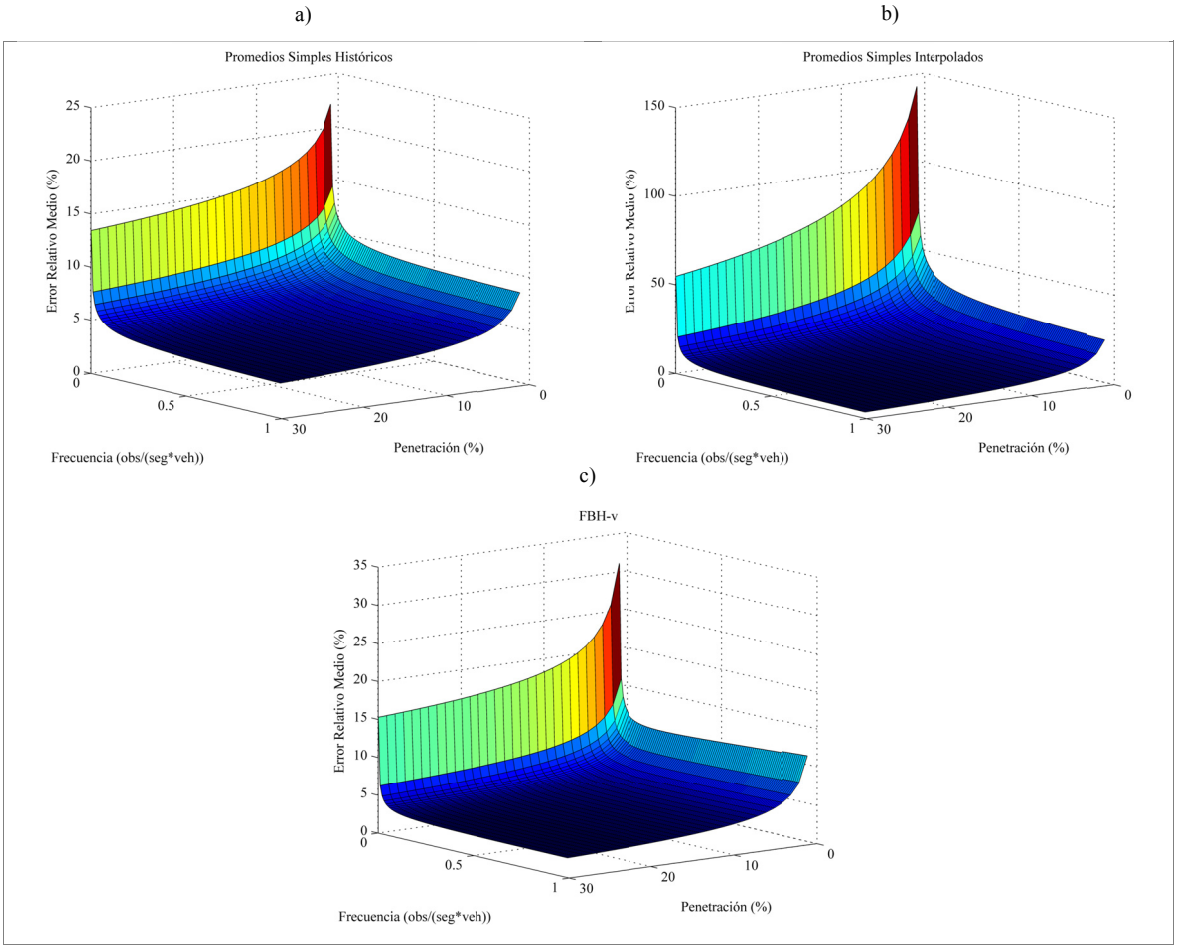

Figura 4-18: Superficies de error estimadas

a) Promedios simples históricos; b) Promedios simples interpolados y c) FBH-v [diferentes escalas]

Fuente: Elaboración propia

Para una máxima cantidad de datos, todos los métodos convergen a un error similar, puesto que la información es redundante. En el caso de menos información, promedios simples históricos siempre supera a promedios simples interpolados. Por otro lado, promedios simples históricos entrega mejores resultados para penetraciones o frecuencias de muestreo muy bajas. En otro caso, es de conveniencia utilizar FBH-v.

4.2.7 El efecto de la composición de información

Las curvas de indiferencia, determinadas para cada método de estimación utilizando la ecuación (4-4), se presentan en la Figura 4-19.

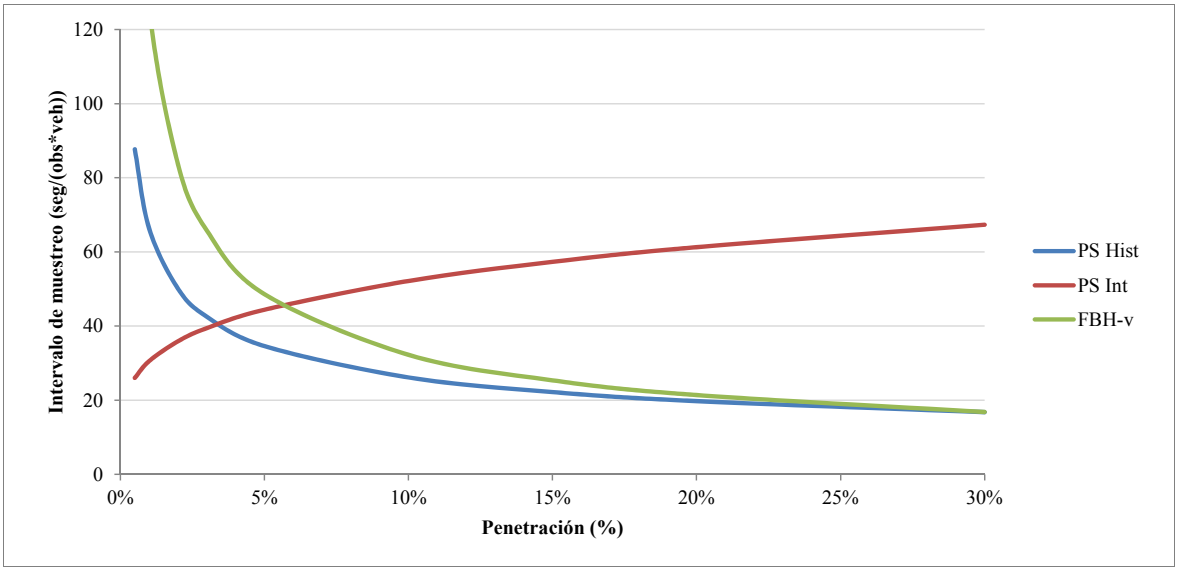

Figura 4-19: Curvas de indiferencia de aumento de observaciones [AIMSUN]

Fuente: Elaboración propia

Para promedios simples históricos y FBH-v, la curva de indiferencia de aumento es decreciente con respecto la penetración, hasta alcanzar valores cercanos a los 20 segundos. Por otro lado, la curva asociada a promedios simples interpolados crece al incrementar la penetración.

El cociente de elasticidades $E_{\varepsilon,P}/E_{\varepsilon,z}$, para cada escenario y método de estimación, se presenta en la Tabla K-8, Tabla K-9 y Tabla K-10 del Anexo. Sus valores se mueven desde 0,2 hasta 27, dependiendo del método de estimación y escenario. En la Figura 4-20 se presenta gráficamente el cociente para los distintos métodos. Presenta una escala de valores entre 0,5 y 1,5.

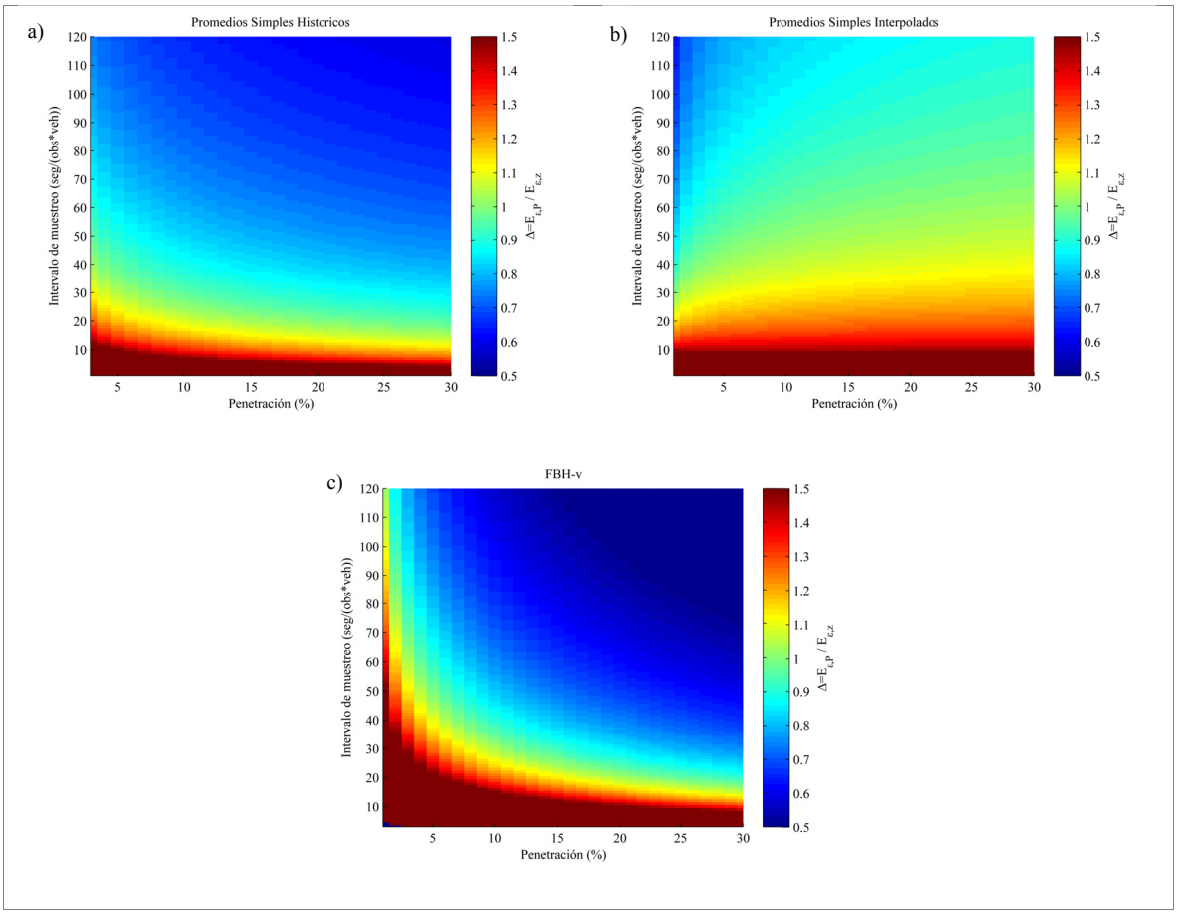

Figura 4-20: Cociente Δ para los diferentes métodos de estimación [AIMSUN]

a) Promedios simples históricos; b) Promedios simples interpolados y c) FBH-v.

Valores mayores o menores que el límite de la escala se presentan del color del valor límite.

Fuente: Elaboración propia

4.3 Comparación y discusión de resultados de NGSIM y AIMSUN

La presente sección tiene como objetivo comparar y discutir los resultados obtenidos y analizados para las bases de datos NGSIM y AIMSUN.

En cuanto a la cuantificación de las observaciones mediante la ecuación (3-2), la estimación del número esperado de observaciones por escenario entrega errores relativos inferiores al 6% para ambas bases de datos en caso de conocer los valores correctos del tiempo de viaje y número de vehículos. La precisión del número promedio de observaciones según escenario tiene un valor máximo de un 5% en el caso de NGSIM, y

de un 7% para AIMSUN, ambos en escenarios de baja cantidad de datos. Para escenarios de alta cantidad de datos, la precisión es cercana al 1% en ambas bases de datos. La calidad de la estimación de cantidad de observaciones parece ser independiente del largo de carretera y matriz origen-destino de los vehículos.

En relación al efecto de la anonimidad de sensores para la estimación de velocidades para celdas-intervalo con información, ambas bases de datos entregan resultados similares. Cuando el intervalo de muestreo es menor o igual a 10 segundos, definiciones generalizadas estima levemente mejor que promedios simples. Para intervalos de muestreo mayores, promedios simples estima notablemente mejor. Cabe notar que este resultado es constante para ambas bases de datos, a pesar de la diferente discretización de la vía ocupada para NGSIM (Δx=30 s, Δt=50 m) y AIMSUN (Δx=5 s, Δt=200 m). Además, depende principalmente del intervalo de muestreo y no de la penetración.

Tres métodos de estimación de estados de tráfico fueron comparados: dos simples (promedios simples interpolados e históricos) y una heurística más avanzada (FBH-v). En el caso de NGSIM, FBH-v siempre entregó mejores resultados, y promedios simples históricos el peor. Para AIMSUN, promedios simples interpolados siempre entregó peores resultados porque las condiciones de borde son menos relevantes (sección más larga). Para baja penetración o frecuencia de muestreo, promedios simples históricos entrega mejores resultados que FBH-v, y para los otros escenarios, sucede lo contrario. También se encontró que la densidad de observaciones no es un buen indicador para el error estimado para cada método y base de datos. Si comparamos escenarios con igual densidad de observaciones entre NGSIM y AIMSUN (Tabla 4-4 y Tabla 4-24) con los errores de, por ejemplo, promedios simples históricos (Tabla 4-13 y Tabla K-5) se puede notar bastantes diferencias. Esto significa que la densidad de observaciones no es un buen predictor del error, dada las diferencias en la composición de estas observaciones.

Se propuso una métrica para estimar errores, utilizando una función Translog, en base a la penetración y a la frecuencia de muestreo. Esto permitió analizar el efecto de la composición de información, diferenciando penetración y frecuencia. La curva de indiferencia de aumento de observaciones es parecida para el caso de promedios simples

históricos para NGSIM y AIMSUN, y diferente en el caso de FBH-v y promedios simples interpolados. Se cree que las diferencias que existen entre estas curvas surgen a raíz de la utilización de las condiciones de borde. El método de promedios simples históricos no las utiliza, y las curvas son parecidas. Los últimos dos métodos las utilizan, y existe un cambio en esta curva. Cabe recordar que para AIMSUN, las condiciones de borde siempre están a flujo libre, mientras que en NGSIM se encuentran principalmente congestionadas. El cociente de elasticidades tiene variaciones parecidas en los escenarios analizados, donde puede ser dos o tres veces más conveniente un aumento de penetración que de frecuencia (o bien, frecuencia en vez de penetración).

5. CONCLUSIONES

A partir de este estudio, se pudo examinar en detalle la relación existente entre los datos recolectados a partir de sensores móviles y la forma de estimar estados de tráfico. Esto se logró aplicando la metodología propuesta a dos bases de datos distintas: **NGSIM** y **AIMSUN**. NGSIM posee trayectorias de todos los vehículos de una porción de vía cercana a 600 m, con una rampa de entrada y salida. Por otro lado, AIMSUN posee trayectorias de una porción de vía de 5 km, con dos rampas de entrada y dos de salida. Se simuló un incidente que reduce la capacidad de la autopista, que es removido luego de 20 min. Se analizaron tres métodos para la estimación de estados de tráfico: promedios simples históricos, promedios simples interpolados y una heurística más avanzada, FBH-v. Los dos primeros son métodos simples, mientras que el tercero incorpora la teoría de tráfico vehicular a la de estimación.

En cuanto a la **cuantificación de observaciones**, se propuso una medida que permite cuantificar, de manera a-priori, la cantidad de observaciones que enviarán vehículos equipados en cierta porción de vía con una frecuencia de muestreo fija. En caso de conocer los parámetros correctos, la estimación del número esperado de observaciones por escenario entrega errores relativos menores al 6% para ambas bases de datos. Esto se cumple si se conocen los valores reales de los parámetros en la ecuación. La variabilidad del número de observaciones en cada escenario depende de la cantidad de datos que se envían: un escenario de baja penetración y frecuencia de muestreo presenta una variabilidad más alta que para una alta penetración y frecuencia.

Se cuantificó el **efecto de la anonimidad** en la estimación de estados de tráfico. Se utilizaron dos métodos para estimar velocidades en celdas con información: promedios simples (asegurando la anonimidad) y definiciones generalizadas (necesita identificar trayectorias, por lo que se trabaja con escenarios pseudoanónimos). Los resultados indican que para intervalos de muestreo menores a 10 segundos, definiciones generalizadas puede estimar levemente mejor que promedios simples. Para mayores intervalos, promedios simples estima notablemente mejor, al no poder reconstruirse la

trayectoria correctamente. Esto sucede en ambas bases de datos. Así, la pérdida de precisión en la estimación al preservar la privacidad es baja. A la luz de los resultados, y considerando que la anonimidad de sensores móviles es de importancia si se desea alcanzar un alto nivel de penetración de sensores en la población, se recomienda mantener un esquema anónimo. Así, este sistema se hace más atractivo y podría aumentar el número de participantes, con lo cual se hace cada vez menos necesario re-identificar vehículos. Esto no quiere decir que una etiqueta que permita re-identificar el vehículo no sea de utilidad: puede que sea útil para otro propósito, como obtener tiempos de viaje de manera directa o inferir orígenes y destinos de viajes. O puede haber otro método (diferente a definiciones generalizadas) que pueda usar mejor esta información extra para estimar velocidades en celdas-intervalo.

En cuanto a la **capacidad de estimación de los diferentes métodos de estimación de estados de tráfico**, se pudo verificar resultados diferentes para cada base de datos. En el caso de NGSIM, FBH-v superó la capacidad de estimación de los otros dos métodos. Promedios simples históricos siempre estimó de peor manera. En el caso de AIMSUN, promedios simples interpolados lo hizo peor. Para escenarios de baja penetración o frecuencia de muestreo, promedios simples históricos entrega mejores que resultados que FBH-v. En el resto de los escenarios FBH-v lo hizo mejor. La heurística, al no detectar físicamente el incidente, no logra retener la congestión y la disipa. También fue posible identificar escenarios donde debido a la saturación de información es indiferente el método de estimación a utilizar. Lo primero que se desprende de esto es que no existe un método que sea siempre conveniente a utilizar. Este dependerá del nivel de datos con que se cuenta, el escenario de muestreo, y la vía donde se realizarán las estimaciones. Las condiciones de borde en una porción corta de vía (NGSIM) son sumamente útiles para estimar lo que sucede al interior. En el caso de una autopista más larga, la estimación se torna más difícil. En caso de FBH-v, y en escenarios de baja cantidad de datos, el método no logra replicar el incidente al no notar una baja en la capacidad de cierta celda. Es por esto que la fusión de datos, a partir de las redes sociales, puede volverse conveniente. Esto permitirá mejorar la capacidad de estimación de los modelos.

Se obtuvo una **métrica para inferir el error** en la estimación de estados de tráfico. Esta métrica permite identificar el error en la estimación, en base a información de la penetración de sensores y la frecuencia de muestreo. Además permite la distinción entre los efectos que provoca un aumento de la penetración en contra de un aumento de la frecuencia de muestreo. Esto ya que se verificó que la cantidad o densidad de observaciones no era un buen predictor del error.

Se determinó el **efecto de la composición de datos** en la estimación de estados de tráfico, dado que una misma cantidad de datos puede obtenerse de diferentes maneras, y esto incide en la estimación. Para cada método de estimación, se encontró una curva de indiferencia de aumento de observaciones. Esta curva determina escenarios donde un aumento marginal de observaciones proveniente de un aumento de la penetración es equivalente al mismo aumento de observaciones proveniente del aumento de la frecuencia. Esta curva es diferente para cada base de datos y forma de estimación. Para escenarios sobre esta curva, es más conveniente un aumento de la frecuencia de muestreo que un aumento de penetración. Bajo esta curva sucede lo contrario. Para cuantificar los beneficios de un aumento de penetración o de frecuencia, se analizó el cociente de elasticidades $E_{\varepsilon,P}/E_{\varepsilon,z}$. Este cociente permitió identificar los beneficios, en términos de reducción del error, de un aumento de observaciones a través de un aumento de penetración, en contraposición a un mismo aumento de observaciones a través de un aumento de la frecuencia. El cociente indica que dependiendo del escenario, un aumento de la penetración puede ser hasta dos o tres veces más conveniente que un aumento de la frecuencia, y viceversa. Estos resultados permiten a un ente que realiza las estimaciones entender los efectos de cambios de estas variables. Adicionalmente, se verificó que la curva de indiferencia de aumento de observaciones coincide con la solución al problema de minimización del error, sujeto a un número de observaciones constante. Esto significa que se encontró la combinación óptima de la penetración y frecuencia para cualquier número de observaciones. Esto puede ser de interés en caso de que un sistema no pueda procesar más que cierta cantidad de datos, principalmente por alguna restricción de

costo. Esta puede ser por una capacidad de hardware, o bien, en caso de que se deba pagar a participantes por el envío de datos.

Como **investigación futura**, se propone realizar un análisis similar para diferentes estrategias de muestreo. En específico, la estrategia de muestreo accionado por aceleraciones mayores o menores a cierto umbral definido. En esta estrategia, los vehículos envían información ante cambios bruscos de velocidad. Se cree que esta información puede ser tan o más valiosa como aquella enviada de manera temporal, puesto que esta información está ligada a las ondas de choque generadas por el tráfico. Por lo tanto, la identificación de estados de tráfico se puede facilitar. También puede ser interesante analizar el efecto de estas variables en la estimación de estados de tráfico en arterias urbanas. Asimismo, es interesante desarrollar metodologías para incorporar la información recopilada a partir de las redes sociales en la estimación de fenómenos asociados al tráfico vehicular. Finalmente recordar que, al trabajar con personas, siempre es importante integrar formalmente el concepto de privacidad en los modelos, ya que "en la ausencia del derecho a la privacidad, no puede existir una verdadera libertad de expresión y opinión, y por lo tanto, no existiría una democracia efectiva"[15].

[15] Dilma Rousseff en la Asamblea General de las Naciones Unidas. Septiembre 24 de 2013, Nueva York.

BIBLIOGRAFÍA

Aw, A., & Rascle, M. (2000). Resurrection of "second order" models of traffic flow. *SIAM Journal on Applied Mathematics*, *60*(3), 916–938.

Blandin, S., Argote, J., Bayen, A. M., & Work, D. B. (2013). Phase transition model of non-stationary traffic flow: Definition, properties and solution method. *Transportation Research Part B: Methodological*, *52*, 31–55. doi:10.1016/j.trb.2013.02.005

Cassidy, M. (2003). Traffic Flow and Capacity. *Handbook of Transportation Science*, 1–36.

Chen, C., Kwon, J., & Rice, J. (2003). Detecting errors and imputing missing data for single-loop surveillance systems. *Transportation Research Record*, (03), 160–167.

Chu, L., Oh, S., & Recker, W. (2005). Adaptive Kalman filter based freeway travel time estimation. In *84th TRB Annual Meeting* (pp. 1–21).

Colombo, R. (2002). A 2 × 2 hyperbolic traffic flow model. *Mathematical and Computer Modelling*, *35*(5-6), 683–688. doi:10.1016/S0895-7177(01)00190-X

Courant, R., Friedrichs, K., & Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, *11*(2), 215–234.

Daganzo, C. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, *28*(4), 269–287.

Daganzo, C. (1995a). Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B: Methodological*, *29B*(4), 277–286.

Daganzo, C. (1995b). The cell transmission model, part II: network traffic. *Transportation Research Part B: Methodological*, *29B*(2), 79–93.

Daganzo, C. (1999). The lagged cell-transmission model. In *14th International Symposium on Transportation and Traffic Theory* (pp. 81–104).

Edie, L. (1963). Discussion of traffic stream measurements and definitions. *Port of New York Authority*, 56.

FHWA. (2007). NGSIM US Highway 101 Dataset.

Frank, R., Mouton, M., & Engel, T. (2012). Towards collaborative traffic sensing using mobile phones (Poster). *2012 IEEE Vehicular Networking Conference (VNC)*, 115–120. doi:10.1109/VNC.2012.6407419

Gazis, D., & Knapp, C. (1971). On-line estimation of traffic densities from time-series of flow and speed data. *Transportation Science*, *5*(3), 283–301.

Gelb, A. (1974). *Applied optimal estimation* (p. 382). The MIT Press.

GMSA. (2011). *Observatorio Móvil de América Latina 2011*.

Godunov, S. K. (1959). A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Mat. Sb.*, *47*(89), 271–306.

Greenshields, B. D. (1935). A Study of Traffic Capacity. In *Highway Research Board* (Vol. 14, pp. 448–477).

Hayashi, F. (2000). *Econometrics* (p. 712). Princeton University Press.

Herrera, J. C. (2009). *Assessment of GPS-enabled smartphone data and its use in traffic state estimation for highways*. PhD. Thesis, University of California, Berkeley.

Herrera, J. C., & Bayen, A. M. (2010). Incorporation of Lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B: Methodological*, *44*(4), 460–481. doi:10.1016/j.trb.2009.10.005

Herrera, J. C., Work, D. B., Herring, R., Ban, X., Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, *18*(4), 568–583. doi:10.1016/j.trc.2009.10.006

Herring, R., Hofleitner, A., Amin, S., Abou Nasr, T., Khalek, A. A., Abbeel, P., & Bayen, A. M. (2010). Using Mobile Phones to Forecast Arterial Traffic through Statistical Learning, 1–22.

Hoh, B., Gruteser, M., Herring, R., & Ban, J. (2008). Virtual trip lines for distributed privacy-preserving traffic monitoring. In *6th International Conference on Mobile Systems, Applications, and Services* (pp. 15–28). ACM.

Hoh, B., Gruteser, M., Hong, H., & Alrabady, A. (2006). Enhancing Security and Privacy in Traffic-Monitoring Systems. *IEEE Pervasive Computing*, *5*(4), 38–46. doi:10.1109/MPRV.2006.69

Hoh, B., Iwuchukwu, T., Jacobson, Q., Work, D. B., Bayen, A. M., Herring, R., Herrera, J. C., Gruteser, M., Annavaram, M., & Ban, J. (2012). Enhancing Privacy and Accuracy in Probe Vehicle-Based Traffic Monitoring via Virtual Trip Lines. *IEEE Transactions on Mobile Computing*, *11*(5), 849–864.

Jeske, T. (2013). Floating Car Data from Smartphones : What Google and Waze Know About You and How Hackers Can Control Traffic. In *Black Hat Europe 2013*.

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

Klein, L., Mills, M., & Gibson, D. (2006a). *Traffic Detector Handbook: Third Edition–Volume I* (Vol. I).

Klein, L., Mills, M., & Gibson, D. (2006b). *Traffic Detector Handbook: Third Edition–Volume II* (Vol. II).

Kmenta, A. J. (1967). On Estimation of the CES production Function. *International Economics Review*, *8*(2), 180–189.

Krause, A., Horvitz, E., Kansal, A., & Zhao, F. (2008). Toward community sensing. In *International Conference on Information Processing in Sensor Networks* (pp. 481–492). Ieee. doi:10.1109/IPSN.2008.37

Krumm, J. (2007). Inference Attacks on Location Tracks, *10*(Pervasive).

Lau, L. J., Jorgenson, D. W., & Christensen, L. R. (1973). Transcendental Logarithmic Production Frontiers. *The Review of Economics and Statistics*, *55*(1), 28–45.

Lebacque, J. (1996). The Godunov scheme and what ot means for first order traffic flow models. In *International Symposium on Transportation and Traffic Theory* (pp. 647–677).

Lesort, J., Bourrel, E., & Henn, V. (2003). Various Scales for Traffic Flow Representation: Some Reflections. In *Traffic and Granular Flow '03* (pp. 125–139).

Lighthill, M., & Whitham, G. (1955). On kinematic waves. II. A theory of traffic flow on long crowded roads. In *Proceedings of the Royal Society of London. Series A, Mathematical and Physical* (Vol. 229, pp. 317–345).

Liu, H., van Lint, H., van Zuylen, H., & Salomons, M. (2006). Neural Network Based Traffic Flow Model for Urban Arterial Travel Time Prediction. *Transportation Research Record: Journal of the Transportation Research Board*, *1968*(1), 99–108. doi:10.3141/1968-12

Long, J. S., & Ervin, L. H. (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician*, *54*(3), 217–224.

Maybeck, P. S. (1979). *Stochastic models, estimation and control. Volume 1. New York* (Vol. 1, p. 423). Academic Press, Inc.

Mihaylova, L., & Boel, R. (2004). A particle filter for freeway traffic estimation. In *43rd IEEE Conference on Decision and Control* (pp. 2106–2111).

Mihaylova, L., Boel, R., & Hegiy, A. (2006). An Unscented Kalman Filter for Freeway Traffic Estimation. In *11th IFAC Symposium on Control in Transportation Systems* (pp. 31–36).

Ministerio de Transportes y Telecomunicaciones. (2013). Ministro Errázuriz y fundador de Waze acuerdan desarrollo de nuevas aplicaciones tecnológicas para el transporte público. Retrieved May 22, 2013, from http://www.subtel.gob.cl/?option=com_content&view=article&id=3204

Mohan, P., Padmanabhan, V. N., & Ramjee, R. (2008). Nericell : Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones. In *SenSys '08 Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems* (pp. 323–336). doi:10.1145/1460412.1460444

Muñoz, L., Sun, X., Horowitz, R., & Alvarez, L. (2006). A Piecewise-Linearized Cell Transmission Model and Parameter Calibration Methodology. *Transportation Research Record: Journal of the Transportation Research Board*, *52*(5), 183–191.

Nanthawichit, C., Nakatsuji, T., & Suzuki, H. (2003). Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *Transportation Research Record: Journal of the Transportation Research Board*, *1855*, 49–59.

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, *27*(4), 289–303. doi:10.1016/0191-2615(93)90039-D

Nicholson, W., & Snyder, C. (2008). *Microeconomic Theory: Basic Principles and Extensions* (10th ed., p. 740). Cengage Learning.

Ortúzar, J. de D., & Willumsen, L. G. (2011). *Modelling Transport* (4th ed.). John Wiley & Sons, Ltd.

Papageorgiou, M. (1998). Some remarks on macroscopic traffic flow modelling. *Transportation Research Part A: Policy and Practice*, *32*(5), 323–329.

Payne, H. (1971). Models of freeway traffic and control. *Mathematical Models of Public Systems 1*, *1*(1), 51–6.

Pfitzmann, A., & Shostack, A. (2000). Anonymity, Unobservability, and Pseudonymity — A Proposal for Terminology. In *International Workshop on Design Issues in Anonymity and Unobservability Berkeley, CA, USA, July 25–26, 2000 Proceedings* (pp. 1–9).

Richards, P. (1956). Shock waves on the highway. *Operations Research*, *4*(1), 42–51.

Rose, G. (2006). Mobile Phones as Traffic Probes: Practices, Prospects and Issues. *Transport Reviews: A Transnational Transdisciplinary Journal*, *26*(3), 275–291. doi:10.1080/01441640500361108

Sanwal, K., & Walrand, J. (1995). Vehicles as probes. *California PATH Working Paper UCB-ITS-PWP-95-11*.

Smulders, S. (1990). Control of freeway traffic flow by variable speed signs. *Transportation Research Part B: Methodological*, *24B*(2), 111–132.

Strub, I. S., & Bayen, A. M. (2006). Weak formulation of boundary conditions for scalar conservation laws: An application to highway traffic modelling. *International Journal of Robust and Nonlinear Control*, *16*(September), 733–748. doi:10.1002/rnc

Sun, X., Muñoz, L., & Horowitz, R. (2004). Mixture Kalman filter based highway congestion mode and vehicle density estimator and its application. In *Proceedings of the 2004 American Control Conference* (pp. 2098–2103).

Sun, Z., Zan, B., Ban, X. (Jeff), & Gruteser, M. (2013). Privacy protection method for fine-grained urban traffic modeling using mobile sensors. *Transportation Research Part B: Methodological*, *56*, 50–69. doi:10.1016/j.trb.2013.07.010

Szeto, M., & Gazis, D. (1972). Application of Kalman filtering to the surveillance and control of traffic systems. *Transportation Science*, *6*(4).

Szeto, W. Y. (2008). Enhanced Lagged Cell-Transmission Model for Dynamic Traffic Assignment. *Transportation Research Record: Journal of the Transportation Research Board*, *2085*, 76–85. doi:10.3141/2085-09

Treiber, M., & Helbing, D. (2002). Reconstructing the Spatio-Temporal Traffic Dynamics from Stationary Detector Data. *Cooper@tive Tr@nsport@tion Dyn@mics*, *1*(3), 1–24.

Wang, H., Ni, D., Chen, Q., & Li, J. (2013). Stochastic modeling of the equilibrium speed – density relationship. *Journal of Advanced Transportation*, *47*, 126–150. doi:10.1002/atr

Wang, Y., & Papageorgiou, M. (2005). Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. *Transportation Research Part B: Methodological*, *39*(2), 141–167. doi:10.1016/j.trb.2004.03.003

Wang, Y., Papageorgiou, M., & Messmer, A. (2008). Real-time freeway traffic state estimation based on extended Kalman filter: Adaptive capabilities and real data testing. *Transportation Research Part A: Policy and Practice*, *42*(10), 1340–1358. doi:10.1016/j.tra.2008.06.001

Work, D. B., Blandin, S., Tossavainen, O. P., Piccoli, B., & Bayen, A. M. (2010). A Traffic Model for Velocity Data Assimilation. *Applied Mathematics Research eXpress*, (1), 1–35. doi:10.1093/amrx/abq002

Work, D. B., Tossavainen, O. P., Blandin, S., Bayen, A. M., Iwuchukwu, T., & Tracton, K. (2008). An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices. In *Proceedings of the 47th IEEE Conference on Decision and Control* (pp. 5062–5068).

Yeon, J., Elefteriadou, L., & Lawphongpanich, S. (2008). Travel time estimation on a freeway using Discrete Time Markov Chains. *Transportation Research Part B: Methodological*, *42*(4), 325–338. doi:10.1016/j.trb.2007.08.005

Zhang, H. M. (2002). A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research Part B: Methodological*, *36*(3), 275–290.

Zhang, X., & Rice, J. a. (2003). Short-term travel time prediction. *Transportation Research Part C: Emerging Technologies*, *11*(3-4), 187–210. doi:10.1016/S0968-090X(03)00026-3

ANEXOS

Tabla K-1: Error relativo de la estimación de número de observaciones (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	1	1	1	1	1	1	1	2	1	6	4
2	3	3	3	3	3	3	3	3	4	3	5
3	1	1	1	1	1	1	1	2	1	6	4
4	1	1	1	1	1	1	1	1	0	2	2
5	0	0	0	0	0	0	1	1	1	1	4
10	2	2	2	2	2	2	2	2	1	4	3
15	1	1	1	1	1	1	1	1	1	3	3
20	1	1	1	1	1	1	1	1	1	4	3
25	1	1	1	1	1	1	1	2	1	3	3
30	1	1	1	1	1	2	1	2	1	4	3

Fuente: Elaboración propia

Tabla K-2: Precisión del número promedio de observaciones según escenario (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	3,9	4,1	4,1	4,2	3,9	3,4	5,3	4,6	5,3	4,3	4,3
2	3,2	3,0	3,0	2,5	2,9	3,5	4,2	6,0	5,7	6,0	2,7
3	0,9	1,0	0,9	0,8	1,2	1,9	2,4	3,5	4,8	5,0	5,7
4	1,1	1,1	1,0	0,9	1,0	1,3	2,4	4,6	6,6	6,0	3,6
5	1,8	1,7	1,7	1,6	1,5	1,4	2,1	3,7	4,9	4,0	4,3
10	1,2	1,1	0,9	0,6	0,7	1,1	1,3	1,3	5,0	2,0	3,9
15	1,5	1,3	1,2	0,5	0,8	0,7	1,2	1,3	2,1	2,5	2,6
20	0,9	0,8	0,8	0,8	0,5	1,0	0,9	1,1	3,0	1,9	4,2
25	1,0	1,0	0,9	0,5	0,4	0,5	0,6	1,1	2,0	1,3	2,6
30	1,0	0,9	0,9	0,8	0,7	0,5	0,9	1,0	2,4	1,5	2,8

Fuente: Elaboración propia

Tabla K-3: Razón intervalo de confianza y error relativo medio para promedios simples (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	2,5	2,7	2,6	3,0	3,7	4,0	4,2	4,1	7,3	7,6	12,8
2	1,7	1,7	1,9	1,6	1,4	2,1	3,3	3,7	8,2	6,7	12,5
3	1,4	1,2	1,5	2,4	2,8	2,9	3,3	3,7	5,8	6,6	10,7
4	1,0	1,1	1,2	1,6	2,1	1,7	2,6	4,4	5,2	4,8	7,9
5	1,6	1,5	1,4	1,1	1,2	1,5	2,0	2,9	4,3	4,6	5,8
10	0,9	0,9	0,8	0,9	1,1	1,0	1,0	1,0	1,5	3,5	3,4
15	1,4	1,2	1,2	0,5	0,7	0,8	1,0	1,9	2,6	2,5	2,7
20	0,7	0,6	0,7	0,9	0,9	1,1	1,0	1,5	2,1	2,9	3,0
25	1,1	1,1	0,9	0,6	0,5	0,8	0,7	0,9	1,0	1,7	1,5
30	1,0	0,9	0,9	0,8	0,7	0,8	0,9	1,0	1,6	1,2	2,1

Fuente: Elaboración propia

Tabla K-4: Razón intervalo de confianza y error relativo medio para definiciones generalizadas (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	2,7	2,8	2,9	3,5	4,1	3,8	3,6	5,2	11,8	23,8	105,1
2	1,8	1,9	2,0	1,9	2,3	1,9	3,6	2,7	5,5	6,0	72,2
3	1,3	1,2	1,3	2,3	2,7	3,2	2,1	3,0	4,1	7,2	41,4
4	1,1	1,2	1,4	1,5	1,7	2,0	2,1	3,1	2,6	5,4	20,5
5	1,9	1,9	1,9	1,5	1,7	2,0	2,1	2,0	5,3	3,7	21,9
10	0,9	0,9	0,7	1,0	1,2	1,4	0,9	1,9	2,8	3,7	6,7
15	1,4	1,4	1,3	0,5	0,7	1,1	1,2	1,3	1,8	1,5	4,6
20	0,8	0,7	0,7	1,0	0,9	1,1	0,8	1,1	1,4	1,3	4,1
25	1,0	1,0	1,0	0,7	1,0	0,6	0,7	1,1	1,8	1,4	4,8
30	0,8	0,8	0,8	0,6	0,7	0,5	0,7	0,9	1,1	1,8	5,7

Fuente: Elaboración propia

Tabla K-5: Error relativo medio de Promedios Simples Históricos (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	17	17	18	18	20	22	27	39	58	72	79
2	14	14	14	15	17	18	21	27	41	54	67
3	13	13	13	14	15	16	19	23	35	48	60
4	12	12	13	14	15	16	18	21	31	48	57
5	11	11	12	13	14	15	17	20	30	43	55
10	9	9	9	11	13	14	15	17	24	36	46
15	7	7	8	10	12	13	14	16	21	33	43
20	6	6	7	9	11	12	14	15	21	33	42
25	5	5	6	8	11	12	13	15	20	32	41
30	4	5	5	7	10	11	13	15	20	32	39

Fuente: Elaboración propia

Tabla K-6: Error relativo medio de Promedios Simples Interpolados (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	27	27	28	42	59	70	84	93	97	100	100
2	20	20	20	28	43	52	70	84	92	96	98
3	16	16	17	22	33	41	60	76	88	93	96
4	14	14	15	20	27	35	51	70	84	90	93
5	12	13	13	18	24	30	46	65	80	87	91
10	9	9	10	13	18	21	30	46	64	75	80
15	7	7	8	11	15	18	24	35	53	64	71
20	6	6	7	9	13	16	21	30	45	56	65
25	5	5	6	8	12	14	19	27	40	51	59
30	4	5	5	8	11	13	18	24	35	45	54

Fuente: Elaboración propia

Tabla K-7: Error relativo medio de FBH-v (%) [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	26	27	27	29	33	38	54	79	92	95	96
2	16	16	17	18	20	23	31	52	80	90	93
3	11	12	12	13	15	16	22	37	67	83	90
4	10	10	11	12	13	15	19	30	53	74	85
5	9	9	10	10	12	13	17	25	46	69	82
10	7	7	7	8	9	11	13	18	32	51	68
15	6	6	6	7	8	9	12	16	27	45	61
20	5	5	5	6	7	8	11	15	25	42	59
25	4	5	5	6	7	8	10	14	24	40	56
30	4	4	4	5	7	7	10	14	23	40	55

Fuente: Elaboración propia

Tabla K-8: Cociente de elasticidades para promedios simples históricos [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	-	2,7	2,0	1,7	1,5	1,3	1,0	0,8	0,7	0,6	0,6
2	-	2,3	1,7	1,5	1,3	1,2	0,9	0,8	0,6	0,6	0,5
3	-	2,1	1,6	1,4	1,3	1,1	0,9	0,7	0,6	0,5	0,5
4	-	2,0	1,5	1,3	1,2	1,1	0,9	0,7	0,6	0,5	0,5
5	-	1,9	1,5	1,3	1,2	1,0	0,9	0,7	0,6	0,5	0,5
10	-	1,7	1,3	1,2	1,1	1,0	0,8	0,7	0,5	0,5	0,4
15	-	1,6	1,3	1,1	1,0	0,9	0,8	0,6	0,5	0,5	0,4
20	-	1,5	1,2	1,1	1,0	0,9	0,7	0,6	0,5	0,5	0,4
25	-	1,5	1,2	1,1	1,0	0,9	0,7	0,6	0,5	0,4	0,4
30	-	1,4	1,2	1,0	1,0	0,9	0,7	0,6	0,5	0,4	0,4

Fuente: Elaboración propia

Tabla K-9: Cociente de elasticidades para FBH-v [AIMSUN]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	-	27,7	5,1	3,3	2,6	2,0	1,4	1,0	0,8	0,7	0,6
2	-	10,0	3,6	2,6	2,1	1,6	1,2	0,9	0,7	0,6	0,5
3	-	7,1	3,0	2,2	1,8	1,5	1,0	0,8	0,6	0,5	0,5
4	-	5,8	2,7	2,0	1,7	1,3	1,0	0,7	0,6	0,5	0,4
5	-	5,0	2,5	1,9	1,6	1,3	0,9	0,7	0,5	0,5	0,4
10	-	3,5	1,9	1,5	1,3	1,0	0,8	0,6	0,4	0,4	0,3
15	-	2,9	1,7	1,3	1,1	0,9	0,7	0,5	0,4	0,3	0,3
20	-	2,5	1,5	1,2	1,0	0,8	0,6	0,5	0,4	0,3	0,3
25	-	2,3	1,4	1,1	1,0	0,8	0,6	0,4	0,3	0,3	0,2
30	-	2,1	1,3	1,1	0,9	0,8	0,6	0,4	0,3	0,3	0,2

Fuente: Elaboración propia

Tabla K-10: Cociente de elasticidades para promedios simples interpolados [NGSIM]

P (%) \ τ (seg/(obs*veh))	1	3	5	10	20	30	60	120	240	360	480
1	3,5	1,8	1,4	1,3	1,1	1,0	0,8	0,6	0,5	0,4	0,4
2	3,1	1,8	1,4	1,3	1,2	1,1	0,9	0,7	0,6	0,5	0,5
3	2,9	1,8	1,5	1,3	1,2	1,1	0,9	0,7	0,6	0,5	0,5
4	2,8	1,7	1,5	1,3	1,2	1,1	0,9	0,8	0,6	0,6	0,5
5	2,8	1,7	1,5	1,3	1,2	1,1	0,9	0,8	0,7	0,6	0,5
10	2,6	1,7	1,5	1,3	1,2	1,1	1,0	0,8	0,7	0,6	0,6
15	2,5	1,7	1,5	1,3	1,3	1,1	1,0	0,9	0,7	0,7	0,6
20	2,4	1,7	1,5	1,3	1,3	1,2	1,0	0,9	0,8	0,7	0,7
25	2,4	1,7	1,5	1,3	1,3	1,2	1,0	0,9	0,8	0,7	0,7
30	2,4	1,7	1,5	1,3	1,3	1,2	1,0	0,9	0,8	0,7	0,7

Fuente: Elaboración propia