

Teoría de Flujos Vehiculares
Apuntes de Clase

Profesor Jaime Gibson

Universidad de Chile
División Ingeniería de Transporte

Edición Mauro Huenupi

Versión Agosto de 2001

Índice General

1	Introducción	2
1.1	Presentación del curso	2
1.2	Conflictos y regulación	3
1.3	Modelos elementales	5
1.3.1	Circulación ininterrumpida	5
1.3.1.1	Observación temporal	6
1.3.1.2	Observación espacial	7
1.3.1.3	Distribución de velocidades	8
1.3.1.4	Ecuación fundamental del tráfico	8
1.3.1.5	Teoría del Seguimiento Vehicular	12
1.3.1.6	Inestabilidad del tráfico	15
1.3.1.7	Relaciones velocidad-flujo o flujo-demora	16
1.3.2	Circulación interrumpida	19
1.4	Variables básicas	20

2	Capacidad	22
2.1	Definiciones	22
2.2	Capacidad de intersecciones no semaforizadas	23
2.2.1	Modelo de aceptación de brechas	24
2.2.2	Modelo lineal	28
2.2.3	Modelo lineal aplicado a rotondas	31
2.3	Capacidad de intersecciones semaforizadas	32
2.4	Heterogeneidad del tráfico	36
2.4.1	Heterogeneidad en el flujo de saturación	37
2.4.1.1	Factores de equivalencia para determinar el flujo de saturación	38
2.4.1.2	Duración del verde efectivo	42
2.4.2	Heterogeneidad en intersecciones no semaforizadas . . .	43
2.4.2.1	Tipos de vehículos	43
2.4.2.2	Tipo de movimiento	43
2.4.2.3	Elección de pista	44
3	Formación y disipación de colas	46
3.1	Definición del proceso	46
3.2	Colas estacionarias	48
3.2.1	Sistemas de solución conocida	50
3.3	Colas dependientes del tiempo	51

3.3.1	Transformada de Whiting	53
3.3.2	No transitividad de la función transformada	55
3.3.3	Funciones transformadas para intersecciones semaforizadas	59
3.3.3.1	Funciones estocásticas estacionarias	62
3.3.3.2	Funciones determinísticas dependientes del tiempo	65
3.3.3.3	Funciones transformadas	69
4	Estimación de demoras y detenciones	74
4.1	Periodización y demoras en períodos punta	74
4.1.1	Modelo de flujo variable	75
4.2	Demoras peatonales	81
4.2.1	Intersecciones semaforizadas	81
4.2.2	Intersecciones no semaforizadas	82
4.3	Demoras en paraderos de buses	83
4.4	Demoras totales	84
4.4.1	Intersecciones no semaforizadas	88
4.4.2	Intersecciones semaforizadas	90
4.5	Detenciones	90
4.5.1	Intersecciones semaforizadas	91
4.5.2	Intersecciones no semaforizadas	93

5 Modelos de circulación en redes	94
5.1 Introducción	94
5.2 Modelación de elementos de redes	96
5.2.1 Modelación de arcos	96
5.2.2 Modelación de nodos	96
5.2.3 Modelación de cuellos de botella	96
5.2.4 Modelación de flujos	97
5.2.5 Extensión de la red	97
5.3 Histogramas cíclicos de flujo y dispersión	97
5.4 Indicadores de rendimiento	102

Capítulo 1

Introducción

1.1 Presentación del curso

El objetivo de este curso es analizar el proceso de la circulación de vehículos sobre vías urbanas como fenómeno físico, manifestación de una estructura o patrón de viajes conocido y fijo. Se excluirán por lo tanto aspectos relativos a la generación, distribución, partición modal y asignación de los viajes que realizan los individuos.

Una vez identificado el patrón de viajes, estudiaremos cómo éste se materializa en las vías, qué fenómenos tienen lugar durante su manifestación y cómo éstos pueden modelarse para predecirlos y estimar sus impactos.

Las motivaciones para utilizar esta forma de abordar el problema se pueden resumir en:

- En el proceso de circulación surgen interacciones entre los vehículos que condicionan la calidad de los viajes; un índice tradicional de calidad es el tiempo de viaje. Sabemos que esta calidad influirá en las decisiones de movilidad de los individuos, como por ejemplo en la elección de modo de viaje. Son estas decisiones las que finalmente producen un patrón de viajes. Por lo tanto, es indispensable entender la circulación para comprender el fenómeno del transporte en su conjunto.

- Es posible actuar sobre el proceso de circulación para producir cambios en las características de los viajes de manera de optimizar el uso de recursos que estos involucran. Los recursos son en parte recursos privados, como el tiempo de los individuos, o públicos como las vías. Esta última característica hace de éste un problema relevante ya que da origen a impactos negativos o externalidades de circulación, como la congestión vial. El conjunto de acciones que buscan optimizar el uso de los recursos corresponde al campo específico de la ingeniería de transporte, denominado gestión de tránsito, y es materia del curso Ingeniería de Tránsito.

En resumen, en este curso nos concentraremos en analizar y luego modelar los fenómenos que se producen durante la circulación de vehículos, principalmente sobre vías urbanas, con el propósito de predecir sus impactos y diseñar mecanismos para optimizar el uso de los recursos involucrados.

1.2 Conflictos y regulación

Desde el punto de vista de la circulación de vehículos un viaje es una sucesión de etapas de movimiento y reposo. Para un vehículo aislado esto podría ser descrito por las leyes de la mecánica. Pero cuando múltiples y distintos vehículos participan en el proceso surgen interacciones de otra naturaleza que condicionan la circulación y que requieren una teoría especializada.

Estas interacciones tienen lugar en los elementos de la infraestructura vial: vías, intersecciones, estaciones de transferencia. Derivan de un hecho físico básico: tanto los vehículos como la infraestructura tienen dimensiones finitas. De allí los conflictos: la presencia de un vehículo condiciona la circulación de los otros.

Son estos conflictos entre vehículos los que deben ser analizados para describir la circulación. No obstante, los vehículos no son entes autónomos sino que son comandados por personas.

Luego, los aparentes conflictos entre vehículos son en realidad entre personas. De allí que la descripción de la circulación no es sólo una extensión de las

leyes de la mecánica. Aspectos como la conducta y características físicas de los individuos juegan un rol fundamental.

A continuación estudiaremos los conflictos, clasificándolos en tres tipos para su estudio más detallado: concurrentiales, direccionales y funcionales.

Conflictos concurrentiales: Se producen entre los vehículos que circulan por la misma vía y en el mismo sentido, pero con distintas velocidades (p. ej. tramos de calles, caminos, etc.). La superación de este tipo de conflicto se logra ajustando la velocidad de los vehículos o permitiendo el adelantamiento del vehículo más lento.

Conflictos direccionales: Se dan entre vehículos cuyas trayectorias se cruzan en una misma vía o en vías distintas. Su solución se obtiene dando prioridad a uno de los movimientos sobre el otro. Se producen típicamente en intersecciones.

Conflictos funcionales: Ocurren entre vehículos que hacen uso antagónico de un mismo sector de la vía: movimiento y reposo. Se superan a través del adelantamiento del vehículo detenido, la detención del vehículo en movimiento o el ajuste de su velocidad hasta que el conflicto desaparece (por ejemplo, paraderos de buses y estacionamientos en la calzada).

Los conflictos en la circulación difícilmente pueden ser eliminados del todo, lo que da origen a eventuales accidentes. Entonces, si el conflicto no puede eliminarse debe regularse. La regulación debe reunir dos características esenciales:

- eficiencia: permitir un adecuado desarrollo o solución del conflicto.
- seguridad: prevenir las ocurrencia de accidentes.

Sin embargo, la regulación no se impone sobre los vehículos. Es, en realidad, una regulación de la conducta de los usuarios, donde el actor principal es el conductor. Es aquí donde la ingeniería juega su rol a través de su actividad por excelencia: el diseño (entendido como concepción de soluciones) de vías, intersecciones, estaciones de transferencia y sistemas de control o regulación.

Los conflictos más simples de estudiar son los concurrentes. Para su descripción basta considerar un proceso de circulación ininterrumpida. Hablaremos de circulación ininterrumpida cuando analicemos una sola vía, de un solo sentido y de longitud ilimitada (tramo de carretera).

El siguiente caso corresponde a los conflictos direccionales. En esta situación un vehículo podrá usar el área común primero y otro después, requiriéndose la detención de este último. Su estudio requiere observar un proceso de circulación interrumpida. Hablaremos de circulación interrumpida cuando analicemos más de una vía, diferentes sentidos, limitado a un área (intersección). Este mismo análisis es factible para estudiar los conflictos funcionales, como los generados en los paraderos de buses.

En este curso revisaremos primero la teoría relativa a la circulación ininterrumpida, por simplicidad y tradición histórica, para luego concentrarnos en la modelación de la circulación interrumpida, como ampliación de la primera y por su relevancia en los procesos de circulación en redes viales densas como las urbanas.

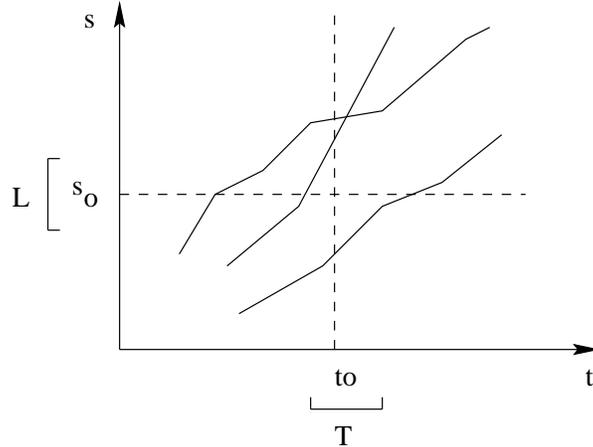
1.3 Modelos elementales

1.3.1 Circulación ininterrumpida

El desplazamiento de vehículos se verifica en un ambiente físico que comprende tanto dimensiones espaciales como temporales en forma simultánea. No obstante, a diferencia del transporte marítimo o aéreo donde la dimensión espacial posee dos o tres componentes, el tráfico urbano tiene una componente espacial que puede considerarse cuasi-unidimensional: la vía.

Por otra parte, en la circulación vehicular los objetos en movimiento no sólo obedecen a leyes físicas sino que son conducidos y ocupados por personas. Luego, están dotados de una dimensión humana: la voluntad. Así, la circulación vehicular es un problema complejo, pero a la vez muy interesante de analizar.

Figura 1.1: Diagrama Espacio-Tiempo



En este caso la circulación se representa a través de un diagrama espacio-tiempo, como el de la Figura 1.1. En este diagrama se muestra la posición en cada instante de los vehículos.

El diagrama espacio-tiempo se puede analizar desde dos puntos de vista: temporal y espacial.

1.3.1.1 Observación temporal

La recta horizontal de la Fig. 1.1 representa la observación temporal de la circulación en una sección transversal de la vía, determina los vehículos que pasan por la sección s_o durante un período de tiempo T , el cual denominamos m . Esta visión da origen a las siguientes variables :

Flujo o intensidad del tráfico: es el número de vehículos que atraviesan una determinada sección de la vía por unidad de tiempo: $q = \frac{m}{T}$. Esta variable corresponde al flujo promedio en el punto s_o durante el período T . Por su definición el flujo describe macroscópicamente la circulación.

Intervalo: tiempo que transcurre entre el paso de dos vehículos sucesivos por una sección de la vía. Se suele asignar al primero de estos vehículos y se le denomina h_i . El intervalo bruto se mide entre las partes homólogas

(delantera o trasera) de los vehículos. El intervalo neto o brecha se mide entre la parte trasera de un vehículo y la delantera del siguiente. El intervalo describe microscópicamente la circulación. Considerando un período de observación suficientemente largo y definiendo adecuadamente el comienzo y el final de dicho período, se tiene $T = \sum_{i=1}^m h_i$ y, por lo tanto, $q = \frac{m}{T} = \frac{m}{\sum_{i=1}^m h_i} = \frac{1}{\bar{h}}$. Luego

$$\bar{h} = \frac{1}{q} \quad (1.1)$$

Es decir, el intervalo medio es el inverso del flujo. Por ejemplo, un flujo de 1.800 veh/h equivale a un intervalo promedio de 2 seg.

1.3.1.2 Observación espacial

La recta vertical de la Fig. 1.1 representa la observación espacial de la circulación desde un punto de vista elevado sobre la vía. Determina el número de vehículos presentes en un tramo de la vía de longitud L en un cierto instante t_o , variable que denominamos n . Se tienen las siguientes variables:

Concentración o densidad de tráfico: número de vehículos que están presentes en un tramo de vía en un cierto instante de tiempo $k = n/L$. La concentración describe a nivel macroscópico la circulación.

Espaciamiento: distancia entre dos vehículos sucesivos circulando por una misma pista en un cierto instante. Generalmente el espaciamiento es bruto. Se asocia al primer vehículo y se denomina s_j . El espaciamiento es una variable que describe microscópicamente la circulación. Análogamente, el espaciamiento medio corresponde al inverso de la concentración ya que $k = \frac{n}{L} = \frac{n}{\sum_{j=1}^n s_j} = \frac{1}{\bar{s}}$. Luego,

$$\bar{s} = \frac{1}{k} \quad (1.2)$$

1.3.1.3 Distribución de velocidades

Además de las dos visiones recién mencionadas, que son estáticas en el tiempo o el espacio, podemos definir una no-estática. Esta visión es la que lograría un observador en movimiento. En este caso se tienen dos grupos de vehículos que pueden ser analizados: aquellos que pasan por un punto o los que están presentes en un tramo. Con estas agrupaciones es posible definir dos funciones de distribución para cualquier variable que se quiera estudiar. Estas funciones se conocen como distribución espacial y temporal.

Distribución espacial: La función de distribución espacial se obtiene al considerar la proporción de vehículos de tipo j presentes en un tramo de longitud L en un cierto instante: $p_{sj} = n_j/n$. Utilizando la variable concentración podemos calcular la distribución espacial como:

$$p_{sj} = \frac{k_j L}{kL} = \frac{k_j}{k} \quad (1.3)$$

Así entonces el promedio espacial de la variable z se calcula como $\bar{z}_s = \sum_j p_{sj} z_j = \sum_j k_j z_j / k$.

Distribución temporal: La función de distribución temporal se obtiene al considerar la proporción de vehículos de tipo j que cruzan una sección de la vía durante un período de duración T : $p_{tj} = m_j/m$. Utilizando la variable flujo podemos calcular la función de distribución temporal como:

$$p_{tj} = \frac{q_j T}{qT} = \frac{q_j}{q} \quad (1.4)$$

Por lo tanto, el promedio temporal de la variable z se calcula como $\bar{z}_t = \sum_j p_{tj} z_j = \sum_j q_j z_j / q$.

1.3.1.4 Ecuación fundamental del tráfico

Algunas de las variables recién definidas están relacionadas. La relación más importante involucra el flujo, la concentración y la velocidad media espacial.

Si suponemos una corriente vehicular j homogénea y estacionaria (igual velocidad y constante en el tiempo) en condiciones de circulación ininterrumpida entonces geoméricamente se tiene

$$s_j = v_j h_j \quad (1.5)$$

Si el flujo y la concentración de la corriente se miden sobre un período que tengan muchos vehículos, la ecuación anterior se puede reescribir como

$$\frac{1}{k_j} = \frac{v_j}{q_j} \Leftrightarrow k_j = \frac{q_j}{v_j}$$

Pero sabemos que

$$\bar{v}_s = \frac{\sum_j k_j v_j}{\sum_j k_j} = \frac{\sum_j \frac{q_j}{v_j} v_j}{\sum_j k_j} = \frac{\sum_j q_j}{\sum_j k_j}$$

En consecuencia,

$$q = k \bar{v}_s \quad (1.6)$$

De la ecuación anterior se tiene $\frac{\partial q}{\partial k} = \bar{v}_s \geq 0$, es decir, el flujo aumenta a medida que aumenta la concentración, debido a que los vehículos se desplazan más juntos. Además $\frac{\partial q}{\partial \bar{v}_s} = k \geq 0$, es decir, el flujo es más alto a medida que la velocidad aumenta, porque los vehículos cruzan más rápido la sección de la vía bajo análisis. Sin embargo, es evidente que existe una relación entre velocidad y concentración ya que el aumento de la concentración implica una mayor probabilidad de conflictos concurrenciales, lo que obligará a los conductores a desplazarse con mayor precaución para evitar accidentes.

Lo anterior se traduce en que $\bar{v}_s = \bar{v}_s(k)$ con $\frac{\partial \bar{v}_s}{\partial k} \leq 0 \forall k$. Reemplazando esta relación en la ecuación fundamental se tiene

$$q = k \bar{v}_s(k) \quad (1.7)$$

Ecuación que se conoce como "Ecuación fundamental del tráfico".

A partir de ésta se tiene

$$\frac{\partial q}{\partial k} = \underbrace{\bar{v}_s(k)}_{\geq 0} + k \underbrace{\frac{\partial \bar{v}_s(k)}{\partial k}}_{\leq 0}$$

$$\frac{\partial q}{\partial k} \leq \geq 0$$

Es decir, bajo ciertas condiciones el flujo aumentará con la concentración, pero bajo otras podría disminuir.

El punto de quiebre se produce cuando $\frac{\partial q}{\partial k} = 0 \Leftrightarrow -\frac{\partial \bar{v}_s}{\bar{v}_s} = \frac{\partial k}{k}$ que es una ecuación en k .

En la solución de la ecuación anterior es necesario considerar que la concentración está acotada inferior y superiormente. Su valor mínimo es cero, es decir, ningún vehículo en el tramo de vía bajo análisis. Su valor máximo es igual al número máximo de vehículos que se pueden almacenar en el tramo ubicados uno detrás del otro y se denomina concentración de embotellamiento k_e . Si conocemos el largo promedio de los vehículos \bar{l} entonces $k_e = 1/\bar{l}$. Los límites de la concentración definen valores extremos de la velocidad media espacial.

Cuando la concentración es cercana a cero existe una baja probabilidad de conflictos concurrentes y, por lo tanto, la velocidad puede alcanzar su máximo valor, el cual denominamos velocidad a flujo libre v_l . Esta es la velocidad a la cual circularía un conductor si no existiese ningún otro vehículo presente en el tramo de vía.

Por el contrario, cuando la concentración tiende a k_e los conflictos son frecuentes y la velocidad disminuirá hasta llegar a cero en el límite.

Existen muchas funciones que podrían representar la relación entre velocidad media espacial y concentración recién planteada. Una alternativa sencilla es una función lineal. Aplicando las condiciones de borde y obviando los subíndices se tiene

$$v = v_l \left(1 - \frac{k}{k_e}\right) \quad \forall k \in [0..k_e] \quad (1.8)$$

Reemplazando esta relación en la ecuación fundamental del tráfico, obtenemos

$$q = v_l k \left(1 - \frac{k}{k_e}\right) \quad (1.9)$$

Esta ecuación representa una parábola que tiene un máximo en el punto $k^* = \frac{k_e}{2}$ y cuyo valor es $Q = \frac{k_e v_l}{4}$. Antes de k^* el flujo crece con q , lo contrario sucede después de k^* .

El valor máximo del flujo Q se denomina capacidad de la sección de vía y representa el máximo número de vehículos que puede cruzar la sección bajo todos los supuestos antes mencionados. Vemos entonces que la capacidad depende de:

- Concentración de embotellamiento: función de la longitud promedio de los vehículos.
- Velocidad a flujo libre: función de características de los vehículos y la vía y del comportamiento de los conductores.

La existencia de la capacidad permite definir dos regímenes para la circulación. Para esto se define el grado de saturación x como el cociente entre el flujo y la capacidad, $x = q/Q$. Luego, se tiene

- régimen no saturado: $x < 1$
- régimen saturado: $x \geq 1$

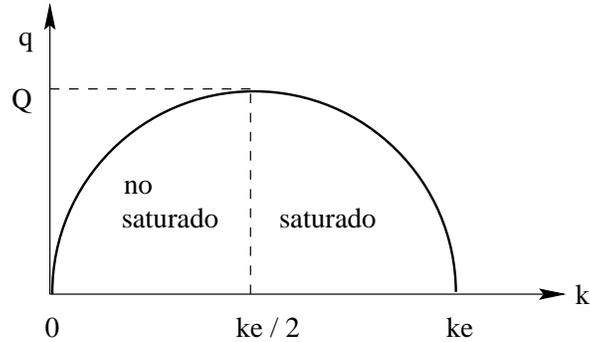
La Figura 1.2 muestra los dos regímenes, suponiendo una relación lineal entre la velocidad espacial y la concentración.

Nótese que en la definición del grado de saturación q es el flujo que desea cruzar por la sección demanda que puede ser distinto del flujo que efectivamente cruza la sección, que es el que se aprecia en la Fig. 1.2 y que siempre es menor que Q . Esta diferencia en el caso $x \geq 1$ se traduce en la formación de colas.

El grado de saturación es un concepto fundamental en este curso ya que, como veremos más adelante, de él depende la magnitud de los impactos de la circulación vehicular.

El análisis anterior se basa en una elección arbitraria de la relación entre velocidad media espacial y concentración (lineal). Es posible obtener expresiones más realistas y mejor justificadas de esta relación haciendo uso de la Teoría del Seguimiento Vehicular.

Figura 1.2: Regímenes de circulación



1.3.1.5 Teoría del Seguimiento Vehicular

La Teoría del Seguimiento Vehicular es un enfoque microscópico para modelar el comportamiento de conductores que circulan por un tramo de vía. Esta teoría se basa en la ley psicológica de Webster que establece lo siguiente: **la reacción de un individuo frente a un estímulo es directamente proporcional a su sensibilidad y a la magnitud del estímulo.**

Lo importante para nosotros es identificar la respuesta, la sensibilidad y el estímulo en el caso de la circulación ininterrumpida. Las condiciones bajo las cuales se hizo esto son las siguientes:

- una sola vía
- un solo sentido de circulación
- dos vehículos sucesivos
- no hay posibilidad de adelantamiento, corriente confinada.

Tenemos entonces

$$\text{Respuesta} = \text{Sensibilidad} * \text{Estímulo}$$

Sea $x_n(t)$ la posición del vehículo n en el instante t . Se define:

Respuesta: aceleración del vehículo n en el instante $t + T$, $\frac{dv_n(t+T)}{dt}$

La variable T corresponde al tiempo de respuesta y se origina en el tiempo necesario para realizar las siguientes acciones: Percepción del estímulo, interpretación del estímulo, determinación de la acción necesaria, ejecución de la acción necesaria y transmisión mecánica de la acción.

Estímulo: velocidad relativa entre el vehículo n y el $n - 1$ en el instante t

$$v_{n-1}(t) - v_n(t) = \frac{d[x_{n-1}(t) - x_n(t)]}{dt} = \frac{ds_n(t)}{dt} \quad (1.10)$$

Sensibilidad: función de la velocidad y el espaciamiento en el instante t .

En general, la sensibilidad es proporcional a la velocidad e inversamente proporcional al espaciamiento. Esto se debe a que mientras más alta o más baja sea la velocidad, mayor o menor será la atención que debe prestar el conductor a las condiciones de circulación. Por otro lado, mientras mayor sea el espaciamiento, menor será la probabilidad de conflictos y así el individuo podrá prestar menos atención a la conducción; lo contrario sucede cuando el espaciamiento es pequeño. Una alternativa de modelación es considerar

$$\lambda = \lambda_o \frac{v_n^m}{s_n^l} \quad (1.11)$$

donde λ_o , m y l son parámetros positivos.

Por lo tanto, el modelo general de seguimiento vehicular queda representado por la siguiente ecuación

$$\frac{dv_n(t+T)}{dt} = \lambda_o \frac{v_n^m}{s_n^l} \frac{ds_n(t)}{dt} \quad (1.12)$$

Se han propuesto diferentes valores para λ_o , m y l , lo que implica distintos modelos para la sensibilidad y desde luego un modelo distinto para la circulación. Por sus características estos modelos los denominamos modelos

microscópicos de circulación ininterrumpida. Ejemplo: Sensibilidad constante. Se obtiene suponiendo $m = l = 0$. Tenemos entonces $S = \lambda_o$ y por lo tanto

$$\frac{dv_n(t+T)}{dt} = \lambda_o \frac{ds_n(t)}{dt}$$

Si despreciamos T e integramos la ecuación anterior se obtiene $v_n = \lambda_o s_n + C$. Aplicando la condición de borde $v_n \rightarrow 0$ si $s_n \rightarrow s_{min}$, resulta $v_n = \lambda_o(s_n - s_{min})$. Agregando esta relación sobre un grupo de vehículos:

$$\bar{v} = \lambda_o(\bar{s} - \bar{s}_{min})$$

o bien

$$\bar{v} = \lambda_o \left(\frac{1}{k} - \frac{1}{k_e} \right) \quad (1.13)$$

Esta ecuación se conoce como modelo de *Herman*. Vemos que cuando la concentración tiende a cero la ecuación anterior predice una velocidad infinitamente grande. Obviamente, esto es incorrecto y nos advierte que el modelo de Herman (sensibilidad constante) no es adecuado para describir la velocidad cuando la concentración es baja.

Utilizando la ecuación fundamental del tráfico obtenemos un modelo macroscópico de la circulación ininterrumpida

$$q = \lambda_o \left(1 - \frac{k}{k_e} \right) \quad (1.14)$$

En consecuencia, el flujo disminuye linealmente desde un máximo igual a λ_o cuando la concentración es cero, hasta un mínimo de cero cuando se alcanza la concentración de embotellamiento. El hecho de que el flujo no sea cero cuando la concentración es cero es otro inconveniente del modelo de Herman. Recuérdese que la concentración es el número de vehículos presentes en el tramo de vía.

Otros modelos para la sensibilidad se muestran en la Tabla 1.1.

Tabla 1.1: Modelos para la sensibilidad

Autor	Sensibilidad
Herman	λ_o
Greenberg	λ_o/s
Greenshields	λ_o/s^2
Underwood	$\lambda_o v/s^2$

1.3.1.6 Inestabilidad del tráfico

La idea es estudiar el efecto que tiene sobre una corriente vehicular, bajo condiciones de circulación ininterrumpida, un cambio repentino en la velocidad de un determinado vehículo. Si consideramos este cambio como un estímulo para el conductor del vehículo siguiente, la reacción o respuesta de éste como un estímulo para el próximo conductor, y así sucesivamente; podemos distinguir dos situaciones:

- Si la respuesta de los conductores se amortigua, es decir, disminuye hasta desaparecer aguas arriba de la corriente entonces hablaremos de estabilidad asintótica en la circulación.
- Si la respuesta de los conductores se amplifica aguas arriba de la corriente entonces hablaremos de inestabilidad asintótica en la circulación.

Cuando en la circulación existe inestabilidad asintótica hay una mayor probabilidad de accidentes porque las respuestas de los conductores frente a los cambios de velocidad se vuelven cada vez violentas, por ejemplo expresadas en frenajes más bruscos. Por el contrario frente a condiciones de inestabilidad local, los cambios en la velocidad son adecuadamente enfrentados por los conductores, lo que provoca una disminución de la probabilidad de accidentes.

Diversos autores han estudiado este fenómeno y concluyen que si se cumple la siguiente condición la circulación presenta estabilidad asintótica:

$$\lambda T \leq \frac{1}{2} \quad (1.15)$$

donde λ es la sensibilidad y T el tiempo de respuesta.

A modo de ejemplo, utilizando la especificación de la sensibilidad del modelo de Greenshield se obtiene la condición $\frac{\lambda_o}{s^2}T \leq \frac{1}{2}$. De aquí podemos encontrar el valor crítico de la concentración bajo el cual hay estabilidad:

$$k_c = \sqrt{\frac{1}{2\lambda_o T}}$$

se puede demostrar que para este caso $\lambda_o = \frac{v_l}{k_e}$, luego

$$k_c = \sqrt{\frac{k_e}{2v_l T}}$$

Este valor no es absoluto, no sólo depende de características físicas de la vía y los vehículos sino que también de los conductores, principalmente a través de T .

A partir de valores típicos para los parámetros de k_c se puede mostrar que $k_c \leq k^*$. Luego, dentro del régimen no saturado habrá una zona, previa a la capacidad, en la que las interacciones entre vehículos comienzan a producir inestabilidades que se propagan a toda la corriente y vuelven insegura la circulación.

Nuestro interés final reside en las trayectorias espacio-tiempo. Ellas dependen de la velocidad con que se recorre un tramo y ahora sabemos que ésta depende de los conflictos concurrentes. Esto se traduce en una relación $v - k$. Sin embargo, no suele disponerse de la concentración como dato. Aprovechando la ecuación fundamental del tráfico podemos convertir una relación $v - k$ en una $v - q$. El flujo sí es una variable fácil de obtener.

1.3.1.7 Relaciones velocidad-flujo o flujo-demora

Un concepto tradicional en teoría de tráfico es que la velocidad o su inverso, la demora, a través de un dispositivo vial es función del flujo circulante. Este concepto fue originalmente desarrollado para caminos rurales, túneles o autopistas para el estudio de la congestión. Posteriormente, los métodos de

asignación de tráfico utilizaron relaciones que ligaban los atributos de un arco o ruta con los tiempos de viaje, demoras o costos resultantes de de viajar por ellos. Esto se puede escribir como

$$t^k = t^k(\{q\}) \quad (1.16)$$

donde t^k es el tiempo medio de viaje en el arco k expresado en unidades de tiempo por unidad de distancia y $\{q\}$ el vector de flujos en toda la red expresado en vehículos por unidad de tiempo. La relación anterior puede ser simplificada si se consideran arcos suficientemente largos, es decir, donde la mayor parte del tiempo de viaje ocurrirá a lo largo del arco, más que en el nodo aguas abajo. En este caso el tiempo de viaje sólo dependerá del flujo en el arco, en tal caso hablaremos de una relación $t - q$ separable: $t^k = t^k(q^k)$.

Otra razón para desarrollar relaciones $v(q)$ o $t(q)$ es su aplicación en gestión de la demanda de tráfico.

Sea T el tiempo de viaje total para todos los usuarios del arco k : $T = q^k t^k(q^k)$. Luego, el tiempo marginal de viaje, es decir, la contribución al tiempo total que producirá un vehículo adicional circulando por el arco es:

$$t_{ma}^k = \frac{\partial T}{\partial q^k} = t^k(q^k) + \frac{\partial t^k}{\partial q^k} q^k \quad (1.17)$$

El primer término del lado derecho de la ecuación se interpreta como el tiempo medio de viaje en el arco, el segundo término es la contribución marginal del vehículo adicional. Este último corresponde a una externalidad, no observada por el que la produce quien sólo percibe el tiempo medio. Esta situación provoca un uso excesivo de las vías. La contribución marginal puede ser internalizada mediante la tarificación vial cobrando un peaje equivalente al valor de la externalidad para que se haga un uso eficiente de las vías.

Otro uso de las relaciones demora-flujo se puede apreciar en el siguiente ejercicio. Sea la siguiente relación $t(q)$:

$$t = t_o (1 + \alpha x^\beta) \quad (1.18)$$

Tabla 1.2: Relaciones $v(q)$ y $t(q)$

Autor	Relación
Estudio Detroit - 1962	$t = t_o e^x$
Bureau of Public Road (BPR) - 1964	$t = t_o (1 + \alpha x^\beta)$
Overgaard - 1967	$t = t_o \alpha^{\beta x}$
UK Department of Transport - 1985	$v = \begin{cases} v_o & q \leq q_1 \\ v_o - \frac{v_o - v_1}{Q - q_1} (q - q_1) & q_1 \leq q \leq Q \\ \frac{v_1}{1 + \frac{v_1}{8d} (\frac{q}{Q} - 1)} & q \geq Q \end{cases}$

donde x es el grado de saturación y t_o , α y β son parámetros. Sea P la demanda de viajes en el arco y k la tasa de ocupación de los vehículos. Por lo tanto, $q = P/k$ y $x = P/(kQ)$. Así,

$$t = t_o \left(1 + \alpha \left(\frac{P}{kQ} \right)^\beta \right) \quad (1.19)$$

Observando esta ecuación cabe un par de preguntas:

- ¿Qué pasa si P crece con k constante?
- ¿Qué pasa si P crece con k creciente?

En el primer caso q crece y se debería aumentar Q para mantener el tiempo de viaje en el valor original. Esto corresponde a la forma más tradicional de enfrentar un aumento del número de viajes, por ejemplo a través del aumento de la infraestructura de transporte. En el segundo, q no necesariamente crece y no sería obligatorio aumentar Q . Esto último es lo que persigue la gestión de la demanda de tráfico.

Las relaciones $v(q)$ o $t(q)$ utilizadas en la práctica, por lo general, descansan en el supuesto que se está modelando condiciones estacionarias y algún tipo de comportamiento promedio. Algunas de las relaciones más conocidas son que aparecen en la Tabla 1.2.

Las variables no definidas hasta ahora son: t_o tiempo medio de viaje a flujo libre, v_o velocidad a flujo libre, v_1 velocidad a capacidad, q_1 máximo flujo para el cual prevalece v_o y d la longitud del arco.

1.3.2 Circulación interrumpida

Estudiamos la circulación ininterrumpida para caracterizar cuantitativa y cualitativamente los conflictos concurrentiales. El análisis cuantitativo se hizo a través de la ecuación fundamental del tráfico, la cual nos indica la cantidad de vehículos (q) que pueden circular bajo diferentes niveles de ocupación de la vía (k). El análisis cualitativo se hizo mediante modelos que predicen la velocidad de circulación para distintos niveles de ocupación. Al unir ambos puntos de vista podemos a través de una sola ecuación modelar cuantitativa y cualitativamente la circulación ininterrumpida: $v(q)$.

Al estudiar la circulación interrumpida los conflictos relevantes son los direccionales: una misma área de la vía pretende ser usada por vehículos con diferentes direcciones. Por lo tanto, ahora lo relevante es la capacidad de esta área común. Así, un vehículo deberá pasar primero por el área común y el otro después, lo cual podría obligar al segundo a detenerse para darle paso al primero. Por eso se habla de circulación interrumpida.

La consecuencia inmediata es que el uso compartido de áreas comunes impone una restricción de capacidad tan fuerte que ésta es nula para los vehículos que se detienen para dar paso a otros. Esto es un nuevo concepto de capacidad. Si la circulación por el área común fuera ininterrumpida tendría una capacidad que depende de características de una sola corriente vehicular. Ahora el área ofrece capacidad para más de una corriente y, lógicamente, la capacidad para cada una será inferior a la de circulación ininterrumpida y dependerá de la forma de regular el uso del área común.

Luego, la aparición de conflictos direccionales no sólo genera variaciones en la velocidad sino que también detenciones, las que su vez se transforman en demoras: tiempo de espera para cruzar la intersección. Lo que se requiere para caracterizar la circulación interrumpida es modelar la capacidad del área común y después relacionarla con las demoras que experimenta cada corriente. No obstante lo anterior, siguen existiendo conflictos concurrentiales, pero su impacto es más limitado y se remite a restringir la velocidad de circulación en los tramos entre intersecciones.

Como ya se mencionó, la capacidad del área común es en general menor que su capacidad como tramo de vía. Por lo tanto, en redes viales densas, como

las urbanas, la capacidad global de la red estará determinada principalmente por la capacidad de las intersecciones y no por la capacidad de los tramos de vías. De allí la importancia de profundizar en el estudio de las intersecciones.

Debe notarse que en la circulación interrumpida pueden aparecer grados de saturación altos durante períodos de tiempo muy cortos, por lo cual una frontera entre régimen saturado y no saturado deja de ser un concepto sencillo ya que puede haber alternancia constante, lo que contradice la idea de régimen.

En la realidad urbana predominan las condiciones de circulación interrumpida como forma de interacción, pero lógicamente ésta incluye etapas de circulación ininterrumpida entre los lugares donde se producen conflictos direccionales o funcionales. Para abordar esta situación se parcelan las trayectorias $s - t$ en etapas de movimiento (relación $v - q$) y fuentes de demora caracterizados por relaciones flujo-capacidad y las demoras que se originan en ellas.

1.4 Variables básicas

Del análisis previo es claro que la circulación puede ser descrita por un número reducido de variables. Desde una perspectiva macroscópica basta con q , v y Q . Lo importante y complejo son las relaciones entre estas variables.

La complejidad no radica sólo en la dinámica de los conflictos que subyace a esas relaciones. También es relevante que estas variables tienen dos características especiales. Están sujetas a variabilidad temporal y contienen dimensiones de heterogeneidad. Esto último se refiere a que vehículos, intersecciones y vías son conceptos abstractos, pero sus manifestaciones reales son bastante diversas.

La variabilidad temporal es muy importante, tanto en la definición de las variables individuales como en el análisis de sus interrelaciones. Siempre hemos de pensar en estas variables como multidimensionales y dependientes del tiempo.

Para incorporar la variabilidad temporal debemos distinguir primero variación temporal aleatoria y sistemática. Por ejemplo, si se mide el flujo en

un mismo punto en diferentes momentos siempre se encontrará fluctuaciones. Estas variaciones dependen de tantos factores que se comportan como si fueran aleatorias. Sin embargo, se puede apreciar que existen ciertos patrones estables de variación que responden a tendencias bien definidas de la demanda, a éstas se refiere la variación sistemática. Este tipo de variación se manifiesta en la presencia de horas punta durante el día, diferencias según época del año, etc.

La variabilidad temporal se analizará distinguiendo períodos, bajo el supuesto de que *toda variación al interior de un período es de carácter aleatorio y entre períodos es sistemática*. Luego, durante el curso cuando se hable de flujo, velocidad, capacidad se estará hablando de su valor medio dentro de un período dado.

Para abordar la heterogeneidad se buscará homogeneizar mediante una unidad de referencia y factores de conversión o equivalencia de otras unidades a ella. La unidad y los factores dependen del fenómeno particular de que se trate. Por ejemplo, la unidad de referencia del flujo vehicular es el auto particular.

Capítulo 2

Capacidad

2.1 Definiciones

En el Capítulo 1 vimos que los conflictos en la circulación se producen porque tanto los vehículos como las vías tienen dimensiones finitas, a raíz de lo cual una porción del espacio vial sólo puede ser ocupada por un vehículo a la vez.

Son estos conflictos los que dan origen a la variable capacidad.

Definición:

La capacidad (Q) de una sección de vía para una cierta corriente vehicular es el máximo número de vehículos de dicha corriente que puede atravesarla por unidad de tiempo, en las condiciones prevalecientes de circulación.

Las condiciones a las que se hace referencia en esta definición son entre otras:

- Tipo de conflictos existentes;
- Regulación de ellos;
- Características de la(s) corriente(s) vehiculares involucradas;
- Características del tramo de vía y factores ambientales.

Por ejemplo, la expresión de la capacidad en un proceso de circulación ininterrumpida deducida con la relación lineal $v - k$ es $Q = \frac{k_e v_l}{4}$. La concentración de embotellamiento depende del largo promedio de los vehículos y que la velocidad a flujo libre, entre otros factores, es función de:

- Potencia, capacidad de frenado de los vehículos;
- Aversión al riesgo, pericia de los conductores;
- Pendiente, curvatura, carpeta de rodado de la vía;
- Condiciones climáticas, día-noche.

Así entonces Q depende de todas estas variables.

Por otra parte podemos decir que Q representa la oferta vial existente en un dispositivo, que q representa la demanda por ese dispositivo vial y que x (grado de saturación) refleja el equilibrio demanda-oferta. El grado de saturación es una variable relevante ya que como veremos más adelante los impactos que produce la circulación de los vehículos son función de él.

En resumen, la capacidad es un límite conceptual absoluto. No obstante, dada la multiplicidad y variedad de factores involucrados en su definición, su valor puede variar para una misma vía y corriente de tráfico. Si además agregamos el factor humano, parece más adecuado suponer que la capacidad es una variable aleatoria que puede ser estimada bajo distintas condiciones.

A continuación veremos cómo se estima la capacidad de intersecciones y estaciones de transferencia.

2.2 Capacidad de intersecciones no semaforizadas

En una intersección de prioridad se distinguen dos tipos de corrientes: prioritarias y secundarias. Para las corrientes prioritarias podemos decir que la

intersección *no existe* y, por lo tanto, pueden ser analizadas utilizando el enfoque de circulación ininterrumpida. Para las corrientes secundarias es necesario redefinir el concepto de capacidad ya que estamos frente a un proceso de circulación interrumpida.

Definición:

La capacidad de un acceso secundario corresponde a la tasa de descarga de una cola de vehículos suficientemente larga formada en el acceso.

Para estimar la capacidad en este caso existen dos tipos de modelos: aceptación de brechas y lineal.

2.2.1 Modelo de aceptación de brechas

El desarrollo del modelo de aceptación de brechas se fundamenta en la siguiente premisa:

Para cruzar una vía prioritaria o para incorporarse a ella (virar) desde una vía secundaria, los conductores de la rama secundaria observan las brechas (intervalos) que se presentan en el flujo prioritario esperando que se produzca una suficientemente grande como para realizar el movimiento deseado.

La variable **brecha** corresponde a la percepción que tienen los conductores de la rama secundaria del intervalo en la rama principal.

El modelo se basa en los siguientes supuestos:

- Regulación se respeta;
- La función de distribución de brechas e intervalos es la misma (no hay sesgo en la percepción);
- El intervalo mínimo para cruzar, denominado brecha crítica (τ), es igual y constante para todos los conductores (homogeneidad y consistencias de los conductores);
- Los vehículos de la rama principal llegan a la intersección de dos formas:

1. En pelotón, o restringidos por la capacidad de la vía prioritaria, con intervalo constante Δ que es igual al intervalo mínimo, $\Delta = 1/Q_p$
2. Libres, con intervalos distribuidos según una función de distribución exponencial negativa desplazada, $h \geq \Delta$

- Proporción de vehículos en pelotón igual a θ ;
- Tiempo de avance de un lugar en la cola de la rama secundaria igual a β .

Obviamente, la brecha crítica debe ser mayor que el intervalo mínimo porque de otro modo los vehículos de la rama secundaria no tendrían problema para cruzar la rama prioritaria.

Bajos estos supuestos tenemos que el evento

$$\tau + (n - 1)\beta \leq h \leq \tau + n\beta \quad n \in Z^+ \quad (2.1)$$

equivale a que crucen o viren n vehículos desde la cola de la vía secundaria.

Luego, la probabilidad que crucen n vehículos es igual a

$$f(n) = Prob(h) = F_h(\tau + n\beta) - F_h(\tau + (n - 1)\beta) \quad (2.2)$$

donde F_h es la función de distribución del intervalo.

Pero de los supuestos sabemos que la distribución de h para los vehículos libres es

$$F_h(x) = \begin{cases} 0 & x \leq \Delta \\ 1 - (1 - \theta)e^{-\gamma(x-\Delta)} & x \geq \Delta \end{cases} \quad (2.3)$$

La capacidad de la rama secundaria se obtiene como

$$Q_s = \frac{E(n)}{E(h)} \quad (2.4)$$

Pero $E(h) = \bar{h} = 1/q_p$. Luego, $Q_s = q_p E(n)$. Además, $E(n) = \sum_{n=1}^{\infty} n f(n)$.

Considerando a los vehículos en pelotón y los libres se puede demostrar que

$$\bar{h} = \Delta + \frac{1 - \theta}{\gamma} = \frac{1}{q_p} \quad \Leftrightarrow \quad \gamma = \frac{q_p(1 - \theta)}{1 - q_p\Delta}$$

Reemplazando este valor en F_h , evaluando en los límites adecuados y ordenando los términos se llega al modelo de Plank y Catchpole (ver Plank y Catchpole, 1984)

$$Q_s = \frac{q_p(1 - \theta) \exp\left(\frac{-q_p(1 - \theta)(\tau - \Delta)}{1 - q_p\Delta}\right)}{1 - \exp\left(\frac{-q_p(1 - \theta)\beta}{1 - q_p\Delta}\right)} \quad (2.5)$$

En este modelo las variables temporales se miden en segundos y en consecuencia los flujos y la capacidad se expresan en *veh/seg* para mantener la consistencia.

Si en el modelo anterior suponemos que $\theta = \frac{q_p}{Q_p} = q_p\Delta$ y reemplazamos esta relación en la expresión de la capacidad se obtiene el modelo de Tanner

$$Q_s = \frac{q_p(1 - q_p\Delta) \exp(-q_p(\tau - \Delta))}{1 - \exp(-q_p\beta)} \quad (2.6)$$

De aquí es fácil ver que

$$\lim_{q_p \rightarrow \frac{1}{\Delta}} Q_s = 0$$

$$\lim_{q_p \rightarrow 0} Q_s = \frac{1}{\beta}$$

Cuando la corriente secundaria debe enfrentar a más de una corriente prioritaria el modelo resultante es el siguiente

$$Q_s = \frac{q_p [\prod_{i=1}^n (1 - q_{p_i}\Delta_i)] \exp(-q_p(\tau - \bar{\Delta}))}{1 - \exp(-q_p\beta)} \quad (2.7)$$

donde n es el número de corrientes prioritarias, $q_p = \sum_{i=1}^n q_{p_i}$, $\bar{\Delta} = \sum_{j=1}^n \frac{q_{p_j}}{q_p} \Delta_j$ y además se supone $\theta = q_p\bar{\Delta}$.

Un supuesto fuerte del modelo cuando existe tráfico mixto es el de homogeneidad de los conductores. Para levantarlo admitiremos que existen varios tipos de unidad conductor-vehículo, cada uno con su propia brecha crítica (τ_j). Estas unidades pueden diferir por clase de vehículo (auto, bus, etc) o por tipo de movimiento (cruce, viraje, etc). El modelo que se obtiene es

$$Q_s = \frac{q_p(1-\theta)\exp\left(\frac{q_p(1-\theta)\Delta}{1-q_p\Delta}\right)}{\left(1 - \exp\left[\frac{-q_p(1-\theta)\beta}{1-q_p\Delta}\right]\right) \sum_{j=1}^n p_{s_j} \exp\left(\frac{q_p(1-\theta)\tau_j}{1-q_p\Delta}\right)} \quad (2.8)$$

donde $p_{s_j} = q_{s_j}/q_s$ es la proporción en el flujo secundario de la clase j . Si se hace $\theta = q_p\Delta$ el modelo se transforma en

$$Q_s = \frac{q_p(1-q_p\Delta)\exp(q_p\Delta)}{(1 - \exp(-q_p\beta)) \sum_{j=1}^n p_{s_j} \exp(q_p\tau_j)} \quad (2.9)$$

Por lo tanto, la capacidad total corresponde a la media armónica de las capacidades de cada clase.

Otro supuesto importante es el de consistencia de los conductores. Se ha observado que éstos tras rechazar varias brechas se vuelven más osados y están dispuestos a aceptar brechas menores. Esto se puede tratar haciendo depender τ del proceso de generación de brechas.

Otro supuesto criticable es la constancia del tiempo de avance. Hay indicios que β crece cuando el flujo prioritario disminuye. Lo cual podría ser modelado haciendo $\beta = \beta(\theta)$. Por otro lado, se ha comprobado empíricamente que el modelo de aceptación de brechas sobreestima la capacidad para valores bajos del flujo prioritario y que la subestima cuando el flujo prioritario es alto (Kimber, 1989).

El modelo de aceptación de brechas no incluye explícitamente variables de diseño de la intersección: visibilidad, ancho de pista, pendiente de la vía, etc; lo cual hace difícil su uso con propósitos de evaluación de proyectos de rediseño geométrico.

Además, existe el problema de obtener valores adecuados para los parámetros del modelo, especialmente τ y β . Es razonable pensar que ambos dependen

de la geometría de la intersección. Pero es necesario destacar que la brecha crítica no es observable, sólo se sabe que si el conductor cruza, la brecha observada es igual o superior a la brecha crítica.

Los problemas mencionados previamente condujeron al desarrollo de un modelo alternativo, de orientación más empírica, dada la dificultad de formular un modelo teórico capaz de incorporar todas las particularidades del fenómeno.

2.2.2 Modelo lineal

El modelo lineal para estimar la capacidad surge como una alternativa para superar algunos de los inconvenientes del modelo de aceptación de brechas. Se basa en suponer una relación lineal entre la capacidad de la rama secundaria y un conjunto de variables explicativas: flujos prioritarios y geometría de la intersección. En forma genérica,

$$Q_s = Q_o - \sum_{i \in I} \alpha_i q_{p_i} \quad (2.10)$$

donde I es el conjunto de flujos prioritarios (q_{p_i}) para el movimiento de la rama secundaria bajo análisis y Q_o y $\alpha_i \geq 0$ son parámetros de modelación.

Numerosas observaciones mostraron que son factores relevantes para determinar la capacidad, además de los flujos prioritarios, las siguientes variables:

- ancho vía principal
- ancho vía secundaria
- visibilidad desde vía secundaria

Las siguientes variables mostraron no ser relevantes para determinar la capacidad:

- pendiente de las vías

- radios de giro
- ángulo de la intersección
- velocidad de los vehículos de la vía principal

A partir de numerosas observaciones se estableció el siguiente modelo lineal para la capacidad de un movimiento desde un acceso secundario

$$Q_s = X \left[Q_o + Z - Y \sum_{i \in I} \alpha_i q_{p_i} \right] \quad (2.11)$$

donde X, Y y Z son parámetros que dependen explícitamente de la geometría de la intersección. Específicamente,

$$\begin{aligned} X &= X(w, V) \\ Y &= Y(W) \\ Z &= Z(w_b) \end{aligned} \quad (2.12)$$

donde w es el ancho de pista en el acceso secundario, V es la visibilidad desde el acceso secundario, W es el ancho de la vía prioritaria y w_b el ancho del bandejon central de la vía prioritaria (ver Kimber y Coombe, 1980; Schumilo y Coeymans, 1987).

IMPORTANTE:

- En el modelo lineal los flujos deben expresarse en unidades de vehículos equivalentes (*veq*) por hora. De hecho la capacidad Q_s tiene unidades de *veq/h*. Con esto se recoge la heterogeneidad de los tipos de vehículos que circulan por la intersección.
- Para cada movimiento desde la rama secundaria existe un conjunto propio de parámetros Q_o y α_i y obviamente un conjunto específico de flujos prioritarios I . Con esto se recoge la heterogeneidad de los movimientos que realizan los vehículos en la intersección.

Para transformar flujos expresados en vehículos (*veh*) a flujos medidos en vehículos equivalentes (*veq*) se utilizan factores de equivalencia. Para esto se establece como referencia a los vehículos livianos (autos particulares). Así por ejemplo, si un bus tiene factor de equivalencia 2 entonces el paso de un bus equivale al paso de 2 autos particulares, en relación al fenómeno de la capacidad.

Sea f_i el factor de equivalencia de un vehículo de tipo i y sea q_i el flujo de vehículos de ese tipo entonces el flujo total expresado en *veh* es $\sum_i q_i$ y el flujo expresado en *veq* será $\sum_i f_i q_i$. Luego, es posible definir el factor de composición del flujo f_c , como

$$f_c = \frac{\sum_i f_i q_i}{\sum_i q_i} \left[\frac{veq}{veh} \right] \quad (2.13)$$

que permite por ejemplo transformar la capacidad entregada por el modelo lineal en (*veq/h*) a (*veh/h*):

$$Q_s[veh/h] = \frac{Q_s[veq/h]}{f_c} \quad (2.14)$$

Cuando más de un movimiento utiliza una misma pista se debe calcular una capacidad compuesta para la pista. Sea q_m y Q_m el flujo y la capacidad según el modelo lineal del movimiento m , entonces se define la capacidad compuesta (\tilde{Q}) como

$$\tilde{Q} = \frac{q}{\sum_m \frac{q_m}{Q_m}} \quad (2.15)$$

donde q es el flujo total en la pista. Lo anterior equivale definir \tilde{Q} de manera que el grado de saturación de la pista sea igual a la suma de los grados de saturación de los movimientos que se producen en ella: $x = \sum_m x_m$ con $x = q/\tilde{Q}$ y $x_m = q_m/Q_m$.

Este procedimiento se conoce como composición de movimientos y es válido para pistas analizadas con modelos lineales o de aceptación de brechas.

2.2.3 Modelo lineal aplicado a rotondas

A diferencia de las intersecciones de prioridad reguladas por señal ceda el paso o pare, donde un movimiento es siempre prioritario o secundario, en una rotonda la prioridad es variable. Los movimientos que entran a la rotonda son secundarios, pero una vez dentro de ella se vuelven prioritarios.

Lo relante en este caso es conocer la capacidad de cada acceso o entrada de la rotonda, el cual podría estar compuesto por varias pistas. Al igual que en las intersecciones se busca relacionar linealmente dicha capacidad con los flujos prioritarios. En una rotonda los flujos prioritarios están representados por el flujo circulante, q_c . Es decir, se plantea una ecuación del tipo:

$$Q_e = Q_o - \alpha q_c$$

Del mismo modo que antes se plantea que la geometría de la intersección influye en la determinación de la capacidad. El análisis empírico de un conjunto de rotondas en Inglaterra (Kimber, 1980) permitió especificar y estimar el siguiente modelo para la capacidad de cada entrada a la rotonda:

$$Q_e = k(F - f_c q_c) \quad (2.16)$$

$$\begin{aligned} k &= 1 - 0,00347(\phi - 30) - 0,978\left(\frac{1}{r} - 0,05\right) \\ F &= 303x_2 \\ f_c &= 0,210t_D(1 + 0,2x_2) \\ t_D &= 1 + \frac{0,5}{1 + \exp\left(\frac{D-60}{10}\right)} \\ x_2 &= v + \frac{e-v}{1+2S} \\ S &= \frac{e-v}{l} \end{aligned}$$

Las variables e , v , l , D y r se miden en metros, ϕ en grados, y q_c y Q_e en vehículos equivalentes por hora. La definición conceptual de éstas y otras variables es la siguiente.

- e : ancho de la entrada a la rotonda;
- v : mitad del ancho de la calzada de aproximación;
- l : longitud efectiva del ensanchamiento de la entrada.

D : diámetro del círculo inscrito en la rotonda;
 r : radio de la curva de entrada;
 ϕ : ángulo de entrada;
 S : agudeza del ensanche.

El modelo de capacidad permite analizar los accesos de la rotonda en forma separada. Sin embargo, ya que la capacidad de un acceso depende del flujo circulante frente a él y este flujo se origina en los accesos previos, el balance total de flujos en la rotonda se vuelve en algunas circunstancias complejo. Existen dos casos extremos.

- Si todos los accesos están saturados la capacidad de cada acceso se vuelve completamente dependiente del resto ya que la magnitud del flujo circulante depende del flujo entrante en los accesos previos y estos están definidos por las respectivas capacidades. Luego, para encontrar la capacidad se debe utilizar un método iterativo en el cual se determina el flujo circulante frente a cada acceso de forma consistente con la capacidad de todos ellos.
- Si ninguno de los accesos está saturado el flujo circulante frente a cada acceso se obtiene de la demanda de los accesos previos en concordancia con las proporciones de viraje.

En una situación intermedia sólo algunos de los accesos estarán saturados. También será necesario utilizar un algoritmo iterativo para encontrar la capacidad.

2.3 Capacidad de intersecciones semaforizadas

En este tipo de intersecciones existe un dispositivo, el semáforo, que alterna cíclicamente la prioridad en el uso del área común. Dentro de un ciclo del semáforo se distinguen fases donde se autoriza el paso de ciertas corrientes dándoles luz verde. Al completarse el ciclo todas las corrientes que llegan a la

intersección habrán recibido un período de verde. Cuando una fase termina aparece la luz roja que impide el paso de las corrientes respectivas.

Si la regulación del semáforo se respeta, cada corriente atraviesa por una etapa en la que goza de prioridad absoluta sobre otras corrientes (período verde) y otra en que no puede pasar (período rojo). Por motivos de seguridad existe una transición entre ambas, que es el período amarillo.

En una intersección semaforizada, la descarga de una cola formada en alguno de los accesos se modela de la siguiente forma. Durante el período de rojo la tasa de descarga es nula, al comenzar el período de verde los vehículos comienzan a salir de la cola y rápidamente se establece una tasa aproximadamente constante, finalmente al comenzar el período amarillo la tasa de descarga disminuye hasta llegar a cero al iniciarse el rojo. Este proceso se repite cíclicamente de la misma manera mientras exista cola en el acceso.

La tasa de descarga durante el período de verde se denomina flujo de saturación, S .

Cuando el período de verde comienza los vehículos en cola demoran algunos segundos en alcanzar la velocidad de desplazamiento normal. Luego de esta transición inicial la cola se descarga a tasa S . Mientras el semáforo esté en verde y exista cola, los vehículos continúan pasando a una tasa igual al flujo de saturación. Al término del período de verde algunos vehículos harán uso del período de amarillo para cruzar y la tasa de descarga disminuirá finalmente a cero.

Sea

- λ_1 : pérdida de verde
- λ_2 : ganancia de amarillo
- V_i : tiempo de verde de la fase i
- R_i : tiempo de rojo de la fase i
- A_i : tiempo de amarillo de la fase i
- C : tiempo de ciclo del semáforo

Por construcción $C = V_i + A_i + R_i$. Estas variables representan la programación del semáforo.

resto de los accesos. Por ejemplo, si el flujo aumenta por uno de los accesos el semáforo debería asignarle mayor tiempo de verde y si el tiempo de ciclo permanece constante la capacidad de este acceso aumentará mientras que la del resto disminuirá. Si la programación no es dinámica, igualmente existe una dependencia indirecta de las condiciones en los otros accesos ya que la programación se realiza para condiciones de demanda representativas de cada período en todos los accesos.

El modelo de capacidad recién descrito puede ser interpretado en términos físicos. En efecto, cuando un vehículo no alcanza a cruzar debe detenerse al igual que todos los que vienen a continuación, al finalizar el período de verde los vehículos salen del reposo acelerando hasta alcanzar una velocidad y concentración equivalentes al flujo de saturación. Bajo estas condiciones, la definición de la variable pérdida de verde (λ_1) equivale a suponer que los vehículos salen del reposo y alcanzan de inmediato la velocidad de régimen.

A su vez, la variable ganancia de amarillo corresponde al promedio de tiempo, medido desde el comienzo del período de amarillo, que tarda en cruzar frente a la línea de detención el último vehículo que alcanzó a salir de la cola.

Luego, se debe cumplir

$$V + \lambda_2 = \lambda_1 + n\beta$$

donde $\beta = 1/S$ es el tiempo promedio que demora un vehículo en cruzar frente a la línea de detención y n es el número de vehículos que logran salir de la cola durante el ciclo.

La variable β es de similar naturaleza conceptual que la utilizada en el modelo de aceptación de brecha. Su valor depende de una gran cantidad de factores: ambientales, geométricos y de composición del tráfico. Nótese que a partir de observaciones de V y n es posible obtener mediante una regresión lineal el valor de β .

El modelo anterior supone una cola compuesta por vehículos homogéneos. Sin embargo, en un problema real pueden existir en la misma cola varios tipos de vehículos y varios movimientos. Esto hace necesario analizar e incorporar la heterogeneidad en la estimación de la capacidad. Algo de esto vimos a propósito de los modelos de intersecciones de prioridad.

2.4 Heterogeneidad del tráfico

Al hablar de heterogeneidad del tráfico nos referimos al hecho de que en condiciones reales existen varios tipos de vehículos circulando por las vías de la ciudad, los que además realizan distintos movimientos en cada intersección. El problema de esta situación es que cada uno de estos tipos podría tener un comportamiento distinto al del resto. Obviamente, también podrían exhibir características físicas (longitud) u operacionales (velocidad, tasa de frenado) distintas. Luego, en condiciones reales es poco probable que exista homogeneidad en la circulación y en consecuencia parece más razonable reconocer explícitamente que existe heterogeneidad.

Para incorporar explícitamente la heterogeneidad la estrategia será encontrar factores de equivalencia que permitan traducir corrientes de distintas características en corrientes homogéneas. Ahora bien, es importante mencionar que el patrón de homogeneidad no es único, es decir, estamos hablando de homogeneizar bajo ciertas condiciones.

La unidad de referencia para expresar el flujo será el veq/h por pista o bien el ADE/h por pista.

- veq : corresponde a un auto particular. Su nombre significa Vehículo EQUIvalente.
- ADE : es un veq cuyo movimiento en la intersección es continuar directo (no virar). Su nombre es una abreviación de Auto Directo EQUIvalente.

Estas unidades adquieren ciertos valores básicos de acuerdo con las dimensiones y características mecánicas de los vehículos, factores ambientales y características de los conductores.

El problema de la heterogeneidad se presenta en todas las facetas de la circulación, por lo que puede haber distintos criterios de equivalencia según el fenómeno que se trate. En nuestro caso el criterio es la capacidad, de modo que los factores de equivalencia antes mencionados traducen una corriente heterogénea de vehículos a una homogénea de capacidad equivalente, compuesta por ejemplo sólo de $ADEs$.

La heterogeneidad se debe analizar para cada uno de los modelos de capacidad analizados: aceptación de brechas, lineales y modelo para intersecciones semaforizadas; ya que cada uno de ellos se fundamenta en distintos supuestos sobre homogeneidad en el comportamiento de los vehículos (conductores).

2.4.1 Heterogeneidad en el flujo de saturación

Recordemos que la capacidad de una pista de una rama de una intersección semaforizada es dada por $Q = Su$. La variable u corresponde a la razón de verde efectivo y S al flujo de saturación. La razón de verde efectivo está determinada por la programación del semáforo y para nosotros es un dato. El flujo de saturación es el máximo número de vehículos por unidad de tiempo que salen de la cola durante el período de verde efectivo. Observamos entonces que el flujo de saturación podría ser distinto en distintas condiciones de circulación. Así por ejemplo, si en la cola existe una alta proporción de buses el flujo de saturación expresado en *veh* será menor que si en la cola existieran sólo autos; debido a que las dimensiones de un bus son mayores que las de un auto, a que el bus tiene una menor capacidad de aceleración que el auto, etc.

Para abordar este problema trataremos de encontrar el flujo de saturación a partir de un flujo de saturación elemental o básico más sencillo de calcular y que luego se corrige para encontrar el flujo de saturación real. Se define entonces

Flujo de saturación básico (S_b): tasa de descarga de una cola compuesta sólo por *ADEs* que circulan por una pista utilizada sólo por *ADEs*.

El flujo de saturación básico según estudios realizados recientemente en Santiago varía entre 1.933 y 2.292 *ADE/h – pista*. Las variaciones se producen entre pistas y entre distintos períodos del día, tal como se muestra en la Tabla 2.1.

El flujo de saturación para una corriente real se obtiene a partir del básico, tal como se muestra a continuación:

$$S[\text{veh}/h] = f_k[\text{veh}/\text{ADE}] S_b[\text{ADE}/h] \quad (2.20)$$

Tabla 2.1: Flujo de saturación básico ($ADE/h - pista$)

Período	Derecha	Central	Izquierda
Mañana	2.055	2.121	2.292
Otro	1.933	1.992	2.141

donde f_k es un factor de corrección que recoge las diferencias entre la corriente real y la básica, en relación a tipos de vehículos y/o pistas. Estas diferencias están asociadas a:

- ancho de pista
- vehículos: tipo (composición del flujo); movimientos (virajes, interferencias)

Todos los desarrollos que se muestran a continuación se refieren a una pista, pero para simplificar la notación no se colocará explícitamente esta dimensión en las unidades del flujo.

2.4.1.1 Factores de equivalencia para determinar el flujo de saturación

Factor por ancho de pista

$$f_a = 1 + 0,058(A - 3,0)D_{PE} \quad (2.21)$$

donde A es el ancho de pista en metros y D_{PE} es una variable muda asociada a la ubicación de la pista, vale 1 para pista izquierda o derecha (pista extrema).

Factores asociados a los tipos de vehículos y movimientos

a) Factor por tipo de vehículo

Los factores de equivalencia por tipo de vehículos resultan de comparar la descarga de vehículos de algún tipo con la descarga de un flujo equivalente compuesto sólo por $ADEs$, suponiendo además que no hay virajes, efecto que se analiza después.

Tenemos entonces que bajo estas condiciones el factor de equivalencia para vehículos de tipo i está dado por:

$$f_i = \frac{\beta_i}{\beta_1} [ADE/veh] \quad (2.22)$$

donde β_i es el intervalo promedio de salida a saturación para un vehículo de tipo i , $i = 1$ representa al ADE , nuestra referencia.

Sabemos que para una corriente compuesta sólo de $ADEs$ debe cumplirse $V + \lambda_2 = \lambda_1 + n\beta_1$ o equivalentemente, si se define $T = V + \lambda_2 - \lambda_1$,

$$T = n\beta_1$$

en consecuencia $S_b = n/T = 1/\beta_1$ es el flujo de saturación de dicha corriente.

Si tenemos ahora una corriente heterogénea, se deberá cumplir

$$V + \lambda_2 = \lambda_1 + \sum_i n_i \beta_i$$

o bien $\tilde{T} = \sum_i n_i \beta_i$ donde n_i es el número de vehículos de tipo i que salen de la cola y $\tilde{T} = V + \lambda_2 - \lambda_1$. Utilizando el factor de equivalencia, la ecuación anterior se transforma en $\tilde{T} = \beta_1 \sum_i f_i n_i$. Por lo tanto,

$$S_b = \sum_i f_i \frac{n_i}{\tilde{T}} [ADE/h]$$

Si pensamos ahora en la corriente heterogénea como si ella fuera la de referencia, su flujo de saturación será

$$S = \sum_i \frac{n_i}{\tilde{T}} [veh/h]$$

Resulta entonces

$$\frac{S}{S_b} = \frac{\sum_i n_i / \tilde{T}}{\sum_i f_i n_i / \tilde{T}} = \frac{\sum_i n_i}{\sum_i f_i n_i}$$

dividiendo arriba y abajo por la duración del período de análisis, en este caso C , se llega finalmente al factor de composición del flujo f_c

$$f_c = \frac{\sum_i f_i q_i}{\sum_i q_i} [ADE/veh] \quad (2.23)$$

Tabla 2.2: Factor de equivalencia para *ADM*

%TP	Derecha		Izquierda		Central	
	Mañana	Otro	Mañana	Otro	Mañana	Otro
≤ 48	$\leq 1,119$	$\leq 1,112$	$\leq 1,123$	$\leq 1,115$	$\leq 1,133$	$\leq 1,124$
> 48	1,120	1,113	1,124	1,116	1,134	1,125

Tabla 2.3: Factor de equivalencia para bus

Período	Derecha	Izquierda	Central
Mañana	$1,784f_a$	$1,463f_a$	$1,580f_a$
Otro	$1,678f_a$	$1,373f_a$	$1,476f_a$

o bien $f_c = \sum_i p_i f_i$ donde p_i es la proporción de vehículos de tipo i en el flujo.

Este factor permite transformar el flujo de saturación básico en uno específico para una corriente heterogénea compuesta por distintos tipos de vehículos

$$S(veh/h) = S_b(ADE/h) \frac{1}{f_c} (veh/ADE) \quad (2.24)$$

Nótese que hemos supuesto que los vehículos no viran, de ahí que f_c tenga unidades de ADE/veh ya que si existieran virajes este factor tendría unidades de veh/h .

Un primer tipo de vehículo distinto del *ADE* es el auto directo en una pista donde además circulan buses. Este tipo de vehículo se denomina *ADM* por Auto Directo en pista con tráfico Mixto. Al comparar el intervalo de descarga del *ADM* con el *ADE* se obtiene su factor de equivalencia f_{ADM} . De acuerdo a las mediciones realizadas en Santiago, el factor f_{ADM} depende del porcentaje de transporte público en el flujo de la pista, del período del día y de la ubicación de la pista. Los valores obtenidos se muestran en la Tabla 2.2. Otro tipo de vehículo distinto del *ADE* es el bus. Nuevamente, al comparar los intervalos de descarga de ambos tipos de vehículos se obtiene el factor de equivalencia. Los valores obtenidos en Santiago son los que aparecen en la Tabla 2.3.

Ahora incorporaremos en el análisis el efecto de los distintos movimientos que realizan los vehículos en la intersección.

b) Factor por tipo de movimiento

i) Viraje sin oposición

Para los virajes sin oposición se ha demostrado empíricamente que el factor de equivalencia está determinado por el radio de giro (r) del movimiento. Este factor tiene unidades de ADE/veg .

$$f_{vs} = \begin{cases} 1 + 1,5/r & r < 10 \\ 1 + 150/r^3 & r \geq 10 \end{cases} \quad (2.25)$$

En esta ecuación el radio de giro se expresa en metros. Los vehículos que siguen directo tienen factor igual a uno, valor que puede obtenerse a partir de la ecuación anterior con $r \rightarrow \infty$. Se tendrá entonces que el factor de equivalencia f_i en este caso está dado por $f_i f_{vs}$. El factor f_i transforma veh a veg , lo denominaremos entonces f_{T_i} , y f_{vs} transforma veg en ADE , lo llamaremos f_{M_i} . En consecuencia, el factor de equivalencia

$$f_i = f_{T_i} f_{M_i} \quad (2.26)$$

transforma el tipo de vehículo-movimiento i a auto directo equivalente ADE . Luego, el factor de composición se transforma en $f_c = \sum_i p_i f_i = \sum_i p_i f_{T_i} f_{M_i}$

En definitiva, se tendrá

$$S = f_k S_b \quad (2.27)$$

$$f_k = f_a / f_c \quad (2.28)$$

ii) Viraje con oposición

Cuando el viraje es con oposición, ya sea vehicular o peatonal, surge un proceso regulado por prioridad al interior de la intersección semaforizada. Los vehículos que intentan virar deben esperar una brecha en el flujo que los interfiere. Para encontrar el factor de equivalencia debemos calcular el flujo de saturación del movimiento con oposición que, tal como se indica, corresponde a la capacidad de un movimiento secundario. Esta capacidad puede obtenerse con un modelo de aceptación de brecha o un modelo lineal.

Al comenzar el verde no es posible pasar pues se está descargando la cola del movimiento opositor formada en el período de rojo. Esta cola se descarga a intervalo mínimo, por lo cual no se puede encontrar una brecha adecuada. Terminada esta etapa empieza la descarga del viraje de forma análoga a un movimiento secundario en una intersección de prioridad. Esta etapa se extiende hasta el final del verde efectivo a tasa equivalente a Q_s . Sin embargo, se observa que al final de esta etapa se produce la descarga de un número pequeño (n_f) de vehículos.

Podemos construir entonces un modelo binario, descomponiendo $V_e + \lambda_1$ en un período de bloqueo (V_b) y un período de descarga por brecha (\tilde{V}_e), es necesario extender este último para considerar n_f . Tenemos entonces un período de verde efectivo corregido V_o dado por:

$$V_o = \tilde{V}_e + \frac{n_f}{Q_s} = V_e + \lambda_1 - V_b + \frac{n_f}{Q_s} \quad (2.29)$$

Por lo tanto, la capacidad para el viraje con oposición será $Q_{vo} = Q_s \frac{V_o}{C}$.

La mayoría de los parámetros que aparecen en el modelo puede considerarse conocidos, la excepción es el tiempo de bloqueo, el cual depende de la duración del período de rojo y de cuántos sean los movimientos opositores. Si comparamos el modelo obtenido con la expresión típica de la capacidad de un movimiento con flujo de saturación \tilde{S} , se tiene

$$\tilde{S} \frac{V_e}{C} = Q_s \frac{V_o}{C} \Leftrightarrow \tilde{S} = Q_s \frac{V_o}{V_e}$$

En consecuencia,

$$\tilde{S} = Q_s \left(1 - \frac{V_b - \lambda_1 - n_f/Q_s}{V_e} \right) \quad (2.30)$$

Por lo tanto, el flujo de saturación depende de la programación del semáforo. Por esta razón no es conveniente en este caso utilizar factores de equivalencia ya que tendrían la misma dependencia, lo que es poco práctico.

2.4.1.2 Duración del verde efectivo

Para encontrar la capacidad es necesario también calcular la duración del verde efectivo. Por esta razón se utilizaron las mediciones que llevaron a

los valores de las Tablas 2.2 y 2.3 para determinar el valor de λ_1 y λ_2 . Sin bien se obtuvieron valores distintos para los diferentes períodos del día y la distintas ubicaciones de la pista, la diferencia $\lambda_1 - \lambda_2$ tiene aproximadamente el mismo valor en todos los casos. El resultado obtenido fue

$$\lambda_1 - \lambda_2 = 1,4 \quad [seg] \quad (2.31)$$

Esto indica que la duración del verde efectivo es 1,4 seg. menor que el verde del semáforo.

2.4.2 Heterogeneidad en intersecciones no semaforizadas

2.4.2.1 Tipos de vehículos

En el modelo de aceptación de brechas los parámetros que definen el modelo deben ser medidos para cada caso y, por lo tanto, resultarán influenciados por condiciones ambientales y de circulación. No existen factores de equivalencia ya que las brechas se producen, rechazan o aceptan por vehículos reales, no por *veq* o *ADE*.

En el modelo lineal los flujos se miden en *veq/h*, al existir varios tipos de vehículos se deberán utilizar factores de equivalencia análogos a los empleados en intersecciones semaforizadas. En este caso se definen dos tipos de vehículos: livianos y pesados. Luego, para transformar una corriente vehicular heterogénea a una corriente equivalente compuesta sólo por *veq*, se utiliza el siguiente factor de composición del flujo:

$$f_c = \frac{q_{vl} + 2q_{vp}}{q_{vl} + q_{vp}} \quad [veq/veh]$$

donde q_{vl} y q_{vp} es el flujo de vehículos livianos y pesados, respectivamente.

2.4.2.2 Tipo de movimiento

La heterogeneidad por movimientos se trata estimando la capacidad de cada uno utilizando parámetros específicos y luego combinando los resultados. Si

se trata de una rotonda no existe heterogeneidad de este tipo ya que en una rotonda existe sólo un movimiento posible en cada acceso.

Supongamos entonces una cola heterogénea con n_i vehículos que realizan el movimiento secundario i y sea $N = \sum_i n_i$. Por definición, la capacidad de la pista será $Q_s = N/T$ donde T es el tiempo demoran en descargarse los N vehículos de la cola. El tiempo de descarga individual será $T_i = n_i/Q_{s_i}$ donde Q_{s_i} proviene del modelo de capacidad lineal y ha sido convertido a *veh/h*. Entonces

$$\begin{aligned} \frac{N}{Q_s} = T &= \sum_{i=1}^N T_i = \sum_{i=1}^N \frac{n_i}{Q_{s_i}} \\ \Rightarrow Q_s &= \frac{N}{\sum_i n_i/Q_{s_i}} \end{aligned}$$

dividiendo arriba y abajo por la duración del período de análisis se obtiene finalmente

$$Q_s = \frac{q_s}{\sum_i q_{s_i}/Q_{s_i}} \quad (2.32)$$

o bien

$$Q_s = \frac{1}{\sum_i p_{s_i}/Q_{s_i}} \quad (2.33)$$

donde $p_{s_i} = q_{s_i}/q_s$ es la proporción del movimiento i en el flujo total. La ecuación anterior puede ser escrita de una manera más fácil de recordar. En efecto, Q_s es solución de la ecuación

$$\frac{q_s}{Q_s} = \sum_i \frac{q_{s_i}}{Q_{s_i}} \quad \Leftrightarrow \quad x_s = \sum_i x_{s_i}$$

lo que equivale a que **el grado de saturación de la pista es la suma de los grados de saturación de los movimientos que se producen en la pista.**

2.4.2.3 Elección de pista

Vemos que tanto para el factor de composición del flujo (f_c), como para la composición de capacidades en intersecciones de prioridad, se requiere conocer los flujos en la pista bajo análisis. Cuando un movimiento dispone de

más de una pista se presenta un problema pues lo que se conoce normalmente es el flujo del movimiento en su conjunto y no por pista.

Hay que considerar entonces la elección de pista por parte del conductor. Parece razonable suponer que lo hacen tratando de minimizar su demora. Operacionalmente esto se hace imponiendo que el grado de saturación en todas las pistas utilizadas por el movimiento es igual. De no serlo existiría un incentivo para que los usuarios utilicen en mayor grado la pista menos saturada, lo que al final se traducirá en un aumento de su saturación. Este criterio se denomina *equisaturación*. Por ejemplo, si un movimiento utiliza dos pistas entonces

$$\begin{aligned}x_1 &= x_2 \\ \frac{q\theta}{Q_1(q, \theta)} &= \frac{q(1 - \theta)}{Q_2(q, \theta)}\end{aligned}$$

donde q es el flujo total del movimiento, θ la proporción del flujo que utiliza la pista 1 y Q_1 y Q_2 es la capacidad de la pista 1 y 2, respectivamente.

Capítulo 3

Formación y disipación de colas

3.1 Definición del proceso

Como se vio anteriormente existen dos regímenes básicos de circulación: no saturado y saturado. En el primero, el fenómeno más importante es la variación de la velocidad con el flujo $v(q)$ ya analizada. En el segundo, además de lo anterior se presentarán colas de vehículos.

Por definición, la saturación aparece cuando $q > Q$. Ambas variables serán tratadas como dadas para un cierto período, correspondiendo a condiciones medias. Esto puede ser razonable para tramos de vía, pero ya no lo es para intersecciones. Nuestro estudio de capacidades demuestra que ésta varía en el tiempo, aún en cortos períodos; incluso en intersecciones semaforizadas sistemáticamente existen períodos en que $Q = 0$. Así que, a pesar de que en un período dado en promedio se cumpla $q < Q$, puede existir saturación en algún instante dentro del período.

Por otra parte, aunque en promedio $q < Q$ y suponiendo que Q fuera constante, sabemos que q varía en torno al promedio y bien puede suceder que en algunos momentos $q > Q$.

En definitiva $q = q(t)$ y $Q = Q(t)$ y, por lo tanto, $x = x(t)$ sujeto además a variaciones aleatorias.

¿Por qué importarían estas fluctuaciones sistemáticas o aleatorias en torno al promedio? Porque no hay compensación. Los excesos de capacidad disponibles cuando $q < Q$ no sirven para los períodos posteriores en que ocurre lo inverso: el exceso de capacidad no se puede almacenar. La ausencia de compensación implica que las fluctuaciones tienen un impacto negativo en términos de generación de colas. Sólo se pueden usar holguras posteriores a la existencia de cola y hasta la capacidad únicamente puesto que $x = x(t)$. La presencia de colas debe verse como un fenómeno dinámico, en evolución. No basta entonces con decir: si hay cola entonces se descarga a tasa Q . Es necesario estudiar la formación y disipación de éstas, sabiendo lo anterior.

Ahora, ¿qué importancia tiene que se formen colas? Ellas implican la negación de la circulación que es el deseo de los usuarios. Tenemos entonces detenciones involuntarias. A estas detenciones se asocia consumo de recursos valiosos: tiempo, combustible, contaminación, desgaste mecánico y de neumáticos, etc. Incluso, la probabilidad de accidente (también existe en movimiento, claro). Si bien al viajar se consume siempre este tipo de recursos, las detenciones aparecen como pérdidas netas, subjetiva y objetivamente. El tiempo que duran se llama demora e incluye el tiempo empleado en procesos de aceleración y deceleración, con respecto a viajar a velocidad constante.

Las colas son producto de que el flujo de unidades en un sistema es superior a su capacidad en un cierto instante. El sistema en estudio se puede definir por:

- Un proceso de llegada de las unidades (vehículos);
- Un mecanismo de atención de ellos, caracterizado por la cantidad de estaciones de servicio que posee el sistema (número de pistas) y el proceso mismo de atención.

Cada proceso se define por la ocurrencia de un determinado suceso, interesando su cantidad en un período dado y el intervalo entre sucesos. Según la ley que rige la ocurrencia, el suceso puede ser:

- Estocástico o determinístico;

- Estacionario (probabilidad de ocurrencia no varía con el tiempo) o dependiente del tiempo.

A continuación analizaremos dos tipos de sistemas: estocásticos-estacionarios y determinísticos-dependientes del tiempo.

3.2 Colas estacionarias

Entre los procesos estocásticos estacionarios son particularmente conocidos los Markovianos, cuya característica es que la probabilidad de un cierto resultado del proceso sólo depende del estado anterior y no de su historia (estados previos al anterior). Dentro de esta familia de procesos se encuentran los de Poisson: tasa de ocurrencia de eventos es constante en el tiempo y en un intervalo pequeño sólo puede ocurrir o ningún suceso (o sea, éstos ocurren uno a uno). Esto parece muy apropiado para describir llegadas de vehículos.

En nuestro caso entenderemos que los sucesos relevantes son la llegada y la salida de un vehículo del sistema, cuyas tasas de ocurrencia son q y Q , respectivamente. Estas dos variables son variables aleatorias con funciones de distribución que dependen de la clase de proceso que rige a cada una. Sus valores medios son constantes en el período de análisis.

La teoría de colas establece relaciones, dadas las características de los procesos de llegada y salida, donde ahora la variable relevante es la longitud de la cola así como la permanencia de los vehículos en ella (demoras). Típicamente, el sistema se describe por $A/B/N$, donde A y B son el tipo de proceso de llegada y salida, respectivamente, y N es el número de estaciones de atención que se suponen idénticas e independientes. La nomenclatura usual para los procesos es:

- M: Markoviano
- D: Determinístico
- G: General o cualquiera

Para nosotros las estaciones de servicio están representadas por las pistas disponibles para la corriente vehicular analizada. Por las características del proceso de formación y disipación de colas de vehículos en vías urbanas, el fenómeno queda bien representado por un sistema con sólo una estación de atención, es decir, las colas se analizan para cada pista.

Se define:

L = longitud media de cola incluyendo al vehículo en servicio (*veh*);

L' = longitud media de cola excluyendo al vehículo en servicio (*veh*);

D = tasa media de demora incluyendo al vehículo en servicio (*veh-UT/UT*);

D' = tasa media de demora excluyendo al vehículo en servicio (*veh-UT/UT*);

d = demora media por vehículo incluyendo al vehículo en servicio (*UT*);

d' = demora media por vehículo excluyendo al vehículo en servicio (*UT*);

$x = q/Q$: grado de saturación de la pista cuyo flujo q y capacidad Q son conocidos.

En términos simples D corresponde al promedio de la longitud de cola en el período de análisis y d al tiempo promedio que tarda un vehículo en salir de la cola existente en el dispositivo vial estudiado.

Si el sistema es estacionario, se cumple que:

- $L' = qd'$, ecuación de *Little*
- $L = L' + x = qd$
- $D' = L'$: ya que la cola no varía en el tiempo
- $D = L$: idem
- $d = d' + \frac{1}{Q}$: demora promedio en cola más tiempo promedio de atención

Se han estudiado diversos sistemas algunos de los cuales conducen a fórmulas analíticas especificadas para L' y a partir de ésta a las otras variables. Hay casos en los cuales la solución analítica es aún desconocida.

3.2.1 Sistemas de solución conocida

a) Caso $M/M/1$

Se puede demostrar en este caso que la probabilidad de que existan k vehículos en la cola está dada por $p_k = (1-x)x^k$ con $x < 1$. En consecuencia, la longitud media de la cola está dada por

$$\begin{aligned} L = \sum_{k=0}^{\infty} k p_k &= (1-x)x \frac{\partial}{\partial x} \sum_{k=0}^{\infty} x^k \\ &= (1-x)x \frac{\partial}{\partial x} \frac{1}{1-x} \\ &= \frac{x}{1-x} \end{aligned}$$

Por lo tanto, de las relaciones básicas se obtiene

$$L' = \frac{x^2}{1-x} \quad L = \frac{x}{1-x} \quad (3.1)$$

$$d' = \frac{1}{Q} \frac{x}{1-x} \quad d = \frac{1}{Q} \frac{1}{1-x} \quad (3.2)$$

En principio, este caso puede corresponder a una pista en una intersección de prioridad aislada con poca congestión.

b) Caso $M/G/1$

Según el desarrollo de *Pollaczek-Khinchin*

$$L' = C_p \frac{x^2}{1-x} \quad L = L' + x \quad (3.3)$$

$$d' = \frac{C_p}{Q} \frac{x}{1-x} \quad d = d' + \frac{1}{Q} \quad (3.4)$$

con $C_p = \frac{1}{2}(1 + \frac{1}{K_s})$ y donde K_s es el número de *Erlang* para el proceso de servicio, $K_s = \left(\frac{T_s}{\sigma_s}\right)^2$ donde T_s y σ_s son la esperanza y desviación estándar del tiempo de servicio.

Este caso puede extenderse razonablemente a uno $G/G/1$ haciendo que C_p sea sustituido por un parámetro C dado por:

$$C = \frac{1}{2}\left(\frac{1}{K_{LL}} + \frac{1}{K_s}\right) \quad (3.5)$$

con $K_{LL} = \left(\frac{T_{LL}}{\sigma_{LL}}\right)^2$ donde T_{LL} y σ_{LL} son los parámetros de la distribución del tiempo entre llegadas. El interés de este caso radica en que no se fuerza aleatoriedad en los procesos de salida y llegada. Esto es especialmente apropiado en la circulación vial donde por la existencia de semáforos, vehículos muy lentos u otras particularidades, dichos procesos no son estrictamente Markovianos.

Estos casos estacionarios nos permiten apreciar el efecto de las fluctuaciones aleatorias del flujo y la capacidad. Vemos que aunque $x < 1$ puede haber colas muy largas si la condición permanece largo tiempo, lo cual también está influenciado por el parámetro C_p .

Por ejemplo, si las salidas fueran determinísticas entonces $K_s = \infty$ y, por lo tanto, $C_p = 1/2$. Luego la longitud de cola se reduce a la mitad con respecto a un sistema con salidas aleatorias. Si adicionalmente se supone llegadas determinísticas (sistema $D/D/1$) se tendrá $K_a = K_s = \infty \Rightarrow C = 0$ y $L' = 0$. Es decir, en este sistema no hay cola. Esto porque igualmente se supone $x < 1$. Para que exista cola bajo el supuesto de llegadas y salidas determinísticas es imprescindible que $x > 1$. Esta situación no puede ser permanente en el tiempo y nos lleva al caso $x = x(t) > 1$. La condición no es permanente porque si la longitud de cola aumenta demasiado, los conductores evitarán circular por el lugar, disminuyendo así el flujo y generando en consecuencia una disminución del grado de saturación.

3.3 Colas dependientes del tiempo

En este caso el sistema relevante es $D/D/1$ con $x = x(t)$. Para simplificar el análisis de la evolución de la cola, la variación del grado de saturación en el

tiempo se asume discretizada mediante la definición de períodos en los cuales se cumple:

- $x_i = cte$ al interior de cada período i
- $x_{i-1} \neq x_i \neq x_{i+1} \quad \forall i$

Además, en al menos algún período i , $x_i \geq 1$.

A diferencia del caso estacionario, aquí la historia del proceso es decisiva para estimar la longitud de la cola. . . . sta se estima período a período, siendo la final de un período la inicial del siguiente.

Entonces en un período $0 - t$:

$$L'(t) = L'_0 + (x - 1)Qt \quad (3.6)$$

$$L(t) = L_0 + (x - 1)Qt \quad (3.7)$$

donde $L_0 = L(t = 0)$ y $L'_0 = L'(t = 0)$.

El cálculo de las demoras requiere definir una nueva variable, la demora total en el período $0 - t$: Δ_{0-t}

$$\Delta_{0-t} = \int_{z=0}^t L(z)dz \quad (3.8)$$

Entonces la tasa media de demora se define como $D_{0-t} = \Delta_{0-t}/t$.

Se tienen entonces dos valores de la tasa media de demora, dependiendo si se incluye o no el vehículo en servicio en el cálculo de la demora total. El resultado que se obtiene en cada caso se muestra a continuación:

$$D'_{0-t} = L'_0 + (x - 1)Qt/2 \quad (3.9)$$

$$D_{0-t} = L_0 + (x - 1)Qt/2 \quad (3.10)$$

La demora media por vehículo ($d = \Delta/N$) no puede obtenerse directamente a partir de D . Debe calcularse detalladamente ya que al existir un grado de saturación mayor que uno, los vehículos que llegan en el período (N_{LL}) no son los mismos que salen en ese período (N_S). Por convención se acostumbra referir la demora media a los *vehículos que llegan en el período 0 – t*. El resultado final es

$$d'_{0-t} = \frac{L_0}{Q} + (x - 1)\frac{t}{2} \quad (3.11)$$

$$d_{0-t} = \frac{L_0 + 1}{Q} + (x - 1)\frac{t}{2} \quad (3.12)$$

3.3.1 Transformada de Whiting

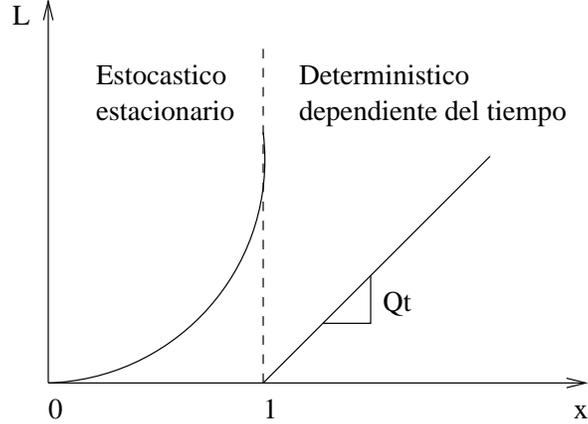
De las aproximaciones tradicionales al problema, hay dos que satisfacen parcialmente los requerimientos de modelación del fenómeno de formación y disipación de colas vehiculares. Una, la teoría estocástica en estado estacionario; otra, la determinística dependiente del tiempo. La primera reconoce la aleatoriedad, pero es estacionaria. Luego, solamente tiene sentido para $x < 1$ y no puede representar la evolución de la cola en períodos cortos. La segunda incorpora el efecto de la variación en el tiempo del flujo y la capacidad, pero al ser determinística sólo surgen colas cuando $x > 1$; el impacto de las fluctuaciones aleatorias es ignorado.

Las predicciones de ambos enfoques sólo coinciden cuando $t \rightarrow \infty$, ver Figura 3.1. En ese caso, la función determinística dependiente del tiempo coincide con la asíntota $\{x = 1\}$ de la función estocástica estacionaria.

El modelo correcto debe provenir de una teoría estocástica dependiente del tiempo. Sin embargo, ésta conduce a expresiones muy complejas, poco tratables.

Whiting propuso una solución práctica, basada en que para grados de saturación bajos, el efecto del tiempo es poco significativo y para grados de saturación altos el efecto de las fluctuaciones aleatorias es despreciable. Entonces tenemos teorías sencillas que reproducen bien la realidad para los

Figura 3.1: Evolución longitud de cola



extremos del rango de variación del grado de saturación. Entonces, **trasladar la curva estacionaria de manera que tenga como asíntota la función determinística y no $\{x = 1\}$, es una forma razonable de introducir la variación aleatoria y la variación en el tiempo del flujo y la capacidad.**

Sea f_D y f_E la expresión determinística y estacionaria de alguna de las variables de interés (L, D, d) y sea f_T la nueva función estacionaria cuya asíntota es f_D , y que denominaremos función transformada. Entonces, a partir de un análisis geométrico, es posible demostrar que la curva transformada es el resultado de la siguiente transformación de coordenadas:

$$f_T = f_E(x + 1 - f_D^{-1}(f_T)) \quad (3.13)$$

Dada la forma funcional de f_E y f_D , la expresión anterior conduce a una ecuación de segundo grado en f_T cuya forma genérica es

$$f_T^2 + U f_T - \frac{V}{4} = 0$$

y cuya única solución con sentido físico está dada por

$$f_T = \frac{1}{2} (\sqrt{U^2 + V} - U) \quad (3.14)$$

Para cada variable de interés, U y V adoptan valores distintos. Ambas son función explícitas de x y t , luego $f_T = f_T(x, t)$.

Tabla 3.1: Parámetros de la función transformada

Variable	U	V
L	$\frac{(1-x)(Qt)^2+(1-L_0)Qt-2(1-C)(L_0+Qxt)}{Qt+1-C}$	$\frac{4(L_0+Qxt)[Qt-(1-C)(L_0+Qxt)]}{Qt+1-C}$
D	$\frac{(1-x)(Qt/2)^2+(1-L_0)Qt/2-2(1-C)(L_0+Qxt/2)}{Qt/2+1-C}$	$\frac{4(L_0+Qxt/2)[Qt/2-(1-C)(L_0+Qxt/2)]}{Qt/2+1-C}$
d	$\frac{1}{2}[t(1-x) - 2\frac{L_0+1}{Q}] + \frac{C-1}{Q}$	$\frac{2}{Q}[t(1-x + xC) + 2(C-1)\frac{L_0+1}{Q}]$

Ahora bien, desde el punto de vista espacio-tiempo las variables que interesan en nuestro caso son las que incluyen al vehículo en servicio. Las otras tienen interés en sistemas en los que hay una clara distinción física entre las unidades en cola y las que están siendo atendidas (por ejemplo, en un puerto). En lo que sigue nos referiremos sólo a las primeras. El valor de U y V para dichas variables se muestra en la Tabla 3.1.

Lo atractivo de la técnica de transformación de coordenadas de Whiting es que logra incorporar en una sola expresión efectos aleatorios y de variación sistemática en el tiempo. Claro que este beneficio se obtiene mediante una ingeniosa aproximación en vez de una formulación exacta.

Un análisis sencillo de las expresiones de U y V muestra que éstas son sensibles a las unidades en que se mide el flujo y la capacidad. Como las colas están compuestas por vehículos y no por vehículos equivalentes o ADEs, ambas variables se deben expresar en unidades de veh/h . Por otro lado, como las expresiones estocásticas-estacionarias corresponde a un sistema con una estación de atención, el análisis de colas y demoras debe realizarse a nivel de pista.

3.3.2 No transitividad de la función transformada

Por el hecho de ser una aproximación y no una fórmula exacta, la función transformada carece de transitividad con respecto al tiempo. Esto significa que la elección del origen temporal de análisis determina la evolución de la longitud de cola o las demoras.

La no-transitividad tiene un efecto despreciable cuando el grado de saturación es moderado. Sin embargo, cuando el grado de saturación es alto

($x \geq 0,85$) el efecto es significativo. Esto es un problema serio pues la variación sistemática de q y/o Q se recoge identificando períodos en que estas variables se mantienen como tasa promedio internamente y sólo están sujetas a variación aleatoria. Si la forma de definir los períodos influye en los resultados, se introduce una inestabilidad indeseable.

Para resolver este problema, *Kimber y Hollis (1979)* proponen un método secuencial de corrección. La cuestión es elegir un nuevo origen temporal consistente que evite la no-transitividad. El método se basa en la siguiente reflexión: ¿Cómo puedo saber algo sobre la historia de una cola observando solamente lo que sucede en un período determinado?

Supongamos que se conocen, en cierto período, q , Q y L_0 . Por la forma en que se construye la función transformada, si en el pasado siempre hubieran existido las mismas condiciones se cumplirá $L_0 = \lim_{t \rightarrow \infty} L_T(t, x) = L_E(x) = x + Cx^2/(1 - x)$. En la práctica se pueden dar varias situaciones:

1. $L_0 = L_E(x)$, la cola está en equilibrio y es aceptable la hipótesis de que las condiciones han permanecido iguales a las actuales;
2. $L_0 < L_E(x)$, el grado de saturación previo tiene que haber sido menor y por lo tanto la cola estará creciendo en el período de análisis;
3. $L_0 > L_E(x)$, el grado de saturación previo tiene que haber sido mayor y por lo tanto la cola será decreciente en el período de análisis.

Por lo tanto, comparando la cola inicial con la cola estacionaria podemos tener un apoyo para elegir un origen temporal consistente: inventar un pasado que explica bien lo que observamos aunque no sea la historia "verdadera".

Entonces sea un cierto período $(0, t)$ en que el grado de saturación es x y la longitud de la cola inicial es L_0 . Bajo estas condiciones se pueden dar los siguientes casos:

1. Cola en equilibrio: $x < 1$ y $L_0 = L_E(x)$
No es necesario corregir el origen temporal ya que

$$L_T(x, t) = L_E(x) \tag{3.15}$$

2. Cola creciente: $x < 1$ y $L_0 < L_E(x)$ ó $x \geq 1$
 El origen temporal se desplaza hacia la izquierda en una magnitud τ dada por

$$\begin{aligned} L_T(x, t = 0) &= 0 \\ L_T(x, t = \tau) &= L_0 \end{aligned}$$

Resolviendo se obtiene

$$\tau(L_0) = \frac{L_0 \left[L_0 + 1 - 2(1 - C)x + \sqrt{(L_0 + 1)^2 - 4L_0(1 - C)} \right]}{2Q \left[x - x^2(1 - C) - L_0(1 - x) \right]} \quad (3.16)$$

$$L_T(x, t) = L_T(x, t = t + \tau(L_0), L_0 = 0) \quad (3.17)$$

3. Cola decreciente: $x < 1$ y $L_0 > L_E(x)$
 Se supone que la formación y la dispación de la cola es simétrica. Luego, la dispación hacia la longitud de equilibrio ($L_E(x)$) se trata como un reflejo del caso anterior. Se tienen dos casos:

(a) $L_E(x) < L_0 \leq 2L_E(x)$

$$L_T(x, t) = 2L_E(x) - L_T(x, t = t + \tau(2L_E - L_0), L_0 = 0) \quad (3.18)$$

donde $\tau()$ corresponde a la expresión para la corrección del origen temporal del caso anterior evaluada en $2L_E - L_0$.

(b) $L_0 > 2L_E(x)$

En este caso el grado de saturación previo ($L_E^{-1}(L_0)$) era mucho mayor que el actual. Se asume entonces un proceso de descarga determinístico a tasa Q hasta llegar a $2L_E(x)$, después se aplica el mismo proceso anterior.

- i. La primera etapa tiene una duración τ' dada por

$$\tau'(L_0) = \frac{L_0 - 2L_E(x)}{Q(L_E^{-1}(L_0) - x)} \quad (3.19)$$

$$L_T(x, t) = L_0 - (L_E^{-1}(L_0) - x)Qt \quad 0 \leq t \leq \tau' \quad (3.20)$$

- ii. En la segunda etapa se tendrá

$$L_T(x, t) = 2L_E(x) - L_T(x, t = t - \tau'(L_0), L_0 = 0) \quad t > \tau' \quad (3.21)$$

De esta manera la predicción de la longitud de cola puede corregirse por el efecto de no-transitividad. Como se trabaja para períodos sucesivos, cada cambio de período implica redefinir el origen temporal (a menos que la cola esté en equilibrio) basándose en la longitud estimada para el instante final del período anterior, que constituye el punto inicial para el nuevo período.

El cambio de origen temporal también tiene consecuencias en la estimación de demoras. En efecto, por definición la tasa media de demora en el período $0-t$ es $D_{0-t} = \int_{z=0}^t L(z)dz/t$, pero con el cambio de origen ésta se transforma en $D_{0-t} = \int_{z=\tau}^{t+\tau} L_T(z)dz/t$.

Para resolver esta integral ya no nos sirve la fórmula que vimos anteriormente. Es necesario integrar numéricamente. Para la demora media es complicadísimo desarrollar expresiones transitivas.

Sin embargo, se ha encontrado que se obtiene una buena aproximación de ambas variables si se emplea el método transitivo para estimar la longitud de cola (y de ahí L_0 en cada período) y para las demoras (tasa media o demora media) se usan las fórmulas no-transitivas con la secuencia de longitudes iniciales corregidas. Una ventaja adicional de este procedimiento es que evita recurrir a integración numérica.

Esto vale en la medida que se usen períodos razonablemente cortos, especialmente para $x \geq 0,85$. Cuando el grado de saturación es alto se recomienda usar períodos de 5 a 10 minutos de duración, pudiendo extenderse a 15 y hasta 30 minutos para grados de saturación bajos, $x < 0,5$ por ejemplo.

La aplicación secuencial de este método para períodos de diferente grado de saturación permite representar los procesos de formación y disipación de las colas.

Nótese que seguimos requiriendo pocos datos:

1. $q(t)$ y $Q(t)$;
2. C (parámetro proceso llegada y salida), puede suponerse igual a uno si no hay semáforos próximos u otras particularidades, tales como indicios de apelonamiento;

3. L_0 para el primer período (recomendación: iniciar predicción en período con bajo grado de saturación previo de modo que sea razonable suponer $L_0 = L_E(x)$).

Desde el punto de vista práctico es importante insistir en que este método permite predecir L , D y d por pista. Como normalmente no se tienen datos de flujo por pista sino por movimiento, hay que definir primero los períodos para los flujos globales (para toda la intersección) y luego estimar q y Q por pista en cada período. Ahora puede comprenderse la importancia de disponer de métodos para calcular la capacidad por pista.

Por otro lado, es necesario recordar que la demora media se refiere a los **vehículos llegados** en $0 - t$ y D al período $0 - t$ mismo, por lo cual su evolución temporal muestra rezago; así mismo con respecto a $L(t)$.

3.3.3 Funciones transformadas para intersecciones semaforizadas

El método que hemos visto permite abordar cualquier fenómeno de colas en un sistema con una estación de atención si conocemos el flujo, la capacidad y el parámetro C del proceso de llegada y salida. Conocer este parámetro en el caso de una intersección semaforizada es difícil pues la varianza del intervalo de salida depende, entre otros factores, de la programación del semáforo. Luego, si se quiere modificar la programación, no se podrá conocer C a priori para estudiar los impactos de la modificación. Por esta razón el análisis del proceso de formación y disipación de colas en intersecciones semaforizadas se realiza de forma distinta, evitando la necesidad de conocer C .

El método de análisis se basa en una formulación explícita del funcionamiento del semáforo y en el supuesto de que las llegadas de los vehículos son aleatorias (intersección aislada). Se configura así un sistema $M/S/1$ donde ' S ' se refiere a la distribución del proceso de salida de los vehículos en un acceso semaforizado. Para este sistema vamos a encontrar funciones estocásticas-estacionarias y determinísticas dependientes del tiempo específicas.

Una particularidad relevante del funcionamiento de un semáforo es la existencia cíclica de períodos de rojo, en que la capacidad es nula, y verde, en que

$Q = S$. Por ello en cada ciclo se forman una cola durante el período de rojo que se descarga completamente en el período de verde, si existe capacidad suficiente. De no ser así, quedarán vehículos en cola para el ciclo siguiente.

Vemos entonces que en la evolución de la cola es posible distinguir dos procesos:

1. Uno de carácter cíclico, en el cual se forma cola durante el período de rojo que luego se descarga completamente durante el período de verde, siempre que no haya saturación ($q < Q$);
2. Otro que representa la evolución de la cola de un ciclo a otro si es que hay saturación ($q > Q$).

Sabemos que la saturación se puede producir por dos razones:

1. Debido a fluctuaciones aleatorias del grado de saturación. En este caso sólo en algunos ciclos se producirá saturación y en otros no;
2. Debido a sobresaturación sistemática. En este caso la cola crecerá ciclo a ciclo.

Debido a estas distinciones se definen dos componentes aditivas de la cola y las demoras:

1. Uniforme: esta componente recoge la variación de la cola y las demoras dentro de un ciclo, la definición de esta componente supone llegadas determinísticas y grado de saturación menor que uno: $D/S < 1$ con $x < 1$. La componente uniforme siempre es no nula;
2. Excedente: esta componente recoge la variación de la cola y las demoras producto de:
 - (a) Fluctuaciones aleatorias
 - (b) Sobresaturación sistemática

Corresponde a un sistema $M/S'/1$ con $x \leq 1$. Esta componente puede ser nula cuando el grado de saturación es bajo.

La componente uniforme evoluciona de la misma forma dentro de cada ciclo, creciendo y disipándose ya que el grado de saturación es determinístico y menor que uno. La componente excedente evoluciona de ciclo en ciclo y se define para un instante preciso de él: final del verde efectivo (o inicio del rojo efectivo; es igual). Si en ese instante queda algún vehículo retenido, como la cola uniforme se ha disipado en ese mismo instante, por definición ese vehículo forma parte de la cola excedente.

En resumen para predecir la evolución de la cola en una intersección semaforizada, el marco será:

- Distinguir períodos donde el flujo y la capacidad son constantes;
- En cada uno de estos períodos se analizan las componentes uniforme y excedente de la cola y las demoras
- Sólo podrían transferirse efectos de un período a otro a través de la componente excedente, pero al final del período de análisis se supone que no queda cola de tipo determinístico.

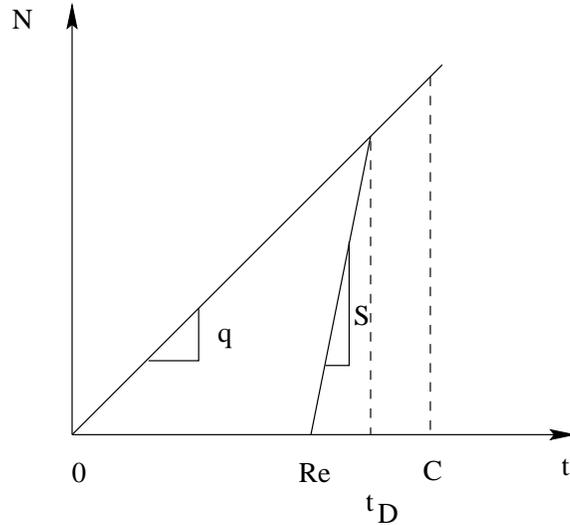
Por su definición, la componente uniforme corresponde a un sistema determinístico sin sobresaturación y que sólo depende del tiempo dentro de un ciclo. Tal sistema no requiere una función transformada. Sin embargo, la componente excedente es estocástica y depende del tiempo, por lo tanto, requiere de una función transformada similar a la vista anteriormente.

Para obtener las funciones transformadas necesitamos las expresiones estocásticas estacionarias y determinísticas dependientes del tiempo para cada variable, notando que la componente uniforme corresponde a un sistema diferente del de la excedente.

Las expresiones que veremos a continuación son válidas cuando se trata de:

- Semáforos aislados: es razonable suponer llegadas aleatorias;

Figura 3.2: Llegadas y salidas en una intersección semaforizada no saturada



- Los semáforos operan con una programación que se mantiene constante ciclo a ciclo dentro del período de análisis (planes prefijados);
- El flujo y la capacidad son constantes dentro del período y sólo están sujetos a variaciones aleatorias.
- La pista analizada tiene un sólo período de verde en cada ciclo.

3.3.3.1 Funciones estocásticas estacionarias

Las funciones estocásticas estacionarias corresponden en este caso a un sistema $M/S/1$ con $x < 1$ y M : Poisson. Como el grado de saturación es menor que uno, sólo puede haber componente excedente por fluctuaciones aleatorias. La situación se ilustra en la Figura 3.2. Tenemos entonces:

1. Longitud de cola

- (a) Componente uniforme: $D/D/1$ con $x < 1$

Esta componente solamente varía dentro del ciclo. Por definición es nula al comienzo del rojo efectivo, a partir de ese instante crece

hasta que se inicia el verde efectivo, momento en el comienza a descargarse la cola hasta que se disipa. Nótese que este sistema no corresponde a uno $D/D/1$, ya que en ese tipo de sistema cuando $x < 1$ la cola es nula. En este caso se tiene en cambio (ver fig. 3.2):

$$LU_E = \begin{cases} qt & \text{si } 0 \leq t \leq R_e \\ \max\{0; qR_e - (S - q)(t - R_e)\} & \text{si } R_e < t \leq C \end{cases} \quad (3.22)$$

(b) Componente excedente: $M/S/1$ con $x < 1$

Esta componente se calcula para el final del verde efectivo. En ese instante, por los supuestos utilizados, nadie está en servicio. Como ya explicamos anteriormente, un semáforo es un dispositivo muy particular en el que es difícil encontrar una expresión teórica estricta que corresponda al sistema $M/S/1$ con $x < 1$. Diversos autores han desarrollado expresiones aproximadas para la longitud de cola en este contexto. La forma genérica de estas expresiones puede escribirse como:

$$N_E = \begin{cases} 0 & \text{si } x \leq x_0 \\ \frac{k(x-x_0)}{1-x} & \text{si } x > x_0 \end{cases} \quad (3.23)$$

donde x_0 y k son parámetros de modelación cuyos valores se muestran a continuación para los modelos más conocidos:

Autor	x_0	k
Webster	0	$x/2$
Mc Neil	0	$1/2$
Akcelik	$0,67 + SV_e/600$	$3/2$
Rouphail	0,5	$1,22(SV_e)^{-0,22}$

2. Tasa media de demora

(a) Componente uniforme: $D/D/1$ con $x < 1$

Por definición $D = \Delta/C$ donde Δ es la demora total y C es el tiempo de ciclo del semáforo, luego es necesario calcular primero la demora total. En la fig. 3.2 se puede ver que la demora total

en un ciclo cualquiera está dada por $\Delta = R_e(qt_D)/2$, donde t_D corresponde al tiempo que tarda en descargarse la cola formada durante el rojo efectivo. Este tiempo se calcula imponiendo la condición: $LU_E(t = t_D) = 0$, el resultado es $t_D = SR_e/(S - q) = (C - V_e)/(1 - q/S)$. Con este valor, la demora total resulta igual a:

$$\Delta U_E = \frac{qC^2(1 - u)^2}{2(1 - y)} \quad (3.24)$$

donde $y = q/S = ux$ es el factor de carga. Por lo tanto, se obtiene

$$DU_E = \frac{qC(1 - u)^2}{2(1 - y)} \quad (3.25)$$

(b) Componente excedente: $M/S'/1$

La demora total excedente se obtiene simplemente multiplicado la longitud promedio de cola por el tiempo de descarga: $\Delta N_E = N_E t_D$, es decir,

$$\Delta N_E = \frac{C - V_e}{1 - y} N_E \quad (3.26)$$

de donde resulta una tasa media de demora igual a

$$DN_E = \frac{1 - u}{1 - y} N_E \quad (3.27)$$

3. Demora media por vehículo

Como estamos analizando un sistema estacionario es válida la relación: $d = D/q$. Por lo tanto, se obtiene a partir de las expresiones anteriores:

(a) Componente uniforme: $D/D'/1$ con $x < 1$

$$dU_E = \frac{C(1 - u)^2}{2(1 - y)} \quad (3.28)$$

(b) Componente excedente: $M/D'/1$ con $x < 1$

$$dN_E = \frac{1 - u}{q(1 - y)} N_E \quad (3.29)$$

Es interesante mencionar que como $y = ux$, la componente uniforme es la misma si se calcula por pista o por grupos de pistas

si se aplica el criterio de equisturación para encontrar los flujos y la capacidad a nivel de pista. Es decir, la componente uniforme es válida para un sistema $D/S'/n$ cuando las n pistas satisfacen las condiciones referidas. Sin embargo, la componente excedente depende de q y, por tanto, su valor para cada pista (q por pista) o para el conjunto (q agregado) será distinto. La componente excedente sólo empieza a tener valores significativos a partir de $x \geq 0,85$. Por lo tanto, estimar la demora para un conjunto de pistas de uso compartido en vez de una a una conduce a errores de importancia sólo a partir de grados de saturación altos.

3.3.3.2 Funciones determinísticas dependientes del tiempo

Al ser un sistema dependiente del tiempo debemos discretizar la variación del flujo y capacidad definiendo períodos en los que estas variables son constantes para simplificar el tratamiento analítico de las colas y demoras. Dado que los semáforos operan cíclicamente, los períodos habrán de referirse a un cierto número de ciclos. El cambio de período estará ligado entonces a una variación del valor medio del flujo o la capacidad, pero además, como la programación del semáforo se hace considerando el nivel de flujo, el cambio de período estará asociado también a una variación de la programación.

Existiendo sobresaturación sistemática ($x \geq 1$) es evidente que los vehículos que llegan en un período no se podrán descargar en él: habrá cola excedente que hereda el período siguiente.

Sin embargo, en el caso de intersecciones semaforizadas esta es una situación poco probable en la realidad. De haber sobresaturación, seguramente será muy transitoria pues, producto de ella, disminuirá el flujo por reasignación de éste. Por otro lado, típicamente existe un programa que se mantiene durante todo un período punta (en planes prefijados) y al terminar éste, la cola se ha disipado o, cuando menos, reducido significativamente.

Estas consideraciones nos llevan a modificar el marco temporal. **En vez de pensar en períodos breves (5 a 10 minutos como vimos antes) estaremos hablando de lapsos del orden de 1 hora o más.** Luego, la asíntota determinística sobresaturada juega más bien un papel de referencia

antes que una asíntota propiamente tal, porque es difícil que se mantenga $x \geq 1$ durante toda una hora.

Es decir, se aplicará la función transformada pero para $x < 1$, aunque posiblemente cercano a 1. Por idénticas consideraciones, en las aplicaciones reales no habrá colas heredadas (o serán muy pequeñas) y el proceso de estimación de las características de la cola no se hará encadenando los resultados de cada período, cada uno se analizará por separado. Esto evita también la aparición del problema de no transitividad de la función transformada. Sin embargo, queda un tema a analizar en mayor profundidad: el efecto de que el grado de saturación pueda variar dentro del período de análisis, comportamiento típico de las *horas punta*.

Entonces para las asíntotas usaremos un período sin pasado (cola inicial nula) ni futuro (no llegan nuevos vehículos a su término). Pero ya que no hay futuro, la cola acumulada (excedente) al final tenemos que considerarla para una contabilidad apropiada de las demoras. Es decir, tomaremos los vehículos llegados en $0 - t$ y extenderemos ficticiamente el período hasta que se descarguen los vehículos en cola en el instante t , sin llegadas adicionales. En ese instante se habrán acumulado $(q - Q)t$ vehículos que para descargarse a tasa Q necesitan un lapso de tiempo $(q - Q)t/Q = (x - 1)t$. Considerando el tiempo discretizado en ciclos los vehículos llegados en un período de t/C ciclos requerirán adicionalmente $(x - 1)t/C$ ciclos para salir. La situación se muestra en la Figura 3.3.

A continuación se muestran las expresiones correspondientes a este caso.

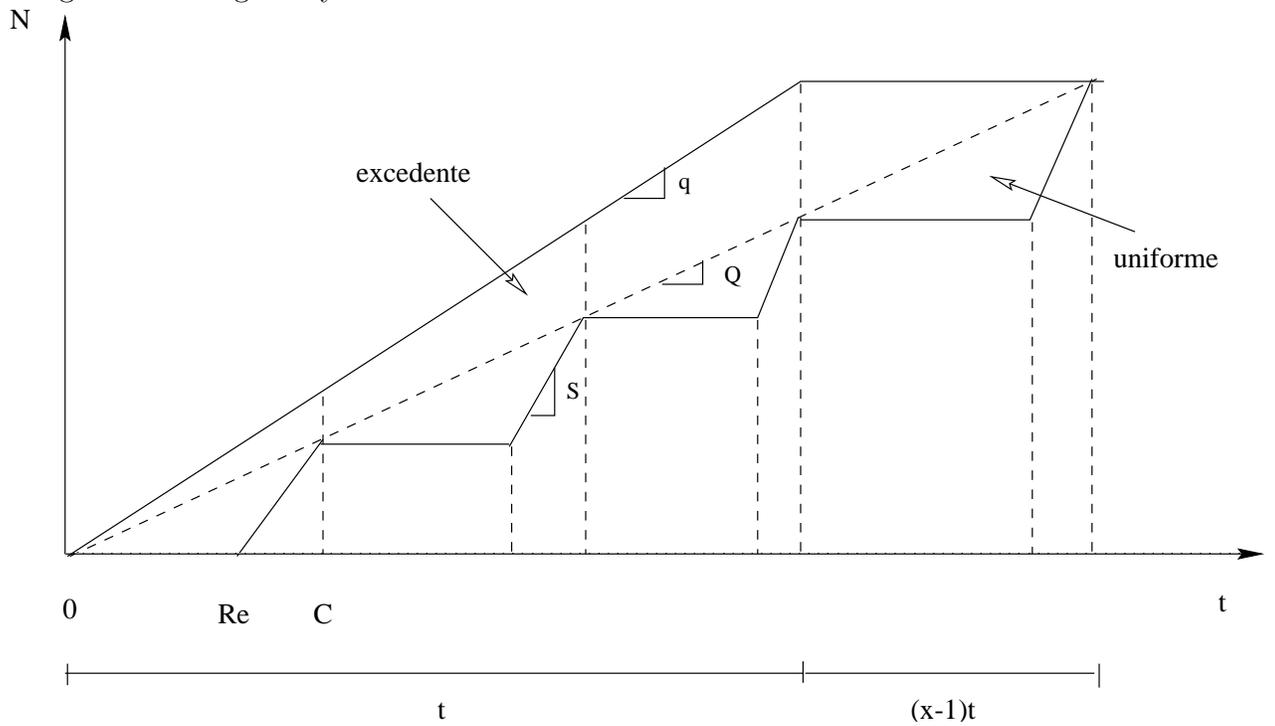
1. Longitud de cola

(a) Componente uniforme: $D/S'/1$ con $x = 1$

Por definición la componente uniforme sólo se hace cargo de la evolución de la cola hasta $x = 1$. Se utiliza entonces la expresión vista anteriormente, pero imponiendo $x = 1 \Leftrightarrow q = uS$:

$$LU_D = \begin{cases} uSt & 0 \leq t \leq R_e \\ S[R_e - t(1 - u)] & R_e < t \leq C \end{cases} \quad (3.30)$$

Figura 3.3: Llegadas y salidas en una intersección semaforizada saturada



(b) Componente excedente: $D'/D'/1$ con $x > 1$

Esta componente aumenta constantemente de ciclo en ciclo. Se calcula para el inicio del rojo efectivo. Para un ciclo i la cola excedente está dada por $N_i = N_0 + i(qC - SV_e)$. Como valor representativo del período $0 - t$ se utiliza el valor promedio sobre t/C ciclos, como suponemos que no existe cola heredada ($N_0 = 0$) el promedio resulta igual a

$$N_D = \frac{1}{2}(q - Q)t = \frac{x - 1}{2}Qt \quad (3.31)$$

2. Tasa media de demora

Por el tratamiento dado al tiempo, se calcula la tasa media de demora para los vehículos que llegan en $0 - t$ y no la producida en el período.

(a) Componente uniforme: $D'/S'/1$ con $x = 1$

$$DU_D = \frac{qC(1 - u)}{2} \quad (3.32)$$

(b) Componente excedente: $D'/D'/1$ con $x > 1$

$$DN_D = N_D x \quad (3.33)$$

3. Demora media por vehículo

(a) Componente uniforme: $D'/S'/1$ con $x = 1$

$$dU_D = \frac{C(1 - u)}{2} \quad (3.34)$$

(b) Componente excedente: $D'/D'/1$ con $x > 1$

$$dN_D = \frac{N_D}{Q} \quad (3.35)$$

Nuevamente se verifica aquí que la componente uniforme no es sensible a ser calculada para cada pista o para un grupo de ellas que tiene uso compartido bajo el supuesto de equisaturación, pero sí lo es la componente excedente.

3.3.3.3 Funciones transformadas

Como vimos en la sección anterior, cada variable tiene una componente uniforme y una excedente. Por su definición, la componente uniforme no requiere de una función transformada. Por el contrario, la componente excedente presenta un comportamiento en x que hace necesaria la especificación de una curva transformada. En resumen, se tienen las siguientes expresiones para la componente uniforme según el valor del grado de saturación.

Componente uniforme

Variable	$x \leq 1$	
LU	qt	$0 \leq t \leq R_e$
	$\max\{0; qR_e - (S - q)(t - R_e)\}$	$R_e < t \leq C$
DU	$qC(1 - u)^2 / (2[1 - y])$	
dU	$C(1 - u)^2 / (2[1 - y])$	

Variable	$x > 1$	
LU	uSt	$0 \leq t \leq R_e$
	$\max\{0; uSR_e - S(1 - u)(t - R_e)\}$	$R_e < t \leq C$
DU	$qC(1 - u) / 2$	
dU	$C(1 - u) / 2$	

Para la componente excedente tenemos que la función estocástica estacionaria es asintótica a $\{x = 1\}$ y la determinística se hace distinta de 0 a partir de $x = 1$. Por lo tanto, podemos usar la técnica de transformación de coordenadas de Whiting. Los resultados son los siguientes:

1. Longitud de cola

Se puede demostrar que la función transformada para la longitud de cola, a partir de la función estocástica y la determinística, está dada por:

$$N_T(x, t) = \begin{cases} 0 & x \leq x_0 \\ \frac{Qt}{4} \left[(x - 1) + \sqrt{(x - 1)^2 + \frac{8k(x - x_0)}{Qt}} \right] & x > x_0 \end{cases} \quad (3.36)$$

Esta función representa el valor promedio de la cola excedente en el período $0 - t$.

2. Tasa media de demora

Vimos que la componente estocástica de esta variable es $DN_E = \frac{1-u}{1-y}N_E$ y que la componente excedente es $DN_D = N_Dx$. Si hacemos la aproximación $\frac{1-u}{1-y} \simeq x$, ambas componentes tendrán la forma $D = Nx$ y en consecuencia la función transformada estará dada por:

$$DN_T = N_Tx \quad (3.37)$$

3. Demora media por vehículo

Utilizando el resultado anterior, se obtiene fácilmente:

$$dN_T = \frac{N_T}{Q} \quad (3.38)$$

En síntesis, tenemos que la componente uniforme tiene un cierto valor para $x < 1$ que a partir de $x = 1$ se estabiliza en la cantidad que se obtiene al sustituir q por uS en la expresión correspondiente. La cola uniforme varía al interior de un ciclo, creciendo y disipándose.

La componente excedente de la longitud de cola y las demoras es nula hasta que $x = x_0$ y a partir de ahí crece asintóticamente a N_D , variando de un ciclo a otro. En dicha componente puede distinguirse una parte asociada a sobresaturación sistemática (N_s) y otra saturación de origen aleatorio (N_a). La primera está dada por:

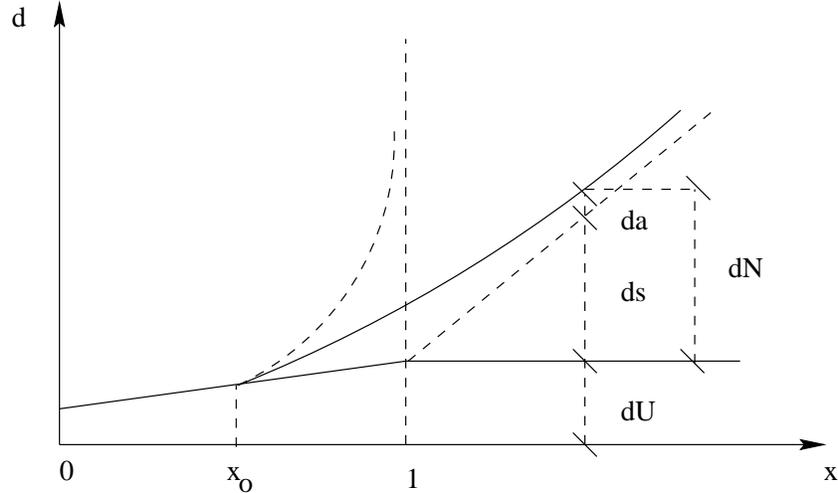
$$N_s = \begin{cases} 0 & x \leq 1 \\ N_D = \frac{x-1}{2}Qt & x > 1 \end{cases} \quad (3.39)$$

La componente aleatoria será entonces:

$$N_a = \begin{cases} 0 & x \leq x_0 \\ N_T & x_0 < x \leq 1 \\ \frac{Qt}{4} \left[-(x-1) + \sqrt{(x-1)^2 + \frac{8k(x-x_0)}{Qt}} \right] & x > 1 \end{cases}$$

La evolución de estas componentes para la demora media se muestra en la Figura 3.4.

Figura 3.4: Componentes de la demora media por vehículo



Vemos que hasta x_0 sólo hay demora uniforme, la que crece de manera casi lineal con x . A partir de x_0 aparece una demora por fluctuaciones aleatorias (d_a) que crece rápidamente, haciéndose mayor que dU para x en el entorno de $0,9$. Para $x \geq 1$ aparece la componente por sobresaturación d_s , la que aumenta fuertemente con x , en tanto d_a va disminuyendo y dU se mantiene constante en el valor correspondiente a $x = 1$. La componente excedente dN reúne a d_s y d_a , y por eso algunos la llaman demora aleatoria y por sobresaturación dAS . El comportamiento descrito explica que la llamada congestión empiece a percibirse para $x \geq 0,9$ y se agudice enormemente para x mayores donde pequeñas variaciones del flujo tienen gran impacto en la demora.

Revisemos ahora los supuestos básicos que impusimos para estudiar la formación y disipación de colas en intersecciones semaforizadas:

1. Llegadas aleatorias: este supuesto no se cumple si hay semáforos cercanos aguas arriba, caso en que las funciones estocásticas que utilizamos pierden validez. La cercanía de otro semáforo es de gran importancia para la cola y la demora uniforme, cuya variación dentro del ciclo es muy sensible a la forma de las llegadas. En cambio, la componente excedente, que varía de ciclo en ciclo, es casi insensible a ese factor.

Para esta componente es determinante el grado de saturación. Entonces, si hay semáforos en red la coordinación entre ellos puede reducir sustancialmente dU , pero ello sólo constituye un beneficio considerable cuando el grado de saturación es bajo. Metodológicamente el análisis de una red de semáforos se realiza de la siguiente forma:

- (a) La congestión uniforme se obtiene por simulación (*TRANSYT*, por ejemplo) ya que no hay métodos analíticos todavía para el complejo problema de patrones de llegadas interdependientes (en EE.UU hay fórmulas para intersecciones aisladas con llegadas en pelotón, usando un factor de progresión tabulado que varía con x , tipo de control y grado de apelsonamiento).
 - (b) Para la componente excedente, se aplican las mismas fórmulas, pero se admite que la existencia de coordinación reduce la varianza de las llegadas y se recomienda reemplazar $8k$ por $4k$ en la fórmula de N_T .
2. Operación con planes prefijados: si los semáforos funcionan activados por el tránsito la capacidad podrá variarciclo a ciclo. No obstante se ha encontrado que no se comete mayor error si se estiman las características de la cola como si fueran de planes prefijados, si el plan respectivo se determina para los flujos promedio del período. De hecho, al acercarse a grados de saturación altos se tenderá a una operación análoga a la de planes prefijados. Eso sí, puede lograrse ahorro de demoras significativo si se usa buena lógica dinámica, que reduzca aleatoriedad para semáforos aislados.
 3. Un período de verde en cada ciclo: si hay dos o más, la evolución de la cola dentro del ciclo se modifica, pero no cambia la evolución entre ciclos. Es decir, sólo es afectada la componente uniforme. Es sencillo extender el modelo para esta componente al caso de dos períodos de verde disjuntos.
 4. No hay herencia entre períodos y en cada período, q y Q son constantes: como vimos, para que este supuesto sea sostenible hay que definir los períodos adecuadamente, de modo que en lo posible la punta quede en su interior. Exige un tratamiento especial que veremos en el Capítulo 4, pues hay que reconocer que la punta implica que q en realidad no es constante en el período.

El método visto para analizar intersecciones semaforizadas tiene varias ventajas:

- Permite evitar el problema de obtener el valor del parámetro C del sistema asociado a la llegada y salida de los vehículos.
- Permite un análisis independiente de evolución de la cola en cada período, obviando correcciones por no transitividad aunque se utilice una función transformada para la componente excedente.

Una debilidad del método es que no es aplicable para la componente uniforme en redes de semáforos. Este problema lo estudiaremos en el Capítulo 5.

Capítulo 4

Estimación de demoras y detenciones

4.1 Periodización y demoras en períodos punta

El método visto para estimar la demora en intersecciones semaforizadas se basa en un período de cierta duración, en el cual se supone que el flujo y la capacidad son constantes en promedio, sólo sujetos a fluctuaciones aleatorias. Este supuesto es razonablemente realista, salvo en las horas punta, en que se puede dar una variación sistemática de la demora en corto tiempo. Otro supuesto del método visto es que cada período se inicia sin cola remanente del anterior, lo que también puede ser violado si hay saturación.

Akcelik y Roupail (1993) muestran que en presencia de flujo variable y alto grado de saturación, las fórmulas vistas subestiman sustancialmente la demora media. Los autores sugieren evaluar dichas fórmulas pero referidas a un subperíodo de mayor flujo que el promedio. Si bien este nuevo enfoque mejora la predicción de la demora media, presenta algunos inconvenientes que pueden ser superados con el método desarrollado por Gibson (1995).

4.1.1 Modelo de flujo variable

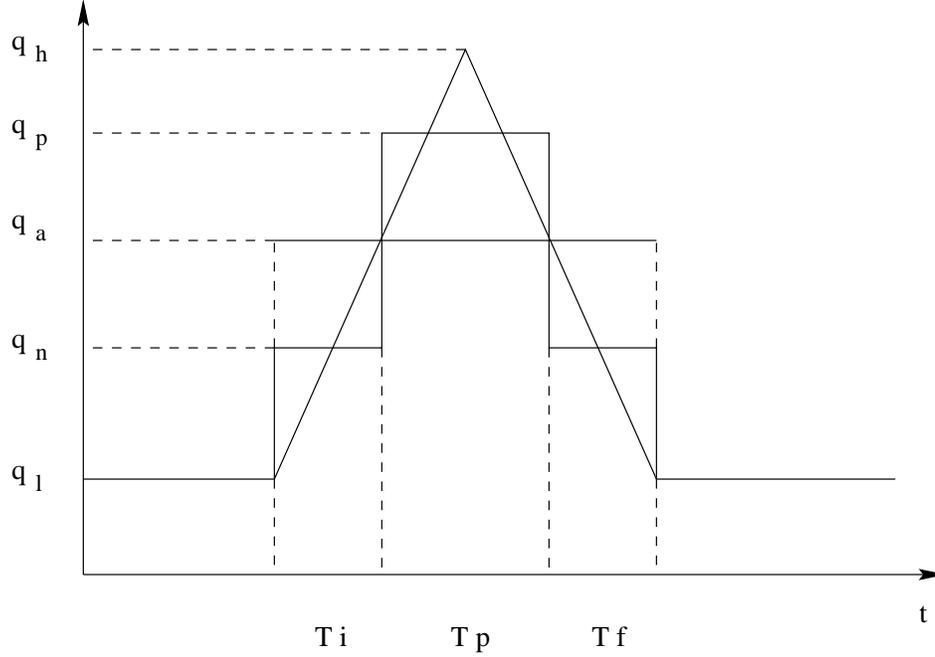
Sabemos que variables como el flujo y la capacidad están sujetas a variaciones aleatorias y sistemáticas. Esta última da origen a la diferenciación de períodos que implica el supuesto de constancia del flujo promedio al interior de cada uno de ellos. No obstante, se suele imponer a los períodos una duración mínima (típicamente, 1 hora) por razones prácticas de variada naturaleza. Si hay un patrón de variación sistemática dentro de este módulo mínimo, la periodización no dará cuenta de él. Cuando tal variación está acompañada de un grado de saturación alto, ignorarla lleva a subestimar la demora. Estas dos condiciones se dan a menudo en los períodos punta: el problema de la subestimación sólo es importante en ellos.

Como el funcionamiento del semáforo define una variación discreta de la capacidad, la variación sistemática del flujo en el período punta se modelará también de forma discreta. Se definirán entonces subperíodos dentro del período punta donde el flujo permanece constante en promedio. Para definir estos subperíodos, el patrón real de variación del flujo lo aproximaremos por una función triangular simétrica y ésta a su vez por una función escalón. Por lo tanto, se asume que antes y después del período punta, el promedio del flujo es el mismo. La situación se muestra en la Figura 5.1.

Como veremos a continuación, el método de Gibson consiste en calcular la demora media del período punta como una función de las demoras de los subperíodos definidos. En los subperíodos se aplica la fórmula tradicional de la demora media (uniforme y excedente), con algunos cuidados especiales para respetar los supuestos detrás de dicha fórmula y también para generar las condiciones que aseguren su cumplimiento en el período siguiente.

Supongamos entonces que tenemos un período de duración T , con flujo promedio q_a , pero sujeto a una variación sistemática dentro del período (una punta). Supondremos también que hay una capacidad constante Q en el período y que $q_a < Q$. Es decir, no habrá herencia determinística para el período siguiente. Si quedara en el último ciclo una cola excedente de origen aleatorio, no es necesario darle una consideración especial pues la teoría estocástica-estacionaria (que comanda el valor de la demora para $x < 1$) supone la existencia de una cola excedente promedio en todos los ciclos incluyendo el primero.

Figura 4.1: Aproximación del patrón de demanda variable



Como se mencionó anteriormente, aproximaremos el patrón real de flujo en el período punta T por una función escalón, que conduce a tres subperíodos internos. Para fines de generalidad, supondremos que el patrón real se ajusta bien a una función triangular simétrica. En relación a la Figura 4.1, podemos ver que la aproximación debe cumplir las siguientes condiciones básicas:

$$T_i + T_p + T_f = T \quad (4.1)$$

$$q_n T_i + q_p T_p + q_n T_f = q_a T \quad (4.2)$$

A partir de estas condiciones básicas y la simetría de la aproximación, se deducen las siguientes relaciones.

$$q_n = (1 - z/4)q_a \quad (4.3)$$

$$q_p = (1 + z/4)q_a \quad (4.4)$$

$$T_p = T/2 \quad (4.5)$$

$$T_i = T_f = T/4 \quad (4.6)$$

$$z = (q_h - q_l)/q_a \quad (4.7)$$

$$= 2(1 - q_l/q_a) \quad (4.8)$$

donde q_n y q_p es el flujo promedio en el subperíodo inicial o final, y en el subperíodo final, respectivamente; q_h , q_l es el flujo máximo y mínimo en el patrón real; T_i , T_f y T_p es la duración del subperíodo inicial, final y punta, respectivamente; y z es una medida de la intensidad de la punta.

Dependiendo de las condiciones prevalecientes en el subperíodo punta, la forma en que se calcula la demora media en el período T es distinta.

1. Subperíodo punta no saturado: $x_p \leq 1$

En este caso es válida la metodología que vimos anteriormente. Se deduce fácilmente lo siguiente:

$$d_i = d_f = dU(x_n) + dN_T(x_n, T/4) \quad (4.9)$$

$$d_p = dU(x_p) + dN_T(x_p, T/2) \quad (4.10)$$

2. Subperíodo punta saturado: $x_p > 1$

En este caso, la cola excedente acumulada al final del subperíodo punta deberá descargarse en el subperíodo final y los vehículos llegados en éste deberán esperar que termine esa descarga para empezar a salir, con el consiguiente efecto en su demora. Esto hace necesario diferenciar dentro de T_f el período de descarga, que estará saturado. Su duración (T_D) está dada por $T_D Q = (q_p - Q)T_p + q_n T_D$ con $T_p = T/2$, es decir

$$T_D = \frac{x_p - 1}{1 - x_n} \frac{T}{2} \quad (4.11)$$

Bajo ciertos supuestos, la demora media de los vehículos llegados en el período de descarga es igual a la demora media de los llegados en el subperíodo punta y la de los que llegan en el resto del subperíodo final es igual a la de los llegados en el subperíodo inicial. Entonces se tiene:

$$d_i = d_{f''} = dU(x_n) + dN_T(x_n, T/4) \quad (4.12)$$

$$d_p = d_{f'} = dU(x_p) + dN_T(x_p, T/2) \quad (4.13)$$

$$\begin{aligned} T_{f'} &= T_D \\ T_{f''} &= T_f - T_D \end{aligned}$$

Para que la cola se descargue dentro del período final, debe cumplirse $T_D \leq T_f = T/4$, o bien

$$\frac{x_p - 1}{1 - x_n} \frac{T}{2} \leq \frac{T}{4} \Leftrightarrow x_p \leq \frac{3 - x_n}{2}$$

Además, de las expresiones de los flujos promedio se puede mostrar que

$$\begin{aligned} x_p &= (1 + z/4)x \\ x_n &= (1 - z/4)x \end{aligned}$$

donde $x = q_a/Q$ es el grado de saturación promedio. Al reemplazar estas expresiones en la desigualdad anterior, se concluye que las condiciones básicas para que este método sea válido se requiere que

$$z \leq 12 \frac{1 - x}{x} \quad (4.14)$$

$$x < 1 \quad (4.15)$$

Con las expresiones de la demora en cada subperíodo, la demora media en el período T se calcula como

$$d = \frac{d_i q_n T_i + d_p q_p T_p + d_f q_n T_f}{q_a T} \quad (4.16)$$

si llamamos $d_n = d_i = d_f$ y reemplazamos las expresiones de los flujos promedios en cada subperíodo, la ecuación anterior se transforma en las siguientes expresiones dependiendo del grado saturación en la punta.

1. Subperíodo punta no saturado: $x_p \leq 1$

$$d = d_p - (d_p - d_n) \frac{4 - z}{8} \quad (4.17)$$

2. Subperíodo punta saturado: $x_p > 1$

$$d = d_p - (d_p - d_n) \frac{4 - z}{4 + z \frac{x}{1-x}} \quad (4.18)$$

y en ambos casos la tasa media de demora resulta ser

$$D = dq_a \quad (4.19)$$

Para determinar más fácilmente cuándo es necesario aplicar este método detallado para estimar la demora media y cuál de las fórmulas anteriores utilizar, Gibson propone los siguientes criterios basados en el grado de saturación del período completo ($x = q_a/Q$) y derivados a partir de simulaciones:

1. Si $x \leq \frac{3,6}{4+z}$ utilizar la fórmula tradicional de d con el flujo promedio q_a
2. Si $\frac{3,6}{4+z} < x \leq \frac{4}{4+z}$ utilizar la ecuación del caso $x_p \leq 1$
3. Si $\frac{4}{4+z} < x < 1$ utilizar la ecuación del caso $x_p > 1$

La longitud de cola evoluciona dentro del período T . En el subperíodo inicial, la cola uniforme y excedente corresponden a x_n y $(x_n, T/4)$; en el subperíodo punta, corresponden a x_p y $(x_p, T/2)$; en el subperíodo final, dependen del valor de x_p . No tiene sentido pensar en la longitud promedio de cola en el período T , es más razonable pensar en su valor máximo, al final del período punta,

$$L_{max} \approx q_p R_e + 2N_T(x_p, Q, T/2) \quad (4.20)$$

En resumen, en cualquier caso la demora media en el período completo puede ser aproximada por la demora media en el subperíodo punta, que evidentemente es mayor que la que resulta de estimar la demora para el período en su conjunto con el flujo promedio. Es decir, ignorar la presencia de una punta dentro del período conduce a una subestimación de la demora, que puede ser muy importante. Ahora bien, si la punta es poco notoria ($z \rightarrow 0$) y/o el grado de saturación promedio es bajo, $x < 0,8$, se comete muy poco error con la estimación tradicional basada en un único período.

Finalmente, es importante mencionar algunas consideraciones propias de la aplicación de este nuevo método.

Sea $q(i)$ el patrón de flujo observado en un intervalo de duración t (máximo 15 minutos). El intervalo de comienzo i_0 y final i_f del período de análisis se define de forma que la punta quede en su interior. Si $n \geq 4$ es el número de intervalos en el período de análisis, entonces las variables básicas del método son

$$T = nt \quad (4.21)$$

$$q_a = \frac{1}{n} \sum_{i=i_0}^{i_f} q(i) \quad (4.22)$$

$$q_l = \frac{1}{6} \left[\sum_{i=i_0-3}^{i_0-1} q(i) + \sum_{i=i_f+1}^{i_f+3} q(i) \right] \quad (4.23)$$

$$z = 2(1 - q_l/q_a) \quad (4.24)$$

$$x = q_a/Q \quad (4.25)$$

El resto de las variables necesarias se obtienen a partir de las anteriores.

A continuación es necesario verificar si se cumplen las condiciones $x < 1$ y $z \leq 12\frac{1-x}{x}$. Si ambas se cumplen, se acepta la definición del período y se procede a calcular las demoras.

Si $x \geq 1$ y $z \leq 12\frac{1-x}{x}$ hay que extender la duración del período moviendo simultáneamente los límites i_0 e i_f . La extensión incorpora nuevos intervalos de menor flujo, por lo tanto, q_a disminuye y, con ello, lo hace x . Nótese que frente a este cambio, q_h se mantiene constante y q_l y q_a disminuyen, lo que genera un aumento de z que podría violar la cota máxima de esta variable. Sin embargo, al disminuir x aumenta dicha cota. Entonces no puede asegurarse que al extender el período, la condición sobre la agudeza de la punta se cumpla, afortunadamente es difícil que en casos reales ocurra este problema.

Si $x < 1$ y $z > 12\frac{1-x}{x}$ la solución puede ser reducir la duración del período (crece x y disminuye z).

Estos posibles cambios en la definición del período de análisis, afectarán a los períodos previo y posterior. En consecuencia, la periodización debe comenzar por los períodos punta.

Si bien el método descrito permite estimar adecuadamente la demora media en períodos punta, tiene el inconveniente de ser particularmente sensible al valor de z . Este parámetro es función de q_l , q_a y q_h , los dos primeros se calculan como promedios sobre cierta cantidad de intervalos, sin embargo, q_h es uno solo y por su carácter está sujeto a errores de medición mayores que cualquier otro flujo.

4.2 Demoras peatonales

En esta sección estudiaremos las demoras que experimentan los peatones al enfrentarse a una intersección. Al igual que en el caso de los vehículos, la intersección “ofrece” una capacidad ($peatones/h$) a una corriente peatonal que desea cruzar una de las ramas de la intersección. Esta capacidad y la demora asociada es distinta según el tipo de regulación existente. Estudiaremos entonces las demoras que experimentan los peatones en intersecciones semaforizadas y no semaforizadas.

4.2.1 Intersecciones semaforizadas

En un cruce peatonal en una intersección semaforizada los peatones cruzan cuando el semaforo asigna luz verde al movimiento peatonal. Haciendo una analogía con un movimiento vehicular es posible derivar el siguiente modelo simplificado para calcular el flujo de saturación del movimiento peatonal:

$$S = 3.600Wvd \quad [peat/h] \quad (4.26)$$

donde W es el ancho del cruce peatonal [m], v es la velocidad promedio de los peatones al cruzar [m/seg] y d es la densidad peatonal de la corriente [$peat/m^2$]. Valores típicos de estas variables son $W = 3[m]$, $v = 1,2[m/seg]$ y $d = 1,5[peat/m]$, por lo tanto, $S = 19.500[peat/h]$.

Se aprecia entonces que la capacidad, $Q = uS$, de un cruce peatonal es muy alta. De hecho, es difícil encontrar en la práctica flujos peatonales cuya magnitud sea del orden de la capacidad. En consecuencia, el grado de saturación del cruce será bastante bajo. Si para calcular la demora utilizamos el mismo enfoque que en el caso vehicular, entonces podemos limitarnos a calcular sólo la componente uniforme de dicha variable; la excente será nula o despreciable:

$$d = dU = \frac{C(1-u)^2}{2(1-ux)} \quad (4.27)$$

Pero si $x \rightarrow 0$ ($q \leq 500[\text{peat}/h]$) se tendrá finalmente

$$d = \frac{C(1-u)}{2} = \frac{R_e}{2} \quad (4.28)$$

En esta expresión, u es la razón de verde efectivo para los peatones, en la fase en que pueden cruzar. Se supone que disponen de un solo período de verde en el ciclo.

4.2.2 Intersecciones no semaforizadas

Cuando los peatones no disponen de un cruce formal con prioridad y deben atravesar una corriente vehicular, estimar capacidad es prácticamente imposible. En cambio, se estima directamente la demora, basándose en un modelo de aceptación de brechas.

Igual que en el Capítulo 2, suponemos que la función de distribución del intervalo en la corriente vehicular prioritaria está dada por:

$$F(h) = \begin{cases} 0 & \text{si } h < \Delta \\ 1 - (1 - \theta)\exp(-\gamma(h - \Delta)) & \text{si } h \geq \Delta \end{cases}$$

con $\gamma = q_p(1 - \theta)/(1 - \theta q_p)$. Bajo este supuesto es posible demostrar que la demora media por peatón es:

$$d = E(d) = (\exp(\gamma\tau) - 1)\left(\Delta\mu + \frac{1}{\gamma}\right) - \tau + \frac{q_p\Delta^2}{2}\left(\mu + \frac{\sigma^2}{\mu}\right)$$

donde τ es la brecha crítica de los peatones, y μ y σ^2 son la esperanza y la varianza del tamaño del pelotón (número de vehículos en él) en la vía

prioritaria. Para desarrollar un modelo más práctico, Cowan (ver referencias del curso) asume que el tamaño del pelotón (m) sigue una distribución geométrica: $f(m) = (1 - \theta)\theta^{m-1}$, luego $\mu = E(m) = 1/(1 - \theta)$ y $\sigma^2 = \mu^2 - \mu$. Por lo tanto, la demora media se transforma en

$$d = (\exp(\gamma\tau) - 1)\left(\frac{\Delta}{1 - \theta} + \frac{1}{\gamma}\right) - \tau + \frac{q_p\Delta^2}{2}\left(\frac{2}{1 - \theta} - 1\right) \quad (4.29)$$

Se pueden dar algunos casos particulares:

1. Capacidad de la vía no acotada, ausencia de pelotones: $\theta = \Delta = 0$ (Modelo de Adams)

$$d = \frac{\exp(\tau q_p) - 1}{q_p} - \tau \quad (4.30)$$

2. Modelo lineal para la proporción de pelotones: $\theta = x_p = q_p\Delta$

$$d = \frac{\exp(\tau q_p) - 1}{(1 - \Delta q_p)q_p} - \tau + \frac{q_p\Delta^2}{2} \frac{1 + \Delta q_p}{1 - \Delta q_p} \quad (4.31)$$

Estas fórmulas, excepto la de Adams, sobreestiman la demora cuando el flujo peatonal se aproxima a la capacidad ya que son de carácter estacionario. Si el grado de saturación peatonal es alto, conviene estimar la capacidad y usar colas dependientes del tiempo. Aunque habrá poca precisión, el error cometido será muy inferior al asociado al uso de estas fórmulas.

En todos los casos, la tasa media de demora es aproximadamente, $D = qd$, donde q es el flujo peatonal.

4.3 Demoras en paraderos de buses

[Rodrigo Fernández]

4.4 Demoras totales

Con lo que hemos visto hasta aquí, podemos estimar las demoras que experimentan los usuarios al producirse colas en cualquier elemento individual de un sistema de transporte.

En efecto, tenemos que pueden surgir colas cuando:

1. En estaciones de transferencia con múltiples sitios, por fluctuaciones aleatorias de la tasa de ocupación (equivalente al grado de saturación) ya que es improbable que ellas experimenten sobresaturación sistemática. Disponemos de una teoría estocástica estacionaria para n sitios. El único problema puede ser en instalaciones de este tipo cuando hay alto x , en que aparece dependencia del tiempo. Caso significativo son los paraderos de buses, las plazas de peaje o aeropuertos con más de una pista, también puede haber algún puerto congestionado. En tales casos, no disponemos de modelos analíticos confiables y operacionales. Caben dos opciones: uso de modelos econométricos (útiles sólo para fines específicos por ser dependientes de condiciones de calibración) o de modelos de simulación, recurso mucho más empleado pero costoso.
2. En vías urbanas o carreteras, donde haya estrangulamientos o intersecciones no semaforizadas, podemos usar el método simplificado de la función transformada (que corrige la no-transitividad sólo de la longitud de cola), haciendo análisis pista a pista ya que la función se genera con sistemas $G/G/1$ y $D/D/1$. Hemos discutido cómo periodizar y el problema práctico puede ser determinar el valor apropiado del parámetro C del sistema, para el cual también propusimos una aproximación mediante valores característicos de las distribuciones de los intervalos de llegada y salida.
3. Si se trata de intersecciones semaforizadas, disponemos de un método apropiado que reconoce incluso la existencia de períodos punta. Está asociado a un método de definición de los períodos y a condiciones para la programación del semáforo y debe aplicarse pista a pista en la estimación de demoras.

4. En los cruces peatonales en que los peatones no tienen prioridad, tenemos también métodos especiales.

Es decir, está cubierto el espectro de situaciones y necesitamos muy pocos datos:

- Flujos, que pueden ser heterogéneos y sabemos tratar la heterogeneidad;
- Capacidades, para lo que disponemos de métodos que normalmente requieren de información sobre flujos y características geométricas y funcionales del elemento;
- Duración del período, y hemos visto como proceder a definir períodos;
- En ciertos casos, datos específicos: salvo semáforos, L_0 y el parámetro C , para lo que se han recomendado métodos de estimación; en un semáforo, su programación, que como dato es fácil de obtener en terreno. La cuestión real es que la programación se determina en función de las demoras y para ello hay métodos que se ven en el curso siguiente.

El gran obstáculo remanente es qué hacer cuando hay semáforos cercanos que inducen llegadas en pelotón, lo que afecta sensiblemente a la componente uniforme de las demoras en semáforos y al parámetro C en otro tipo de elementos. Esto es, cuando hay que mirar la circulación en una red no como una suma de elementos aislados, nuestros métodos son insuficientes. Pronto veremos cómo tratar este caso. No obstante, la componente excedente, la principal para grados de saturación altos, puede ser bien estimada con los métodos vistos.

Ahora bien, nuestro centro de atención son las trayectorias espacio-tiempo de los vehículos, los que hemos considerado compuestas por dos clases de etapas: en movimiento, caracterizada por una velocidad media dependiente del "propio" flujo (salvo en caminos de dos pistas) y detenido, caracterizado por una demora. Lo que hemos estudiado son las demoras generadas por colas pero no son éstas las únicas que se producen: puede haber detención y demora en ausencia de cola. Ejemplos sencillos: disco PARE o detención en

terminal para transferir carga o pasajeros. Necesitamos estudiar este tema para estimar realmente las demoras totales.

Nótese que el concepto de demora que usamos tiene connotaciones engañosas. Se podría pensar que la demora es el "tiempo perdido" en viajar. No es así pues:

- También se "gasta" tiempo en movimiento;
- Es discutible que el tiempo que permanece un vehículo haciendo labores de transferencia en un terminal sea "perdido".

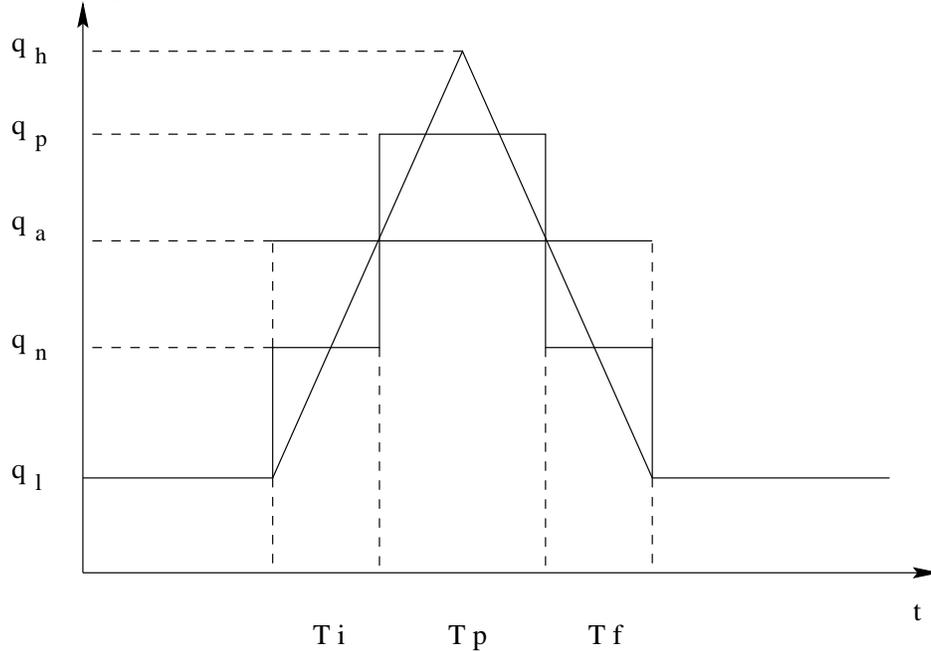
Simplemente, para nosotros demora tiene un significado operacional claro: es la duración de las etapas de detención en un viaje. Lo que le importa al usuario es el tiempo total empleado en viajar, que incluye tiempo en movimiento y detenido. Cuánto de él es "perdido" es una cuestión subjetiva. La forma en que se trata usualmente es considerar las diversas opciones que hay para hacer viajes y calcular las diferencias de tiempo empleado entre ellas. Así se puede hablar de "ahorros" de tiempo de manera objetiva. Estos ahorros pueden obtenerse en el tiempo en movimiento (mayor velocidad promedio: \bar{v}) y/o en demoras y pueden depender de longitud del recorrido y del número de detenciones, no sólo de \bar{v} y la demora en cada detención individual.

Volvamos ahora a las demoras en ausencia de cola. Como muestran los ejemplos mencionados, pueden existir en intersecciones o estrangulamientos y en estaciones de transferencia. Veremos cada caso separadamente.

En lo que se refiere a intersecciones, hay dos fuentes de demora en ausencia de cola:

- Detención, total o parcial, ante la presencia de un vehículo que goza de prioridad (o para verificar si él existe).
- Recorrido adicional por cambios geométricos de la intersección, que tiene sentido al comparar distintas maneras de organizarla. Es el típico caso de usar una rotonda en vez de una intersección convencional, de prioridad o semaforizada, o bien de hacer una intersección a desnivel.

Figura 4.2: Trayectoria espacio-tiempo en una intersección



El proceso general de un vehículo al atravesar una intersección puede esquematizarse mediante la Figura 4.2.

La demora que experimente el vehículo en referencia es el tiempo adicional respecto de si no hubiera intersección, es decir, pudiera recorrer el segmento AD a la velocidad que traía en A . Eso corresponde al segmento XY . Podemos distinguir diversas componentes de esta demora total:

- d_a : demora por desaceleración
- \tilde{d}_q : demora en cola
- d_m : tiempo de avance en la cola
- d_c : demora por chequeo de presencia de vehículo prioritario
- d_e : espera por presencia de vehículo prioritario, encabezando la cola
- d_a : demora por aceleración

A ella podría añadirse demora por recorrido excesivo si la trayectoria entre A y D se alarga por rediseño de la intersección.

4.4.1 Intersecciones no semaforizadas

La forma en que determinamos nuestra demora por cola (usando L y no L') nos asegura que en ella incorporamos \tilde{d}_q , d_m y d_e . El chequeo (d_c), propio de intersecciones no semaforizadas, está incluido sin duda si se usan modelos lineales de capacidad por ser éstos empíricos (tasa de salida observada). Es decir, la demora media estimada con la función cubre, en esas intersecciones: $d_q = \tilde{d}_q + d_m + d_c + d_e$.

Falta incorporar d_d y d_a . Además, debe notarse que los vehículos que no encuentran cola también tienen una demora por chequeo. Entonces, tenemos una demora media en cola, calculada como media para todos los vehículos que llegan en un período, a la que debe añadirse d_d y d_a para todos (haya o no cola, en la rama secundaria hay que asegurarse que no vienen vehículos prioritarios) y d_c , sólo para los vehículos que no encuentran cola. Se define entonces la demora geométrica como

$$d_g = d_d + d_a \quad (4.32)$$

Esto corresponde a una aproximación pues ignora d_c para aquella parte del flujo que no encuentra cola al llegar a la intersección, pero esto se compensa con una simplificación que se hace más adelante.

Luego, la demora media total se calcula como

$$d = d_q + d_g \quad (4.33)$$

y la tasa media de demora resulta ser

$$D = qd \quad (4.34)$$

Tabla 4.1: Demora geométrica para vehículos livianos (seg)

	Movimiento					
	Viraje derecha		Viraje izquierda		Directo	
Desde rama	Secundaria	Prioritaria	Secundaria	Prioritaria	Secundaria	Prioritaria
d_g	7,8	5,7	10,6	6,5	12,2	0

Para estimar d_g hay que considerar los procesos de frenado y aceleración. Esto implica una velocidad antes de llegar a la intersección; una que se logra superado el efecto de la intersección (que pueden ser iguales) y velocidades intermedias durante el cruce.

Ciertamente, si hay un proceso de chequeo antes de ingresar al cruce, se producirá un efecto sobre la referida velocidad intermedia, no imputable estrictamente a la demora geométrica ya que está más bien relacionada con fenómenos propios del proceso de colas. no obstante, si se ignora este hecho y se asocian simplemente las consecuencias de toda variación de velocidad a la demora geométrica, eso compensa no haber considerado d_c para los vehículos que no encuentren cola.

En la práctica, la estimación de d_g puede orientarse por dos vías:

1. Establecer modelos para calcularla como diferencia entre el tiempo real empleado en el tramo AD y un tiempo de referencia para recorrerlo, como si no hubiera intersección;
2. Análisis estadístico.

En Gran Bretaña se hizo un extenso análisis estadístico del fenómeno (ver Mc Donald, Hounsell y Kimber, 1984; Report SR 810, TRRL) y no se sabe de otro equivalente. A partir de sus resultados se recomiendan los valores que aparecen en la Tabla 4.1 para la demora geométrica de los vehículos livianos en intersecciones en T o en Cruz.

Los autores recomiendan algunas correcciones cuando las condiciones en la intersección no coinciden con las del estudio. A saber:

1. Si la velocidad de aproximación a la intersección es mayor que 65 km/h, agregar 2 segundos excepto al movimiento directo desde rama principal;
2. Si la visibilidad es pobre, agregar 1,4 excepto al movimiento directo desde rama principal;
3. Para vehículos pesados, la demora geométrica es un 25% más alta que la de los vehículos livianos para el mismo movimiento, excepto para el movimiento directo desde rama prioritaria.

Para otro tipo de intersecciones, rotondas e intersecciones a desnivel por ejemplo, puede consultarse la referencia citada.

La demora geométrica en intersecciones no semaforizadas es la componente dominante en la demora total para grados de saturación bajos ($x < 0,7 - 0,8$). En este sentido, es análoga a la demora uniforme en intersecciones semaforizadas, pero no varía con x .

4.4.2 Intersecciones semaforizadas

En el caso de intersecciones semaforizadas, la aceleración y deceleración están consideradas pues tanto en la capacidad (modelo binario) como en los modelos de demoras, hemos trabajado con trayectorias $s - t$ rectificadas (frenado y aceleración instantáneos). Sólo podría haber demora geométrica por recorrido excesivo, lo que parece improbable. Por lo tanto,

$$d = d_q \tag{4.35}$$

Entonces, en ausencia de cola, los vehículos no sufren demoras en una intersección semaforizada.

4.5 Detenciones

Las detenciones se originan en la existencia de cola. Para calcular su número el proceso es aparentemente sencillo: sumar las llegadas en los períodos en

que hay cola. Sin embargo, esto no considera la existencia de detenciones múltiples de un mismo vehículo. La utilidad de conocer la cantidad de detenciones es que a través de esta variable es posible estimar el consumo de combustible, frenos y otros recursos, adicional a la demora misma. También es un índice de peligrosidad. Veremos a continuación como se obtiene el número de detenciones en intersecciones semaforizadas y no semaforizadas.

4.5.1 Intersecciones semaforizadas

Aquí se identifican dos tipos de vehículos que deben detenerse: los que llegan mientras hay cola debida al período de rojo efectivo y los que quedan formando parte de la cola excedente. En el caso de semáforos aislados, para los que llegan durante el rojo efectivo la situación es la siguiente:

Luego, el número de vehículos que tiene que detenerse al menos una vez (componente uniforme) está dado por:

$$\begin{aligned} N_u &= qt_D \\ &= \frac{qC(1-u)}{1-y} \end{aligned}$$

Para los que quedan formando la cola excedente, el número de vehículos que se detiene es simplemente N_T . Despreciemos el efecto de la cola excedente en el siguiente ciclo, pero no fijemos la componente uniforme para $x > 1$, compensando.

Luego, el número de detenciones que se produce en un ciclo es $N_u + N_T$. Ahora, la tasa de detenciones por unidad de tiempo H :

$$\begin{aligned} H &= N_u + N_T \\ &= q \left(\frac{1-u}{1-y} + \frac{N_T}{qC} \right) \end{aligned}$$

y el número medio de detenciones h resulta

$$\begin{aligned} h &= \frac{H}{q} \\ &= \frac{1-u}{1-y} + \frac{N_T}{qC} \end{aligned}$$

Se ve que h puede ser mayor que 1 si N_T es grande (saturación alta). Ahora bien, es posible que por nuestra rectificación ideal de las trayectorias de los vehículos estemos suponiendo que se detienen completamente vehículos que sólo frenan sin llegar a parar.

Ocurre que a las detenciones les asignamos importancia por el consumo de combustible que implican los procesos de aceleración y deceleración. Al no ser paradas "completas" este consumo es menor y con las fórmulas anteriores lo sobrestimaríamos.

Se hace una conversión a "paradas completas", teniendo en cuenta que su probabilidad de ocurrencia está relacionada con la magnitud de la demora. Para condiciones típicas urbanas, un factor de reducción 0,9 es apropiado. Luego:

$$H = 0,9q \left(\frac{1-u}{1-y} + \frac{N_T}{qC} \right) \quad (4.36)$$

$$h = 0,9 \left(\frac{1-u}{1-y} + \frac{N_T}{qC} \right) \quad (4.37)$$

En el caso de semáforos en red coordinados, la simulación mediante histogramas cíclicos de flujos permite contar los vehículos que se detienen por cola uniforme y habrá que agregar N_T con su propia expresión. Vale aquí también el factor de reducción.

4.5.2 Intersecciones no semaforizadas

El problema de determinar el número de detenciones en este caso es más complejo. La cola no se forma y disipa con la regularidad característica de las intersecciones semaforizadas. Hay un avance discontinuo de la cola que probablemente conducirá a numerosas detenciones fraccionales.

Además, por la necesidad de comprobar si hay vehículos prioritarios, igual se producen detenciones (al menos fraccionales) en ausencia de cola.

Ambos fenómenos son difíciles de modelar y no existen hasta hoy métodos formales para estimar h y H en este caso. Entretanto haya investigación al respecto, se recomienda el procedimiento siguiente.

Los vehículos que no encuentran cola, sufrirán detención por comprobación de prioridad. Los que la encuentren, deberán además realizar detenciones durante el avance en la cola. Para los primeros, se puede suponer que:

$$h_g = \begin{cases} 1 & \text{si la regulación es PARE} \\ h_g(v, d_g) & \text{si la regulación es CEDA EL PASO} \end{cases} \quad (4.38)$$

donde h_g es una relación empírica para calcular el número medio de detenciones a partir de la velocidad y la demora geométrica. Para los que encuentran cola se puede utilizar como aproximación

$$h_q = h_g(v, d_q) \quad (4.39)$$

El promedio del número de detenciones se obtiene a partir de la probabilidad de ocurrencia de cada caso: encontrar o no encontrar cola. Sea $P = Prob(\text{encontrar cola}) = P(h \leq \tau) = F(\tau)$ donde τ es la brecha crítica de los conductores y F la función de distribución del intervalo en la rama prioritaria, entonces

$$h = (1 - P)h_g + Ph_q \quad (4.40)$$

Capítulo 5

Modelos de circulación en redes

5.1 Introducción

En los capítulos anteriores hemos estudiado la capacidad y los fenómenos de formación y disipación de colas en dispositivos viales AISLADOS, es decir, hemos supuesto que lo que sucede en otros dispositivos cercanos NO los afecta. Sin embargo, en realidad es muy probable que los fenómenos producidos en intersecciones o paraderos cercanos determinen en parte, por ejemplo, la longitud de cola en la intersección bajo análisis. Naturalmente esto ocurre con todos los dispositivos viales de una determinada área.

Por lo tanto, cuando se requiere predecir impactos globales derivados de acciones que modifiquen ciertas características de los dispositivos, es necesario una perspectiva más amplia que el simple análisis de conflictos aislados. La visión que integra lo que ocurre con el tráfico en un área es el análisis de circulación en redes.

En términos generales una red es un conjunto de líneas (arcos) que se conectan entre si en determinados puntos (nodos) y por donde circula algún elemento (flujo).

En nuestro caso las redes que representan la circulación vehicular describen los siguientes elementos:

- la infraestructura vial, a través de arcos que representan tramos de vía y de nodos que corresponden a las intersecciones;
- los flujos, mediante parámetros que caracterizan el comportamiento de los usuarios;
- la forma de operación, por parámetros que indican el funcionamiento de arcos y nodos.

La representación de la vialidad puede variar de profundidad según el objetivo que se persiga.

En general, los modelos de circulación en redes consideran las siguientes variables para caracterizar los distintos elementos de la vialidad:

- Nodos:
 - forma de regulación
 - capacidad
 - número de viajes generados/atraídos (para nodos centroides)
- Arcos:
 - flujo
 - velocidad/tiempo de viaje
 - longitud
 - flujo de saturación

Debido a la complejidad y magnitud de los problemas matemáticos involucrados en el análisis de redes se han desarrollado diversos programas computacionales para ayudar en su estudio. Algunos ejemplos son TRANSYT, SATURN, SCOOT, que consideran la red en su totalidad, y SIDRA, PICADY, ARCADY, IRENE que están orientados a la modelación de nodos (intersecciones, rotondas, paraderos). En general, los programas para redes analizan con mayor detalle los fenómenos que se producen en los arcos, por lo tanto, pueden ser completados con los programas orientados a nodos para conseguir un estudio mas preciso.

5.2 Modelación de elementos de redes

5.2.1 Modelación de arcos

Como las características de los arcos no son las mismas para todos los vehículos de un tramo se requiere tomar una decisión:

- promediar las características según la composición de tráfico
- definir arcos distintos para distintos flujos vehiculares

Por otra parte, al llegar a la intersección ciertos vehículos tienen comportamientos distintos que originan conflictos concurrenciales o direccionales. También existen ciertos vehículos no interactúan con el resto del tráfico. Para incorporar estas situaciones a la modelación se define un nuevo elemento, denominado línea de detención del arco (LD). La línea de detención actúa como una interfaz entre arcos y nodos con un flujo de saturación o capacidad distinta. Cada arco puede tener su propia línea de detención, pero también varios arcos pueden compartir una misma línea de detención. Mientras mayor es el número de arcos y líneas de detención, la modelación es más detallada, pero también es más compleja y costosa.

5.2.2 Modelación de nodos

Definidos los arcos y sus líneas de detención, al nodo le corresponde la modelación de la forma de regulación del dispositivo que representa. En el caso de nodos de tipo centroides, su objetivo es representar la demanda total de viajes en una determinada área.

5.2.3 Modelación de cuellos de botella

Un cuello de botella es un estrangulamiento de un tramo de vía producto de variaciones del ancho de la vía, estacionamientos, paraderos de buses, bloqueo temporal de pistas, etc.

Una forma de modelar los cuellos de botella es con un nodo que representa una intersección semaforizada con 100% de verde, pero con un flujo de saturación menor al del tramo.

5.2.4 Modelación de flujos

Los flujos de vehículos en la red se modelan desde dos perspectivas:

- desde el punto de vista de la conectividad de la red, especificando las contribuciones de movimientos al flujo total aguas abajo en un arco
- desde el punto de vista del nivel de flujos asociado a cada arco, en el sentido si son fijos o variables

5.2.5 Extensión de la red

La definición de la red a modelar tiene implícita la pregunta de hasta dónde considerar el área de influencia de acciones e impactos, y cómo incorporar los efectos de red mas allá de la zona de influencia.

La primera pregunta cae en el arte de la modelación de redes y se relaciona con los objetivos de cada modelación en particular. La segunda se resuelve definiendo nodos frontera y arcos de entrada a la red más allá de los cuales se supone que los efectos de red (por ejemplo, bloqueos) no se propagan.

5.3 Histogramas cíclicos de flujo y dispersión

Cuando existe una intersección semaforizada aguas arriba de una determinada intersección, también semaforizada, se produce una variación sistemática en el flujo que llega a esta última. Dicha variación se produce cíclicamente debido a la forma de operación del semáforo. Esto invalida el uso de las fórmulas anteriormente vistas para el cálculo de la componente uniforme de

la longitud de cola, ya que éstas se basan en el supuesto que el flujo sea constante durante cada ciclo. La componente excedente no se ve afectada por esta variación temporal del flujo.

Para modelar esta situación se utiliza la técnica de los histogramas cíclicos de flujo (HCF). Los histogramas son una representación detallada del flujo en el tiempo. Su objetivo es permitir el cálculo correcto de la componente uniforme de la longitud de cola cuando el flujo varía dentro de un ciclo típico del semáforo. Una vez calculada esta componente se puede obtener fácilmente la componente de los otros indicadores, tales como demora y detenciones.

A continuación se definen los tipos de histogramas que existen.

- Histograma de llegada: representa el flujo vehicular, en función del tiempo, que llega a una línea de detención durante un ciclo típico del semáforo. Se denomina $q_{LL}(t)$
- Histograma de largada: representa el máximo flujo, en función del tiempo, que puede salir de la línea de detención durante un ciclo típico del semáforo. Corresponde a una representación de la capacidad de la línea de detención. Se denomina $q_L(t)$.
- Histograma de salida: representa el flujo, en función del tiempo, que efectivamente sale de la línea de detención durante un ciclo típico del semáforo. Se denomina $q_S(t)$.

En todos los histogramas el tiempo de ciclo se divide en intervalos de corta duración (1 a 2 seg.) para así captar mejor la variabilidad del flujo.

De los tres tipos de histogramas, el de largada es el fácil de construir ya que basta con conocer el flujo de saturación asociado a la línea de detención y la programación del semáforo. Si por el momento suponemos conocido el histograma de llegada entonces la componente uniforme de la longitud de cola en el intervalo i está dado por:

$$LU(i) = \max \{LU(i) + I [q_{LL}(i) - q_L(i)]; 0\} \quad (5.1)$$

donde I es la longitud de los intervalos. El primer intervalo para el cual se calcula la longitud de cola es aquel donde comienza el rojo efectivo, para esto se supone que al final del verde efectivo la cola es nula. Lo anterior significa además que suponemos que no existe sobresaturación sistemática ($x > 1$). Si existiera sobresaturación entonces el histograma de llegada se corrige de la siguiente forma:

$$\tilde{q}_{LL}(i) = \alpha q_{LL}(i) \quad (5.2)$$

donde

$$\alpha = \frac{\sum_i q_L(i)}{\sum_i q_{LL}(i)} < 1 \quad (5.3)$$

El histograma de salida se obtiene como:

$$q_S(i) = \min \left\{ q_{LL}(i) + \frac{LU(i-1)}{I}; q_L(i) \right\} \quad (5.4)$$

Si existe sobresaturación entonces se utiliza \tilde{q}_{LL} en vez de q_{LL} .

Algunos descriptores interesantes que se obtienen a partir de las expresiones anteriores son:

- Tasa media de demora uniforme

$$DU = \frac{\sum_{i=1}^n LU(i)}{n} \quad (5.5)$$

donde n es el número de intervalos dentro del ciclo.

- Demora media uniforme

$$dU = \frac{DU}{\sum_{i=1}^n q_{LL}(i)} \quad (5.6)$$

- Número de detenciones uniforme

$$HU = \sum_{i \in A} q_{LL}(i) \quad (5.7)$$

donde A es el conjunto de intervalos en cuyo comienzo la longitud de cola es distinta de cero.

Para obtener el histograma de llegada a la línea de detención bajo análisis se necesita conocer el histograma de salida en la línea de detención aguas arriba y la forma en que se dispersa el pelotón de vehículos. Esto se realiza a través de los modelos de dispersión de pelotones. La dispersión de los pelotones se debe a las diferencias de velocidad que existen entre los vehículos: hay vehículos lentos y rápidos. Luego, para modelar la dispersión es necesario conocer la velocidad de circulación en el arco de los distintos vehículos, lo que es equivalente a conocer el tiempo de viaje ya que la longitud de recorrido es fija y conocida. Para incorporar la variabilidad de la velocidad o el tiempo de viaje se trabaja con funciones de distribución más que con valores determinísticos. La relación más elemental que describe la dispersión es entonces:

$$q_{LL}(i) = \sum_{j=1}^i q_S(j) P(t = i - j) \quad (5.8)$$

donde P es la función de distribución del tiempo de viaje en el arco. La ecuación anterior no considera el carácter cíclico de los histogramas, de ahí que una expresión más adecuada de la dispersión es:

$$q_{LL}(i) = \sum_{j=i-n+1}^i q_S(j) \sum_{z=0}^{\infty} P(t = i - j + zn) \quad (5.9)$$

Dependiendo de la forma de P se obtienen distintos modelos de dispersión.

1. Modelo de dispersión de Robertson

$$q_{LL}(1+T) = \frac{F}{1 - (1-F)^n} \sum_{j=2}^{n+1} q_S(j)(1-F)^{n+1-j} \quad (5.10)$$

$$q_{LL}(i+T) = Fq_S(i) + (1-F)q_{LL}(i+T-1) \quad i = 2, \dots, n \quad (5.11)$$

donde

$$F = \frac{1}{1 + \bar{t} - T} \quad (5.12)$$

2. Modelo de dispersión de Gibson y Witik

$$q_{LL}(1+T) = F \left\{ \sum_{i=1}^P (P+1-i)[q_S(i-P) + q_S(n+1-P-i)] + \sum_{i=1}^{P+1} q_S(n+i-P) \right\}$$

$$q_{LL}(i+T) = F \left\{ \sum_{j=0}^P q_S(n+i-j) - \sum_{j=P+1}^{2P+1} q_S(n+i-j) \right\} + q_{LL}(i+T-1) \quad i = 2, \dots, n$$

$$F = \frac{1}{(1 + \bar{t} - T)^2} \quad (5.13)$$

$$P = \bar{t} - T \quad (5.14)$$

F se conoce como factor de dispersión y T y \bar{t} corresponden al tiempo mínimo y promedio de viaje en el arco, respectivamente. El modelo de Robertson proviene de suponer una distribución del tiempo de viaje geométrica truncada y el modelo de Gibson-Aguirre de suponer una distribución triangular simétrica.

En ambos modelos se tiene la relación

$$T = \lfloor \beta \bar{t} + 0,5 \rfloor \quad (5.15)$$

para arcos de tráfico general, donde β es un parámetro a calibrar.

Para arcos de buses la relación es

$$T = \lceil \gamma \bar{b} + \sigma \bar{t}_r + 0,5 \rceil \quad (5.16)$$

donde \bar{b} es el tiempo neto de parada y \bar{t}_r el tiempo de recorrido: tiempo total de viaje $-\bar{b}$, y γ y σ son parámetros.

Valores típicos de los parámetros son

$$\beta = 0,60 - 0,80$$

$$\sigma = 0,80 - 0,85$$

$$\gamma = 0,30 - 0,60$$

Por otro lado, el tiempo se mide en intervalos (valores enteros) y los histogramas se construyen en unidades de vehículos equivalentes. Además, siempre existe continuidad en el flujo.

5.4 Indicadores de rendimiento

Los indicadores de rendimiento corresponden a medidas cuantitativas del consumo de recurso involucrado en la circulación de los usuarios por la red. Se construyen a partir de las variables estudiadas durante el curso. Los indicadores son importantes porque permiten evaluar el impacto de modificaciones en la red producto de proyectos de infraestructura y/o gestión de tránsito o bien ayudan en el diseño de dichos productos.

Conviene entonces distinguir dos tipos de indicadores: para diseño y para evaluación.

Indicadores para diseño:

- Longitud de cola: L
- Grado de saturación: x

- Demora media por vehículo: d
- Número medio de detenciones por vehículo: h

Indicadores para evaluación:

Lo primero es identificar cuales son los recursos que se desea cuantificar y luego asociarlos a variables físicas. Los recursos más comunes son tiempo de los usuarios y consumo de combustibles, otros menos utilizados son emisiones de contaminantes y accidentes.

En el caso del tiempo de los usuarios tenemos dos componentes:

- Tiempo en movimiento: $t_r \times T.O.$
- Tiempo detenido: $D \times T.O.$

donde $T.O$ es la tasa de ocupación por tipo de vehículo. Vemos aquí otra razón para definir arcos separados por tipo de vehículo.

En el caso del consumo de combustible tenemos tres componentes:

- Consumo en movimiento: $f(v)$ por tipo de vehículo
- Consumo en ralentí: $f(D)$ por tipo de vehículo
- Consumo por detenciones: $f(v, H)$

Estas tres componentes deben ser multiplicadas por el flujo respectivo para obtener el consumo total.

Vemos entonces que el consumo de recursos se realiza para cada arco y nodo de la red. Para obtener una única medida del consumo debemos agregar los consumos individuales. Esto se realiza mediante un indicador de rendimiento (IR) como el siguiente:

$$IR = \sum_i \sum_j V F_{ij} \left(\sum_k F C_{ijk} P R_{jk} \right) \quad \left[\frac{U.M}{U.T} \right] \quad (5.17)$$

donde

- VF_{ijk} : valor de la variable física i (demora, detenciones, velocidad, etc.) en el arco j
- FC_{ijk} : factor de consumo del recurso k por consumo de la variable i en el arco j
- PR_{jk} : precio del recurso k en el arco j

En el caso del tiempo de los usuarios el cálculo del consumo es elemental: está totalmente contabilizado y el problema es asociar una tasa de ocupación a cada arco. En el caso del combustible o emisiones de contaminantes el problema es más complejo. Sus tasas de consumo dependen de la operación real y nosotros utilizamos trayectorias idealizadas. Además, estas tasas varían con el tipo de vehículo, cuya diversidad puede ser amplia. Entonces hay que tener mucho cuidado con dos aspectos:

- Modelación agregada acorde con el tipo de trayectorias con que trabaja el modelo de red
- Encontrar valores representativos para los parámetros en arcos en que hay vehículos diversos

Construido el indicador de rendimiento, su utilidad es doble:

- Para evaluar económicamente un proyecto y comparar alternativas;
- Para influir directamente sobre el diseño de cada alternativa.

Ciertamente la red es una abstracción imperfecta de la realidad, hay que considerarla una ayuda a la toma de decisiones, pero no el instrumento para decidir.

Cuando no hay efectos de red, se realiza un análisis acumulativo, por separado, de tramos de vías e inte