

Formulación matemática de TRANUS

FORMULACIÓN MATEMÁTICA Y ALGORÍTMICA DE TRANUS	1
EL MODELO DE ACTIVIDADES	1
<i>Conceptos básicos</i>	<i>1</i>
<i>Demanda y distribución de la producción.....</i>	<i>4</i>
<i>Estructura del modelo de actividades.....</i>	<i>5</i>
<i>Incremento y localización de variables exógenas.....</i>	<i>5</i>
<i>Cálculo de atractores para la producción inducida.....</i>	<i>8</i>
<i>Generación de la demanda inducida.....</i>	<i>8</i>
<i>Localización de la producción inducida.....</i>	<i>10</i>
<i>Costo y desutilidad de consumo.....</i>	<i>12</i>
<i>Verificación de restricciones y ajuste de precios.....</i>	<i>13</i>
<i>Convergencia.....</i>	<i>13</i>
INTERFAZ ACTIVIDADES-TRANSPORTE.....	14
<i>Ecuación de transformación de los flujos.....</i>	<i>15</i>
EL SISTEMA DE TRANSPORTE	17
<i>Conceptos básicos</i>	<i>17</i>
<i>Estructura de los costos de transporte.....</i>	<i>24</i>
<i>Estructura del modelo de transporte</i>	<i>28</i>
<i>Construcción de pasos.....</i>	<i>29</i>
<i>Desutilidades y probabilidades.....</i>	<i>34</i>
<i>Restricción de capacidad.....</i>	<i>40</i>
<i>Oferta de Transporte Público.....</i>	<i>46</i>
<i>Convergencia.....</i>	<i>48</i>

Formulación matemática y algorítmica de TRANUS

Este capítulo está dedicado a presentar la formulación matemática de TRANUS como material de referencia para quienes deseen investigar su estructura interna de manera detallada.

Los principales componentes del sistema son los siguientes:

- Modelo de Actividades
- Interfase Actividades-Transporte
- Modelo de Transporte
- Interfase Transporte-Actividades
- Procedimiento de Evaluación

El modelo de actividades

El objetivo del modelo de localización de actividades es simular un sistema económico-espacial. Dada una región o ciudad dividida en zonas, el modelo simula las actividades que se localizan en cada zona y las interacciones que se generan entre ellas.

Conceptos básicos

El elemento central del modelo de actividades lo constituye un modelo de insumo-producto espacial definido por sectores económicos y sus relaciones de producción y consumo.

La estructura clásica de un modelo de insumo-producto, con sus elementos de demanda final, demanda intermedia e insumos primarios, se toma como punto de partida. El vector de demanda final representa a los destinatarios últimos de la producción. Usualmente en los modelos de insumo-producto la demanda final incluye al consumo privado, al consumo de gobierno, las exportaciones y las inversiones. El sistema económico debe producir las cantidades demandadas de cada sector, y para ello requiere insumos o demandas intermedias, lo que genera una cadena de producción y consumo. Además de los insumos intermedios, están los insumos primarios, formados normalmente por salarios, importaciones, ganancias e impuestos. La suma de la demanda final y las demandas intermedias es igual a la producción total del sistema, y de manera similar, la suma de las producciones intermedias y los insumos primarios debe ser igual a la producción total.

En TRANUS, los conceptos básicos del modelo de insumo-producto se han generalizado y espacializado. El término *sectores* es más amplio que en la versión tradicional, ya que puede representar a los clásicos sectores en que se divide la economía (agricultura, industria, minería, gobierno, etc.), factores de producción (capital, tierra y trabajo), como también grupos de población, empleo, construcciones, suelo, etc., que forman parte del sistema espacial. El número y tipo de sectores se define de acuerdo a los requerimientos de cada aplicación específica del modelo, así como las unidades más convenientes para medirlos, lo que permite adaptarlo a situaciones urbanas o regionales.

Una primera distinción entre sectores es la característica de ser *transportables* y *no-transportables*. La principal diferencia es que los sectores transportables pueden ser consumidos en lugares diferentes a donde se producen. Una demanda de carbón para una industria siderúrgica, por ejemplo, puede ser satisfecha por industrias mineras ubicadas en otras regiones. La demanda de mano de obra en un sector de una ciudad puede ser satisfecha por población residente en otras áreas de la ciudad. El ejemplo más típico de un sector no-transportable es el suelo o

las edificaciones; éstos deben ser consumidos en el mismo lugar en donde son producidos. En consecuencia, los sectores transportables generan flujos o desplazamientos, ya sea de mercancías o personas, y estos flujos se pueden transformar luego en demanda de transporte. A su vez, para que se puedan dar estos desplazamientos, se requiere de facilidades de transporte, el cual impone costos al sistema productivo. Los sectores no-transportables, en cambio, no requieren transporte ni generan flujos.

Por otra parte, el modelo de localización distingue entre *zonas internas* y *zonas externas*, las cuales reciben un tratamiento diferente. Las zonas internas conforman lo que se denomina 'el área en estudio'; entre ellas se dan todas las relaciones socio-económicas y de transporte. Las zonas externas, en cambio, están restringidas a representar algunas relaciones del área en estudio con otras áreas, como son las importaciones y exportaciones. Estas pueden ser de bienes o de personas, como es el caso de la demanda laboral del área de estudio que se satisface con residentes de fuera. También se pueden definir zonas externas sólo para ser utilizadas en el modelo de transporte en la representación de viajes externos y de paso.

Cada sector localizado en zonas específicas tiene asociadas diversas variables que lo caracterizan, las cuales se describen a continuación.

Producción Exógena

Es la producción no generada por las demandas internas de otros sectores, y es equivalente a la demanda final en los modelos de insumo-producto. Su localización no depende de los mecanismos que simula el modelo, sino de decisiones de política o de elementos no modelados externos al sistema de análisis. Dicha producción, sin embargo, demanda insumos del sistema. La producción exógena no entra en el mecanismo de distribución sectorial-espacial del modelo, sino que se agrega a la producción inducida distribuida. El crecimiento entre períodos se realiza por un modelo incremental, bien sea asignado directamente a sectores y zonas o mediante la distribución espacial de un incremento global dado.

Producción Inducida

Es la producción de un sector generada dentro del área de estudio, destinada a cubrir la demanda interna o externa de otros sectores. Dicha demanda se calcula dentro del modelo a partir de coeficientes técnicos simples (como en el insumo-producto tradicional) o mediante funciones elásticas con respecto al precio. Su distribución espacial-sectorial la realiza el modelo y el crecimiento depende del crecimiento de los sectores que la demandan.

Sólo se considera este tipo de producción en las zonas internas. La parte de ella que se destina a zonas externas, es una exportación. Pero no se considera en el sistema económico analizado, la producción inducida fuera del área de estudio.

Demanda Exógena

En el insumo-producto tradicional, la demanda exógena está constituida por el consumo directo de la población y del gobierno, la inversión y las exportaciones. Es posible incluir demanda adicional a la que generan las actividades del sistema. Si esta demanda proviene del exterior, el modelo la trata como *exportación*, de tal manera que el término *demanda exógena*, en el modelo, aplica exclusivamente a la que ocurre dentro del área de estudio. La demanda exógena se distribuye conjuntamente con la demanda inducida. El crecimiento entre un período y otro lo realiza el modelo incremental.

Demanda Inducida

Corresponde a las necesidades de consumo final o intermedio de las actividades internas al sistema de análisis.

Exportaciones

Las exportaciones se definen como producción interna del área de estudio dirigida al consumo en las zonas externas (demanda exógena en zonas externas). El modelo distribuye la demanda de las zonas de externas entre las zonas internas exclusivamente.

Importaciones

Las importaciones se definen como demanda de las zonas internas que se satisface con producción proveniente de zonas externas. Se distribuye conjuntamente con el resto de la demanda interna. Las importaciones compiten en cierta medida con la producción interna, aunque están limitadas a un dato exógeno que se modifica entre períodos con el modelo incremental.

Capacidad de Producción

La producción total (exógena + inducida) en un sector y zona determinada puede estar limitada a cierta capacidad de producción. Puede haber capacidad máxima, mínima o ambas a la vez. También se denomina *restricción* a la producción.

Costo de Consumo

Es el costo unitario de un sector en la zona de consumo (CIF). Depende de su precio en la zona de producción y del costo de transporte a la zona de consumo. La cantidad consumida de un sector puede provenir de varias zonas de producción, y en cada una el precio puede ser distinto, como lo serán también los costos de transporte. Por lo tanto, el costo de consumo es un costo promedio ponderado por la proporción que se adquiere de cada zona de producción y sus correspondientes costos de transporte.

Costo de Producción

Es el costo unitario de un sector en la zona de producción (FOB); es función del costo de consumo de los insumos más el valor agregado.

Precio de Equilibrio

Es el valor que puede alcanzar un sector en una zona donde su capacidad de producción está restringida. Se trata de un bien escaso cuya oferta está limitada a las restricciones de producción. Si no hubiese restricciones, el precio es igual al costo de producción. Cuando la demanda es mayor que la capacidad de producción, el precio se incrementa, generando una ganancia al productor o renta de escasez. En el caso contrario, cuando la demanda es menor que la capacidad mínima de producción, el precio del sector baja, pudiendo significar pérdida de los productores.

Valor Agregado

Es el valor del capital y trabajo adicional a los insumos necesarios para obtener una unidad de producción. Típicamente se incluyen las remuneraciones al capital y al trabajo, impuestos o subsidios, consumo de capital fijo, amortización de equipos, etc.

Utilidad de Consumo

Es el promedio logarítmico (*log-sum*) de las utilidades utilizadas en la distribución probabilística de la demanda a zonas de producción. Utiliza la desutilidad de transporte en lugar del costo monetario. Ingresa en las funciones de demanda.

Costo de Transporte

Es el gasto monetario de transporte de una unidad de producción de un sector desde la zona de producción a la de consumo, proveniente del modelo de transporte. Para el caso de las mercancías, el costo de transporte es un costo unitario, pero para otros sectores, por ejemplo, residentes, representa el gasto total, dependiendo del número de viajes que realicen en el período de tiempo considerado en el modelo de actividades (mensual o anual). Incide en el costo de producción.

Desutilidad de Transporte

Proviene del modelo de transporte; además del costo monetario de viaje, incluye el valor del tiempo y factores subjetivos. La desutilidad siempre ingresa al modelo de actividades como un valor unitario, es decir, por viaje. Incide en la distribución sectorial-espacial de la producción.

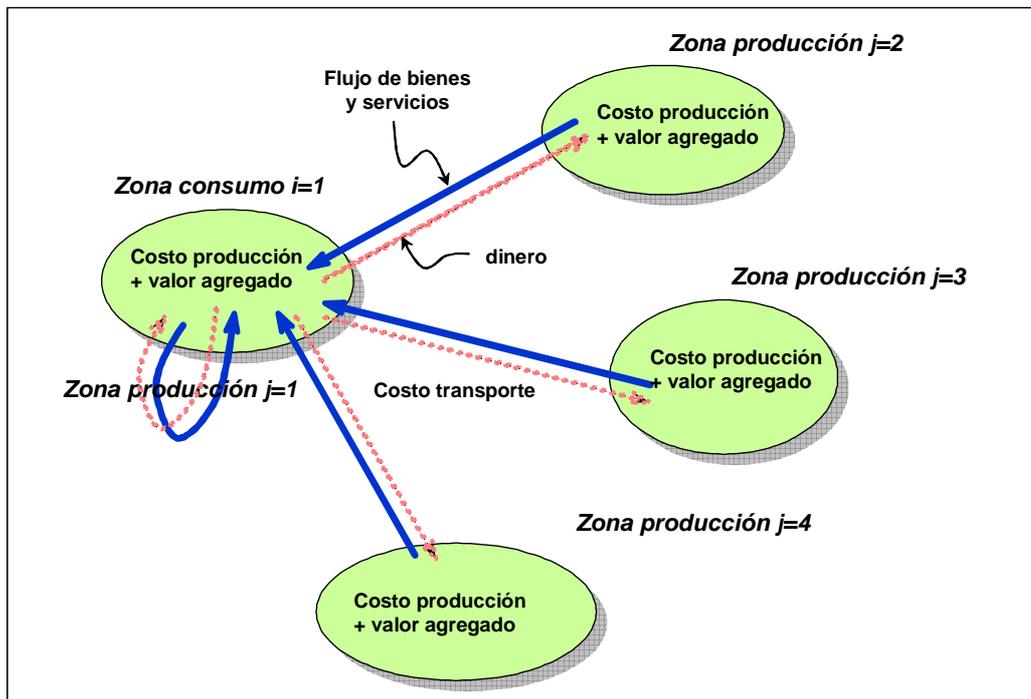
Demanda y distribución de la producción

En principio, todo sector requiere insumos de otros sectores; por tanto, la producción se destina, en parte, al consumo intermedio, y otra parte va directamente al consumo final, bien sea interno o externo (exportaciones). Dada cierta demanda final localizada de uno o más sectores el modelo determina la producción inducida a través de *funciones de demanda*, y la localiza espacialmente mediante funciones de distribución. A la vez, los sectores inducidos demandan insumos, generándose una cadena de producción y de localización de actividades.

De las relaciones antes descritas, se derivan transacciones económicas, que dan origen a *flujos funcionales*, cuando la producción y el consumo ocurren en distintas zonas. Se trata de flujos de personas y de bienes o servicios *transportables*, de los cuales se deriva la demanda de transporte. En algunas transacciones intervienen bienes *no transportables*, como el consumo de suelo o de edificaciones; por tanto, implican un intercambio económico pero no generan flujos funcionales. Cada sector puede efectuar distintas transacciones y dar origen a diversos tipos de flujos; una industria, por ejemplo, requiere consumir suelo y edificaciones (no transportables), lo cual genera sólo un intercambio económico; pero requiere también materias primas (transportables), y mano de obra, cuyo traslado genera demanda de transporte.

Como puede verse en el gráfico de la Figura 1, en una transacción hay una zona de consumo y una o varias zonas de producción, pudiendo ocurrir que la misma zona de consumo sea a la vez zona de producción. El modelo distribuye las compras de manera probabilística entre las zonas que producen el sector demandado. En el diagrama, las flechas indican el sentido en que se mueve la producción, pero el dinero fluye en sentido contrario.

Figura 1: Relaciones de producción y consumo

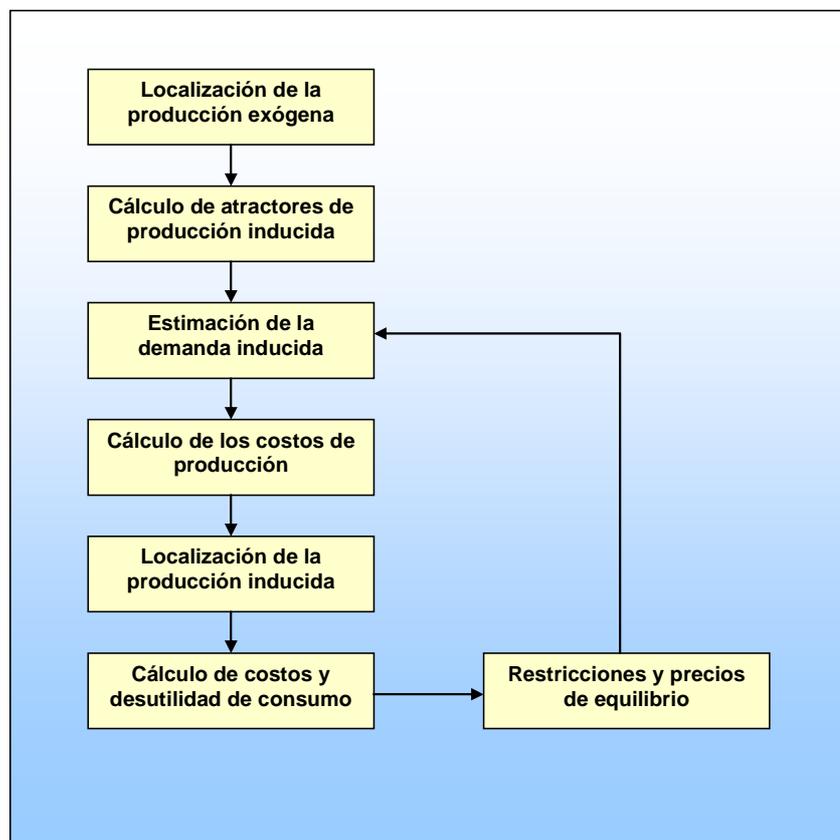


Estructura del modelo de actividades

El modelo de localización de actividades realiza los procesos que se ilustran en la Figura 2.

- localización incremental de las variables exógenas
- cálculo de los atractores para la producción inducida
- estimación de la demanda inducida
- estimación de los costos de producción
- localización de la producción inducida
- cálculo de costos y desutilidades de consumo
- verificación de las restricciones y ajuste de precios de equilibrio

Figura 2: Estructura del modelo de localización de actividades



Incremento y localización de variables exógenas

La primera etapa en la secuencia de cálculo del modelo de localización de actividades consiste en estimar el crecimiento de las variables exógenas en cada zona. Por definición, estas variables dependen de elementos no si-

mulados, por lo que son datos dados exógenamente al modelo. Por tanto, cualquier incremento de estas variables previsto a futuro, debe darse también exógenamente al modelo como dato en el período correspondiente.

Las variables de cada sector n que pueden modificarse entre períodos con el modelo incremental son:

producción exógena
 consumo exógeno
 capacidad de producción
 exportaciones
 importaciones
 atractores iniciales

Llamando de manera global H a cualquiera de las variables exógenas indicadas, su incremento entre los períodos $t-1$ y t (positivo o negativo) en una zona específica i se agrega directamente:

$$H_i^{n,t} = H_i^{n,t-1} + \Delta H_i^{n,t} \quad , \quad (1)$$

Para las tres primeras variables, producción y consumo exógenos, y capacidad de producción, se puede especificar un incremento global al área de estudio. En ese caso se debe indicar un conjunto de variables atractores, en función de las cuales el modelo calcula la proporción de ese incremento que corresponderá a cada zona. A continuación se incluye la ecuación correspondiente al incremento de la producción exógena. Las otras dos variables se manejan con la misma formulación.

Si $X_i^{*n,t}$ es la producción exógena de sector n en la zona i , en el tiempo t , entonces:

$$X_i^{*n,t} = X_i^{*n,t-1} + \Delta X^{*n,t} \rho_i^{n,t} + \Delta X_i^{*n,t} \quad , \quad (2)$$

en donde:

$X_i^{*n,t-1}$ producción exógena del sector n en la zona i en el período anterior $t-1$,
 $\Delta X^{*n,t}$ incremento global de la producción exógena del sector n entre $t-1$ y t ,
 $\Delta X_i^{*n,t}$ incremento de la producción exógena de n específicamente localizada en i , adicional a la anterior,
 $\rho_i^{n,t}$ proporción del incremento n que se localiza en i en el período t .

La proporción que aplica a la distribución del incremento global de variables exógenas, se basa en funciones de atracción, similares a las utilizadas para la distribución de la producción inducida.

$$\rho_i^{n,t} = \frac{A_i^{n,t}}{\sum_i A_i^{n,t}} \quad , \quad (3)$$

donde $A_i^{n,t}$ es una función de atracción de la zona i para la localización del sector n , en el período t . Hay dos

opciones para definirla. La primera forma es usando una función lineal que ya está especificada dentro del modelo con tres variables, de las cuales puede usarse una sola de ellas o combinación de dos o tres.

$$A_i^{n,t} = \sum_k b^{nk} \left(\alpha^{nk} \tilde{X}_i^{k,t-1} + \beta^{nk} p_i^{k,t-1} + \chi^{nk} Q_i^{k,t-1} \right), \quad (4)$$

donde:

b^{nk}	peso del sector k en la función de atracción del sector n ;
$\tilde{X}_i^{k,t-1}$	producción total (exógena + inducida) de k en la zona i en el período $t-1$
$p_i^{k,t-1}$	precio de k en la zona i en el período $t-1$;
$Q_i^{k,t-1}$	capacidad disponible (restricción máxima – producción total) del sector k en la zona i en el período $t-1$
$\alpha^{nk}, \beta^{nk}, \chi^{nk}$	parámetros que regulan la importancia relativa de cada elemento.

Como resultado, varios sectores k pueden formar parte de la función de atracción de un sector n , con pesos regulados por b^{nk} . Para cada sector atractor se puede escoger una o más de las tres variables que formarán parte de la función; por ejemplo, si sólo el precio es relevante, entonces $\alpha^{nk}=0$, $\beta^{nk}>0$, y $\chi^{nk}=0$.

La segunda forma de la función de atracción, es libremente definida por el usuario, quien puede especificar totalmente una función lineal, potencial o logarítmica para cada sector cuyo incremento global desea distribuir entre las zonas internas. El usuario selecciona los sectores atractores, las variables y los parámetros que entrarán en las ecuaciones que definen la función, además de un término constante, pudiendo utilizar hasta cuatro ecuaciones para definirla. Los tipos de funciones son los siguientes:

Lineal:
$$A_i^{n,t} = C + \sum_k \left(a1^{nk} X1_i^{k,t-1} + a2^{nk} X2_i^{k,t-1} + \dots \right) \quad (5)$$

Potencial:
$$A_i^{n,t} = C + \sum_k \left(\left(X1_i^{k,t-1} \right)^{a1^{nk}} \times \left(X2_i^{k,t-1} \right)^{a2^{nk}} \times \dots \right) \quad (6)$$

Logit:
$$A_i^{n,t} = \exp \left(C + \sum_k \left(a1^{nk} X1_i^{k,t-1} + a2^{nk} X2_i^{k,t-1} + \dots \right) \right) \quad (7)$$

Para cada sector n se pueden especificar varios sectores atractores k , indicando cuales variables de dichos sectores ($X1, X2, \dots$) atraen la localización de n y sus correspondientes parámetros ($a1, a2, \dots$). La variables que pueden seleccionarse son las siguientes: Producción Exógena, Producción Total (exógena + inducida), Demanda Exógena, Restricción Mínima, Restricción Máxima, Precio y Capacidad (restricción máxima - producción total), de las cuales el modelo tomará las cantidades del período anterior. Sin embargo, si en algunas zonas se han indicado incrementos puntuales de algunas variables para el período actual, esas cantidades incidirán en la distribución del incremento global.

Cálculo de atractores para la producción inducida

Para la localización de la producción inducida, la función de atracción se define de la siguiente forma:

$$A_i^{n,t} = \left(\sum_k b_k^n (\tilde{X}_i^{k,t-1}) \right) W_i^{n,t} \quad (8)$$

donde:

$\tilde{X}_i^{k,t-1}$ producción total (exógena + inducida) del sector k atractor de n en la zona i ;

b_k^n peso relativo del sector k como atractor del sector n ,

$W_i^{n,t}$ atractor inicial de la zona i que toma en cuenta factores no-modelados que atraen la localización de n .

Generación de la demanda inducida

La cantidad de insumos que una unidad de producción de un sector requiere consumir de otro sector está determinada por una función de demanda. El modelo contempla funciones de demanda fija (coeficiente técnico), demanda variable (elástica) y la posibilidad de incluir bienes sustitutos. El suelo es un típico bien sustitutivo cuando se distinguen varios tipos, como unifamiliar, multifamiliar, rural, etc.

La forma de la función de consumo es la siguiente:

$$a_i^{mn} = \min^{mn} + \left(\max^{mn} - \min^{mn} \right) \cdot \exp\left(-\delta^{mn} U_i^n\right), \quad (9)$$

en donde:

a_i^{mn} cantidad del sector n demandada por una unidad del sector m en la zona i ,

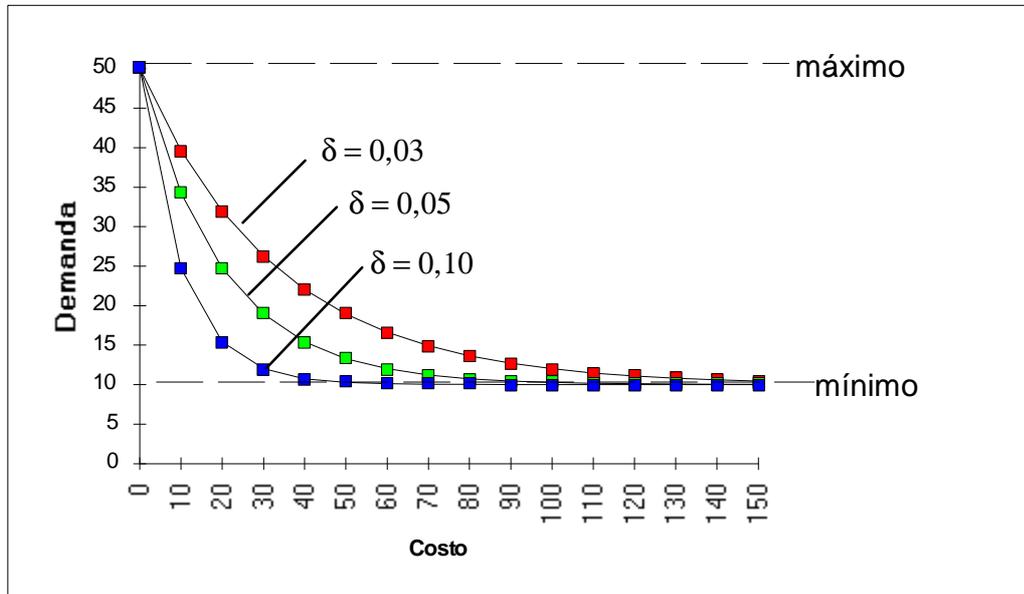
\min^{mn} cantidad mínima de n requerida por una unidad de producción m ,

\max^{mn} cantidad máxima de n que está dispuesto a consumir el sector m ,

δ^{mn} parámetro de elasticidad de m al costo del insumo n ,

U_i^n desutilidad de consumo de n en i ,

La forma resultante de la función es la que se indica en la Figura 3, en que para un consumo máximo de 50 y un mínimo de 10 se aplican tres valores diferentes de.

Figura 3: Funciones de demanda


Luego, la proporción que se aplica a la función de demanda para tomar en cuenta la presencia de sustitutos se estima con un multinomial logit:

$$S_i^{mn} = \frac{\exp(-\delta \tilde{U}_i^{mn})}{\sum_k \exp(-\delta \tilde{U}_i^{mk})}, \forall k, n \in K^n \quad (10)$$

K^n es el conjunto de todos los sectores sustitutos de n . \tilde{U}_i^{mn} es la utilidad escalada en el modelo de sustitución, definido de la siguiente forma:

$$\tilde{U}_i^{mn} = \frac{a_i^{mn} \tilde{c}_i^n \bar{\omega}^{mn}}{\left[\min_k (a_i^{mk} \tilde{c}_i^n \bar{\omega}^{mk}) \right]^{\theta^m}}, \quad (11)$$

El término $a_i^{mn} \tilde{c}_i^n$ es la cantidad de n que el sector m está dispuesto a consumir en la zona i multiplicado por el costo de consumo de n en i , es decir, el gasto de dicho consumo. A su vez, el gasto se multiplica por el factor de penalización $\bar{\omega}^{mn}$. El denominador de la ecuación escala las utilidades al dividir las por la de la mejor opción, la de menor gasto. Finalmente, θ^m establece el nivel de escala; si $\theta^m=1$, la función es totalmente escalada; si $\theta^m=0$, no se escalan las utilidades; es posible definir valores intermedios.

La cantidad de insumo n demandado por el sector m en la zona i será

$$D_i^{mn} = \left(X_i^{*m} + X_i^m \right) a_i^{mn} S_i^{mn} \quad (12)$$

La demanda total del insumo n en una zona i se obtiene sumando el consumo de todos los sectores m que lo demandan, más posibles demandas exógenas:

$$D_i^n = \sum_m D_i^{mn} + D_i^{*n}, \quad (13)$$

en donde:

D_i^n demanda total de n en la zona i ,

D_i^{*n} demanda exógena de n en i .

En la primera iteración sólo hay producción exógena en el sistema y la inducida que se genera por el consumo de ésta. En iteraciones sucesivas se agrega el consumo de los sectores inducidos en la iteración previa.

Cálculo del Costo de Producción

Es el costo de consumo de los insumos necesarios para producir una unidad de m en la zona i , más el valor agregado a la producción:

$$c_i^m = \left(\sum_n D_i^{mn} \tilde{c}_i^n \right) + VA_i^m, \quad (14)$$

En donde VA_i^m es el valor agregado a la producción de m y \tilde{c}_i^n es el costo de consumo del insumo n en la zona i .

Localización de la producción inducida

Una vez determinada la demanda total de cada sector, se procede a distribuir espacialmente las compras desde los lugares de consumo a los de producción. Si un sector no es transportable, toda la producción se asigna a la misma zona en que se demanda. Si el sector es transportable, se realiza el proceso de distribución de la producción con un multinomial logit, utilizando la utilidad marginal de las opciones de localización:

La desutilidad de cada opción de localización de la producción viene dada por:

$$U_{ij}^n = \lambda^n \left(p_j^n + h_j^n \right) + t_{ij}^n, \quad (15)$$

donde:

p_j^n es el precio del sector n en la zona de producción j ,

h_j^n es el precio sombra de producción del sector n en la zona de producción j ,

t_{ij}^n es la desutilidad de transporte para el sector n desde la zona de producción j a la zona de consumo i , y

λ^n es un parámetro que regula la importancia relativa de los precios y la desutilidad de transporte en la función de utilidad.

El precio sombra de producción lo calcula el modelo en la calibración. Los valores resultantes contienen además ajustes por relaciones no modeladas. Por tanto, mientras mejor especificado esté el sistema de actividades, serán menores los precios sombra en valor absoluto.

Los resultados se dividen por la utilidad de la mejor opción para obtener la utilidad escalada

$$\tilde{U}_{ij}^n = \frac{U_{ij}^n}{(\min_j U_{ij}^n)^{\theta^n}}, \quad (16)$$

θ^n es el nivel de escala en el término de utilidad. Si $\theta^n = 1$, se utiliza un modelo totalmente escalado; Si $\theta^n = 0$, no se aplica escala a las utilidades. Estas utilidades escaladas ingresan al multinomial logit para estimar la probabilidad de cada zona j de localizar producción del sector n demandado en i :

$$Pr_{ij}^n = \frac{(A_j^n)^{\alpha^n} \cdot \exp(-\beta^n \tilde{U}_{ij}^n)}{\sum_j (A_j^n)^{\alpha^n} \cdot \exp(-\beta^n \tilde{U}_{ij}^n)}, \quad (17)$$

$$X_{ij}^n = D_i^n Pr_{ij}^n, \quad (18)$$

donde:

X_{ij}^n producción de n que se localiza en la zona de j , inducida por las actividades de i

A_j^n atractor para la producción de n en la zona j ,

α^n es un parámetro que regula la importancia relativa del atractor vs. la función de utilidad en la localización del sector n ,

\tilde{U}_{ij}^n utilidad marginal de localización de la producción de n en la zona j para suplir la demanda de la zona i ,

β^n parámetro que regula la distribución en el multinomial logit.

Si la zona de consumo i es interna, la distribución se realiza para todas las zonas, internas o externas; si la zona de consumo es externa (exportaciones), la distribución se realiza sólo entre las zonas internas. En otras palabras, no se permite que las exportaciones sean satisfechas por importaciones.

Finalmente, la producción inducida total que se localiza en una zona se obtiene sumando respecto a todas las zonas de demanda:

$$X_j^n = \sum_i X_{ij}^n . \quad (19)$$

Costo y desutilidad de consumo

Una vez asignada la demanda de cada sector a las zonas de producción, se recalcula el costo de consumo, es decir, cuánto le cuesta a un sector m localizado en i consumir una unidad del insumo n . Dado que las compras se distribuyeron espacialmente, se estima un promedio ponderado del precio pagado en cada zona de producción más los costos de transporte respectivos:

$$\tilde{c}_i^n = \frac{\sum_j X_{ij}^n (p_j^n + tm_{ij}^n)}{\sum_j X_{ij}^n} , \quad (20)$$

donde:

X_{ij}^n cantidad del sector n demandado en i y producido en la zona j

p_j^n precio de una unidad de n en la zona de producción j ,

tm_{ij}^n costo monetario de transporte de una unidad del sector n entre la zona de producción y la de consumo (distinto de la desutilidad de transporte).

La desutilidad de consumir n en i es el promedio logarítmico de las desutilidades utilizadas en la distribución a las distintas zonas de producción:

$$U_i^n = -\frac{\ln Pg^n}{\beta^n} (\min_j U_{ij}^n)^{\theta^n} , \quad (21)$$

Nótese que se multiplica por la desutilidad mínima, pues en la distribución se utilizó la desutilidad escalada; de esta forma se retorna el valor original para calcular la utilidad compuesta. Pg se define como una serie de la siguiente forma:

$$Pg^n = \sum_{j=1}^z G_j \prod_{h=1}^{j-1} (1 - G_h) , \quad (22)$$

donde la función G_j es el elemento exponencial en el numerador de la probabilidad de localizar en j producción demandada en i , de la ecuación (17):

$$G_j = \exp(-\beta \tilde{U}_{ij}^n) , \quad (23)$$

Verificación de restricciones y ajuste de precios

La producción de un sector en una zona puede estar limitada a la capacidad de producción máxima y/o mínima. Si está dentro del rango establecido por los límites, el precio es igual al costo de producción. Si la producción está por encima del valor máximo o por debajo del mínimo, el precio quedará determinado por la relación oferta/demanda. Al final de cada iteración, el precio de los sectores se ajusta en cada zona; aumenta si se viola la restricción máxima, y disminuye si se viola la restricción mínima. Estas variaciones afectan la distribución de las compras en sucesivas iteraciones, hasta alcanzar un precio de equilibrio.

$$p_j^{n,t} \left\{ \begin{array}{l} < p_j^{n,t-1}, (X_j^{*n} + X_j^n) < R \min_j^n \\ > p_j^{n,t-1}, (X_j^{*n} + X_j^n) > R \max_j^n \\ = c_j^{n,t}, R \min_j^n = 0, R \max_j^n = \infty \end{array} \right\}, \quad (24)$$

donde:

$p_j^{n,t-1}$ precio unitario en la zona j del sector n en la iteración anterior $t-1$

$p_j^{n,t}$ precio unitario en la zona j del sector n en la iteración actual t

$R \min_j^n$ y $R \max_j^n$ son las restricciones, mínima y máxima respectivamente, a la producción del sector n en la zona j

$C_j^{n,t}$ costo de producción de n en la zona j en la iteración actual

$X_j^{*n} + X_j^n$ producción total: producción exógena + producción inducida de n

La diferencia de precios entre una iteración y otra pueden controlarse mediante un parámetro (*smoothing parameter*) que suaviza los cambios entre iteraciones para evitar saltos bruscos y aproximarse de forma más continua a la convergencia.

Convergencia

En cada iteración se evalúa la convergencia de precios y la de producción como la variación porcentual con respecto a la iteración anterior. Los indicadores se calculan para cada sector, y adoptan el valor de la zona que más varía.

$$Cp_j^{n,\tau} = \max_j \left| \frac{p_j^{n,\tau} - p_j^{n,\tau-1}}{p_j^{n,\tau-1}} \right|, \quad CX_j^{n,\tau} = \max_j \left| \frac{X_j^{n,\tau} - X_j^{n,\tau-1}}{X_j^{n,\tau-1}} \right|, \quad (25)$$

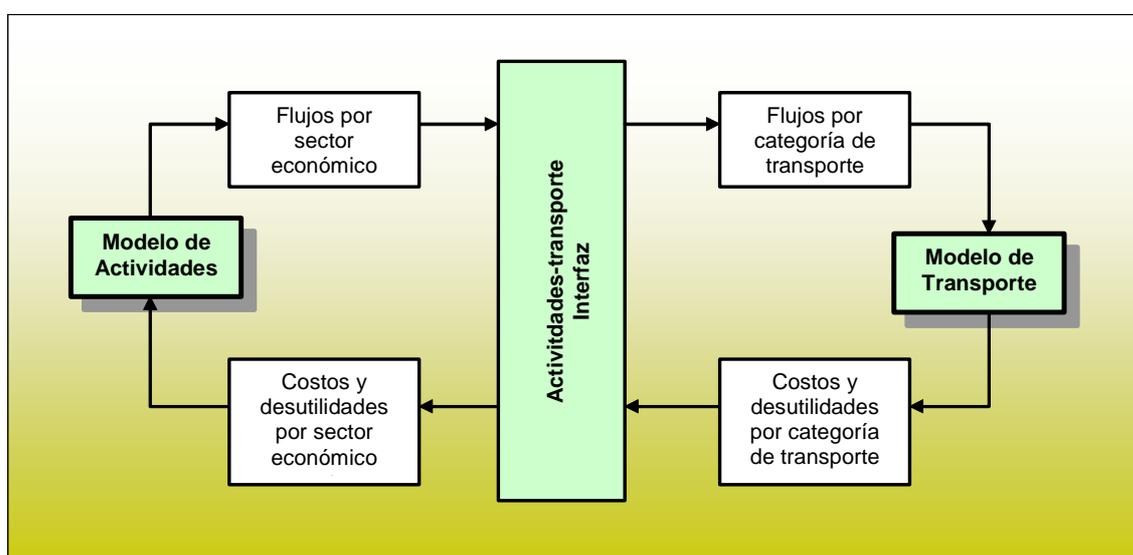
$Cp_j^{n,\tau}$ es el indicador de convergencia de precios, y $CX_j^{n,\tau}$ es el indicador de convergencia de la producción.

El modelo termina las iteraciones cuando ambos indicadores son inferiores al criterio establecido, o cuando se alcanza el número máximo de iteraciones especificado.

Interfaz Actividades-Transporte

El modelo de localización produce como resultado, entre otros, matrices de *intercambio socio-económico* entre sectores. El intercambio entre sectores transportables genera *flujos funcionales*, de los cuales se deriva la demanda de transporte. A su vez, el modelo de transporte estima costos y desutilidades de viaje por categoría, a partir de los cuales es preciso calcular los montos que corresponden a los sectores socioeconómicos que generaron dichos viajes. Por tanto hay dos tipos de transformaciones, desde el modelo de actividades hacia el de transporte, y en sentido contrario, tal como se ilustra en la Figura 4.

Figura 4: Interfaz Actividades - Transporte



El módulo de interfaz actividades-transporte hace compatibles las unidades de tiempo, magnitud y sentido de los flujos para derivar *flujos por categorías de transporte*. Las posibilidades de transformación son las siguientes:

Formación de categorías de transporte

La interacción entre sectores socioeconómicos transportables genera movimientos de bienes o personas de diversa índole. En una aplicación regional, por ejemplo, las interacciones económicas entre sectores como agricultura, minería, industria, etc., generan flujos de carga. En una aplicación urbana las distintas categorías de población generan flujos de pasajeros. El módulo de interfaz genera categorías de transporte a partir de los flujos socioeconómicos, en función de un conjunto de parámetros que definen las relaciones que existen entre ellos.

Factor de tiempo

Generalmente el modelo de actividades está definido en un espacio temporal diferente al de transporte. En una aplicación regional es común que las actividades estén referidas a un período anual (producción sectorial), y el transporte referido a un período diario. En una aplicación urbana, las actividades suelen representarse en períodos mensuales (salarios, rentas), mientras que el transporte se representa diario u hora pico.

Para hacer compatibles las unidades temporales, la interfase reconoce dos tipos de flujos: *normales* y *habituales*. Los flujos normales son típicamente los de carga. Por ejemplo, un intercambio anual entre dos sectores productivos implicará cierto flujo de transporte diario de mercancías; habrá, por lo tanto, un factor de conversión temporal que indique el flujo de transporte diario que se deriva de la transacción anual. En sentido contrario, los costos y desutilidades que el modelo de transporte calcula son unitarias (ej. por tonelada transportada) por lo que no es preciso realizar transformaciones para estimar los costos asociados al sector económico que generó la movilización. El factor Volumen-Valor (descrito abajo) lleva cuenta de las toneladas asociadas a cada sector.

Los flujos habituales, en cambio, son típicamente los movimientos de personas. La interacción entre sectores de empleo y población genera una matriz empleo-hogar, que no es indicativa del flujo, pues sólo representa personas que trabajan en una zona i y viven en una zona j . La estimación de los viajes diarios al trabajo que se realizan es materia del modelo de transporte. Por lo tanto, en la interfase no se aplica el factor tiempo. En sentido contrario, sin embargo, el modelo estima el costo por viaje al trabajo, y es preciso conocer el gasto mensual (o anual) de los sectores que los generaron, el cual forma parte de su "paquete" de consumo, junto con educación, servicios, etc., que incide en la localización.

La siguiente tabla resume la aplicación de los factores de tiempo.

	Actividades→Transporte (flujos)	Transporte→Actividades (costos y desutilidades)
Flujo Normal	Divide por factor tiempo	Ignora factor tiempo
Flujo Habitual	Ignora factor tiempo	Multiplifica por factor tiempo

Factor volumen/valor

Los flujos socioeconómicos en el modelo de actividades pueden generar distintos tipos de flujo por categoría de transporte. El ejemplo más típico son los movimientos de mercancías; el modelo de localización puede representar flujos monetarios, mientras que en transporte se requiere trabajar en toneladas. En cuanto a la población, cuando el sector económico está expresado en personas, no requiere esta transformación, pues es la misma unidad de los pasajeros. Sin embargo, un sector de población puede estar representado en familias en el modelo de actividades y un factor transformaría los flujos a pasajeros.

El factor volumen/valor multiplica los flujos socioeconómicos antes de ingresar al modelo de transporte. Dicho factor divide los costos y desutilidades de viaje para asignarlos a los sectores que los generaron en el modelo de actividades.

Sentido de los flujos

En el modelo de localización, los flujos siempre se producen desde la zona de consumo a la zona de producción (flujos de compra), el cual puede no ser el sentido en que se mueven los bienes o las personas. Por ejemplo, en una aplicación urbana, un flujo de personas generado a partir del empleo tendrá el sentido empleo → población (el sentido en que fluye el dinero); para derivar los viajes al trabajo en la hora pico, se debe invertir la matriz para representar viajes basados en el hogar. Cuando el modelo de transporte trabaja en totales diarios, es necesario considerar los flujos de ida y vuelta.

Ecuación de transformación de los flujos

Todas las transformaciones descritas en los párrafos anteriores se resumen en una sola ecuación de transformación de los flujos:

$$F_{ij}^s = \sum_n \left(X_{ij}^n \frac{vol^{ns} pc^{ns}}{tiem^{ns}} + X_{ji}^n \frac{vol^{ns} cp^{ns}}{tiem^{ns}} \right), \quad (26)$$

donde:

F_{ij}^s flujo de la categoría de transporte s desde el origen i al destino j ,

X_{ij}^n producción del sector transportable n localizada en j y demandada en i ; expresión del flujo socioeconómico entre i y j

vol^{ns} factor de transformación de valor a volumen para el flujo socioeconómico n que forma parte de la categoría de transporte s ,

$tiem^{ns}$ factor de transformación temporal entre el flujo socioeconómico n y el flujo de transporte s ,

cp^{ns} proporción del flujo de bienes o personas que se mueve en el sentido consumo \rightarrow producción,

pc^{ns} proporción del flujo de bienes o personas que se mueve en el sentido producción \rightarrow consumo.

La suma se realiza para todos los flujos socioeconómicos n que forman parte de la categoría de transporte s . Puede ocurrir que un flujo socioeconómico no forme parte de ninguna categoría de transporte. No existen restricciones acerca de los valores que deben tener las proporciones en cada sentido, cp^{ns} y pc^{ns} aunque generalmente suman 1. Por ejemplo, cuando toda la movilización en transporte tiene el mismo sentido de los flujos de compra que vienen del modelo de actividades, entonces:

$$cp^{ns} = 1 \text{ y } pc^{ns} = 0$$

En caso contrario:

$$cp^{ns} = 0 \text{ y } pc^{ns} = 1$$

Si la movilización es en ambos sentidos (por ejemplo, viajes diarios de ida y vuelta al trabajo):

$$cp^{ns} = 1 \text{ y } pc^{ns} = 1$$

El sistema de transporte

Conceptos básicos

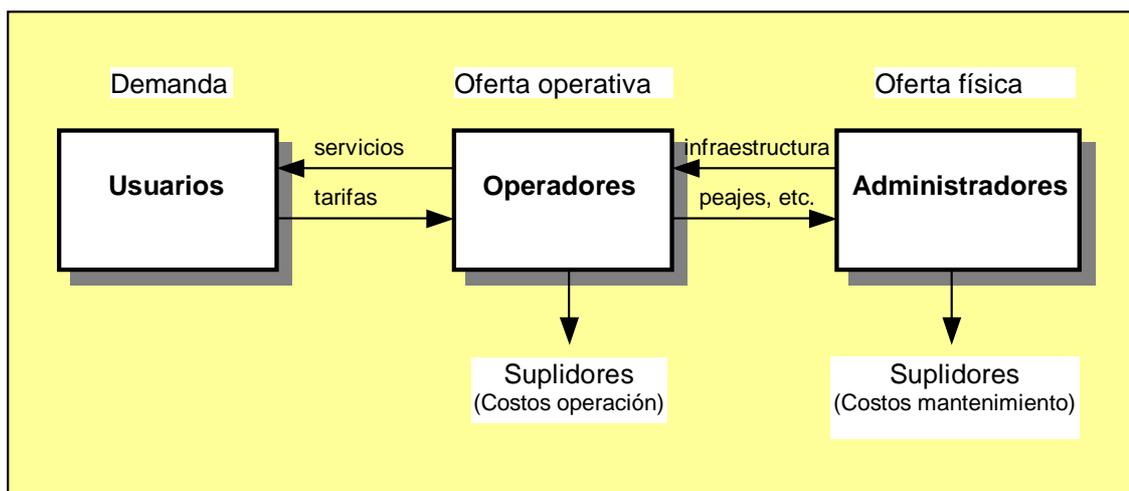
El objetivo principal del modelo es estimar la demanda de transporte y asignarla a la oferta, estableciendo un equilibrio entre ambas. La demanda se estima a partir de los flujos por categoría de transporte resultantes de la interfase. A partir de dichos flujos, el modelo de transporte estima el número de viajes que se realizan en el período considerado, mediante un modelo de generación elástica. Los viajes se asignan a la oferta física y operativa de transporte. Los costos y desutilidades de transporte que resultan del equilibrio demanda/oferta, son utilizados por el modelo de localización de actividades para la simulación de un período siguiente.

En el sistema de transporte se distinguen dos componentes principales: demanda y oferta. La demanda está formada por los usuarios, que son las personas, familias o empresas que demandan el servicio de transporte de carga o de pasajeros. En cuanto a la oferta, se puede distinguir en primer lugar la oferta física (vías, estacionamientos, estaciones, puertos) la cual tiene una administración encargada de su mantenimiento y que puede cobrar por su uso. Por otra parte, está la oferta operativa, representada por los transportistas privados o públicos, que disponen de distintos tipos de vehículo, pueden cobrar tarifas a los usuarios y pagan a los administradores por el uso de la infraestructura.

En la Figura 5 se muestra el esquema general de las relaciones económicas que se dan entre los distintos agentes. Los usuarios demandan servicio de transporte a los operadores, por lo cual les pagan tarifas. Los operadores le cobran a los usuarios y, a su vez, tienen costos de operación y pagan por el uso de la infraestructura. Los administradores de la infraestructura cobran a los operadores y pagan a suplidores por su mantenimiento.

El esquema general planteado puede adoptar formas específicas según los casos. El transporte particular, por ejemplo, es un caso especial en que el usuario de automóvil es su propio operador. El transporte ferroviario es otro caso en que, generalmente, un mismo agente funge de operador y administrador del sistema. Sin embargo, en ambos casos el modelo los considera en la contabilidad como entes separados. Los pagos pueden ser directos o indirectos; los peajes, por ejemplo son pagos directos de los operadores a los administradores de las vías; en algunos casos este cobro puede darse indirectamente como impuesto al combustible.

Figura 5: Elementos del sistema de transporte



Demanda de transporte

Los usuarios se clasifican en categorías de transporte, lo que permite distinguir entre tipos de mercancías y pasajeros. En el caso de los pasajeros, se pueden clasificar, por ejemplo, por niveles de ingreso, propósitos de viaje, o combinaciones de ambos; estas categorías van asociadas a determinada *disponibilidad vehicular* que limita su elección entre transporte público y privado, por el hecho de tener o no un vehículo particular disponible.

De cada opción de viaje que presenta la oferta, el usuario percibe una *desutilidad*, la cual está compuesta por el costo monetario del viaje, por la valoración del tiempo de viaje y de espera, y por elementos subjetivos como grado de confort, confiabilidad, seguridad, etc.

Oferta operativa

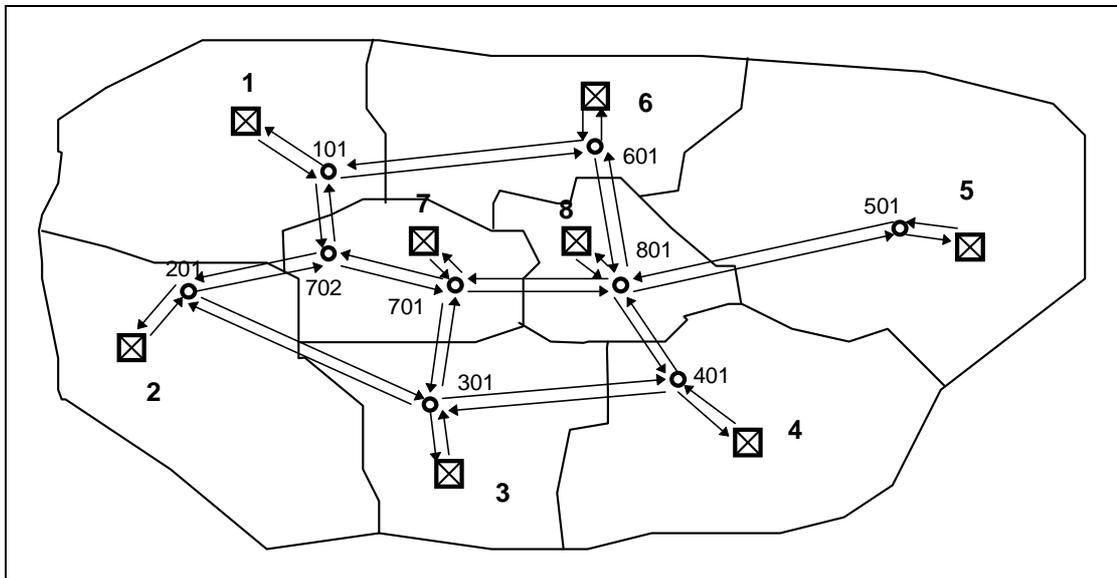
La oferta operativa está organizada jerárquicamente en tres niveles: modos, operadores y rutas. Los modos representan un conjunto de operadores que proveen servicio de un tipo determinado. En Tranus expresan la clasificación más amplia, tales como los Público, Privado, Carga Ligera, Carga Pesada. Cada categoría de demanda puede escoger entre determinados modos, de manera tal que las mercancías sólo pueden escoger entre modos de carga y las personas entre modos de pasajeros. De tal forma que hay un conjunto de modos disponibles para cada categoría.

Cada modo, a su vez, puede tener varios operadores. Un operador presta un servicio con determinadas características tales como tipo de vehículo, tarifa, costo de operación, consumo de energía, etc. Los viajeros pueden realizar transferencias entre los operadores de un modo (transbordos). En una aplicación típica, se consideran habitualmente dos modos: carga y pasajeros; en algunos casos, sin embargo, puede dividirse el transporte de pasajeros en dos modos: transporte público y transporte privado, cuando hay marcada incidencia de la disponibilidad vehicular en el área de estudio. El modo carga puede estar compuesto de operadores tales como camiones y barcos; el transporte público por operadores de autobús, microbús y metro, cada uno de los cuales puede tener rutas y frecuencias específicas; el transporte privado incluye al operador de automóvil.

El sistema urbano de transporte público generalmente tiene una estructura compleja formada por distintos operadores, que el usuario puede combinar para realizar un viaje. El modelo permite simular tarifas integradas entre operadores, común entre el metro y autobuses alimentadores. También es posible prohibir la transferencia entre dos operadores específicos.

Oferta física

La oferta física se representa con la red de transporte, un grafo direccional compuesto por enlaces unidireccionales y nodos, tal como se muestra en la Figura 6. Los nodos pueden representar intersecciones entre vías, puntos en que las características de las vías cambian, estaciones, paradas o cualquier otro punto de interés. Un subconjunto de nodos se denominan *centroides*, que son puntos representativos de cada zona; para el modelo de transporte todos los viajes tienen su origen y destino en centroides. Un centroide puede estar conectado a uno o más puntos de la red. Los enlaces, por su parte, pueden representar calles, carreteras, vías férreas, aéreas o acuáticas, y cualquier otro tipo de infraestructura. Los enlaces tienen características específicas, tales como distancia, capacidad y otros, y características genéricas que se definen por *tipos de enlace*. Los enlaces de un mismo tipo tienen atributos similares tales como: velocidad, costo de mantenimiento, peajes, etc. Los tipos se definen para diferenciar vías tales como carreteras, autopistas, líneas de metro, ferrocarriles o vías acuáticas; por tanto, hay un subconjunto de operadores que puede usar cada tipo de vía. Por ejemplo, una carretera puede ser usada por automóviles, autobuses y camiones, pero no por trenes o barcos.

Figura 6: Red de Transporte


Las características de los arcos de la red son las siguientes:

- nodo de origen
- nodo de destino
- tipo de vía
- longitud
- capacidad física
- rutas de transporte público
- giros prohibidos o demoras

Los atributos comunes a cada tipo de vía son los siguientes:

- Velocidad a flujo libre por operador
- vehículos equivalentes por operador
- costo de operación por unidad de distancia de cada operador
- peajes u otros cargos por operador
- entidad a cargo de la administración
- costo de mantenimiento por unidad de distancia (fijo y marginal)
- funciones de restricción de capacidad

La capacidad física de las vías compartidas se mide en vehículos equivalentes por unidad de tiempo. En las vías exclusivas se puede utilizar otra unidad a conveniencia, como trenes, vagones o toneladas por unidad de tiempo. En vías urbanas, puede utilizarse la capacidad de las intersecciones, la cual es más restrictiva.

Las rutas de transporte público se codifican en cada enlace por el que pasan; el programa hace automáticamente un seguimiento de sus recorridos. Cada ruta tiene una frecuencia de operación (vehículos/unidad de tiempo). El modelo asigna el número de vehículos especificado y evalúa el nivel de utilización de las unidades. Opcionalmente, se pueden especificar rutas con frecuencia indeterminada, y dejar que el modelo estime la frecuencia necesaria para atender la demanda.

Los giros prohibidos codificados en un enlace, indican nodos con los cuales ese arco no puede comunicarse, aunque aparezcan conectados en la red. Evita la codificación de nodos y arcos ficticios para representar los movimientos en las intersecciones. También pueden indicarse demoras en intersecciones, para representar conflictos en los giros izquierdos, ciclos de semáforos, tiempos de carga y descarga, acceso a puertos, etc.

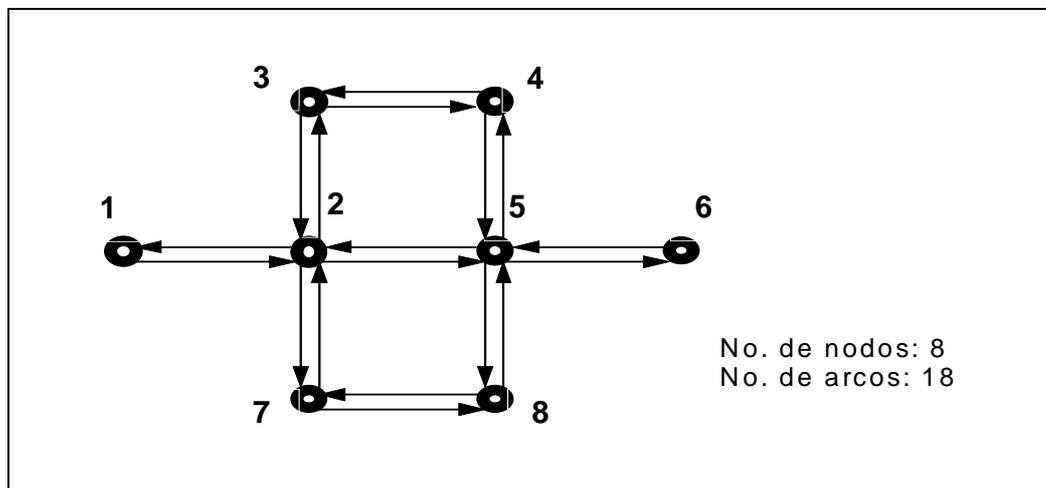
Representación de la red de transporte

El modelo de transporte trabaja sobre una representación interna de la red denominada *grafo dual*. La red se codifica de la manera tradicional (arcos para las vías y nodos para las intersecciones) y el programa transforma en nodos los arcos reales, y en enlaces las conexiones factibles, considerando los giros prohibidos. El método es transparente para el analista, porque los resultados se traducen a la representación original. A continuación se presenta una breve descripción conceptual de la técnica del grafo dual. La descripción completa y su especificación matemática puede encontrarse en la publicación *Dual Graph Representation of Transport Networks*, por T de la Barra, B Pérez y J Añez, *Transportation Research B*, Vol 30, pp 209-216, 1996.

En la Figura 7 se ilustra un ejemplo de red que consta de 8 nodos y 18 enlaces; en la mayoría de los modelos de transporte su codificación se realiza con una lista como la siguiente:

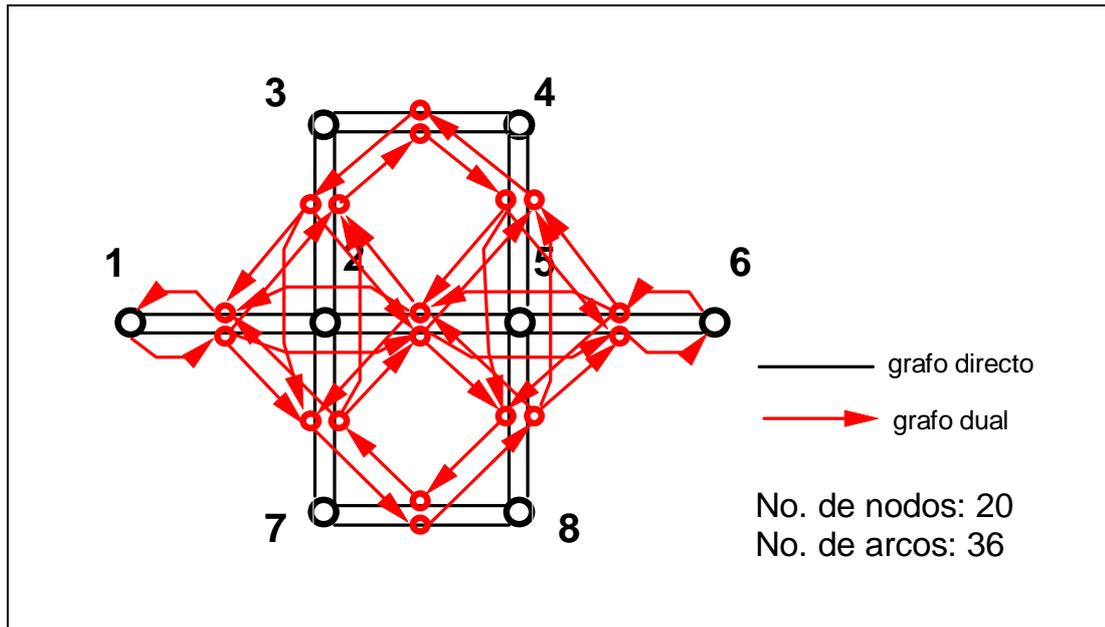
Nodo de origen	Nodo de destino
1	2
2	1
2	3
3	2
2	5
...	

Figura 7: Ejemplo de red en la codificación original



Con la lista, el modelo genera automáticamente el grafo dual correspondiente, en el cual cada enlace original se transforma en vértice o nodo, y las conexiones factibles generan enlaces. La Figura 8 muestra el resultado del grafo dual interno creado por el modelo, el cual tendría 20 nodos duales y 36 enlaces duales, de acuerdo con la siguiente lista:

Nodo de origen	Nodo de destino
1-2	2-3
1-2	2-5
1-2	2-7
2-3	3-4
...	

Figura 8: Ejemplo de un grafo dual


Una de las ventajas de este esquema es la facilidad para definir las prohibiciones de giro y demoras en las intersecciones. Es bien conocida la dificultad de representarlos en los grafos directos, usualmente sustituyendo cada nodo por varios nodos ficticios que deben codificarse, bien manualmente o internamente en algunos programas. Ese proceso conduce a error, agrega nodos y enlaces no deseados y tiene mayor exigencia de cómputo. Con el grafo dual, en cambio, lo único que se debe especificar en el registro de cada enlace es el nodo hacia el cual los vehículos no pueden girar, tal como se muestra arriba en la lista inicial. El modelo generará un enlace dual sólo si el nodo de destino del enlace no se encuentra en la lista de giros prohibidos.

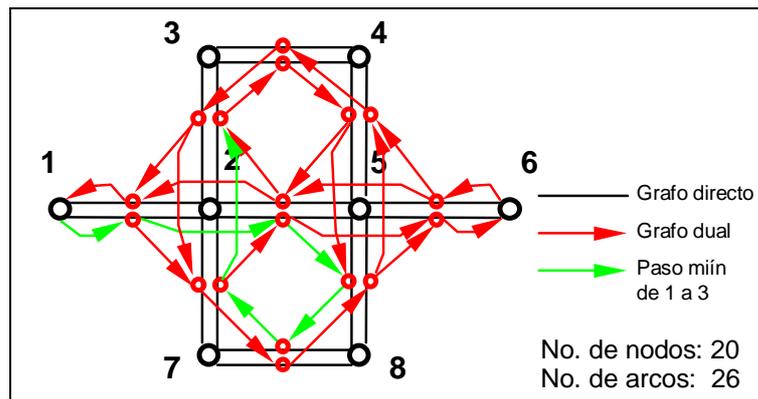
Si en la red considerada arriba se prohíben todos los giros a la izquierda, el código de la red contendrá 26 enlaces en lugar de 36, como se muestra a continuación.

Nodo Origen	Nodo Destino	Giro Prohibido
1	2	3
2	1	
.. etc.	...	
2	5	4
5	2	7

...etc.

Mientras más giros prohibidos existan, la red generada será menor, lo contrario de lo que ocurre con los modelos tradicionales. El modelo generará el grafo dual que se muestra en la Figura 9, donde se ilustra el paso mínimo entre 1 y 3 calculado limpiamente con la prohibición de giro izquierdo en 2 sin que dicho nodo requiera dividirse.

Figura 9: Grafo dual con prohibiciones de giro

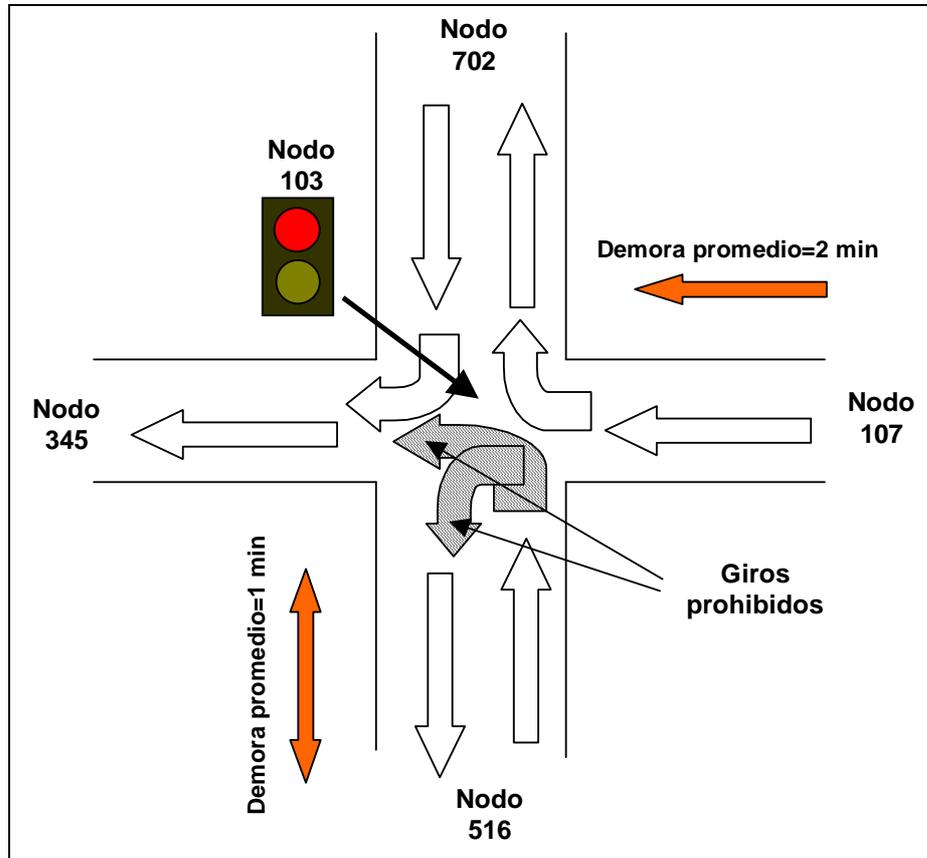


Demoras en intersecciones

Opcionalmente, es posible especificar demoras en determinados giros. Esto es particularmente útil para representar intersecciones con semáforos y otro tipo de demoras como las que se producen en la entrada de los puentes, estaciones de peaje o tiempo de carga y descarga.

En la base de datos de Tranus se ingresa la demora promedio que impone a cada giro la programación de las fases del ciclo del semáforo. En el diagrama de la Figura 10 se muestra una intersección señalizada codificada 103. La demora promedio es de dos minutos (0,033Hr.) para la vía unidireccional Este-Oeste, y un minuto (0,017 Hr.) para la vía bidireccional Norte-Sur. Los giros hacia la izquierda están prohibidos.

La parte inferior del diagrama muestra los giros posibles en la intersección y sus correspondientes demoras. Por ejemplo, desde el enlace 107-103 no se puede girar hacia el nodo 516 (demora infinita); si puede girar al nodo 702 con una espera promedio de 2 minutos. La demora se agrega al tiempo de viaje, y afecta a todos los vehículos (autos, autobuses, camiones) asignados a la red multimodal de Tranus que se describe en la sección siguiente.

Figura 10: Demora en intersecciones


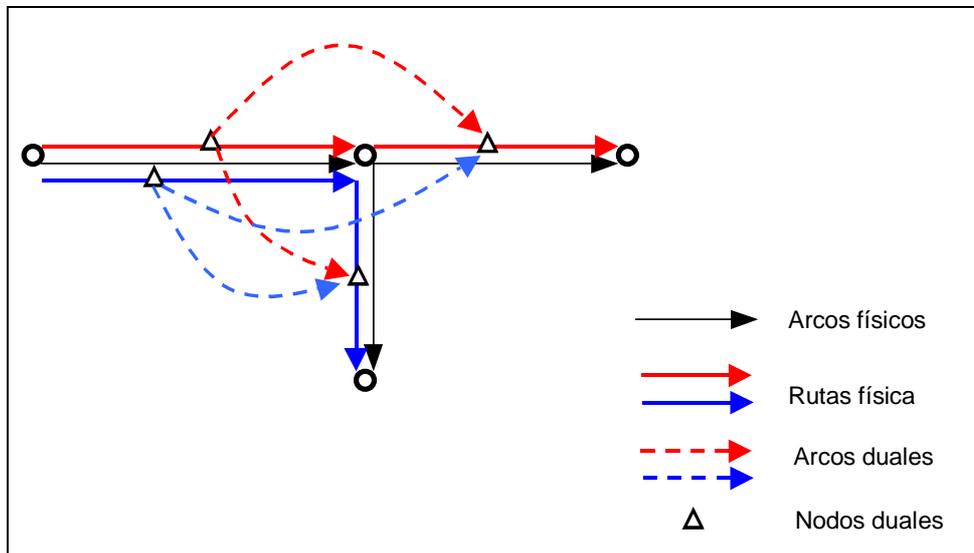
Nodo origen	Nodo Destino	Giro al nodo	Demora (horas)
107	103	702	0.033
107	103	345	0.033
107	103	516	Infinita
516	103	702	0.016
516	103	345	Infinita
702	103	516	0.016
702	103	345	0.016

Red multimodal

El concepto de grafo dual se extiende al de *red multimodal*. Distintos operadores pueden usar un mismo enlace, por ejemplo automóviles, autobuses y camiones. El modelo genera un enlace dual para cada operador que pueda transitarlo, resultando en un conjunto de combinaciones enlace-operador. Igualmente, este proceso es totalmente transparente para el modelista.

La Figura 11 muestra un ejemplo simple de una red multimodal con tres enlaces y dos rutas de autobús. El modelo genera un enlace para cada conexión enlace-ruta factible (considerando las prohibiciones de giro). Si ocurre una transferencia entre las rutas, el modelo considera el tiempo y costo de trasbordo.

Figura 11: Generación de enlaces duales en una red multimodal



Estructura de los costos de transporte

En el modelo se manejan tres tipos de costos, que son los incurridos por los tres agentes del sistema de transporte, usuarios, operadores y administradores, tal como se ilustran en la Figura 5:

- costos al usuario, en los cuales hay componentes monetarios y no monetarios, por lo que se denominan *costos generalizados*; se miden por unidad transportada;
- costos de operación, estrictamente monetarios; se contabilizan por vehículo;
- costos de mantenimiento de la infraestructura de transporte, también monetarios; se contabilizan por unidad de distancia.

Los dos primeros intervienen en la estimación de la demanda y su asignación a la oferta; en cambio los costos de administración sólo intervienen en la contabilidad de ingresos y gastos en la evaluación.

Costo monetario al usuario

El componente monetario del costo al usuario es la *tarifa*, la cual pagan a los operadores y puede ser de dos tipos: dependiente o independiente de los costos de operación. El transporte público suele cobrar tarifas que no tienen correspondencia alguna con sus costos de operación. En cambio, el transporte de carga suele cobrar al usuario una tarifa que cubra sus costos de operación más una ganancia. En el caso del transporte particular, el usuario es el mismo operador, por tanto su costo monetario de viaje está en función del costo de operación; sin embargo, hay evidencia de que los usuarios no perciben el costo total de usar el automóvil cuando toman sus decisiones de viaje.

El término *tarifa* se utiliza en forma general para referirse al costo monetario al usuario, aunque dicho término no sea adecuado al caso del auto particular. Es común que algunas categorías de usuarios paguen tarifas especiales, como es el caso de los estudiantes y ancianos. También el transporte de carga suele cobrar tarifas distintas

en función del tipo de mercancía. El modelo proporciona los siguientes elementos para definir las tarifas en cada aplicación:

$$t_o^s = tp_o^s \left(tf_o + tt_o + td_o + \frac{tc_o c_o}{to_o} \right), \quad (27)$$

donde:

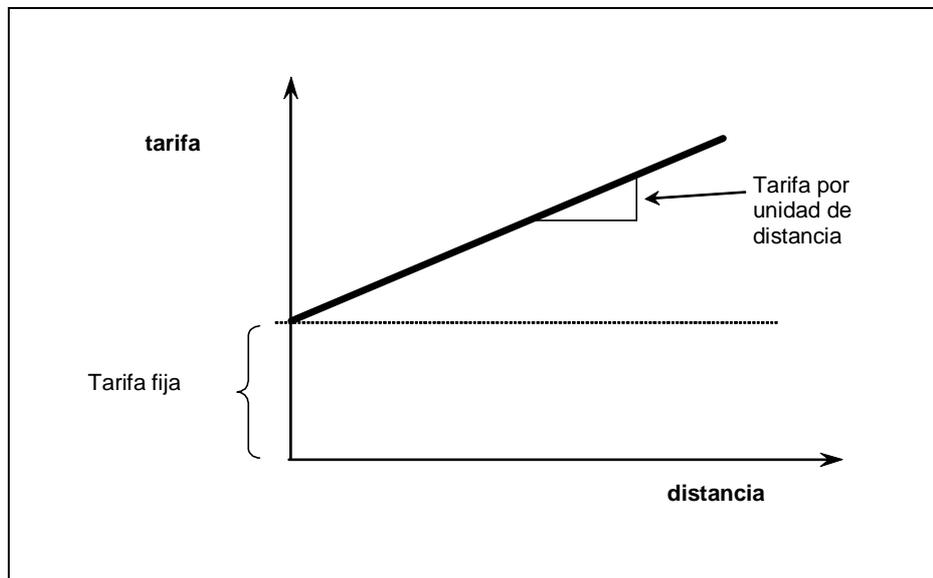
t_o^s	la tarifa que un usuario de la categoría s paga la operador o ;
tp_o^s	proporción de la <i>tarifa normal</i> que un usuario de la categoría s debe pagar al operador o ; el resto de los elementos de la función definen la tarifa normal;
tf_o	tarifa fija que se paga al abordar el operador o ; si hubiese integración da tarifas entre operadores, tf_o depende del operador anterior;
tt_o	tarifa por unidad de tiempo del operador o , aplica al tiempo de viaje en dicho operador;
td_o	tarifa por unidad de distancia del operador o , aplica a la distancia de viaje en dicho operador;
$\frac{tc_o c_o}{to_o}$	c_o es el costo de operación del operador o (por vehículo); tc_o representa la proporción de dicho costo que se transfiere como tarifa al usuario; to_o es la tasa de ocupación del operador o .

Las tarifas que se definen en una aplicación, pueden tener uno o más de los componentes de la función. En algunos casos, el transporte regional de carga establece su tarifa con un elemento fijo (mínimo) más un componente por distancia; en otros casos, cuando algunos recorridos incluyen vías muy congestionadas, prefieren utilizar el componente tiempo en lugar de la distancia. Por lo general, cuando la tarifa se define en función del costo de operación, se asigna exclusivamente al último término de la función, y el resto de los elementos se iguala a cero.

Nota: en la simulación, la congestión afectará las tarifas de un operador cuya función contiene el componente relativo a tiempo. Los usuarios pagarán más en áreas congestionadas.

En la Figura 12 se ilustra una función de tarifa típica del transporte público, donde hay un componente fijo y uno adicional por unidad de distancia

Figura 12: Función de tarifa típica con componentes fijos y por distancia



Costos de operación

Los costos de operación son los incurridos por los operadores que prestan el servicio de transporte, y pueden o no formar parte de la tarifa que ellos cobran a los usuarios. Se calculan por vehículo, y se pueden definir en cada aplicación, con los elementos siguientes:

$$c_o = cf_o + ct_o + cd_o + cho + ce_o, \quad (28)$$

donde

- cf_o costo fijo de operación del operador o ; generalmente referido a costos administrativos y a labores de carga y descarga en el transporte de mercancías;
- ct_o costo en que incurre el operador o por vehículo por unidad de tiempo; generalmente se refiere al salario de los choferes;
- cd_o costo de operación de los vehículos del operador o por unidad de distancia; se refiere al gasto por neumáticos, mantenimiento y reparaciones; puede incluir el consumo de combustible, pero preferiblemente este se computa separadamente como ce_o descrito abajo;
- cho cargos que un vehículo del operador o paga a los administradores de la infraestructura, tales como peaje, costo de estacionamiento y otros;
- ce_o costo de energía de los vehículos del operador o por unidad de distancia; es función de la velocidad y, por ende, varía con la congestión.

El costo de la energía (gasolina, diesel, electricidad), puede incluirse simplemente dentro de cd_o con un estimado del consumo promedio por unidad de distancia y el precio unitario de la energía. Alternativamente, el modelo permite estimar el consumo mediante una función dependiente de la velocidad que se aplica durante la simulación en cada enlace recorrido por el operador; el costo estaría dado por la siguiente ecuación:

$$ce_o = \left[ed_o^{\min} + (ed_o^{\max} - ed_o^{\min}) * \exp(-\delta^o V_o) \right] pe_o, \quad (29)$$

donde:

ce_o costo por consumo de energía por unidad de distancia del operador o en el enlace l

ed_o^{\min} consumo mínimo de energía por unidad de distancia; cuando el vehículo del operador o circula a la velocidad de flujo libre

ed_o^{\max} consumo máximo de energía por unidad de distancia; cuando la velocidad del operador o se acerca a cero

V_o Velocidad a la que circula el operador o en las condiciones de congestión

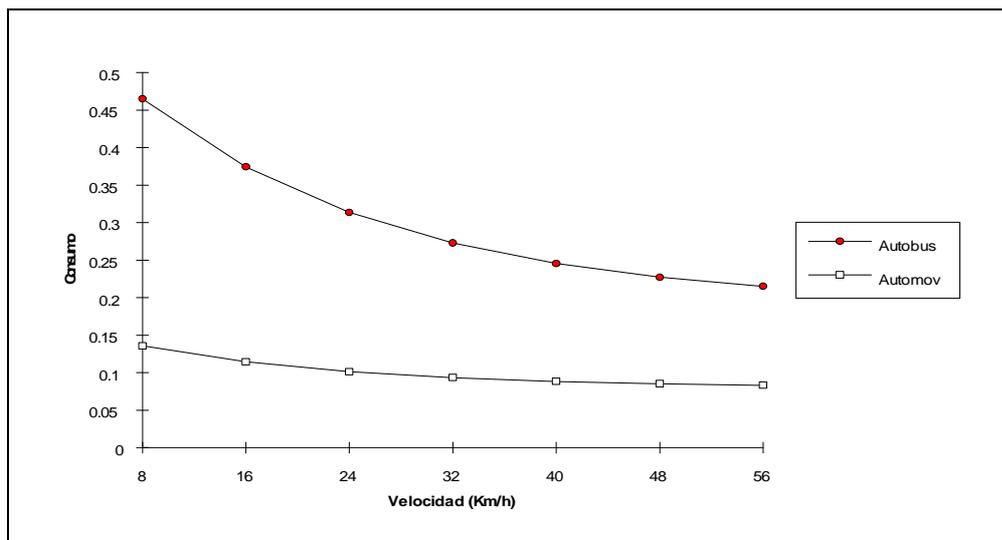
δ^o Parámetro que regula la curva de consumo de energía del operador o ; es negativo, porque a medida que la velocidad aumenta se reduce el consumo

pe_o precio unitario de la energía que consume el operador o

El modelo realiza el cálculo para cada enlace recorrido por el operador, porque la velocidad varía en cada enlace. En la Figura 13 se ilustran las curvas típicas de consumo de energía para dos operadores.

La energía se define en el modelo con las unidades apropiadas a cada operador. A los autos, por ejemplo, se especifica gasolina en litros; diesel para los camiones, y Kw/h para trenes. Cuando no se conoce la función de energía ni es posible estimarla, puede incluirse conjuntamente con los otros costos relativos a distancia.

Figura 13 : Curvas típicas de consumo de energía



Costos de mantenimiento

Los administradores pagan por mantenimiento vial dos tipos de costos. Un costo fijo por unidad de distancia que corresponde al mantenimiento preventivo rutinario de la infraestructura, y un costo marginal dependiente del volumen de vehículos que circula por las vías. Cada operador incide de manera diferente en el deterioro de la vialidad; por tanto, el costo marginal de mantenimiento depende, además del volumen y de la composición de tráfico.

Cada administrador tiene a su cargo determinados tipos de vía. Siendo τ un tipo de vía del conjunto T^a a cargo del administrador a , y siendo l un arco del conjunto L^τ de enlaces tipo τ , el costo de mantenimiento que le corresponde al administrador a viene dado por:

$$cm^a = \sum_{\tau \in T^a} \sum_{l \in L^l} \left[mf_\tau * d_l^\tau + \sum_o ma_\tau^o * Ve_l^o \right], \quad (30)$$

donde:

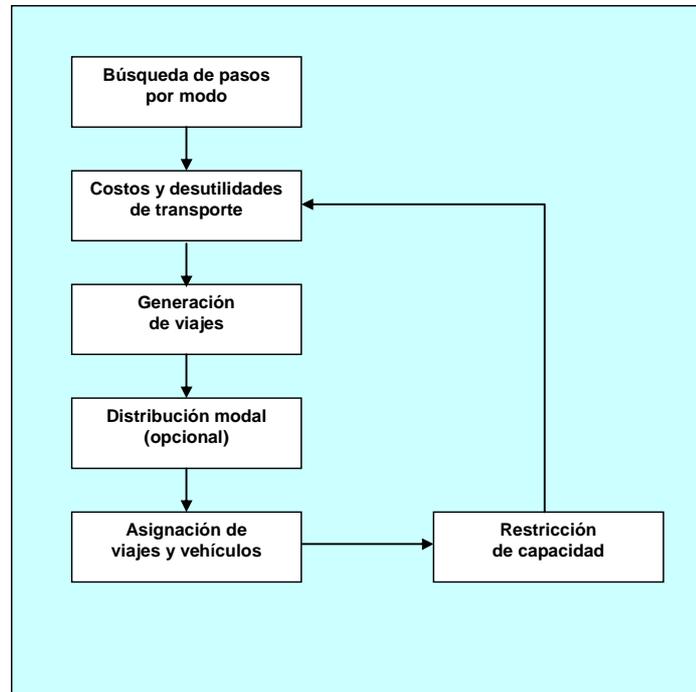
- mf_τ costo de mantenimiento fijo por unidad de distancia de las vías tipo τ
- d_l^τ distancia del enlace l de tipo τ
- ma_τ^o costo marginal de mantenimiento de las vías tipo τ por vehículo del operador o
- Ve_l^o número de vehículos del operador o que circulan en el enlace l

Estructura del modelo de transporte

El proceso de simulación en el modelo de transporte sigue la secuencia que se ilustra en la Figura 14. Se distinguen dos procesos: la búsqueda de pasos y asignación de la demanda. El primero genera opciones de viaje entre cada par de zonas origen-destino para cada modo. Con base en la descripción de la red y en las definiciones de funciones y parámetros de costos y desutilidades, el algoritmo define, no sólo el paso mínimo, sino los primeros n pasos.

El modelo calcula los costos y desutilidades de cada paso para cada categoría de transporte. Por agregación de todos los pasos, estima luego los costos y desutilidades por modo y, por agregación de éstos, se obtienen los costos y desutilidades por categoría de transporte.

En cuanto a la demanda, el primer paso es la estimación del número de viajes por par O-D y categoría, mediante una función de generación elástica. Luego se realiza la separación modal y la asignación; la primera es opcional, pues puede combinarse con la asignación en un solo proceso, como se explica más adelante. Finalmente, se realiza la restricción de capacidad para ajustar las velocidades de circulación y los tiempos de espera en función a la congestión de cada enlace. Lo dicho en este párrafo se repite en un proceso iterativo hasta lograr un equilibrio demanda/oferta, del cual resultan nuevos costos y desutilidades a *red cargada*. Opcionalmente el analista puede ejecutar el algoritmo de pasos con éstos costos

Figura 14: Estructura del modelo de transporte


Construcción de pasos

El procedimiento de construcción de pasos determina las *formas de viajar* desde una zona de origen a una de destino a través de un modo de transporte determinado. Ello significa que cada paso no es simplemente una ruta física en la red de transporte, sino una combinación de enlaces y operadores a través de los cuales es posible realizar un viaje entre un origen y un destino. De tal manera que puede haber dos pasos que sigan una misma secuencia de enlaces, pero en un operador o combinación de operadores distinta.

Un paso, entonces, queda descrito de la siguiente forma:

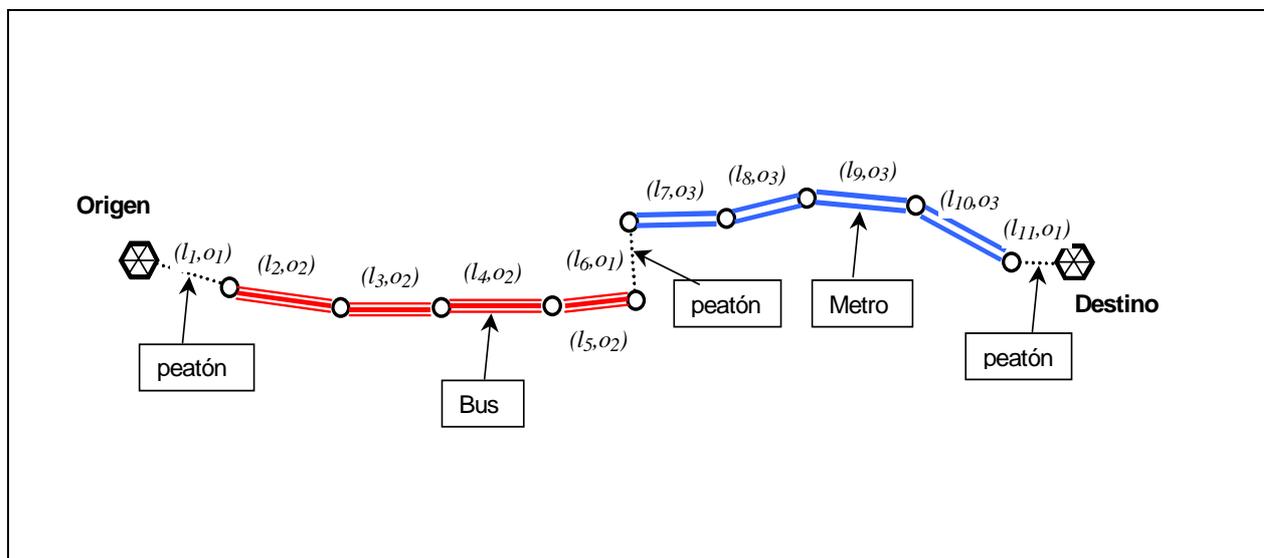
$$m_1, m_2, m_3, \dots, m_z$$

$$m_i = (l_i, o_i)$$

m_i es la combinación de un enlace l_i y un operador (o ruta) o_i en la secuencia del paso. El nodo de origen de l_1 es el centroide de la zona de origen, en la cual se aborda el operador o_1 . En la secuencia del paso puede ocurrir un cambio de operador, lo que introduce los *transbordos* a la búsqueda. El nodo de destino de l_z corresponde a la zona de destino del viaje.

En la Figura 15 se ilustra un ejemplo de paso multimodal. La opción incluye 11 enlaces físicos y tres operadores (peatón, autobús y metro). Los viajeros que seleccionan este paso caminan a la parada para tomar el autobús; luego bajan de éste y caminan a la estación para abordar el metro; finalmente caminan hasta el destino del viaje.

Figura 15: Ejemplo de paso multimodal



En el proceso de búsqueda, el algoritmo calcula el costo generalizado de los pasos, acumulando los siguientes elementos de cada enlace/operador m que forma parte de la secuencia:

$$c_{ijp}^{ks} = \sum_{m=1}^z RT_m^s + RD_m^s + TR_m^s, \quad (31)$$

donde:

c_{ijp}^{ks} costo generalizado del paso p entre i y j por el modo k para la categoría s

RT_m^s tarifa por unidad de tiempo en la combinación enlace/operador $m(l,o)$ para la categoría s

RD_m^s tarifa por unidad de distancia en la combinación enlace/operador $m(l,o)$ para la categoría s

$TR_{m-1,m}^s$ tarifa que se paga al abordar el operador o en el enlace l ; puede ocurrir al inicio del viaje o al realizar una transferencia, cuando $o(m-1) \neq o(m)$.

A continuación se incluye la descripción de cada uno de estos elementos, cuyos componentes, en su mayoría, fueron ya definidos en la descripción de tarifas y costos de operación.

Costos al usuario relativos al tiempo de viaje

Incluye costos monetarios y no monetarios de acuerdo a la siguiente ecuación

$$RT_m^s = tv_m \left(tt_o + \frac{ct_o tc_o}{to_o} \right) pc_o^s + tv_m (vv^s pt_m pg_o pp_o^s), \quad o \in m \quad (32)$$

donde

RT_m^s	costo del tiempo percibido por el viajero de categoría s en la combinación enlace/operador $m(o,l)$
tv_m	tiempo de viaje del usuario en el operador o en el enlace l , el cual es función de la distancia del enlace y la velocidad del operador;
tt_o	tarifa que el operador o cobra por unidad de tiempo;
ct_o	costo de operación por unidad de tiempo del operador o ;
tc_o	proporción del costo de operación que el operador o transfiere al usuario;
to_o	tasa de ocupación del operador o ;
vv^s	valor del tiempo de viaje del usuario de categoría de transporte s ;
pt_m	penalización al operador o en el enlace l
pc_o^s	proporción de los costos que se cargan al usuario de categoría s
pg_o	penalización global por uso del operador o , equivalente a la constante modal
pp_o^s	penalización del operador o asociada a la categoría de demanda s .

La primera parte de la ecuación contiene los elementos monetarios; la segunda tiene los componentes no monetarios, la parte subjetiva o percibida del costo generalizado. Incluye el valor del tiempo y tres penalizaciones, asociadas al tipo de enlace, al operador y a la categoría de usuario.

La penalización asociada al tipo de enlace, se utiliza para reflejar las condiciones de la infraestructura y su calidad en términos de seguridad, disponibilidad de servicios y otros. En algún caso donde tome el mismo tiempo y costo viajar por dos vías alternativas, el usuario preferirá la que esté mejor señalizada e iluminada y tenga servicios de emergencia; estos aspectos no modelables se expresan con una penalización menor para la mejor vía.

De la misma forma, la penalización asociada al operador se utiliza para representar distintas condiciones de los vehículos en cuanto a comodidad y seguridad, así como la regularidad del servicio que presta. En la práctica, se asigna el valor 1 al mejor operador y valores ligeramente mayores al resto de las opciones.

Finalmente, la penalización del operador para una categoría de demanda específica, se utiliza para expresar la preferencia de determinados usuarios. Este parámetro permite, en ciertas aplicaciones, omitir la etapa de separación modal y realizarla conjuntamente con la asignación. Con el uso de esta penalización, la disponibilidad vehicular de las categorías de demanda no sería un dato para el modelo, pasa a ser un resultado de la calibración.

Costos al usuario relativos a la distancia

Tienen dos componentes, la tarifa por unidad de distancia de los operadores y los costos de operación que éstos transfieren a los usuarios:

$$RD_m^s = d_l \left(td_o + \frac{cd_m tc_o}{to_o} \right) pc_o^s, \quad l, o \in m \quad (33)$$

donde

RD_m^s	costo al usuario por unidad de distancia en la combinación enlace/operador m
d_l	distancia del enlace l
td_o	tarifa que el operador o cobra por unidad de distancia
cd_m	costo de operación por unidad de distancia del operador o en el enlace l
tc_o	proporción del costo de operación que el operador o transfiere al usuario
to_o	tasa de ocupación de los vehículos del operador o
pc_o^s	proporción de los costos que se cargan al usuario de categoría s

Costos de Transferencia

Monto que el usuario paga al abordar un vehículo del operador, bien sea al inicio del viaje o al realizar un transbordo. Además de la tarifa, puede incluir costos que el operador traspasa al usuario y el valor del tiempo de espera percibido. Se computa a lo largo de un paso cada vez que hay un cambio de operador o ruta.

$$TR_m^s = \left(tf_o + \frac{cf_o tc_o}{to_o} \right) pc_o^s + te_m ve^s ; \quad o \in m \quad (34)$$

donde

TR_m^s	Costo percibido por el usuario de la categoría s al abordar al operador o en el enlace l ; bien sea su primer abordaje o una transferencia
tf_o	tarifa fija que el operador o cobra al abordar a un viajero proveniente del operador en la combinación $m-l$
cf_o	Costo fijo de operación del operador o
tc_o	Proporción del costo de operación que el operador o transfiere al usuario
te_m	tiempo de espera por una unidad del operador o en el enlace l
ve^s	valor del tiempo de espera de la categoría de transporte s
pc_o^s	proporción de los costos que se cargan al usuario de categoría s

Los costos fijos de abordaje están en una matriz de la forma *desde operador - hacia operador* lo cual permite definir tarifas integradas entre operadores o prohibir la transferencia entre algunos de ellos. Un caso típico de integración ocurre entre el metro y sistemas de autobuses alimentadores. En cuanto a la prohibición de transferencias, un caso ilustrativo ocurre cuando se incluye la bicicleta como operador en la aplicación; si no existen estacionamientos para ellas se prohíbe la transferencia hacia el autobús.

Los tiempos de espera dependen de dos elementos: la frecuencia del servicio y la relación demanda/capacidad de los vehículos. La frecuencia determina el tiempo de espera mínimo por una unidad del operador; a medida que el servicio se congestiona el tiempo de espera se incrementa. A red vacía, y asumiendo una llegada aleatoria de pasajeros, el tiempo de espera promedio viene dada por:

$$te_m = te \min_o + \frac{1}{2f_o}, \quad o \in m \quad (35)$$

donde:

te_m es el tiempo de espera por una unidad de la ruta u operador o ; cuando en determinado enlace l la relación de demanda/capacidad es muy baja, la espera promedio es el inverso de la frecuencia de la ruta; a mayor demanda, la espera varía en la función de restricción de capacidad

te_{min_o} es el tiempo de espera fijo por un vehículo de operador o , adicional al relativo a la frecuencia de la ruta

f_o es la frecuencia de la ruta correspondiente al operador o (vehículos por unidad de tiempo)

El modelo reconoce dos tipos de rutas, dependiendo si cumplen o no un horario fijo (*scheduled*). En el primer caso, solo aplica la espera mínima, dato que debe ingresarse como el promedio de espera por este tipo de servicio, normalmente los interurbanos de baja frecuencia. En caso contrario, aplican ambos términos de la ecuación y la restricción de capacidad. El modelo permite simular operadores de transporte público sin rutas determinadas, en cuyo caso la frecuencia se calcula internamente en función a la demanda.

Adicionalmente, el modelo permite definir un rango de frecuencias para cada operador y estima la frecuencia que maximice el ingreso del operador. f_o es la frecuencia resultante.

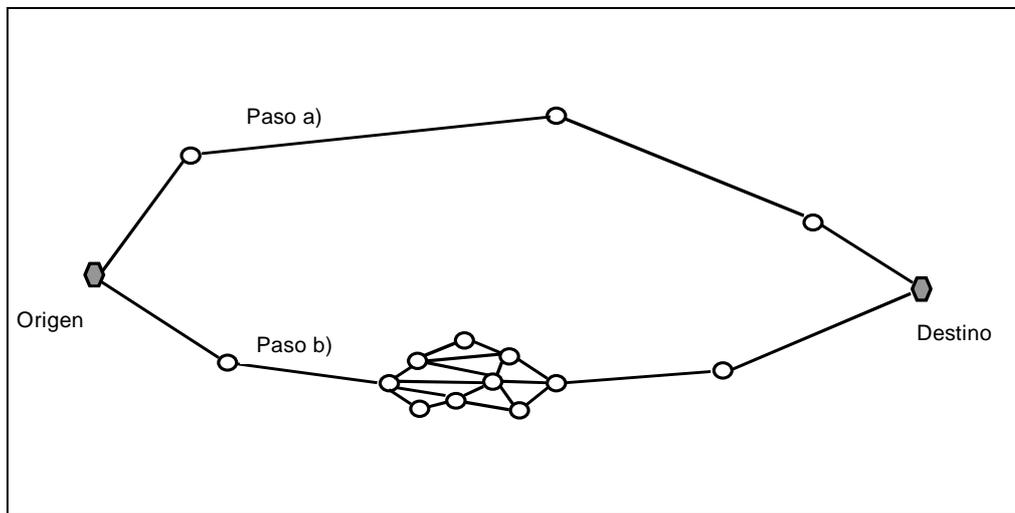
Algoritmo de búsqueda de pasos

En la búsqueda de pasos intervienen los elementos de costo descritos en las secciones previas. Para cada par origen-destino el algoritmo analiza todas las combinaciones enlace/operador o enlace/ruta que constituyan una opción razonable de viaje. En redes muy densas pueden resultar muchas opciones, por lo cual el algoritmo limita la búsqueda a las primeras n , utilizando dos criterios:

- los que tengan el menor costo generalizado
- los que constituyen verdaderas opciones distintas, no pequeñas variantes

El primer criterio es obvio, pero por si solo conduciría a un conjunto de pasos muy similares que pueden considerarse como variantes de una misma opción. En la Figura 16 se ilustra un caso típico que tiene dos opciones reales (a y b) para viajar entre dos zonas, donde la opción b permite ciertas variantes. Si se considera $n=6$ pasos, pueden resultar 6 variantes de la opción b con menor costo generalizado que la opción a, verdadera alternativa que quedaría fuera del conjunto de pasos.

Figura 16: ejemplo de pasos solapados en una red simple



Para controlar la independencia de las opciones y evitar las altamente correlacionadas, el algoritmo de búsqueda de pasos incluye un *control de solapes*, diseñado especialmente para el sistema TRANUS. El método lleva cuenta de las coincidencias (enlace/operador) en la secuencia de cada paso para descartar los que no representan una verdadera opción alternativa claramente diferenciada.

El control de solapes se realiza mediante un factor de penalización denominado *Factor O_z* , que es un número positivo que se define en función del nivel de dispersión requerido para los pasos del modo k en la aplicación (urbana o regional). Con dicho factor, el modelo procede de la siguiente forma:

- busca el paso mínimo de i a j por el modo k y lo guarda;
- penaliza por O_z los componentes de costo al usuario relativos al enlace (excluye los de transferencia) de todas las combinaciones enlace/operador m que forman parte del paso;
- regresa iterativamente al punto a) hasta que:
- el paso mínimo encontrado en a) sea idéntico a alguno de los pasos previamente encontrados; este paso se descarta y termina la búsqueda.

Si $O_z=1$ el paso mínimo encontrado emergerá de nuevo como paso mínimo en la segunda búsqueda, lo que implica obtener un solo paso entre cada par O-D. A medida que O_z es mayor, aumenta el número de pasos diferenciados que pueden surgir. Nótese que un mismo enlace-operador puede emerger varias veces, y cada vez será penalizado de nuevo. Como resultado se obtienen n pasos tales que su costo generalizado es menor que el del mínimo multiplicado por O_z .

Cada categoría de demanda le da un valor distinto a los pasos, pero buscar un conjunto de pasos para cada combinación categoría/modo tiene un alto costo de computación. Por tanto, el algoritmo busca un conjunto de pasos por modo, y el modelista debe asegurarse de que las opciones razonables estén presentes en el resultado. El modelo de transporte aplicará luego a cada opción las variables de costo y desutilidad específicas a cada categoría de demanda.

Desutilidades y probabilidades

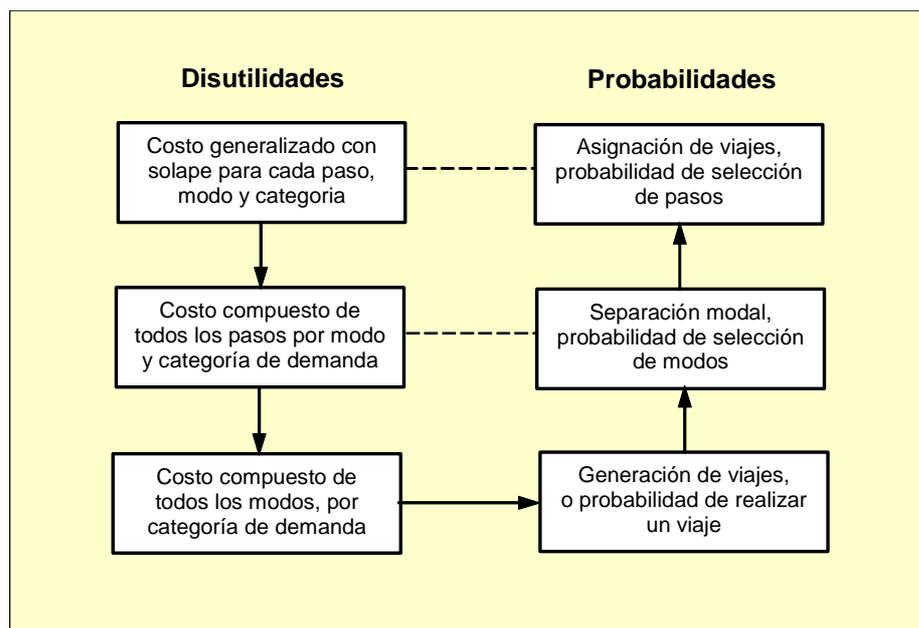
La desutilidad es una medida de accesibilidad que incide tanto en las decisiones de localización como en las de transporte. El costo monetario de viaje, en cambio, forma parte de la estructura de precios de los sectores en el modelo de uso del suelo. El modelo lleva cuenta separada del costo monetario y la desutilidad.

Hay tres niveles de decisión en el modelo de transporte: selección de pasos, selección de modos y del número de viajes a realizar en el período considerado. Las decisiones se simulan con modelos logit multinomial escalado, los cuales se vinculan entre sí por las desutilidades, formando un conjunto jerárquico de logit anidados.

La Figura 17 muestra la secuencia en que se realiza el cálculo de desutilidades. El proceso se inicia con las desutilidades de los pasos que es el nivel inferior en la cadena de decisiones. La desutilidad de un paso es su costo generalizado, cuya forma de cálculo se describe en secciones anteriores. Los costos generalizados de los pasos competitivos se agregan para formar la desutilidad al nivel de modo para cada categoría de demanda, la cual interviene en la selección modal. Finalmente, se realiza una nueva agregación de los modos competitivos para obtener la desutilidad al nivel de las categorías que interviene en la generación de viajes.

La etapa de selección modal es opcional en TRANUS, y puede omitirse totalmente en ciertas circunstancias. Esto es así por la naturaleza multimodal del algoritmo de asignación, el cual combina enlaces físicos con operadores y rutas (oferta operativa). Para determinada categoría de demanda, tal como personas de cierto nivel de ingreso, es posible especificar un solo modo, que puede abarcar diversos operadores de transporte disponibles a los pasajeros: automóviles, autobuses, trenes, etc. Varias de estas modalidades de la oferta operativa (o todas) pueden combinarse en la red física para formar una opción de viaje con transferencias, con ciertas restricciones. Por ejemplo, la transferencia entre automóvil y tren sólo se permite donde hay facilidades para ello (park-and-ride). Este tipo de configuración es muy recomendable cuando la disponibilidad vehicular es alta en el área de estudio.

Figura 17: Desutilidades y probabilidades de decisión



Selección de pasos

Como se mencionó, la desutilidad de cada paso es su costo generalizado, a partir del cual se calcula el costo *con solape*. El control de solapes tiene una primera función en el algoritmo de búsqueda de pasos, para producir un conjunto de opciones diferenciadas. En el modelo de transporte, posteriormente, dicho control compensa por la correlación (o solape) entre las opciones resultantes.

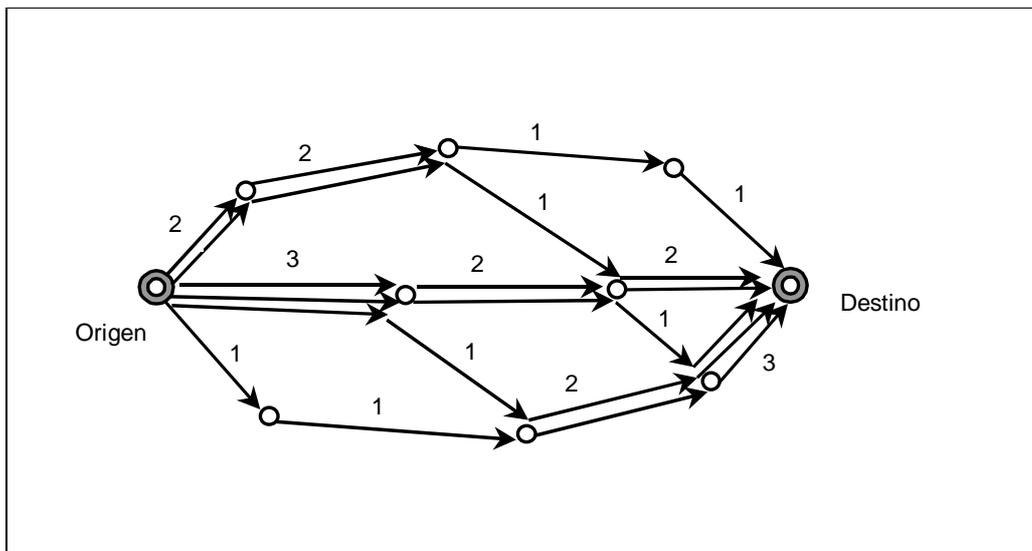
La probabilidad de que viajeros pertenecientes a la categoría s escojan el paso p para viajar entre las zonas i y j por el modo k , viene dado por el siguiente modelo multinomial escalado:

$$P_{ijp}^{ks} = \frac{\exp(-\gamma^s \tilde{c}_{ijp}^{ks})}{\sum_p \exp(-\gamma^s \tilde{c}_{ijp}^{ks})}, \quad (36)$$

Donde \tilde{c}_{ijp}^{ks} es el costo generalizado, *escalado* y compensado por solape, y γ^s es el parámetro de dispersión en el modelo logit de selección de pasos. Si es γ^s grande (en valor absoluto), se escogerá mayoritariamente el paso mínimo; si es muy pequeño, los viajes se repartirán en proporciones similares entre los pasos disponibles.

La compensación por solape en el modelo de asignación resulta de multiplicar los costos (monetarios y por tiempo) de cada enlace/operador por el *factor de solape*, siendo éste el número de enlaces/operadores que comparte el paso con los demás. La Figura 18 muestra un ejemplo simple de los factores de solape en un conjunto de 6 pasos.

Figura 18: Ejemplo de solapes en el proceso de asignación



\dot{c}_{ijp}^{ks} es el costo generalizado del paso p compensado por solape, la desutilidad escalada del paso será:

$$\tilde{c}_{ijp}^{ks} = \frac{\dot{c}_{ijp}^{ks}}{\left(\min_p(\dot{c}_{ijp}^{ks}) \right)^{\theta^s}}, \quad (37)$$

θ^s es el nivel de escala de la función de utilidad. La desutilidad del modo k para los usuarios de la categoría s se define por agregación como el costo compuesto de las opciones de pasos, de la siguiente forma:

$$\tilde{c}_{ij}^{ks} = -\frac{\ln P_g^{ks}}{\gamma^k} \left(\frac{\min(\tilde{c}_{ijp}^{ks})}{p} \right)^{\theta^s}, \quad (38)$$

La ecuación va multiplicada por el costo mínimo para retornar el valor escalado utilizado en la probabilidad al valor original.

P_g^{ks} se define como una serie de la siguiente forma:

$$P_g^{ks} = \sum_p G_p \prod_{h=1}^{p-1} (1 - G_h), \quad (39)$$

Donde la función G_p es el numerador del modelo logit de la ecuación (36).

$$G_p = \exp\left(-\gamma^s \tilde{c}_{ijp}^{nk}\right), \quad (40)$$

Separación modal

Como fue dicho, la separación modal puede omitirse, sustituyéndola por penalizaciones categoría-operador en la etapa de asignación. Pero si en la aplicación se hubiera definido este nivel de decisión, se realiza el siguiente proceso de cálculo:

La probabilidad de que un viajero perteneciente a la categoría s escoja un modo k para viajar entre i y j se calcula con las desutilidades de cada modo \tilde{c}_{ij}^{ks} aplicadas en el siguiente modelo un logit multinomial:

$$P_{ij}^{ks} = \frac{\exp\left(-\lambda^s \left(\tilde{c}_{ij}^{ks} / (\min_k(\tilde{c}_{ij}^{ks}))^{\theta^s} \right)\right)}{\sum_k \exp\left(-\lambda^s \left(\tilde{c}_{ij}^{ks} / (\min_k(\tilde{c}_{ij}^{ks}))^{\theta^s} \right)\right)}, \quad k \in K^s \quad (41)$$

donde

λ^s parámetro de dispersión en la distribución logit del modelo de selección modal para la categoría de transporte s ; Si es muy alto, la categoría adoptará mayoritariamente el modo de menor desutilidad; en cambio, si se acerca a cero, los viajes de la categoría s se distribuirán en proporciones similares entre los modos disponibles

K^s es el conjunto de modos k disponibles a la categoría s ; las mercancías sólo tienen disponibles los modos de carga, y las personas los de pasajeros

\tilde{c}_{ij}^{ks} es la desutilidad del modo k ;

$\min_k(\tilde{c}_{ij}^{ks})$ es la desutilidad del mejor modo;

θ^s es el parámetro que regula el grado de escala de desutilidades para la categoría s

Finalmente, la desutilidad de los viajeros de la categoría s al viajar de i a j se estima agregando sobre todos los modos disponibles a dicha categoría:

$$\tilde{c}_{ij}^s = -\frac{\ln P_g^s}{\lambda^s} \left(\min_k (\tilde{c}_{ij}^{ks}) \right)^{\theta^s}, \quad k \in K^s \quad (42)$$

La ecuación se multiplica por la desutilidad del mejor modo para eliminar la escala

P_g^s se define como una serie de la siguiente forma:

$$P_g^s = \sum_k G_k \prod_{h=1}^{k-1} (1 - G_h), \quad (43)$$

Donde la función G_k es el numerador del modelo logit de la ecuación (41).

$$G_k = \exp\left(-\lambda^s \tilde{c}_{ij}^{ks}\right), \quad (44)$$

Generación de viajes

El objetivo del submodelo de generación es estimar el número de viajes que se derivan de los flujos funcionales provenientes del modelo de localización de actividades, los cuales, individuales o agregados, forman las categorías de transporte. El número de viajes que genera una categoría s entre un par de zonas en el período de análisis, varía en función de la desutilidad, calculada en la ecuación (41). Se estima con un modelo de generación elástico que tiene la forma de una curva de demanda:

$$T_{ij}^s = F_{ij}^s \left[v_{\min}^s + \left(v_{\max}^s - v_{\max-\min}^s \right) \exp\left(-\eta^s c_{ij}^s\right) \right], \quad (45)$$

donde:

F_{ij}^s flujos por entre i y j de la categoría de transporte s

v_{\min}^s cantidad mínima de viajes que realiza la categoría s independiente de la desutilidad

v_{\max}^s cantidad máxima de viajes que realiza la categoría s cuando la desutilidad tiende a cero

η^s elasticidad de la categoría s a la desutilidad del viaje

En cada iteración, la desutilidad aumenta por efecto de la congestión, y el número de viajes entre cada par O-D disminuye, en mayor o menor medida, según sea la elasticidad de la categoría. Cuando el sistema converge a un equilibrio, la diferencia entre el número de viajes estimados en la primera y la última iteración es la demanda reprimida, el número de viajes que dejan de realizarse por efectos de la congestión.

Separación modal

La separación modal consiste en la estimación del número de viajes de la categoría de transporte s que utiliza cada modo k , a partir de las probabilidades de cada modo de la ecuación (46) y los viajes totales por categoría calculados en (45).

$$T_{ij}^{ks} = T_{ij}^s P_{ij}^{ks} \varphi^s + (1 - \varphi^s) B^k, \quad B^k = \begin{cases} 1 & \text{si } k \text{ es publico} \\ 0 & \text{si } k \text{ no es publico} \end{cases} \quad (46)$$

donde:

φ^s tasa de disponibilidad vehicular de la categoría de transporte s

Nótese que la probabilidad se aplica sólo a los que disponen de vehículo; los cautivos del transporte público sólo pueden escoger entre los modos públicos. Se utiliza la *disponibilidad vehicular* y no la *propiedad vehicular*. La propiedad de vehículo no implica su disponibilidad; es posible no ser propietario de auto pero disponer del auto familiar o compartir el viaje con vecinos o compañeros de trabajo; estas personas no demandan transporte público, a pesar de no ser propietarias de automóvil.

Asignación de viajes

En el proceso de asignación, los viajes por categoría y modo, se asignan a los pasos disponibles mediante un modelo multinomial logit, utilizando la probabilidad de cada paso calculada en la ecuación (36) y los viajes por modo de (46), de la siguiente forma:

$$T_{ijp}^{ks} = T_{ij}^{ks} * P_{ijp}^{ks}, \quad (47)$$

Finalizado el proceso de asignación, el modelo calcula y presenta los siguientes resultados:

T_m demanda por la combinación m (arco l operador o), en toneladas o pasajeros

V_m número de vehículos que utiliza la combinación m (arco l operador o); se obtiene aplicando tasas de ocupación de cada operador

VE_l el número de vehículos equivalentes en el enlace l estimado a partir de las tasas de equivalencia de los vehículos de cada operador en el enlace.

$$VE_l = \sum_m V_m eq_m, \quad (48)$$

donde eq_m es la tasa de vehículos equivalentes para la combinación enlace operador m (operador o , enlace l)

Nótese que las tasas de equivalencia de las unidades de cada operador varían según el tipo de enlace, por ello llevan el subíndice m que corresponde a una combinación enlace/operador.

Los operadores del transporte público reciben un tratamiento especial; si se ha especificado una frecuencia de las rutas de un operador, el número de vehículos en las rutas es un dato que el modelo asignará a los enlaces correspondientes, independiente de la demanda. En cambio, si se ha dejado la frecuencia indeterminada, el modelo calcula la frecuencia que requiere la demanda estimada, a partir de las tasas de ocupación to :

$$f_m = \max_l \frac{T_m}{to^o}, \quad \text{si la frecuencia es indeterminada,} \quad (49)$$

Donde to^o es la tasa de ocupación del operador o . La capacidad q de un operador o ruta en el enlace es:

$$q_m = f_m * to^o \quad , \quad (50)$$

El nivel de utilización de las unidades se define como la relación demanda/capacidad de los vehículos del operador en cada enlace:

$$dc_m = \frac{T_m}{q_m} \quad (51)$$

La relación demanda capacidad total en cada enlace se calcula dividiendo los vehículos equivalentes que lo comparten entre la capacidad física, que es un dato.

$$DC_l = \frac{VE_l}{Q_l} \quad , \quad (52)$$

También se presenta como resultado de la asignación, las velocidades y tiempos de espera de los operadores y el nivel de servicio en que funciona cada enlace, los cuales se calculan en el proceso de restricción de capacidad que se describe a continuación.

Restricción de capacidad

La restricción de capacidad tiene por objeto equilibrar la demanda de viajes con la oferta, tanto física como operativa. El elemento principal del equilibrio es la variable tiempo, la cual afecta los costos. En cada iteración, el modelo compara la demanda con la capacidad y ajusta los tiempos, hasta que se logre una condición de equilibrio.

La restricción de capacidad tiene dos componentes principales:

- Restricción de la velocidad y tiempos de circulación de los vehículos en las vías
- Restricción de los tiempos de espera en pasajeros de transporte público

En el primer caso, la velocidad de circulación de los vehículos se ajusta de acuerdo al nivel de congestión de cada enlace de la red. En los enlaces que se hayan definido con capacidad indeterminada no se realiza este ajuste, por tanto, se mantiene constante la velocidad inicial de los operadores y un nivel de servicio A (a flujo libre) en todas las iteraciones del modelo de transporte. Durante el proceso, el modelo estima los vehículos en cola y la propagación de la cola hacia los enlaces precedentes.

En el segundo caso, cuando la demanda de pasajeros en el transporte público se acerca a la capacidad de servicio en determinadas rutas, el modelo estima el incremento correspondiente en el tiempo de espera de los viajeros.

En condiciones de equilibrio, el número de vehículos no sobrepasará la capacidad vial en ningún enlace, y los pasajeros de transporte público no superará la oferta de este servicio.

Restricción de capacidad de los vehículos en las vías

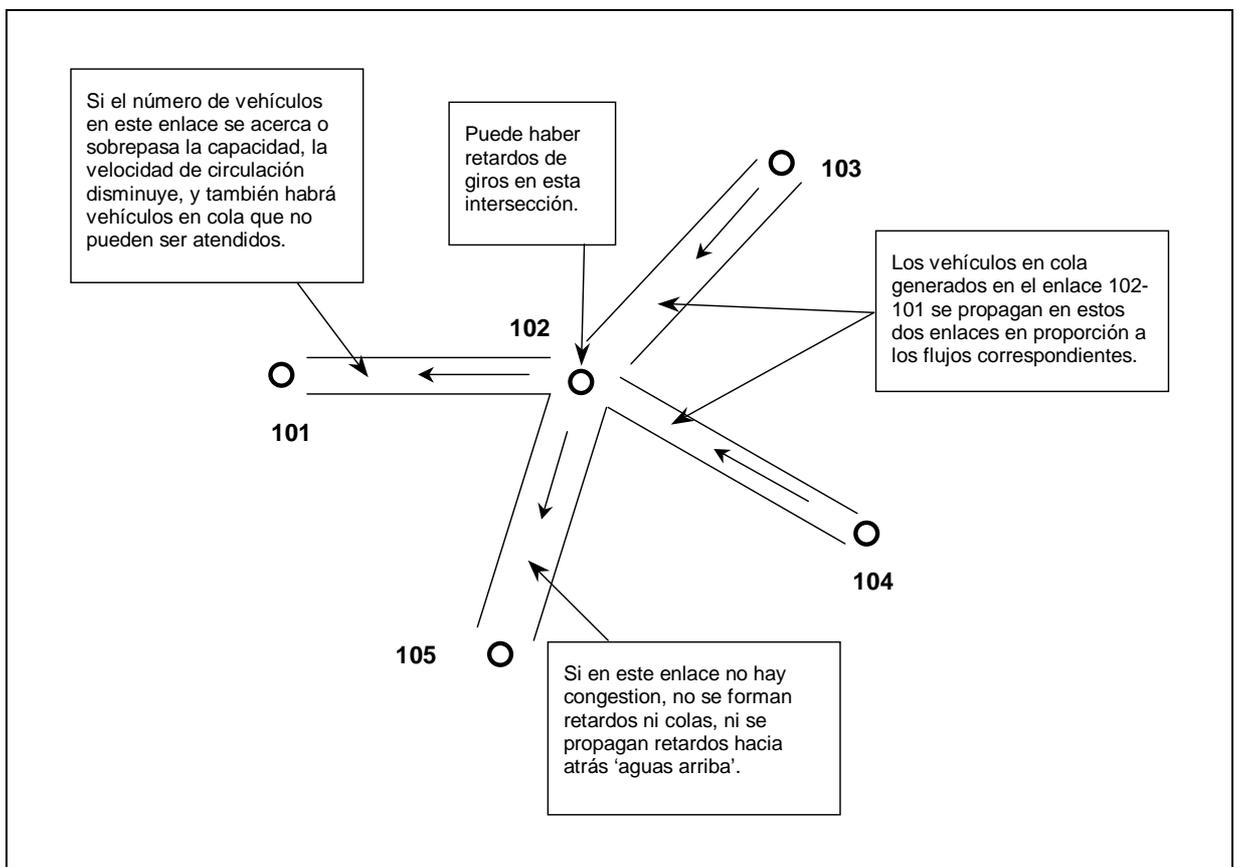
La Figura 19 ilustra los principales componentes de la restricción de capacidad de los vehículos en las vías, con un ejemplo simplificado. Allí se muestran dos enlaces que convergen en la intersección 102 del cual parten otros

dos enlaces, todos unidireccionales. Cuando el volumen asignado al enlace 102-101 se aproxima a la capacidad, se producen las siguientes acciones:

- La velocidad de circulación en el enlace 102-101 disminuye
- Habrá cierto número de vehículos en el enlace 102-101 que están en cola
- La cola en el enlace 102-101 se ‘propaga’ hacia los enlaces que lo preceden, es decir, en los enlaces 103-102 y 104-102

En este último caso, si el flujo de vehículos que transita por el enlace congestionado 102-101 proviene mayoritariamente del enlace 103-102, el retardo propagado será mayor que en el enlace 104-102.

Figura 19: Restricción de capacidad en la vía



Para calcular la velocidad de un enlace l , se usa su volumen asignado, aumentado por los vehículos que están en cola debido a la congestión de enlaces posteriores, pues el efecto de la congestión se transmite hacia atrás:

$$\text{VolAum}(l) = \text{Vol}(l) + \text{Cola}(l), \quad (53)$$

El tiempo de cola de un enlace se calcula aproximando una curva Poisson usando los siguientes parámetros para un enlace l :

Tiempo de Servicio	$S = 1/ \text{VelocidadFlujoLibre}(l)$
Frecuencia de llegada	$F = \text{Flujo}(l)$
Servidores	$C = \text{Capacidad}(l)$

El tiempo de cola es:

$$Q(l) = \text{Poisson}(S, F, C)$$

El tiempo de cola en un paso se multiplica por el valor del tiempo de viaje de la categoría en cuestión:

$$\text{Cost}(P) = \text{Transferencias}(P) + \text{SUM}(\text{Cost}(l) + Q(l) * \text{VOT}(ip)), \text{ para todo } l \text{ en } P.$$

Adicionalmente, se calcula el número de vehículos en cola de la siguiente manera:

$$\text{Congestión}(l) = Q(l) * \text{Flujo}(l)$$

Los valores $Q(l)$ y $\text{Congestión}(l)$ son promediados entre iteraciones para facilitar la convergencia.

Sólo para efectos de la restricción de capacidad, el número de vehículos en cola de un enlace l se suma a los enlaces concurrentes en proporción al flujo que aportan a l . Si se está asignando el volumen $\text{Assig}(P)$ al paso P , entonces:

$$\begin{aligned} \text{Proporción}(k,l,P) &= \text{Assig}(P) / \text{Flujo}(l), \text{ si } l \text{ sigue a } k \text{ en el paso } P \\ &= 0 \text{ si no} \end{aligned}$$

$$\text{Cola}(k) = \text{SUM}(\text{Proporción}(k,l,P) * \text{Congestión}(l)), \text{ para todo paso } P$$

El ajuste de la velocidad de todos los operadores que utilizan un enlace l se realiza con el siguiente grupo de ecuaciones que definen una secante hiperbólica, donde el subíndice m expresa una combinación enlace-operador (lo).

$$V_m^{it} = V_m^0 - \text{sech} \left[\rho (DC_l)^\beta \right], \quad (54)$$

$$\rho = \text{sech}^{-1}(1 - \alpha), \quad (55)$$

$$\beta = \frac{\ln \left[\frac{\text{sech}^{-1}(v)}{\rho} \right]}{\ln \gamma}, \quad (56)$$

donde:

V_m^{it} Velocidad del operador o en el enlace l , en la iteración it

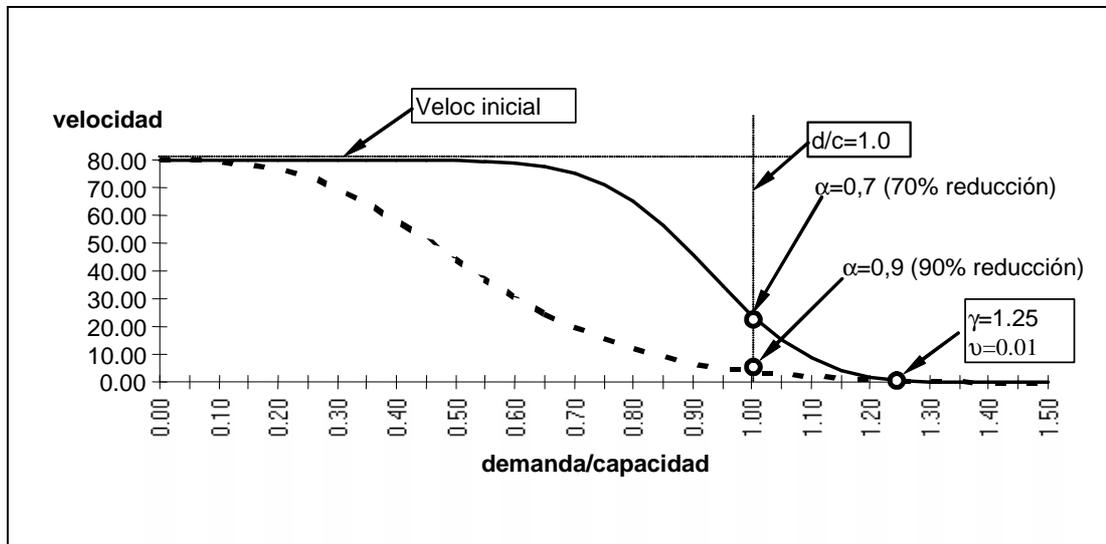
V_m^0 Velocidad inicial del operador o (iteración 0) en el enlace l , a flujo libre

- DC_l Relación demanda/capacidad en el enlace l ; los vehículos en cola están agregados a la demanda
- α proporción en que se reduce la velocidad inicial cuando $DC = 1$
- ν proporción en que se reduce la velocidad inicial cuando $DC = \gamma$
- γ DC en la cual la velocidad se reduce al $\nu\%$ de la velocidad inicial ($\gamma > 1$)

El valor de DC, la relación demanda/capacidad, se calcula dividiendo el número de vehículos equivalentes asignados al enlace (más los vehículos en cola) por la capacidad del enlace, también en términos de vehículos equivalentes. Los parámetros α ν γ son los datos que debe proveerse al modelo para la realización de los cálculos.

La Figura 20 muestra gráficamente la función para un operador cuya velocidad inicial es de 80 Km./hr, habiéndose ingresado $\gamma = 1.25$. La curva de línea continua corresponde a un $\alpha = 0,7$ y la de línea punteada ilustra el cambio en la forma de la función cuando se cambia α a 0,95. En ambos casos ν se fijó en 0.01. Como puede observarse, cuando $DC=0$ se mantiene la velocidad inicial, la reducción es nula. A medida que DC aumenta, la velocidad se reduce y alcanza el valor de $(1 - \alpha)V_m^0$ cuando $DC=1$. Más allá del punto $DC=1$, la reducción continúa, hasta llegar al tercer punto prefijado en la curva cuando $DC = \gamma$ y la velocidad es νV_m^0 . Más allá de este punto la velocidad se hace asintótica a cero.

Figura 20: Función de reducción de velocidad por congestión



Hay dos razones principales para haber elegido esta formulación para la función de restricción de capacidad en Tranus:

- a) Produce resultados realistas según los numerosos estudios de campo disponibles. La curva se puede ajustar de manera muy cercana a las conocidas curvas BPR desarrolladas por el Transport Research Laboratory de Gran Bretaña.
- b) A diferencia de las curvas que se utilizan en diversos modelos, incluidas las curvas BPR, la formulación utilizada aquí se logra con una sola ecuación, y por lo tanto no presenta discontinuidades ni puntos de inflexión, lo cual facilita el proceso de convergencia.

Para facilitar aún más la convergencia, la velocidad que se utiliza en una nueva iteración $\tau+1$ es el promedio ponderado de la obtenida en el proceso de restricción de la iteración actual τ con las de la iteración anterior $\tau-1$.

$$V_m^{\tau+1} = V_m^{\tau-1} + (V_m^{\tau} - V_m^{\tau-1})/(1 + w) \quad , \quad (57)$$

En donde w es el peso dado a la iteración anterior $\tau-1$ con respecto a la iteración actual, $w \geq 0$. Puede verse que en esta ecuación si $w=0$ la velocidad es igual a la velocidad de la iteración actual; si $w=1$ ambas velocidades tienen el mismo peso resultando en un promedio simple; a medida que w crece >1 la velocidad converge a la de la iteración anterior. Si el modelo presenta problemas de convergencia, especialmente en las primeras etapas de calibración, se recomienda darle mayor peso a la iteración anterior utilizando valores relativamente altos de w , por ejemplo 3.

Ajuste de los tiempos de espera

En el proceso de restricción de capacidad también se ajustan los tiempos de espera para los pasajeros abordando unidades. A medida que el número de pasajeros abordando una ruta en determinado enlace, se acerca al número de puestos vacantes, el tiempo de espera se incrementa. Este es un procedimiento muy importante que le permite al modelo equilibrar la demanda con la oferta de los servicios de transporte público. De acuerdo a la teoría de colas, los tiempos de espera tienden a infinito cuando la relación demanda/capacidad tiende a uno. Sin embargo, aquí se utiliza una versión 'suavizada' de este modelo para facilitar la convergencia.

El cálculo se inicia con el tiempo de espera mínimo, que tiene dos componentes: un término constante y la frecuencia de la ruta. El término constante es el requerido para compra de tickets, recorrido en estaciones, etc. La frecuencia incide en el tiempo de espera cuando la ruta no tiene un horario fijo; en ese caso se presume que los pasajeros llegan aleatoriamente a la parada. Cuando la ruta tiene un horario estipulado, no se presume llegada aleatoria de los pasajeros, quienes, al saber exactamente la hora del próximo vehículo, llegarán a la parada justo a tiempo. El tiempo de espera mínimo queda definido de la siguiente manera.

Si el servicio tiene horario fijo:

$$MW_m = CW_m, \quad (58)$$

donde CW_m es el tiempo constante de espera por la ruta o en el arco l ;

Si el servicio no tiene horario fijo:

$$MW_m = CW_m + \frac{1}{2f_o}, \quad (59)$$

donde f_o es la frecuencia de la ruta o . Se presume que los pasajeros llegan aleatoriamente a la parada, por tanto el promedio de ellos debe esperar un tiempo equivalente al inverso de la frecuencia de la ruta, además del tiempo constante.

Cuando en determinado enlace, la demanda se acerca al número de puestos libres que llegan, aumenta la probabilidad de que algunos pasajeros no puedan abordar el próximo vehículo. Con los pasajeros que continúan llegando aleatoriamente se acumula la demanda en la parada y el tiempo de espera se incrementa.

En la teoría de colas este fenómeno se inscribe en el típico caso denominado en inglés *bulk service*, es un servicio prestado a una tasa constante, con capacidad de atender simultáneamente a varios clientes que llegan de manera aleatoria. En Tranus se asume que la tasa de llegada de los pasajeros se distribuye exponencialmente, por lo que la demanda tiene una distribución Poisson. Se ρ_m es la relación demanda/puestos libres de la ruta o en el enlace l , el tiempo de espera total se calcula de la siguiente forma.

$$TW_m = MW_m + \frac{\rho_m}{1 - \rho_m} \frac{1}{f_o} \quad (60)$$

Nótese que a medida que la frecuencia f_o se incrementa, el tiempo total de espera se reduce. A medida que ρ_m se acerca a 1, el denominador $1 - \rho_m$ tiende a cero y la espera tiende a infinito. Esta función se caracteriza por tener una asíntota vertical en $\rho_m = 1$, y queda indefinida para valores de ρ_m superiores a 1, lo cual se puede observar en la Figura 21.

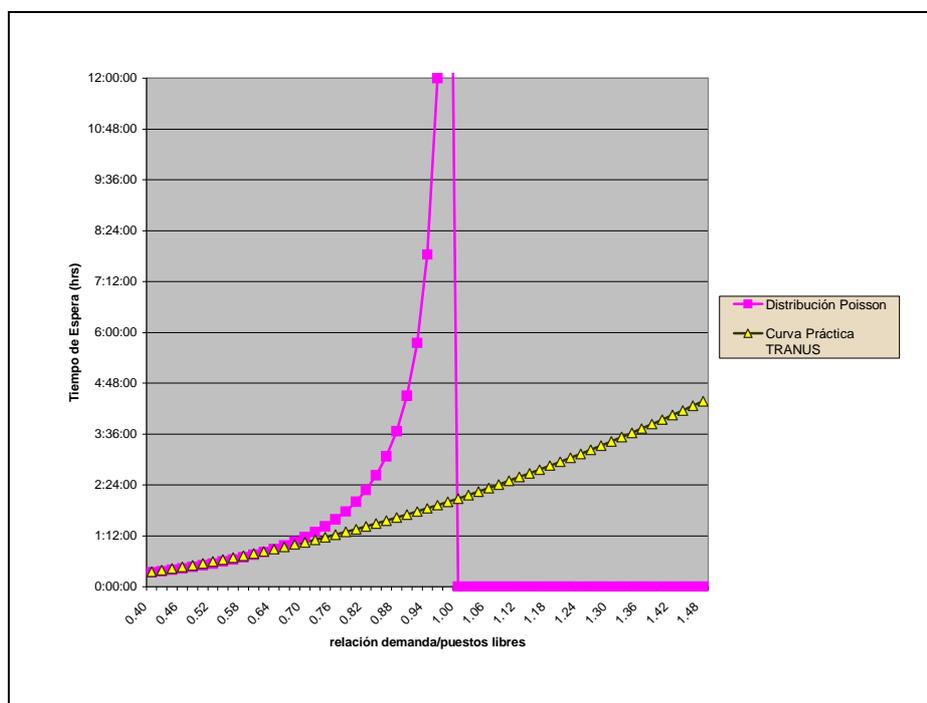
Este comportamiento no puede tratarse numéricamente en un modelo iterativo como TRANUS, ya que ρ_m puede alcanzar valores mayores que 1 en iteraciones intermedias o en una aplicación que aún no esté calibrada. Por razones prácticas, el modelo calcula la ecuación en forma de serie y la trunca. La Figura 21 compara la curva teórica con la aproximación usada por TRANUS. Esta última se acerca a la distribución Poisson hasta un valor de $\rho_m = 0.7$. A partir de ese punto la curva se aleja de la teórica y asume una forma monótonicamente creciente.

Al igual que se aplica a las velocidades, los tiempos de espera resultantes se promedian al final de cada iteración:

$$TW_m^{\tau+1} = \frac{TW_m^{\tau} + TW_m^{\tau-1}}{2} \quad (61)$$

Cada nueva iteración se inicia con nuevas estimaciones de costos y desutilidades, calculados en función a los tiempos ajustados de viaje y de espera. Esto tiene efecto en todas las etapas de la estimación de la demanda: generación de viajes, separación modal y asignación. Al final de la ejecución del modelo de transporte, los costos y desutilidades resultantes afectarán la localización de actividades del período siguiente de simulación.

Figura 21: Distribución de los tiempos de espera

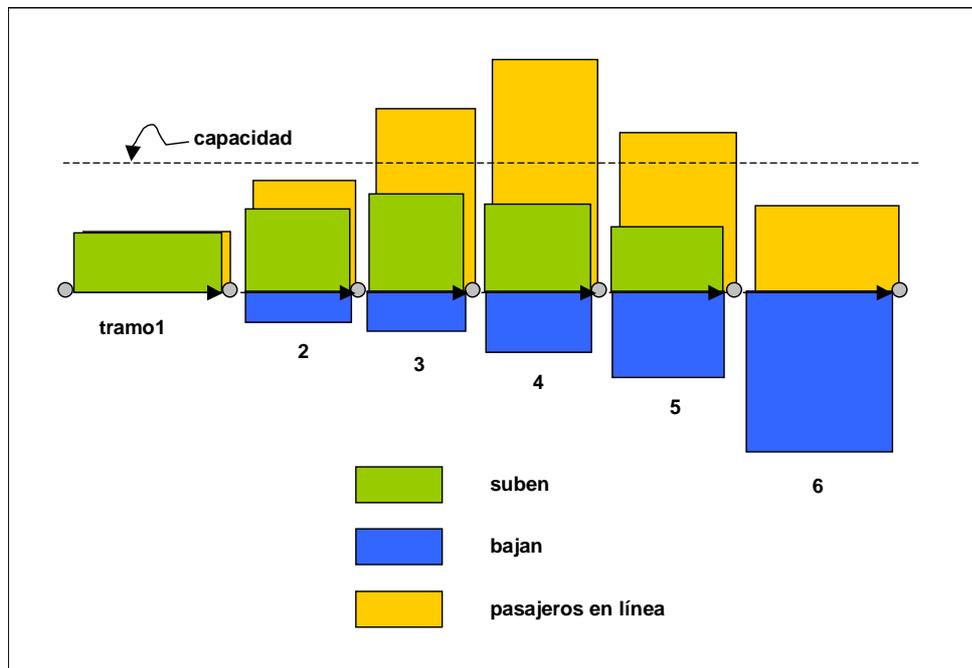


Oferta de Transporte Público

En Tranus, la frecuencia de los servicios de transporte público se define con un rango mínimo y máximo, $freq_{min}$ y $freq_{max}$. Cuando ambos valores son iguales, indica una frecuencia fija que el modelo no puede variar, independiente del nivel de saturación de las unidades. De lo contrario, cuando se especifica un rango, el modelo de transporte ajusta la oferta a la demanda al final de cada ciclo iterativo. Para ello, la frecuencia de servicio se modifica dentro del rango dado para la siguiente iteración. Las demás características de la ruta permanecen inalteradas, tales como el recorrido, tipo de unidad, tarifa, etc.

La frecuencia se expresa en número de unidades por hora. La capacidad de servicio de una ruta se calcula como el producto de la frecuencia por la capacidad máxima de cada unidad vehicular. Por ejemplo, si una ruta tiene una frecuencia mínima de 10 unidades por hora y cada unidad puede llevar 60 pasajeros, entonces su capacidad de servicio es de 600 pasajeros/hora.

En la primera iteración, el modelo de transporte supone que todas las rutas operan con la frecuencia mínima $freq_{min}$. Luego del proceso de asignación, el modelo estimará la demanda de pasajeros en cada tramo de la red para cada ruta, lo cual se denomina *perfil de demanda* de cada ruta. Puede ocurrir que en una ruta específica, la demanda asignada sea inferior a la capacidad de servicio en todos los tramos de la red. El caso contrario, es que la demanda asignada a una ruta supere su capacidad en uno o más tramos, tal como lo indica el ejemplo de la Figura 22. Allí puede verse que la ruta consta de 6 tramos. La figura indica los pasajeros que suben en cada tramo, los que bajan y el saldo que corresponde al número de pasajeros que se encuentran en línea. También se indica la capacidad de la línea, que es constante para todos los tramos. Puede verse que en los tramos 3, 4 y 5, la demanda en línea supera a la capacidad.

Figura 22: Ejemplo del perfil de demanda de una ruta de transporte público


Si la demanda supera la oferta, el modelo estima un incremento en los tiempos de espera de los pasajeros que intentan abordar la ruta en los tramos saturados, tal como se ha explicado en la sección anterior. Además, si existe una diferencia entre la frecuencia mínima y máxima, el modelo estima la posibilidad de que los operadores incrementen la oferta, si es que ello les conviene.

Para estimar la conveniencia para los operadores de incrementar la oferta, el modelo realiza el análisis que se describe a continuación.

- Cálculo del ingreso que percibe el operador de la ruta en cada tramo de la misma. Para ello se multiplica el número de pasajeros que suben por la tarifa de abordaje. Luego se suma el número de pasajeros en línea por la longitud de cada tramo y por la tarifa por distancia de la ruta. Esto da un ingreso total por tramo l , $Ingreso_l$. El ingreso total de la ruta será la sumatoria con respecto a todos los tramos, es decir, $IngresoTotal = \sum_l Ingreso_l$.
- Cálculo de la *frecuencia requerida* en cada tramo, es decir, el número de unidades que deberían pasar por hora para atender completamente la demanda del tramo. Esto se calcula dividiendo la demanda en línea en el tramo por el número de horas que opera el servicio y por la capacidad máxima de las unidades. Por ejemplo, si la demanda en línea es 14500 pasajeros diarios y la ruta opera con unidades de 60 pasajeros durante 12 horas diarias, la frecuencia del servicio en ese tramo deberá ser $14500/(60 \times 12) = 20.14$ unidades por hora.
- Cálculo de la *frecuencia ponderada* en cada tramo. Resulta de multiplicar el ingreso de cada tramo por la frecuencia requerida. Es decir, $FreqPonderada_l = Ingreso_l \times FreqRequerida_l$.
- Cálculo de la *frecuencia promedio ponderada*. Para esto se divide la suma de las frecuencias ponderadas por la suma de los ingresos. Es decir, $FreqPromedioPonderada = \sum_l FreqPonderada_l / \sum_l Ingreso_l$.

- e) La nueva frecuencia de la ruta se calcula dividiendo el resultado anterior por el nivel de ocupación de las unidades que un operador aspiraría lograr para incrementar la capacidad de servicio de la ruta. Dicho nivel se especifica entre 0 y 1 en el parámetro denominado '*Target Occupancy*', dato que debe ser aportado al modelo. La frecuencia resultante es
- $$FreqFinal = FreqPromedioPonderada / TargetOccupancy, \text{ sujeto a las restricciones } freq_{max} \geq FreqFinal \geq freq_{min}.$$
- f) Finalmente, para evitar oscilaciones en el proceso iterativo del modelo de transporte, las frecuencias resultantes en cada iteración se promedian con las de la iteración anterior.

Es importante destacar que en el modelo se presume que todos los servicios de transporte público operan con ganancia (o están cubiertos por subsidios); por ello, los costos de operación de la ruta no entran en el cálculo. El método descrito se basa en el criterio de que los operadores sólo incrementarán su oferta si el perfil de la demanda es conveniente. Por tanto, si la demanda supera a la capacidad en pocos tramos, en relación al total, puede ser no conveniente para el operador incrementar la oferta del servicio. Sin embargo, si hay una tarifa significativa por distancia, puede que incrementar la oferta resulte conveniente.

Convergencia

La convergencia del modelo de transporte se verifica en todos los enlaces por la diferencia porcentual entre una iteración y otra para dos variables: la velocidad de circulación de los operadores y el volumen total asignado. El proceso iterativo termina cuando la máxima diferencia en ambas variables está por debajo de un criterio de convergencia establecido, o cuando se alcanza el máximo número de iteraciones prefijado.

15/03/2007