THE COLOMBIAN STRATEGIC FREIGHT TRANSPORT MODEL BASED ON PRODUCT ANALYSIS

ABSTRACT

Freight transport modelling at interregional scale is relevant for planning issues. However, freight modelling processes are complex because it is not easy to define the relevant variables in the analysis, and to obtain the required information on freight movements through the network. These facts raise the need to adapt the modelling framework to each context.

This paper proposes a strategic national freight transport modelling framework developed as a variant of the traditional four-step modelling process with additional steps to estimate traffic flows from freight flows and to consider empty trips. The country of Colombia is used as the case study to implement and calibrate the proposed model. The data, data sources, and modelling methodologies used for each step are explained. In addition, data limitations and measures taken to complement the available data are discussed. From the implementation, the authors identify a set of advantages derived from the modelling approaches considered and suggestions for improvement.

KEY WORDS

Freight modelling; Colombia; strategic freight model;

1. INTRODUCTION

The study of the freight transport system has received increasing attention from researchers over the last few decades. These efforts have been devoted to fully understand the characteristics and components that make freight a highly complex system, and to develop adequate freight demand models that capture the key dynamics. Unfortunately, this has proven to be a very difficult endeavour. A key limitation is that freight demand modelling has been developed - to a great extent - based on theories from passenger transport modelling; yet, research has shown that relevant modelling attributes are rather different between freight and passenger transport [1]. Therefore, it is important to understand the key characteristics that make freight modelling such a unique task:

- Freight transport is a very dynamic process that changes constantly;
- Commodity flows are directly related to demand; however, freight trips - truck traffic - are an expression of vehicular supply [2, 3];
- There are multiple metrics to define/measure freight (e.g., value, amount transported, vehicle trips produced) [4, 5];
- There are multiple interacting and interrelated agents that define how, when and where the cargo is transported; each agent may only have a partial view of the system;
- Interactions among the agents usually take place in a private context; thus, travel demand data may be commercially sensitive [4];
- There are multiple transport modes (e.g., ground, air, water, pipelines);
- Freight exhibits a very broad range of opportunity costs [6, 7].
Another important factor that has hampered the development of freight demand models is related to the lack of freight data. This is even more crucial in developing countries; where the amount and quality of the available data are limited and scarce resources prohibit comprehensive data collection efforts. Consequently, researchers and practitioners are challenged with the development of freight demand modelling techniques that can use the data available, which would allow them to evaluate infrastructure investments and transport policies.

As a result of the multiple dimensions to measure or define freight, two distinct modelling approaches have been developed: trip-based and commodity-based models [4]. Moreover, each of these have used different modelling techniques including variants of the four-step model, direct demand models, and input-output models, among others. Trip-based models focus on vehicles as the modelling unit, and in general, consist of trip generation/attraction, trip distribution and traffic assignment. In addition, recent research has shown the importance of modelling empty trips, which can account for 30% to 40% of the total truck traffic [8, 9]. One advantage of trip-based models is their ability to use data from traffic volumes, which are easily collected at low cost - a good alternative when there are resource and budget constraints. However, these models have difficulties to characterize the freight flows and to be implemented in multimodal systems. On the other hand, commodity-based models - with freight (tons, weight, volume, cargo) as the modelling unit - are able to better capture the economic behaviours of freight flows and to model multimodal systems. Nevertheless, their high data requirements about commodity flows for model calibration and forecast, usually collected through origin-destination (OD) surveys (expensive, difficult to implement and with a limited ability to capture empty trips), make their application almost impossible when lack of data and collection resources are the pressing factors.

A classification system for trip and commodity-based models in terms of their data requirements and the characteristics for model development have been suggested by [10, 11]; five classes are defined:

Class A: refer to factoring methods, where link flows are obtained by applying growth rates to observed truck traffic volumes.

Class B: include Origin-Destination (O–D) factoring methods. These models also consider modal split and traffic assignment.

Class C: are types of vehicle-based models that generate aggregate truck trips to be assigned to the road network.

Class D: make reference to the Four-Step Commodity Models - commodity-based versions of the Class C vehicle-based models.

Class E: Economic Activity Models are economic land-use-based models that incorporate feedback mechanisms with freight transport costs.

Class F: Logistic models.

Class G: Tour-based models.

In their research, [10, 11] conclude that freight demand models used in practice, heavily rely on aggregate approaches/data and are unable to consider the economic behaviour of the decision makers at the establishment level. As a result, improved and disaggregate models must be developed. Disaggregate models - defined as models using observations at the individual/establishment level or at the production/consumption unit - have several advantages as they are able to consider behavioural theory, do not suffer from aggregation biases, and can include detailed policy relevant variables. However, they are difficult to implement given their high data requirements, rarely available and usually collected using establishment-based surveys, a very expensive effort. To cope with data limitations and take advantage of disaggregate properties, models that combined approaches, aggregate and disaggregate, have been developed [12].

The objective of this paper is to contribute to the modelling practice by proposing a strategic freight modelling framework which was applied to Colombia. The framework is developed as a variant of the four-step modelling process that can be used when freight data available and resources to collect comprehensive data are limited - a common characteristic of developing countries.

2. FREIGHT DEMAND MODELLING FRAMEWORK

In developing countries, the available information for freight demand modelling is limited to most of O-D matrices. Back in 2001, the Colombian Ministry of Transport [13] commissioned the design of a theoretically sound freight demand model, which was developed as an input-output (I-O) model. However, the implementation was not deemed to succeed because of the difficulties to fulfill the data requirements, specifically, the regional technical coefficients and the exchange or trade coefficients matrices. Later on, in order to support the infrastructure investments to enhance the Nation’s competitiveness, the Ministry of Transport requested the development of a new freight demand model that would be able to use the available data. The framework proposed in this paper was developed to meet the Ministry’s requirement [14]. For this purpose, a modified version of the four-step travel model - generation, distribution, modal split, and traffic assignment - was developed. In addi-
tion, the framework is able to adequately model empty trips. Considering the differences between freight generation and freight trip generation [15, 16], the vehicle load factors are used to convert commodity flows to truck trips [17].

The general structure of the proposed framework is depicted in Figure 1. As shown, two main sections can be identified. The first section considers the activities required for the estimation and calibration processes of the main modelling components (i.e., generation, distribution, modal split, traffic flows, empty trips and assignment), which defines the methodological processes to follow. The second section is proposed for forecasting and evaluation purposes. Here, the impact in equilibrium conditions due to future scenarios, projections of demand, and network modifications is modelled and evaluated through a set of system indicators such as total costs, delays, and travel times.

The forecast and evaluation section was intended to evaluate infrastructure investments to increase the competitiveness of the countries’ products to take advantage of recent free trade agreements signed with other nations. The main role of the economic analyses is to guide infrastructure investments on the transport network whose deficiencies are at the root of the country’s high transport costs, specifically between the country’s central region - where the highest demand takes place - and the ports. With the proposed model, projects were evaluated for various future planning horizons. Examples of projects evaluated include: construction of multilane roads, rail networks, airport improvements, improvement in waterways (including an Atlantic-Pacific inter-oceanic route), and pipeline infrastructure investments.

The discussions in this paper will focus on the first section of the proposed model given the ease of application and adaptability to other countries. General descriptions of the components are provided next.

2.1 Freight/Freight trip generation

The objective of generation models is to provide quantitative estimates of the amount of freight - freight generation FG - or the number of trips - freight trip generation FTG - produced or attracted as a function of socio-economic (e.g., employment, population, total sales, industry type, are) and land-used variables [16]. These are estimated using multiple techniques such as [3, 18-20]: Trend and time series models; freight generation and freight trip generation models based on multiple classification analysis and multiple linear regression [2]; fuzzy regression algorithms [21]; dynamic system models; zonal travel rates; and input-output models.

2.2 Distribution

Distribution models are used to estimate the freight flows or freight trips between origins and destinations. Estimating freight flows from freight generation models is straightforward; however, estimating freight trips between origin zone \( i \) and destination \( j \) is especially complicated because there is a significant difference between source/destination and production/consumption relations. For a complete discussion about the differences between FG and FTG the reader is referred to [2]. In order to estimate the freight or freight trip flows between origins and destinations, distribution models use a measure of transport strength expressed as transport cost or generalized travel cost. Techniques such as gravity models based on entropy maximization principles are commonly used [18]. The functional form of a gravity model is shown in the following expression [22]:

\[
V_{ij} = A_i O_i B_j D_j f(c_{ij})
\]  

where \( A_i \) and \( B_j \) are the balancing factors so that they can meet the model constraints, \( O_i \) is the total number of trips generated in zone \( i \), and \( D_j \) is the total number of trips attracted to zone \( j \). In addition, \( f(c_{ij}) \) is a composite cost function between the pair \( ij \) usually expressed as \( f(c_{ij}) = \exp(\cdot\beta c_{ij}) \), where \( \beta \) is a parameter to be estimated.

2.3 Modal split

The most common approach in modal split analyses is the use of discrete choice models based on random utility theory [22]. One type of choice models is the multinomial logit model, where the minimum cost modal choice is selected. In essence, the probability of using mode alternative \( m \) from among those available \( (1, 2 \ldots K) \) for a specific \( i \)-to-\( j \) movement is:

\[
R_{m/i} = \exp(\lambda c_{im}) / \sum_k \exp(\lambda c_{ik})
\]  

where \( c_{ik} \) is the generalized cost from \( i \) to \( j \) of alternative \( k \). Furthermore, given a modal cost sensitivity parameter, \( \lambda \), the composite cost to be used as representative of the cost in the distribution model is given by:

\[
c_{ij} = 1/\lambda \ln(\sum_k \exp(\lambda c_{ik}))
\]  

2.4 Traffic flow estimation

In the absence of freight trip generation models, a conversion methodology from commodity-based models to truck traffic is proposed by [23]. This methodology estimates the OD matrix based on shipment characteristics and has the ability to consider empty and loaded trips. In addition, trip chaining is modelled by analyzing the average payload, the characteristics of
adjacent zones and commodity flow patterns. Alternatively, [24] discuss the estimation of traffic flows from commodity flows using the average payload for loaded trips. Clearly, the latter method is a viable alternative for the freight demand modelling framework to be developed.

2.5 Empty trips estimation

Empty trips cannot be estimated directly from the commodity flows as only loaded trips are included and there are routing patterns to be considered [25]. To overcome these limitations, additional models that represent empty trips as a function of the choice routes for commercial vehicles could be developed. Models of this type found in the literature include [9] and [24, 26], and are based on the following notation: \( m_{ij} = \text{Commodity flow from origin } i \text{ to destination } j; \) \( a_{ij} = \text{Average payload (ton/trip) for loaded trips between } i \text{ and } j; \) \( x_i = m_{ij}/a_{ij} = \text{Estimated number of loaded trips between } i \text{ and } j; \) \( y_i = f(x_i) = \text{Estimated number of empty trips between } i \text{ and } j; \) \( z_{ij} = x_i + y_i = \text{Estimated total number of trips (loaded and empty) between } i \text{ and } j. \)

Additionally, a practical approach was proposed by [27]. They assume that the number of empty trips between \( i \) and \( j \) is a function of the commodity flows in the opposite direction, \( j \) to \( i \), multiplied by a constant \( p_0 \) – empirically estimated. This development, together with the assumption that the average load value for trips from \( i \) to \( j \) is equal to those from \( j \) to \( i \) \( (a_{ij} = a_{ji}) \) leads to:

\[
z_{ij} = \frac{1}{a_{ij}}(m_{ij} + p_0 m_{ji})
\]  

(4)

2.6 Assignment

The assignment process allows determining the flows on the network. A widely used modelling approach is to simultaneously solve the modal split and assignment problems. This is usually done by encoding a multi-modal network that includes connecting links representing the transfer modes; another approach is to maintain independent models. For the inter-regional case, where no congestion problems are considered except in proximity of major urban centres and due to sparse characteristics of the network, additional methods have been required, which has led to the development of stochastic assignment models [27]. These have been suggested as alternative approaches to the ones based on Wardrop’s first principle. In this sense, the use of stochastic user equilibrium (SUE) assignment models, obtained by applying the user equilibrium principle under the assumption of probabilistic path choice behaviour, are other attractive options for modelling the assignment process [28]. In addition, multiclass allocation models have been used when the assignment process depends on commodity type (e.g., to transport perishable goods where the priority is to minimize travel time, emergency vehicles). Moreover, it is also possible to model the assignment process explicitly considering externalities as described in [29].

The next section discusses results from the implementation of the proposed framework using Colombia’s national commodity flows and transport network.

3. CASE STUDY: THE COLOMBIAN STRATEGIC FREIGHT MODEL

This section discusses the implementation of the proposed framework in Colombia. It is important to mention that the model developed is based on commodity flows, measured in metric tons obtained from previous studies, and partial information about truck flows from traffic counts at certain corridors. Additionally, the model considers a multimodal network. For the framework implementation, the first step was to identify a set of modelling parameters such as: base year for analysis, zoning system, freight units, among others. Other characteristics of the transport system used in the analyses include: socio-economic variables, population density, index of transit coverage, and location of counting stations. The analyses are conducted for a time period consisting of a single day. For the sake of brevity and conciseness, the authors decided to limit the discussion of results to the key components of the framework. These are discussed next.

3.1 Zoning system

The development of the zoning system was based on the country’s political-administrative organization (“Municipios” and “Departamentos”, which are equivalent to Counties and States), as most of the socio-economic variables, production and consumption statistics and information from previous studies are available following this system. As a result, a zoning system comprised of 70 internal and eight external zones was considered. The latter represent the freight movements to and from Venezuela, Ecuador, South America Atlantic, South America Pacific, Asia-Pacific, West Coast U.S., East Coast U.S. and Europe-Africa.

3.2 Freight generation and attraction modelling

In Colombia, the key commodities for the economy can be aggregated in 34 different groups, which represent about 83% of total freight flows. The other commodities - representing the remaining 17% of the flows - were included in a group called others. These key groups are: Fertilizers, Oils and animal fats, Food industry products, Pet food, Rice, Sugar, Banana, Beer, Fermented beverages, Bio-fuels, Coffee, Coal, Ce-
COMMODITY ANALYSIS

Figure 1 - Proposed freight demand modelling framework

For each commodity group, a freight generation model was developed. These models estimate the amount of cargo (tons) generated and attracted by each zone as a function of diverse economic variables. During the model estimation process, the authors explored a wide range of possible explanatory variables including commodity characteristics (physical characteristics, modes of transport used, prices, production and consumption, seasonality), logistics (distribution network), projections of production and consumption, and export and import statistics. In doing so, the authors used data sources such as the 2005 truck origin-destination survey; official flow statistics, mostly aggregated, for ports, airports, railways and pipelines; production and consumption statistics made available by the government and other organizations; results from previous studies; traffic counts; and socio-economic information. Major limitations and gaps were found in the available data that needed to be complemented, though limited resources constrained data collection efforts. To overcome this limitation, the authors opted for interviews with different organizations and companies that could provide information about the logistics and distribution patterns. In some cases, this task proved to be difficult as some of the individuals contacted were reluctant to provide the information considered commercially sensitive. Nevertheless, the information provided by the different trade organizations and companies was of great value to understand their logistics and distribution patterns. The interviews provided a good trade-off between the quality of the information and the data collection resources required, and allowed the team to construct a complete picture of the freight flows for the different commodity groups.

In general, available data determined the type of estimation technique to be used. In most cases the analytical techniques were linear regression at the zonal level and generation rates. In other cases, especially for the external zones, the growth factors and time series analysis were used. As an example, the daily potato production at zone \( i \) \( P_{potato, i} \) (tons) is:

\[
P_{potato, i} = 0.016 A_{potato, i}
\]

where \( A_{potato, i} \) is the potato cultivated area in zone \( i \).

### 3.3 Distribution

In order to estimate the internal flow matrices for each commodity groups, doubly constrained gravity models were used [22]. The model parameters were calibrated with special consideration so that the trip length distributions (TLD) obtained in the 2005 truck origin-destination survey and other surveys were maintained. On the other hand, the growth factor methods were used to estimate the external-internal freight flow matrix.

#### 3.4 Modal split modelling

The modal split for the country is shown in Table 1. As shown, about 90% of freight is transported using road and rail modes; however, rail is used almost exclusively (99%) to transport coal from the two most important mines to the ports [30]. Consequently, there is a high dependence on road for freight transport, which is done along a few well-defined corridors. There are only a handful of commodities, mainly coal and oil, for which there is competition between modes (rail and inland waterways).

In this research, the modal split problem was solved by estimating a multinomial logit model from aggregate data. The estimation process considered the modes available for each commodity type. For instance, the modal split model obtained for coal was:

\[
V_q = -0.34486 C_q (0.5360 T_{iq})
\]

where \( V_q \) is the representative utility of mode \( q \) for decision maker \( q \), \( C_q \) is the cost in thousands of Colombian pesos and \( T_{iq} \) is the travel time in days of mode \( i \) for the decision maker \( q \); the parameters were found to be statistically significant as shown by the t statistics values. This model yields a value of time for coal equal to $COL 1,600/ton-day COL (0.84US$/ton-day).

### Table 1 - Freight modal split in Colombia

<table>
<thead>
<tr>
<th>Mode</th>
<th>Metric Tons (10^3)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road</td>
<td>102,961</td>
<td>60.6</td>
</tr>
<tr>
<td>Rail</td>
<td>49,227</td>
<td>29.0</td>
</tr>
<tr>
<td>Inland Waterway</td>
<td>5,045</td>
<td>3.0</td>
</tr>
<tr>
<td>Air</td>
<td>135</td>
<td>0.1</td>
</tr>
<tr>
<td>Coastal</td>
<td>399</td>
<td>0.2</td>
</tr>
<tr>
<td>Oil Pipeline</td>
<td>12,100</td>
<td>7.1</td>
</tr>
<tr>
<td>Total</td>
<td>169,867</td>
<td>100</td>
</tr>
</tbody>
</table>

Sources: Constructed by authors using data from the Ministry of Transport and Ecopetrol.
3.5 Estimation of traffic flows and empty trips

As discussed earlier, there are limitations when estimating freight trip flows from commodity flows, as the latter do not specify the type of commercial vehicles used in freight transport [9, 24-26]. For instance, there are small (2 axles), medium (3-4 axles) and large (5-6 axles) trucks. For the current analysis, the type of truck, distribution strategies and average payloads were considered when estimating truck trips for the commodity flows obtained in the 2005 O-D survey. A similar analysis was performed for rail, inland waterway, coastal and maritime transport where typical configurations and average payload of trains, barges and ships were considered.

In addition, following the approach suggested in equation (4), empty trips models were obtained for small (S), medium (M) and large (L) trucks as a function of the loaded trips between i and j.

\[ y_{ij}^{S} = 0.49 x_{ij}^{S} \quad R^2 = 0.61 \] \hspace{1cm} \text{(8)}

\[ y_{ij}^{M} = 0.35 x_{ij}^{M} \quad R^2 = 0.51 \] \hspace{1cm} \text{(9)}

\[ y_{ij}^{L} = 0.29 x_{ij}^{L} \quad R^2 = 0.62 \] \hspace{1cm} \text{(10)}

Data limitations did not allow the estimation of more advanced empty trip models. However, the results obtained were theoretically sound, and undoubtedly, they are a better option than just using a constant factor of loaded trips - a commonly made assumption in practice.

3.6 Traffic flow assignment

In order to conduct traffic flow analyses, a transport network was required. In this sense, a multimodal strategic network was built by selecting the main routes in different transport modes. The Bureau of Public Roads (31) (BPR) function was used to evaluate link performances. In addition, the parameters for the link performance function where derived considering the network characteristics, specifically, the type of terrain (plain, rolling, mountainous), pavement status and road capacity (single, two-lane or multi-lane) and information from previous research done to calibrate the Highway Development and Management (HDM) manual. Where needed, additional links were used; for instance, to connect the maritime international external zones to a set of centroid connectors representing the internal zones. It is important to mention that ports, logistics platforms and load transfer centres have an important role in the system. Although topologically represented as links, their level of service indicators (travel times, costs) should be evaluated using microscopic simulation techniques and then included as link attributes in the aggregate strategic model. The strategic networks for the different modes are shown in Figure 2 (coastal and international maritime flows are not shown).

Preliminary analyses indicated that congestion conditions were only present at the vicinity of large urban areas. Therefore, the use of multiclass allocation methods (by commodity category) with stochastic equilibrium assignment techniques [32, 33] was appropriate. In addition, available traffic counts from passenger flows (cars and buses) were used as network preloads - fixed background volumes on links. Traffic counts from a considerable number of links allowed network calibration.

4. CONCLUSIONS AND PRACTICAL IMPLICATIONS

This paper proposes a multimodal freight demand modelling framework developed as a variation of a commodity-based four-step travel model, which was applied to Colombia. To explicitly consider freight requirements, additional steps are included to estimate empty trips and freight trips from commodity flows.

When analyzing freight generation, it is important to identify the most representative commodity groups. In Colombia, thirty-four aggregate commodity groups represent 83% of the commodity flows. To avoid aggregation bias the estimation of freight generation models for the different groups required significant amounts of data. The potential sources of partial information available included: commodity flow surveys; truck origin-destination surveys; official flow statistics for ports, airports, railways and pipelines; official or trade production and consumption statistics; previous studies; traffic counts; and socio-economic information. The authors found great value in conducting in-depth interviews with different organizations and companies, to understand their logistics and distribution patterns, and to be able to construct a complete picture of the freight flows. This type of interviews provided a good trade-off between the quality of information gathered and the resources required. Alternatively, a low cost survey of partial coverage could be conducted in order to provide an estimation of the flows. The results show that the most important variables to estimate freight demand were population, GDP, and for agriculture-related commodities, cultivated area. In terms of import and export products, commodity flows were found to be highly related to extremely complex variables such as duty taxes, which depend on commercial agreements, and currency exchange rates. In order to forecast flows from and to the external zones, simulations considering a set of scenarios allowed performing sensitivity analyses.

In terms of freight distribution, doubly constrained gravity and growth factor models were used. Modal split was estimated by analyzing the mode share from
available commodity flow data. Given that conducting
stated preference and revealed preference surveys
were financially infeasible, the known modal shares
were used, as well as estimates of average transport
cost and travel times by mode.

For estimating traffic flows from freight flows, the
average payload factors were used. Although this
method does not allow obtaining traffic flows for truck
categories, it is a good option in cases with limited
data. In addition, modelling empty trips allowed for
a more realistic representation of the network traffic
flow. Results showed the advantages of using this type
of technique, which are better than just using a gener-
al augmenting rate for loaded trips. Though zero order
trip chaining was considered, the evaluation of higher
order chains is recommendable. However, this would
require having reliable O-D traffic flow data.

For traffic assignment, the research shows that it is
indeed possible to construct a multimodal network with
limited resources. For the model proposed, the network
was constructed using data from the national inventory
of infrastructure. The advent of web-based tools and
geographic information systems will likely improve data
availability to construct such type of networks.

In general, the development of the proposed
framework is the result of the efficient use of the avail-
able information. Though there are several limitations
worth discussing such as the lack of formal models to

Figure 2 - Colombian strategic networks
evaluar y adaptar el enfoque de modelación al contexto en análisis.
Este artículo propone un modelo estratégico para el transporte de carga a escala nacional desarrollado como una variante del tradicional modelo de cuatro etapas, al cual se ha agregado una etapa para estimar flujos de tráfico a partir del flujo de productos y otra para considerar viajes vacíos. Se discute la implementación práctica del modelo propuesto al contexto de Colombia.
Para cada etapa se describen los datos, la fuente de información y la aproximación metodológica. Además, se provee una descripción de las limitaciones de información y las estrategias para complementar los datos existentes. De la implementación, los autores identifican un conjunto de ventajas del modelo propuesto y plantean una serie de sugerencias para mejorarla.

PALABRAS CLAVES
Modelación de transporte de carga; Colombia; modelo estratégico de carga;

REFERENCES


