MODE CHOICE MODELLING FOR THE ASSESSMENT OF AN INTERNATIONAL RAILWAY CORRIDOR

Diego Crespo-Pereira[a], Rosa Rios-Prado[b], Oscar de Gregorio-Vicente[c]

[a)(b] University of A Coruña
[c] Consultrans S.A.

[a]diego.crespo@udc.es, [b]rrios@udc.es, [c]oscar.degregorio@consultrans.es

ABSTRACT

When developing a model for multimodal transport planning, the mode choice component is a critical element since it estimates the shares of traffic flow which are absorbed by each competing alternative. This paper describes the steps followed in the mode choice modelling phase in the development of the transport model for the assessment of the Bolivian Central Bioceanic Railway Corridor (CFBC). The model integrates different levels of geographical resolution, alternatives from four competing modes of transport and different categories of freight.

Keywords: Multimodal Transport, Freight Transport, Discrete Choice Model.

1. CASE INTRODUCTION

The CFBC (Central Bi-Oceanic Railway Corridor) is a railway corridor infrastructure project promoted by the Bolivian Government and funded by the Inter-American Development Bank. It will link the Atlantic and Pacific coasts of the central part of South America, from the port of Santos in Brazil to a port in the coast of Peru (Figure 1). Its construction is expected to save both costs and time of transporting cargo and passengers through the great natural barrier which are the Andes. As of nowadays, the only way to cross them is by narrow mountain roads that lead to high logistics costs and difficult the economic development of the region.

Figure 1. Proposed CFBC map.

This paper describes the steps followed for choosing and calibrating the discrete modal choice component of the transport model (Figure 2) developed for strategic planning. The transport model is a M&S tool developed as part of the analysis of prospective trade, market and logistics alternatives. It seeks to forecast the levels of passenger and freight demand and the CFBC's flows absorption among Bolivia, its neighboring countries and the rest of the world. The goal is to analyze its competitiveness compared to other existing alternatives such as the Paraguay-Parana Waterway or the Panama Channel. Import and export trades are regarded as one of the most promising sources of demand and thus accurate international transport modelling is considered a major challenge in the project development.

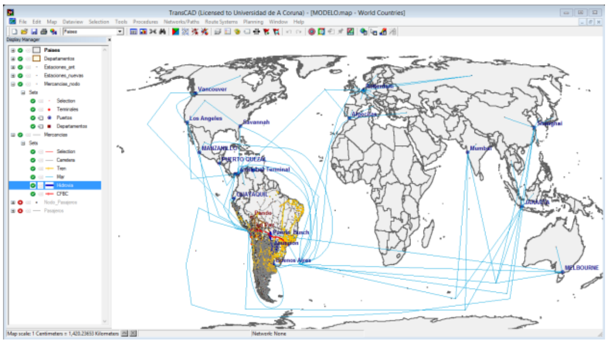

Figure 2. Screenshot of the whole transport model.

The model developed is based on the classical four steps methodology. A full description of the modelling methodology is given by Rios-Prado et al. (2013). The components of the model are:

- Freight generation and consumption. It estimates the total flows of freight produced and demanded in each Traffic Analysis Zone (TAZ).
- Freight generation and distribution. It estimates the flows of freight from each origin to each destination. It is performed for both the flows between regions within Bolivia and for the international import/export flows between Bolivia and the rest of the world.
- Modal choice. It calculates the share of load that uses each transport alternative.
- Network assignment. It estimates the total cargo that uses each link of the road.
- Discrete events simulator of train terminals. This complementary model estimates the

residence times in the terminals of the railway network depending on traffic congestion and terminal design parameters Figure 3.

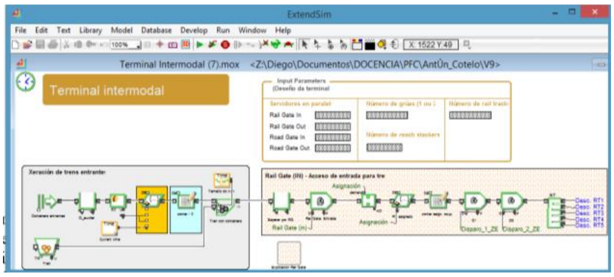

Figure 3. Screenshot of a part of the discrete events model implemented in ExtendSim for the simulation of train terminals.

The transport network included in this model spans:

- Main roads infrastructure in Bolivia and neighboring countries.
- Railway network in Bolivia and neighboring countries.
- Paragay-Parana waterway.
- Major shipping routes departing from the most representative ports in the region.

Although there are many types of cargo that can be considered at the large scale of this model, they were grouped in four categories:

- “Dirty” dry bulk cargo (mainly minerals).
- “Clean” dry bulk cargo (mainly soy).
- Liquid bulk (mainly fuels).
- Containerized cargo (mainly manufactured products).

This paper is focused in the data collection and calibration steps for obtaining the modal choice component of the model. Section 2 presents a brief introduction to discrete choice models and bibliographic review. Section 3 presents the discrete choice model fitted.

2. DISCRETE CHOICE MODELS

One of the key aspects of a transportation model is to estimate how modal choice is made from a set of available alternatives. Discrete choice models aim at capturing how some individuals or groups select a mode weighting some factors that affect the decision (Ortúzar and Willumsen, 2011).

In freight transport, it has been well established in the literature that the best option in terms of a single factor (such as cost or time) is not always chosen by all the users. Hence, probabilistic decision models are preferred to deterministic ones. The main goal of the choice model is to predict the share of trips for each origin and destination pair transported by each mode.

There are some different models that can be adopted such as logistic regression, Bayesian networks or neural networks. The most popular ones applied in practice are the logit models, which can be subdivided in many types. Linear regression models are discarded because the assumptions of ordinary least squares are violated (Aldrich and Nelson, 1995). The main logistical models are Multinomial Logit Model, Binary Logit Model and Nested Logit Model.

Discrete-choice models are based on random utility theory. Its assumptions can be summarized as follows (Ortúzar and Willumsen, 2011):

- There is a homogenous population formed by a set of individuals who have perfect information.
- A set of alternatives is available (it is assumed in general to correspond to each transport mode).
- Each transport alternative (*j*) has a net utility for individual (*p*), U_{jp}. The utility is then assumed to be the sum of two components, a measurable function of certain factors that affect the transport choice (V_{jp}) and a random component which reflects the idiosyncrasies and particular tastes of each individual as well as measurement and observational errors (ε_{jp}).

$$U_{jp} = V_{jp} + \varepsilon_{jp}$$

Logit models are a family of discrete choice models that can be obtained by considering different assumptions about the nature of the stochastic component. In particular, the Multinomial Logit Model (MNL) is a very popular one in practice due to its simple assumptions and robustness (Domencich and McFadden, 1975). It relies on the assumption that the random errors are Gumbel IID distributed, which leads to the well-known equation of a MNL for calculating the share of k^{th} alternative (P_k):

$$P_k = \frac{e^{V_k}}{\sum_j e^{V_j}}$$

The utility function (*V*) is generally modelled as a linear function of a set of alternative dependent factors and/or individual related factors. Two common decision factors considered in practice are the cost and the time incurred by using each transport alternative. These two factors usually determine modal choice (Kreutzberger, 2008). The preference for an alternative given equal cost and time conditions can be captured in the linear model by means of a constant parameter. For instance, a simple utility function form would be:

$$V_k = \beta_k + \beta_C \cdot c_k + \beta_t \cdot t_k$$

Where β_k stands for alternative preference utility, β_C stands for the cost unitary utility (it must be negative since utility is reduced by increases in cost), c_k stands for the cost of choosing alternative k, β_t stands for the time unitary utility and t_k stands for the time of choosing alternative k.

Multinomial logit models (MNL) have found many applications to transportation problems in practice. For instance, they have been applied to analyzing the mode choice behavior in the Oresund region (Rich et al., 2009), for vehicle choice modelling in Colombia (Holguin-Veras, 2002) or joint mode and vehicle choice in the US (Pourabdollahi et al., 2013).

Other logit models of practical interests are the Nested Logit Model (NLM) and the Mixed Logit Model (ML). The NML (Daly and Zachary, 1978) takes into account that when there are groups or hierarchies of the choice alternatives the random errors IID assumption is no longer valid. The paper by Jiang et al (1999) presents a case of application for analyzing the relation between the freight characteristics and the shipper's choice. The model considered two great transport groups such as public and private, and three subgroups in the public one, road, train and multimodal. Another example of application is the analysis of mode choice in India for containerized cargo (Ravibabu, 2013).

The Mixed Logit Model assumes that the parameters of the utility function are not constant but random variables themselves. Mixed logit models (ML) are suitable for cases in which the population of decision makers is very heterogeneous and weights the decision factors differently (Fadden and Train, 2000). This greater model flexibility comes at the expense of greater model complexity since parameters distributions need to be specified.

Another approach considered in this project was the use of neural networks for choice modelling, which is a promising technique for mode choice modelling (Hensher and Ton, 2000; Karlaftis and Vlahogianni, 2011; Nijkamp et al., 2004). Neural networks can fit to complex non-linear data (Karlaftis and Vlahogianni, 2011), although they generally require large datasets, might be subject to overfitting and model parameters have not a direct interpretation. They were discarded in this work because the size of the available datasets was limited and it was desired to provide an interpretation of the model parameters.

3. THE MODEL CALIBRATION

3.1. Alternatives identification

The first step for calibrating the mode choice model was to identify the set of transport alternatives. The relevant alternatives depend on the geographical distances covered. For interregional transport within Bolivia or transport to the neighboring countries, road and train are the only relevant modes.

However, for long distance import/export flows maritime transport is required and the Paraguay-Parana waterway is used as a link to the port of Buenos Aires in Argentina, from where goods can be transported to the rest of the world. Thus, the hardest case for identifying transport alternatives is the export/import to far destinations. All the countries which are not neighbors of Bolivia will be included in the category of "distant TAZs".

The mode choice model will thus operate differently for this two types of transport:

- Short distance transport for interregional flows and import/export to neighbouring countries.
- Long distance transport to "distant" TAZs.

For short distance transport only two options were considered in this case: road and multimodal road-train modes.

For long distance transport, the available alternatives can be grouped following two criteria:

- Alternatives related to the port used for import/export. There are many ports available on both the Atlantic and Pacific coasts. Port selection is important because it is conditioned by ports facilities, connections to the hinterland and it determines the maritime routes followed.
- Alternatives related to the mode of transport employed for reaching the port. There can be used road, multimodal road-train and the waterway along with its combinations with road and train.

Although if all the options of ports and modes combinations were taken into account the number of alternatives would be large, in practice some of the combinations can be disregarded due to their practical infeasibility or lack of use. Also, although the waterway is actually a transport mode that competes with road and train for some of the flows, the only port available is Buenos Aires so other ports get discarded. Thus, the waterway option does not need to be accounted for in both categories of alternatives and it was dealt with as if it was a "port" alternative. Finally, not all the alternatives of transport are valid for all types of cargo, so further constraints can be applied.

Once the relevant alternatives were filtered, the set of following options was established as shown in Table 1.

For different types of cargo, criteria for mode choice are usually different. Thus, at this step of the model development it was decided to calibrate a separate model for each cargo category. For instance, containers have usually higher added value than bulk cargo so travel time is a more important factor for the first category.

Table 1: Available alternatives in the model.

Alternatives			
Port	Mode	Bulk	Containers
Arica	Road	Yes	Yes
Arica	MM	No	No
Peru port	Road	Yes	Yes
Peru port	MM*	Yes	Yes
Iquique	Road	Yes	Yes
Antofagasta	Road	Yes	No
Antofagasta	MM	Yes	No
Pto. Busch (waterway)	Road	Yes	Yes
Pto. Busch (waterway)	MM	Yes	Yes
Pto. Suarez (waterway)	Road	Yes	Yes
Pto. Suarez (waterway)	MM	Yes	Yes
Buenos Aires	Road	Yes	Yes
Santos	Road	Yes	Yes
Santos	MM	Yes	Yes

* This alternative is not available as of nowadays but it is part of the future part of the CFBC so it is considered in the model.

3.2. Dataset

Data for model calibration was mainly gathered from the international trade databases of the Bolivian Ministry of Foreign Affairs. Although interviews with exporters and importers were available, they did not provide useful information for model calibration. The data required for model calibration must have information of costs and times by using each alternative as well as the actual choice made by the shipper or stated preferences. However, the data available only had information about the chosen alternative but not the other options.

On the contrary, the international trade database contains extensive data on the imports, exports and modal choice. The fields of information contained in this database are:

- Origin and destination for each import/export flow.
- Type of cargo.
- Mode of transportation (train, road and waterway).
- Entry/Exit point to Bolivia.
- Year and month.
- Tons of transported flow, along with its CIF and FOB values.

Combining this information with the estimated transport costs of each mode, it was possible to obtain a dataset with the required fields to calibrate the choice model. The process followed for obtaining this dataset can be summarized as follows:

1. Download data from the international trade database. The registries include origin/destination, mode, cargo type and entry/exit point. This data constitutes the "observed flows dataset".
2. Obtain and debug the GIS cartography for the model network.
3. Estimate the average transit unitary costs per mode and average velocities from a market survey.
4. Employ the GIS cartography along with the unitary costs and velocities to estimate the total costs and travel times from each origin to each destination by mode.
5. Add the total costs and times per mode as additional fields to the "observed flows dataset".

However, the maximum likelihood methods employed for calibrating a MNL model require discrete observations of selected mode along with mode attributes. In order to obtain such a sample, Monte Carlo simulation was adopted using the "observed flows dataset" to weight the probability of each simulated trip. The goal of this procedure is to generate a sample of observations that match the observed pattern of mode share present in the data.

A sample of 800 simulated trips was generated for each freight category following the next procedure in order to obtain the "Simulated Trips Survey Table":

1. The probability of generating a trip from an origin to a destination using a certain mode is given by the total freight flow from the origin to the destination divided by the total freight.
2. A sample of 800 combinations of Origin-Destination-Alternative is randomly generated from the "observed flows dataset" using the probabilities calculated in the previous step. A row in the "Simulated Trips Survey Table" is added for each possible alternative and the selected Origin-Destination. The field "Selected Alternative" is set equal to True for the row that matches the observed alternative and False for the others.
3. For each Origin-Destination-Alternative combination generated, the cost and time by each mode are obtained from the "observed flows dataset".

The dataset thus obtained could then be used for model calibration.

3.3. Results

The model was calibrated for the sample obtained as indicated before. R software was employed for this purpose. A separate model was fitted for each cargo type. The variables included in the utility function varied from one type to another depending on the statistical significance of the fitted parameters. Also, some alternatives for cargo transport were not included in the model depending on the actual constraints as explained before.

The values of the fitted parameters are omitted due to confidentiality issues. The presented values are the parameters included in the model along with their p-values for the parameters significance tests. Table 2 shows the models parameters.

Table 2: Utility function parameters along with their statistical significance.

Cargo Type	Parameter	p-value
Container	Arica Preference	0.0000
Container	Cost	0.0000
Container	Time	< 2.2e-16
"Dirty" dry bulk	Cost	0.0000
"Dirty" dry bulk	Time	0.0000
"Clean" dry bulk	Cost	0.5331
"Clean" dry bulk	Time	0.0001
Liquid bulk	Cost	0.0000

The following conclusions are derived from them:

- For containers, both cost and time were significant parameters to be included in the model. It was also found that including an alternative dependent parameter for the Arica Port increased the prediction power of the model. This result can be explained taking into account that the Arica Port offers attractive facilities for containerized freight. Santos port in Brazil is also a port with a large volume of containers, although it is less attractive for Bolivian cargo since it is more far away and it is a highly congested port.
- For "dirty" dry bulk, both cost and time factors were included in the model.
- For "clean" dry bulk, time was found to be the most significant variable. This result makes sense if we take into account that most of this cargo is soy. Soy exports are nowadays conditioned by the seasonal variations of soy prices. Exporters are mainly interested in minimizing travel times to reach their destination markets when prices are high.
- For liquid bulk, it was found that the cost factor could account for the observed share and including the time factor would not contribute to increase the predictive power of the model.

The calibrated MNLs were implemented in the transport model so that assignments of freight flows to each mode and the network could be obtained. Figure 4 shows an example of the kind of results that could be achieved before the CFBC is operating and Figure 5 shows the results once the CFBC is operating.

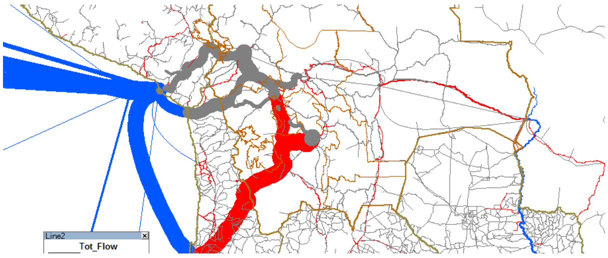

Figure 4. Screenshot of the model results displaying the assignment of freight flows for a certain type of cargo before the CFBC is operating (red is used for rail mode and grey for road).

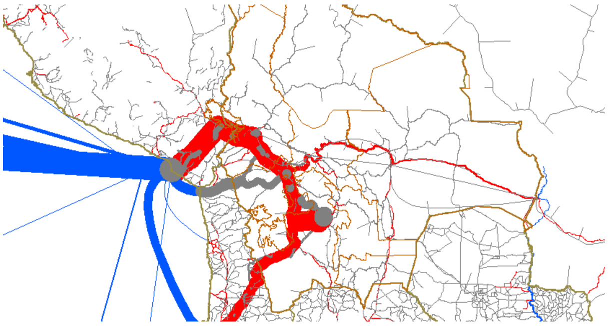

Figure 5. Screenshot of the model results displaying the assignment of freight flows for a certain type of cargo once the CFBC is operating (red is used for rail mode and grey for road).

4. CONCLUSION

This paper describes the process of developing the mode choice component of a large transport model employed for the assessment of a new railway corridor. Multinomial Logit (MNL) models were used for estimating the shares of freight flow that are absorbed by a wide range of transport alternatives which span various ports in the Pacific and Atlantic coasts of South America and combine four different transport modes.

The models were fitted using data from the International Trade Database developed and maintained by the Bolivian government, which provided the widest and most valuable data source available in this project. The model includes the main variables necessary to estimate the distribution of cargo among competing alternatives. The MNLs calibrated were implemented in the transport model so that it was possible to forecast the absorptions of flow by the future CFBC service given the designed conditions and proposed fares.

REFERENCES

Aldrich, J.H., Nelson, F.D., 1995. Linear Probability, Logit, and Probit Models. SAGE Publications, Inc.

Daly, A.J., Zachary, S., 1978. Improved multiple choice models, in: Hensher, D.A., Dalvi, M.Q. (Eds.), Determinants of Travel Choice. Saxon House, Westmead.

Domencich, T., McFadden, D.L., 1975. Urban Travel Demand: A Behavioral Analysis. North-Holland Publishing Co.

Fadden, D.M.C., Train, K., 2000. MIXED MNL MODELS FOR DISCRETE RESPONSE. J. Appl. Econom. 470, 447–470.

Hensher, D.A., Ton, T.T., 2000. A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. Transp. Res. Part E Logist. Transp. Rev. 36, 155–172. doi:10.1016/S1366-5545(99)00030-7

Holguin-Veras, J., 2002. Revealed Preference Analysis of Commercial Vehicle Choice Process. J. Transp. Eng. 128, 336–346. doi:10.1061/(ASCE)0733-947X(2002)128:4(336)

Jiang, F., Jonhson, P., Calzada, C., 1999. Freight Demand Characteristics and Mode Choice: An Analysis of the Results of Modeling with Disaggregate Revealed Preference Data. J. Transp. Stat. 149–158.

Karlaftis, M.G., Vlahogianni, E.I., 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transp. Res. Part C Emerg. Technol. 19, 387–399. doi:10.1016/j.trc.2010.10.004

Kreutzberger, E.D., 2008. Distance and time in intermodal goods transport networks in Europe: A generic approach. Transp. Res. Part A Policy Pract. 42, 973–993. doi:10.1016/j.tra.2008.01.012

Nijkamp, P., Reggiani, A., Tsang, W.F., 2004. Comparative modelling of interregional transport flows: Applications to multimodal European freight transport. Eur. J. Oper. Res. 155, 584–602. doi:10.1016/j.ejor.2003.08.007

Ortúzar, J., Willumsen, L.G., 2011. Modelling Transport, 4th ed. Chichester, West Sussex.

Pourabdollahi, Z., Karimi, B., Mohammadian, A. (Kouros), 2013. Joint Model of Freight Mode and Shipment Size Choice. Transp. Res. Rec. J. Transp. Res. Board 2378, 84–91. doi:10.3141/2378-09

Ravibabu, M., 2013. A nested logit model of mode choice for inland movement of export shipments: A case study of containerised export cargo from India. Res. Transp. Econ. 38, 91–100. doi:10.1016/j.retrec.2012.05.007

Rich, J., Holmblad, P.M., Hansen, C.O., 2009. A weighted logit freight mode-choice model. Transp. Res. Part E Logist. Transp. Rev. 45, 1006–1019. doi:10.1016/j.tre.2009.02.001

Rios-Prado, R., Crespo-Pereira, D., Del Rio-Vilas, D., Rego-Monteil, N., De Gregorio-Vicente, Ó., 2013. MODEL DEVELOPMENT FOR THE ASSESSMENT OF AN INTERNATIONAL RAILWAY CORRIDOR – METHODOLOGICAL OVERVIEW, in: The 15th International Conference on Harbor, Maritime and Multimodal Logistics Modelling and Simulation, Proceedings. pp. 133–139.

AUTHORS BIOGRAPHY

Diego Crespo Pereira, PhD, MSc Industrial Engineer. Diego is Assistant Professor of the Department of Economic Analysis and Company Management of the UDC. He also works in the Integrated Group for Engineering Research (GII) of the UDC as a research engineer and project manager. He is mainly involved in the development and coordination of R&D projects related to industrial and logistic processes optimization.

Rosa Rios Prado, MSc Industrial Engineer. Rosa is Adjunct Professor of the Department of Economic Analysis and Company Management of the UDC as well as a research engineer in the GII of the UDC. Rosa is mainly devoted to the development of transportation and logistic models for the assessment of multimodal networks and infrastructures.

Oscar de Gregorio Vicente holds a MSc degree in Statistics by the Complutense University of Madrid (UCM) and a Master in Transportation, Logistics and Highway Safety by the National Distance Education University of Spain (UNED). He is a PhD student in mathematics (UCM), and Professor of the Master in Auditories in Highway Safety and Mobility of the Politechnic University of Madrid (UPM). He is currently Project Manager in the Transport Department of the Spanish consultancy and engineering company Consultrans S.A. His main area of expertise is transport planning and transportation modeling. His major research areas include discrete choice modeling, Bayesian networks and artificial neural networks.