APROXIMACIÓN BAYESIANA APLICADA AL REPARTO MODAL EN MODELOS DE TRANSPORTE DE MERCANCIAS (CASO PRÁCTICO: CORREDOR FERROVIARIO BIOCEÁNICO CENTRAL)

Óscar DeGregorio-Vicente, Beatriz González-Pérez y Miguel-Ángel Gómez-Villegas

Universidad Complutense de Madrid. Facultad de Ciencias Matemáticas. Dpto Estadística e Investigación Operativa & Instituto de Matemática Interdisciplinar (IMI). Plaza de Ciencias, 3 - Ciudad Universitaria 28040 MADRID. Tfno: +34 91 3944531

Recibido: .../.../2016 -- Aceptado: .../.../2016 – DOI: .....

BAYESIAN APPROACH TO MODEL CHOICE ANALYSIS IN FREIGHT TRANSPORT MODELS (CASE STUDY: CENTRAL BIOCEANIC RAILWAY CORRIDOR)

ABSTRACT:
Transport planning requires tool to model the current and future situation of an infrastructures network. In this way, different scenarios of passenger flows, vehicles or freight can be predicted and serve as information for decision making.
One of these tools are the so called "Demand models", among which the four steps models (Generation/attraction, Distribution, Modal choice, Network assignment) is a remarkable example for its widespread use.
This paper presents a novel Bayesian approach to the third step of a demand transport model. Traditional discrete choice models are the ones most commonly used at this purpose, although other methods such as neural networks have been used by some authors. A Bayesian network is proposed as tool for estimating the decisions made by users when they face the need to choose which transport alternatives to use for sending cargo in a case study corresponding to the Central Bioceanic Corridor in South America.
The results from fitting a logit model and a Bayesian network are compared and show the Bayesian network to be a promising tool to be applied in this kind of applications.

Key Words: Transport Model, Discrete choice model, Logit, Bayesian Network.

RESUMEN:
La Planificación de los Transportes de una determinada zona necesita herramientas que permitan representar o modelizar, de forma analítica, la situación actual y futura de su red de infraestructuras. De esta manera, se pueden estimar, para diferentes escenarios, los flujos de pasajeros, vehículos o mercancías que habrá en dicha red.
Una de estas herramientas son los denominados Modelos de Demanda, de entre los que destaca el modelo clásico de 4 etapas (Generación-Atracción, Distribución, Reparto o elección modal y Asignación).
El presente artículo muestra una novedosa aproximación a la tercera etapa, que es la correspondiente a la decisión del modo de transporte a la que se enfrenta la unidad de decisión dentro de un conjunto discreto de alternativas. Tradicionalmente, los modelos de elección discreta han sido los más utilizados para esto, aunque existen otros métodos, como las redes neuronales, que también han sido utilizadas por otros autores. Las redes bayesianas se proponen como herramienta alternativa para la elección modal, tanto para pasajeros como para mercancías, y para este último caso, se presenta un caso de estudio basado en el Corredor Ferroviario Bioceánico Central en Sudamérica.
Los resultados obtenidos permiten comparar un modelo logit y una red bayesiana, y muestran cómo la aproximación bayesiana surge como una herramienta prometedora en este tipo de aplicaciones.

Palabras Clave: Modelos de Transporte, Modelo de Elección Discreta, Logit, Red Bayesiana.

1. Introduction

The analysis of investment in large infrastructures of transport requires the estimate, with long horizons of analysis, as well as the users of these networks (whether passengers or goods) shall make use of the same, if they choose versus other alternatives and, ultimately, if a new project represents a real improvement for the whole, in such a way as to compensate for the high investment costs to which it has to cope. To achieve this may be subject to various sources of uncertainty.

In this regard, a key aspect in the development of these models is the estimate of the modal split between a number of alternatives. The discrete choice models estimate how different decision-makers choose between a number of options that may correspond to different modes of transport, combinations of modes or settings from one mode of transport on the basis of a set of factors that determine the choice [1].

In the carriage of goods, the two most important factors in the modal choice are travel time and cost for each option, although due to the disparity between the criteria followed by different users, it is considered appropriate to the use of probabilistic models compared to deterministic models, so that the variability in the decisions of the users of the means of transport. Therefore, the model of choice gives the probability of use of an alternative, and that probability is used in an aggregate transport model (usually), such as the proportion of users who take such an alternative.

The models most commonly used in practice are *Multinomial Logit models of type (MNL)*, can be seen in [1], although there are variations that more complex assumptions, such as the nested logit models (*Nested Logit Models*) can be found in [4].

According to the main bibliographic reference on transport modeling [1] , On the one hand we have:

- The choice of transport mode within the network built, represents the most important element in transport planning and decision-making;
- For both models of passenger and freight is the crucial stage in the face of the forward projections;
- Influences the overall efficiency of the transport system, the amount of urban space dedicated to the functions of the transport, as well as in the set of available alternatives;

And on the other hand, we have:

- The likelihood that individuals choose a certain alternative is a function of their socio-economic characteristics and the relative attractiveness of the alternative;
- To represent the attractiveness of the alternative uses the concept of utility (theoretical artifice, which the individual attempts to maximize). The alternatives *per se* does not produce utility, but the utility is derived from the characteristics of the alternatives and of the characteristics of individuals [2]. The observable or measurable utility is usually defined as a linear combination of variables;
- It is necessary to compare the value of the profits of each alternative and transform them into a probability value between 0 and 1, typically using mathematical transformations between the Logit models (Eq. (1)) and Probit (Eq. (2):

$$\text{Logit: } P_1 = \frac{\exp(V_1)}{\exp(V_1)+\exp(V_2)} \tag{1}$$

$$\text{Probit: } P_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{V_1-V_2+x_1} \frac{exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1}{\sigma_1}\right)^2 - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \left(\frac{x_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}}\, dx_2 dx_1 \tag{2}$$

Publications DYNA SL -- c) Mazarredo no.69 - 4° -- 48009-BILBAO (SPAIN)
Tel +34 944 237566 - www.revistadyna.com Email: dyna@revistadyna.com

Being the most commonly used of these the following:

-Multinomial Logit (MNL) can be seen in [1], EC. (3):

$$P_{iq} = \frac{\exp(\beta V_{iq})}{\sum_{A_j \in A(q)} \exp(\beta V_{jq})} \quad (3)$$

- Hierarchical Logit (HL), or nested logit model, which can be seen in [3] and [4], Ec. 4:

$$P(d,m) = \frac{exp\{\beta(V_d+V_v^*)\}\exp(\lambda V_{dm})}{\sum_{d'} exp\{\beta(V_d+V_v^*)\}\sum_{m'} exp\{\lambda V_{dm'}\}} \text{ With } V_d^* = \left(\frac{1}{\lambda}\right) log \sum_{exp}(\lambda V_{dm'}) \quad (4)$$

But, according to [1]:

-"All this theory is based on the assumption that the decision-making unit, the *ideal user* is rational, selfish, and their tastes *never* change, maximizing its utility through careful analysis and thoughtful."
-"However, the *actual user* is partially rational, but it is also emotional and collaborator. You may not use *all* the alternatives, so use heuristic rules to decide:

- You care more for the changes that the absolute values.
- It has a decreasing sensitivity to changes in utility.
- Is the enemy of the losses.
- Does not react immediately."

Having said that, given the uncertainty and the a priori information in decision-making by the maker, why not raise another type of models to analyze the choice of alternatives?, beyond MNL models or HL, optimizing the stage of modal split using Bayesian networks (RB), given the good results of the same in different works for the last years of research, and given the relevance of this stage in transport models.

This article presents a novel approach to the problem of modeling of the modal split through the application of RB although collect all appointments is practically impossible, we must cite to [20], in which are contained the RB associated to the multinomial distribution and [21] where more general distributions can be applied to this type of networks.

2. Methodology AND CASE STUDY

2.1. Methodology

While the usual mathematical models have a long history and a solid basis, to find solutions to the problem of the discrete choice, there are other types of different approaches to the frecuentistas that can lead to better results, and this is the case of the Bayesian approaches.

The RB arise in the decade of the 80, arising from the investigations from the 70 were doing in artificial intelligence (AI) with expert systems, programs capable of simulating and even replace on some occasions to human reasoning.

The term "Bayesian Networks" is attributed in these years, in particular, in the year 1985 to Judea Pearl See [5], To emphasize three fundamental aspects:

- The often subjective of the input;
- The unit of conditioning by the *Bayes theorem* as the basis for the updating of the information;
- The distinction between causal and evidentiary modes of reasoning, which underscores Thomas Bayes in a document published posthumously in 1763 [6].

At the end of 1980 seminal texts *Probabilistic Reasoning in Intelligent Systems* [7] And *Probabilistic Reasoning in Expert Systems* [8] A summary of the Properties of the RB and help to consider the same as a field of study.

Following this line of research, during the decades of the 80 and 90, it became clear that the IA should not only imitate the human rational behavior, but collaborate with these in the taking of decisions through synergies. That is to say, communicating knowledge of the logical process followed to solve a problem and to get a solution.

In this respect, in the year 1993, [9] Directs an appointment to the editors of the journal *Knowledge Acquisition*, those who decide to publish it, which indicates: "The key question is not the IA, but how to improve the natural intelligence with the help of the knowledge based systems."

This shortcut as a precedent, to explain the importance that in the last few decades are taking the RB, these are defined as a probabilistic graphical model, a directed acyclic graph (GAD) that represents: (1) qualitatively, a set of variables, called nodes, and its dependencies conditional probability, codified in its arches, (2) quantitatively, that collect the conditional probability distributions for each node given its parents. Each node can be a parameter, a random variable or a hypothesis.

- For the specification of the qualitative information of the RB is used a DAG, which is denoted D = (V,E), where each of the nodes of D represents the elements of the problem X = {X1, . . . ,Xn}, being therefore V = {X1, . . . ,Xn}; and the edges that are in and show the causal relationships, or the nodes being the father, the cause, and the child nodes, the effect.
- For the specification of the quantitative information is a set of distributions of conditional probability P = {p(x1|pa(x1),..., p(xn|pa(xn)), so that for each variable Xi € X will be the distribution of conditional probability of Xi given the occurrence of their parents pa(Xi) in the graph D, denoted by p(xi|pa(xi)).

In summary, formally, linking the two concepts, a RB is formed by the pair (G,P), where G is a DAG formed by a node for each random variable X = {X1,...,xn}, and arcs that represent the structure of probabilistic dependence between them, P ={p(x1|pa(x1)), ..., p(xn|pa(xn)) is a set of n probability distributions conditional, and pa(xi) is the set of parents of the node Xi in G.

That is to say, a DAG is a RB with respect to a set of variables, if the whole of the probability distribution of the variables node can be written as the product of the local distribution of each node and their parents as the next prime factorization [10], Ec. (5):

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P\big(x_i \big| padres(x_i)\big) \qquad (5)$$

If there is an arc from node A to node B, is called the father of B and B is a son of A. The set of parent nodes of a node xi is denoted as "parents" (xi).

This last equation (5) is a distribution of conditional probability for each random variable. Or what is the same, each node Xi is a variable conditioned by their parents, which establishes a direct relationship between the qualitative and the quantitative part of the network, as it is the GAD which allows you to determine the conditional probability distributions that are considered in the factorization of the joint probability distribution. That is to say, to a GAD is a factorization of the joint probability distribution of a RB.

From this property, there are efficient algorithms that perform inference and learning in RBs can be seen in [11] and [12]. Among the main features of the RB should be highlighted:

1) The properties of *separation*, which determine structures of independence (and dependency). This is, based on the concept of *evidence* or exact value known that takes one of the variables (nodes), such that when you enter this information in the network, affects the uncertainty of the rest of the variables, we study how moves the information of such evidence along a network, i.e., the separation criteria are met in the graph; the relationship of dependence or conditional independence among the variables, which depend on the type of connection that will consider:

a) **Serial connection**, information is transmitted from the evidence, except that such information is contained in the intermediate node, (A and C *separated* given B);

b) **Divergent connection**, the information flows through the network, except if the evidence is located in the parent node, since it is blocked communication between the nodes children, (B) and (C) *separate* given);

c) **Converging connection**, the information can be transmitted over the network only if you have the evidence on the child node, or a descendant of this, (A and B *connected* given C);

2) Comply with the Markov property if and only if each node Xi is conditionally independent of its no descendants, nd(xi), given their parents, pa(xi) [13] .I.e.

$$P(x_i|pa(x_i), nd(x_i)) = P(x_i\ |pa(x_i)) \qquad (6)$$

In the following figure (1) are plotted these properties.

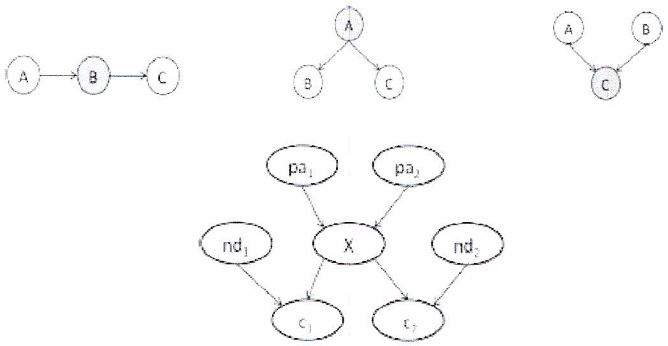

Fig. 1: a) serial connection, (b) connection divergent, convergent connection (c) and (d) Markov Property

This property allows you to define a RB in a manner similar to the Ec. (5), through the following theorems:
- 1 theorem: any two (G,P) that complies with the Markov property, is a Bayesian network.
- 2 Theorem: All Bayesian network formed by the pair (G,P), fulfils the Markov property.

Reciprocal conclusion, considering these two theorems, it can be concluded that all RB defined by the pair (G,P), meets two equivalent properties: the factorization presented in EC. (5) and the Markov presented in EC. (6).

Taking all these considerations into account of the RB, in relation to the transport modeling, it should be noted in recent years, to name a few, the work carried out:

- For the estimation of origin-destination matrices of travel (OD), on the basis of the information plates of enrolment in vehicular flows, or on the basis of partial flows of traffic for a given urban ed. The doctoral thesis [14] contains examples applied to the Nguyen-Dupuis, road network The road network of Ciudad Real and the State of Vermont. Getting that job
- For the estimation of origin-destination matrices of travel (OD), on the basis of the information plates of enrolment in vehicular flows, or on the basis of partial flows of traffic given an urban network the doctoral thesis of Maria Walnut Male [14] Considering applied examples The Nguyen-Dupuis road network, to the Royal City and the State of Vermont. Getting this thesis the 1st International Abertis Award in the year 2012.
- For the analysis of decision at different stages of transport, the work [15], Explains in chapter 3, a postal distribution problem considering a series of variables that are included in a Bayesian network obtaining results enlightening;
- The work [16] contains the construction of a model of prediction of traffic flows in the arches of a given network, to estimate OD matrices, as well as to find the best location for count points of traffic with the aim

Publications DYNA SL -- c) Mazarredo no.69 - 4º -- 48009-BILBAO (SPAIN)
Tel +34 944 237566 - www.revistadyna.com Email: dyna@revistadyna.com

that the Administrations (local, regional and national) achieve a better management of mobility in global terms.

Therefore, although historically it has sought the estimate of flows as a clear direction of research, this article is geared more toward the use of the RB to improve the goodness of the results obtained in the stage of the modal split, classical models of transport of four stages: 1) generation/consumption of Load, Load Distribution 2), 3) 4) modal split and assignment to network.

2.2. Case Study

In order to highlight the application of our method has been used the Central Oceanic Rail Corridor (CFBC) (Source: Transport Model of the study "Analysis of prospective commercial, market and logistic alternatives" with the number of loan BO-L1056-1), Figures 2 and 3, is intended to connect the eastern and western railway networks in Bolivia, which would allow a continued in such a network linking the Atlantic and Pacific Ocean. Bolivian both networks are separated by the cordillera of the Andes and this infrastructure seeks to connect the western Andean high plateau, at an elevation of 4,000m, and the area to the east of Bolivia to approximately 600m of height, connected in turn to Brazil and Argentina. In this way, we can set up a railway corridor from coast to coast of great economic interest.

Especially, it is of utmost importance for the interior regions of South America, since their communications with the nearest ports are slow and costly, and therefore constitute an impediment to the economic development of the region.

Fig. 2: Proposed Map to the CFBC

The work described in this article part of the transport model of 4 stages developed in this project to provide estimates of the flows of goods and passengers that will make use of the corridor in front of the existing alternatives. For goods, the main modes in competition with the railways, by aggregation in the transport chain, are the road and the Paraguay-Paraná . A full description can be read in [17] Although this article focuses on the problem of the transport of goods as it is of greatest interest to the CFBC.

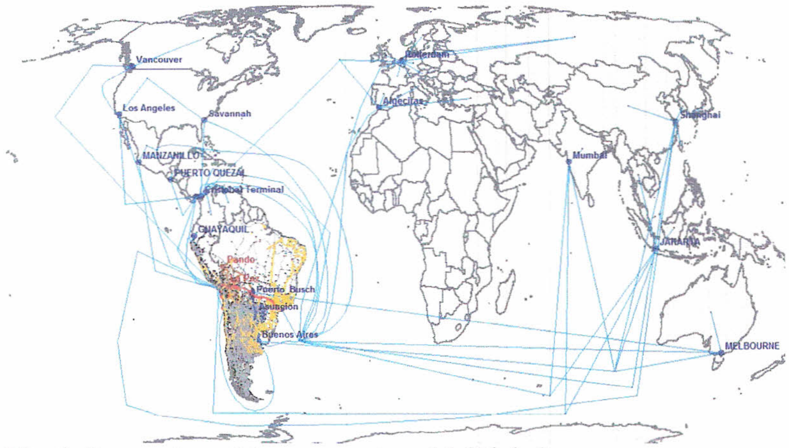

Fig. 3: Screen shot of the transport model (Global)

2.2.1. Alternatives of choice

The problem of modal choice in this project considers how alternative transport combinations of import/export ports, together with the modes of transport to the port. Not all combinations between options are feasible in practice, so that a first step was to identify all the combinations of feasible alternatives.

The following table (I) shows the available alternatives, port combination and so, for the goods in bulk and type for the containerised goods.

Alternatives			
Port	**So**	**Bulk**	**Containers**
Arica	C	If	If
Arica	MM	Not	Not
Ilo	C	If	If
Ilo	***MM****	If	If
Iquique	C	If	If
Antofagasta	C	If	Not
Antofagasta	MM	If	Not
Pto. Busch	C	If	If
Pto. Busch	MM	If	If
Pto. Suarez	C	If	If
Pto. Suarez	MM	If	If
Buenos Aires	C	If	If
Santos	C	If	If
Santos	MM	If	If

* This alternative is not available today, but that will be the result of the construction of the CFBC.
Table I: alternatives available in the model (C=Road, MM=multimodal road/station)

2.2.2. Sample Data

To make the adjustment of the models of choice, the largest sample available that allows to estimate the geographical distribution of cargo movements in Bolivia, which is the database of the National Institute of Statistics (INE) Bolivian foreign trade. This database contains records of exports and imports by identifying the following variables:

- Products grouped by various categories have been used the categories of NANDICA section of four digits and the classification of main products).
- Department of origin or destination.

Publications DYNA SL -- c) Mazarredo no.69 - 4° -- 48009-BILBAO (SPAIN)
Tel +34 944 237566 - www.revistadyna.com Email: dyna@revistadyna.com

- Country of origin or destination of import or export.
- Way out, which makes it possible to identify through that port is sent to the load on shipments to international destinations of long distance.
- Output mode, which allows you to identify if you are using road, rail or waterway.

This information in conjunction with the model of cost and time implemented with the macroscopic modeling software TransCad (Caliper Corp. - www.caliper.com), Allows to analyze the output modes adopted in relation to the cost and time involved in each option. To do this we combined tables of imports and exports Bolivian INE downloaded with the OD of costs and times obtained with the transport model. The export data table used for the adjustment of the models of choice contains a total of 466 records and 4,333 imports.

2.2.3. Calibration of the MNL

Made using the statistical software R as the statistical tool with greater variety of programming libraries and techniques of this type developed by researchers. The values of the parameters are ignored in this article be subject to confidentiality clauses of the draft, but the following table shows the significance of the parameters of the models. We considered three types of parameters in the model:

- A fixed effect parameter associated with the port.
- A parameter of effect of the cost for each type of goods.
- A parameter of effect of time for each type of goods.

The parameters of the utility function that is not shown to be statistically significant in the significance test were eliminated from the model to select the models listed, Table II.

Load Type	Parameter	P-Value
Containers	Puerto Arica	0.0000
Containers	Cost	0.0000
Containers	Time	0.0000
Solid Bulk Dirty	Cost	0.0000
Solid Bulk Dirty	Time	0.0000
Solid Bulk Clean	Time	0.0001
Bulk Liquids	Cost	0.0000

Table II: parameters of the utility function with the p-value of the test of significance

To sum up the previous table, the setting is noted, with the information used, that the parameter on the port of Arica proved to be statistically significant as this port offers facilities suitable for the traffic of containers that are not explained only in terms of time and costs. For solid bulk clean, only the time was found as a significant factor, since it is mainly of agricultural products for which the travel times are very important. For the liquid bulk commodities (mainly fuel) The main factor of decision noted that it was the cost.

Based on this analysis, the MNL choice models were implemented in the transport model and thus were obtained estimates of assignment to network related to the modal estimate. See Figure 4:

Publications DYNA SL -- c) Mazarredo no.69 - 4º – 48009 BILBAO (SPAIN)
Tel +34 944 237566 - www.revistadyna.com Email: dyna@revistadyna.com

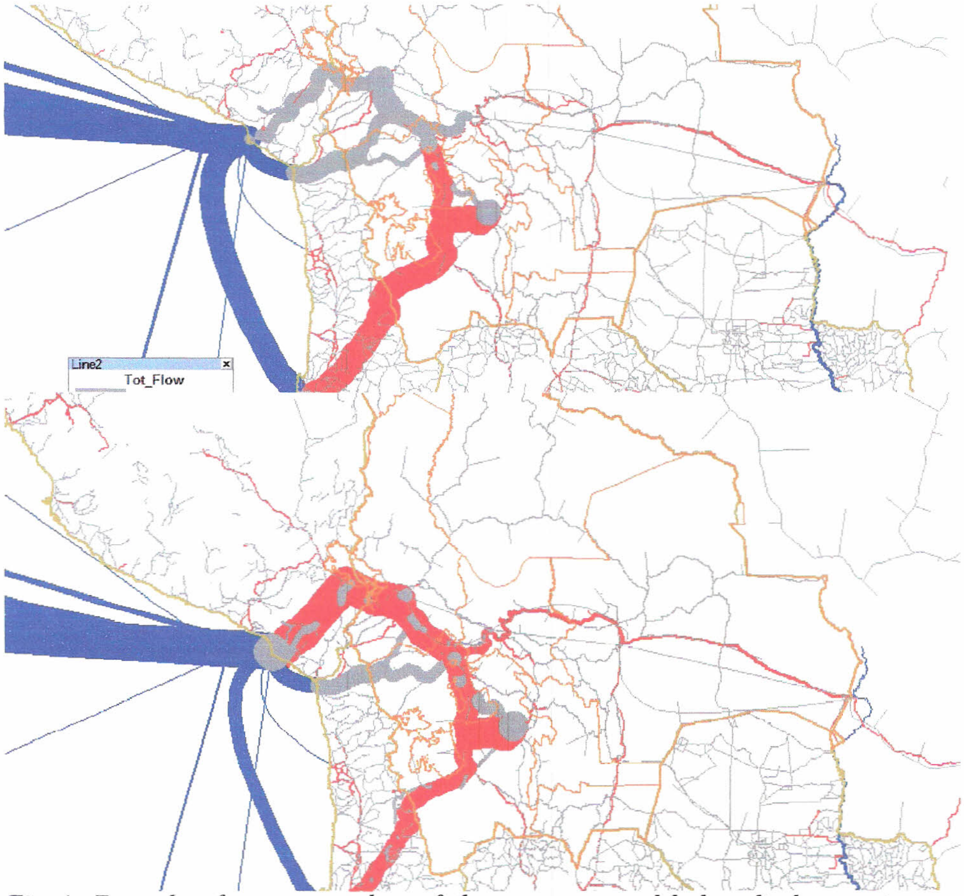

Fig. 4: Example of two screenshots of the transport model that displays assignments for a product, comparison of 2 scenarios-scenarios considered by modes (rail sections: Red; Blue: sea; gray: road)

2.2.4. Calibration of the RB

From the same set of data used for the MNL models in R using the following variables:

- Time
- Cost
- Flow of goods per year
- Price FOB (Free On Board) of the product
- % Of distribution by alternative.

One difference with regard to the encoding used for the adjustment of the MNL models is that they developed a model for each type of goods while for the Bayesian network a single network for all, although it included the FOB price as an indicator of the type of goods (the FOB prices are low bulk and containers high values).

The codification of the variables is explained in the Table (3):

Entry	Variable	Description
Time	$T_k = t_k - \min_k t_k$	Time difference with respect to the time of the best alternative *k*.
Cost	$C_k = c_k - \min_k c_k$	Cost difference with respect to the cost of the best alternative *k*.
Flow of goods per year	$F_{i,j,m}$	Flow of goods from origin *i* to destination *j* for the goods *m*.

FOB Price	$FOB_{i,m}$	FOB price of the product at the point of origin
Proportion of use of the alternative	$p_{m,i,j,k}$	Proportion of merchandise of the type considered between the source and the destination j that make use of the alternative considered k.

Table III: Variables for the Bayesian network

The calibration process followed the following steps:

1. Application of the test of Doornick-Hansen of normality that showed that the variables did not conform to a Normal distribution.
2. Transformation of the data using the Nonparanormal "transformation" proposed by Liu [18].
3. Implementation of new Doornick-Hansen test to the sample transformed to verify that indeed the transformation had provided a sample with a Normal distribution.
4. As a restriction for the calibration of the Bayesian network was included that the nodes corresponding to the proportions $p_{m,i,j,k}$ could not be origins of the links of the graph. This is due to the fact that these variables have to be explained in all cases from the rest of the variables of the network.
5. Implementation of the algorithm max-min hill-climbing (Tsamardinos mmhc) [19], for the calibration of the Bayesian network.

The result of the calibration of the RB provided the results of the table IV (in Spanish)

• 42 Nodes: • Arcs: 61 • Undirected arcs: 0 • Directed arcs: 61 • Average Markov blanket size: 4.86 • Average neighborhood size: 2.90 • Average branching factor: 1.45 • Learning Hill-Climbing Min-max algorithm:	• Constraint-based method: Min-max Parent Children • Conditional indep. Pearson's correlation test: • Score-based method: Hill-Climbing • Score: BIC (Gauss.) • Alpha threshold: 0.05 • Penalization coefficient: 3.061246 • Tests used in the learning procedure: 4504 • Optimized: TRUE

Table IV: Results of the calibration of the RB

On the other hand the following **¡Error! No se encuentra el origen de la referencia.**Figure 5 shows the graph of the RB generated, from where you can see the relationships among nodes of costs (C1…C14) and time (T1…T14) for the different alternatives (E1…E14), and how few of them in conjunction with FOB prices and the flow of goods per annum (mtpa), explain most of percentages of choice observed.

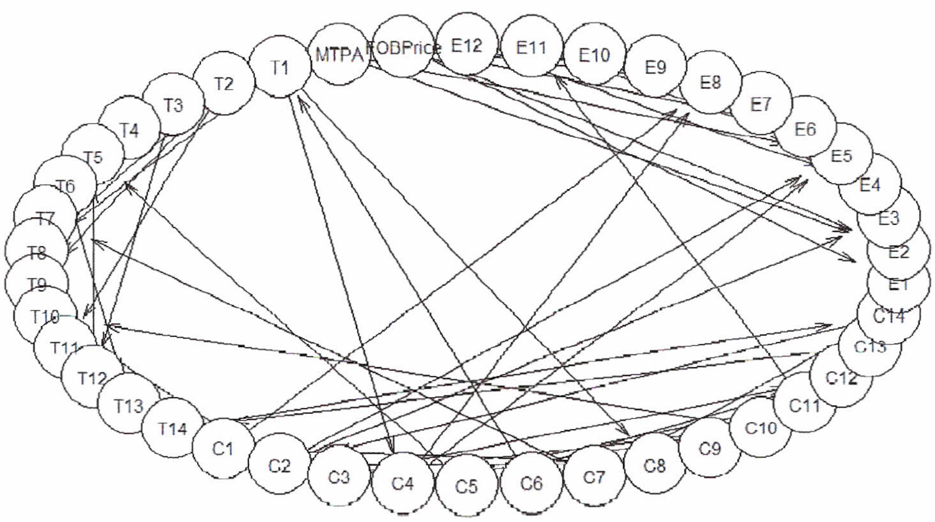

Fig. 5: Graph of the Bayesian network

Publications DYNA SL -- c) Mazarredo no.69 - 4º -- 48009 BILBAO (SPAIN)
Tel +34 944 237566 - www.revistadyna.com Email: dyna@revistadyna.com

3. Results

The results of the Table (V) below show the mean quadratic error (ECM) and the maximum error committed in the estimation of the sample by the MNL model and the Bayesian Network.

Alternative	ECM		Maximum Error	
	MNLogit	RB	MNLogit	RB
ANTOFAGASTA_Road	15.08%	8.50%	99.48%	37.04%
ANTOFAGASTA_ MM	16.26%	7.40%	82.35%	39.53%
ARICA_ Road	*73.71%*	*29.48%*	*100.00%*	*57.00%*
ARICA_ MM	19.67%	4.05%	**41.66%**	**48.81%**
ILO_ Road	14.05%	12.26%	100.00%	40.46%
IQUIQUE_ Road	27.10%	13.27%	100.00%	39.66%
IQUIQUE_MM	5.64%	2.20%	**10.24%**	**46.96%**
BUENOS AIRES_ Road	19.56%	15.83%	100.00%	47.30%
PTO.SUAREZ_ Road	*6.84%*	*3.55%*	*95.49%*	*46.86%*
PTO.SUAREZ_ MM	*5.03%*	*2.20%*	*90.26%*	*46.96%*
SANTOS_ Road	**8.28%**	**10.27%**	100.00%	48.10%
SANTOS_ MM	**3.78%**	**6.30%**	**10.24%**	**50.15%**

Table V: Comparison between the maximum and average quadratic errors committed in the estimate of the proportions of each transport mode for each port considered

You can see that in the majority of transport alternatives, except in the marked in bold, the Bayesian network incurs an average quadratic error and that, furthermore, the maximum errors are also generally lower. This is especially relevant in the case of the option of Arica road and Puerto Suarez which at present are the two points with the greatest volume of import and export of the country.

5. CONCLUSIONS

The results obtained, it was possible to obtain the following conclusions

- Transport models are a powerful tool to help in decision - making related to infrastructure and investments that a country has to carry out in the future.
- The algorithmic advances and processing of large amounts of information are allowing new techniques, such as Rb, as acknowledged alternatives, obtaining promising results.
- This article shows that the results and settings obtained by the application of RB, in comparison with the multinomial logit models, allows you to minimize errors and find alternatives.

This article describes the modal choice models employed in the study of market of a possible new rail corridor between Brazil and Peru (CFBC).

Initially, in the project, it was decided by the use of a model of modal choice (MEM), which estimated the proportions of use of alternatives corresponding to combinations of mode of transport and port, using as factors the times and costs of each alternative, the port in question, and adjusting a model for each type of goods.

As an alternative added presents a Bayesian approach to the problem of the modal choice, which has not been considered by previous authors in this type of problems of transport modeling. The Bayesian network enables us to estimate the proportions of use of the alternatives for their time and costs and adding also the variable FOB price of the product. In the MEM is not used the FOB price of the product but that is a separate model for each product.

The results of the adjustment show that the Bayesian network provides a more accurate estimate both on the basis of the mean quadratic error as to the maximum setting. These findings open up a new line of application for the RB in the problems of modal choice in transportation. On the one hand, offer exciting possibilities in regard to the establishment of causal relationships between the variables of the problem and on the other hand, the results in this particular case, show that the RB offer great potential to be more precise than the classic MEM models that have been used to date.

Publications DYNA SL -- c) Mazarredo no.69 - 4° -- 48009-BILBAO (SPAIN)
Tel +34 944 237566 - www.revistadyna.com Email: dyna@revistadyna.com

Bibliography

[1] J. Ortúzar and L. G. Willumsen, *Modeling Transport*, 4th ed. Chichester, West Sussex, 2011.

[2] K. J. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, vol. 74, no. 2, pp. 132-157, Apr. 1966.

[3] H. C. W. L. Williams, "On the formation of travel demand models and economic evaluation measures of user benefit," *Environ. Plan. To*, vol. 9, no. 3, pp. 285-344, 1977.

[4] A. Daly, S. Zachary, D. Hensher, and Q. Dalvi, "Improved multiple choice models," 1978.

[5] J. Pearl, "Bayesian Networks to Model of Self-Activated Memory for evidential reasoning," *Proceedings of the 7th Conference of the Cognitive Science Society*. Irvine, Calif., pp. 329-334, 1985.

[6] M. Bayes and M. Price, "An Essay Toward Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.," *Philos. Trans. R. Soc. London*, vol. 53, pp. 370-418, Jan. 1763.

[7] J. Pearl and Judea, *Probabilistic Reasoning in Intelligent Systems: networks of plausible inference*. Morgan Kaufmann Publishers, 1988.

[8] R. E. Neapolitan and R. E., *Probabilistic Reasoning in expert systems : theory and algorithms*. Wiley, 1990.

[9] W. J. Clancey, "Notes on 'Heuristic classification,'" *Artificial staple. Intell.*, vol. 59, no. 1, pp. 191-196, 1993.

[10] F. V. Jensen and F. V., *Bayesian networks and decision graphs*. Springer, 2001.

[11] R. E. Neapolitan and R. E., *Learning Bayesian networks*. Pearson Prentice Hall, 2004.

[12] E. Castle, J. M. Gutierrez, and A. S. Hadi, *Expert Systems and Probabilistic Network Models*. New York, NY: Springer New York, 1997.

[13] K. B. Korb and A. E. Nicholson, *Bayesian artificial intelligence*. Chapman & Hall/CRC, 2004.

[14] M. Walnut Male and Mary, *mathematical methods for the prediction of traffic* . University of Cantabria, 2012.

[15] A. Holland, "Bayesian networks for transport decision scenarios," pp. 325-328, 2003.

[16] E. Sánchez-Cambronero, S., Castle and J. M. Menéndez., *Traffic Prediction Models: Using Bayesian Networks and Other Tools.* LAP LAMBERT Academic Publishing, 2011.

[17] R. Crespo-Pereira Rios-Prado, D., D. The Rego-Monteil Rio-Vilas, N., and O. Of Gregorio-Vicente, "MODEL DEVELOPMENT FOR THE ASSESSMENT OF AN INTERNATIONAL RAILWAY CORRIDOR - Methodological Overview," in *The 15th International Conference on Harbor, Maritime and Multimodal Logistics Modeling and Simulation, Proceedings*, pp. 133-139. 2013.

[18] H. Liu, J.Lafferty & L. Wasserman, The Nonparanormal: Semiparametric Estimation ofHigh Dimensional Undirected Graphs, J. Mach. Learn.. Res., 10 pp. 2295-2328, 2009.

[19] Tsamardinos, I., Brown, LE., Aliferis, CF. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. Machine Learning, 65:31-78.

[20] Ricardo A. Daziano, Luis Miranda-Moreno & Shahram Heydari, *Computational Bayesian Statistics in Transportation Modeling: From Road Safety Analysis to Discrete Choice,* Transport Reviews, Vol. 33, 2013.

[21] Tai-Yu Ma, Bayesian Networks for Multimodal Mode Choice Behavior Modeling: A Case Study for the Cross Border Workers of Luxembourg, Transportation Research came, vol. 10, pp. 870-880, 2015.

Acknowledgments

The Research Group in Bayesian methods of the Department of Statistics and Operations Research at the Complutense University of Madrid, for his contributions in regards to different applications and lines of research with RB. To the built-in group of engineering -GII- of the Universidade da Coruña, for his contributions in the field of modeling and simulation of discrete events. At the same time we would like to thank the editor and four anonymous referees for the careful reading of the manuscript and for their suggestions which have led to the improvement of the same.
This research was partially funded by the Ministry of Economy and Competitividadd, Spain, grant TRA2015-65283-R and the Universidad Complutense of Madrid-Banco Santander, Spain, grant PR26/16-20261.