Summary of the Thesis

1. Introduction

The transport sector and particularly the freight transport sector have a negative impact on air pollution and public infrastructure. In this context, the task of freight transport modeling is essential, as it enables to control the positive and negative effects that the freight transport sector can bring. It is all the more relevant in the case of Brazil. Indeed, the research presented in this Master Thesis has been motivated by the fact that the transport flows in Brazil are believed to change quite fast, as its population and structure increase significantly. One of the most crucial requirements for transportation planning consists there in the knowledge of the Origin/Destination pattern, which could help very much decision makers take decisions at the national level.

Of course, various methods already exist that achieve this goal. Some of the most simple and intuitive of all are for example large scale surveys, which do not need any mathematical development but do necessitate the time and budget to work on significant sample of freight travels. However, traffic counting is nowadays a standardized and much more comfortable means of acquisition of real information about circulation. Their use in the construction of Origin/Destination matrices could be useful as they would enable the direct use of some values as input and limit-condition which would be known to be true.

As a consequence, the goal of this Master Thesis is to develop a model that estimates an Origin-Destination matrix specifically based on traffic counts for Southern Brazil. More specifically, the research area consists in the Brazilian states of Rio Grande do Sul (RS) and Santa Catarina (SC).

2. State of the research

The first part of this essay consists in a literature analysis that focuses on models that rely on roadside traffic counts between regions as an input for the estimation of an Origin/Destination matrix. Of course, a brief development on the state of the art of freight transport demand modeling is also provided. Especially, the four-step approach, which constitutes the most common analysis tool sued for macroscopic transportation modeling, is presented. Some models of trip distribution and of route assignment, which form the second and fourth step, are notably presented. Then, their interdependencies with each other as well as with potential traffic volumes data are developed, along with models that already use traffic counts as input for the estimation of an Origin/Destination matrix. The focus is put on the model which was developed by Chen in 1990 **[Yu-Sen Chen, 1990]**, and that serves as basis for the model which this thesis develops afterwards.

3. Model development

Part 3 takes into consideration the advantages and disadvantages of the existing models. Then, explanations on the proper functioning of a new model is developed, as well as a detailed insight in its technical computation. The Software Maple is used, because its language is one of the easiest and most intuitive to understand. Very briefly, the functioning of the model is the following one:

First of all, an algorithm determines the three best paths between all origins and destinations as well as their lengths. The identification and calculation of the first best paths are achieved using the Dijkstra algorithm. Indeed, the central issue of this paper does not involve any negative arcs, so there is no possibility that the Dijkstra algorithm, which gives very interesting results in terms of time and space complexity, may not be applicable. Moreover, the method of identification and calculation of the second and third best paths consist in modifications of previous existing solutions **[Yen, 1970]**. It starts by considering the Dijkstra solution for the first best path. Then it basically considers the nodes and links identified by Dijkstra and computes the algorithm again while forcing to travel through at least a node which had not been visited by Dijkstra. This method offers the advantage of reusing directly the first and simple implementation of the Dijkstra Algorithm.

As a second step, the application of a Logit model with a certain coefficient enables to build an initial traffic allocation. Then, thanks to an initial Origin/Destination matrix $D_{ij}\,(0)$, the virtual volumes on the links of the network can also be determined. This operation is repeated for thirty Logit coefficients, and the Logit coefficient is chosen such that the initial virtual counts match as much as possible the real-world ones. Once this “best Logit-Coefficient” is chosen, iterative operations are realized to modify the best traffic allocation and the initial Origin/Destination matrix, such that the virtual counts C_e^s get closer and closer to the real-world ones C_m^s. The iterative formulas that are used are the following ones:

$$D_{ij}\,(n+1) = \mathrm{r} * \frac{\left[\sum_{s=1}^{S} C_m^s P_{ij}^s(n) - \sum_{s=1}^{S}\; C_e^s(n) P_{ij}^s\,(n)\right]}{\sum_{s=1}^{S}(P_{ij}^s(n))^2} \;+\; D_{ij}\;(\mathrm{n})$$

$$P_{ij}^s\;(n+1) = \mathrm{f} * \frac{(C_m^s - C_e^s\,'\,(n))}{D_{ij}\;(n+1)} + P_{ij}^s(n)$$

$$C_e^s\;(n+1) = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{ij}^s\;(n+1)\;\,D_{ij}\;\,(n+1)$$

With C_m^s = real-world (measured) traffic volume at the station s

C_e^s = estimated traffic volume at the station s
$D_{ij}\;(n)$ = volume travelling from i to j (iteration n)
$P_{ij}^s(n)$ = probability of traveling through s while going from i to j

$$r = \frac{\sum_{s=1}^{S} C_m^s}{\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{j=1}^{N}\; P_{ij}^s\,(n)\;\; D_{ij}'\;(n)}$$

$$f = \frac{\sum_{s=1}^{S} C_m^s}{\sum_{s=1}^{S}\sum_{i=1}^{N}\sum_{j=1}^{N}\; P_{ij}^s\,'\,(n)\;\; D_{ij}\,(n+1)}$$

4. Validity and transferability of the model

Finally, the last part of the Thesis focuses on the accuracy and possible limitations of the method, as well as on recommendations for further improvements. Notably, recommendations to adapt the method to non-Brazilian use-cases are developed, notably for the case of Germany.

This paper first focuses on the stability of the Origin/Destination matrix that is estimated.

A first consideration is to evaluate how different from one another the resultant matrices are if the initial matrices $D_{ij}\,(0)$ vary. In order to determine a quantitative answer to this, the rate of difference between two matrices has to be defined. The measure that this paper uses is inspired on the Root Mean Square Error (RMSE), which was used by Bera and Rao **[Bera and Rao, 2011]**. It is then observable that the model is very much resistant to small variations. Indeed, differences smaller than 2% in the initial matrices change almost nothing in the final estimated matrix: the difference between the final matrices is of 0.08%, that is to say 20 times smaller than the difference between the initial ones. Even if, of course, the final difference between the Origin/Destination matrices tends to increase when the initial one does, the general tendency concerning the effect of modifications in the initial matrices remains stability. Moreover, it is obvious that link volume counts are anyway subject to temporal variations: those can be hourly variations, as well as seasonal variations for example. In order to observe the convergence and resistance of the algorithm towards changes in the traffic volumes too, a second consideration that is made is to observe how different from one another the resultant matrices are if the real-world volumes vary. It can then be observed that up to 2% error in the counts already leads to 2% difference in the final estimated matrix. This means that the system is very sensitive to errors in the counts.

Then, the paper focuses on the validity of the Origin/Destination matrix that is estimated.

One aspect of the study of the validity consists of the comparison of the final estimated matrix with the reference matrix, which was estimated in 2011 by the National Department of Transport Infrastructure of Brazil in 2011 through large scale surveys **[Coppe, 2016]**. Considering once again the definition of Bera and Rao, the structure of the matrix estimated by the model in the third section can be interpreted as being 12% different from the matrix estimated by the Coppe in 2011. This level of validity seems acceptable, particularly as the initial O/D matrix was estimated very easily with a basic gravitational model. It means the correction method through the traffic volumes was efficient.

Another aspect of the study of the validity consists in evaluating whether the virtual traffic volumes converge enough to the real-world ones or not. A method already exists to this effect. It is based on the “GEH Statistic” formula, named after the works of Geoffrey E. Havers in the 70’s. It makes possible to decide whether the modelled data can be considered as a consistent one or not.

The “GEH Statistic” formula is the following one:

$$GEH = \sqrt{\frac{2 * (M - C)^2}{M + C}}$$

where M is the hourly traffic volume which is reconstituted by the model and C is the real (measured) hourly traffic count. Adapted from **[Van Vliet, 2015]**

The UK Design Manual for Roads and Bridges considers the model as acceptable if 85% of the volumes reconstituted by the traffic model have a GEH less than 5 **[UK Highways Agency, 2016]**.

In the use-case of Southern-Brazil that this thesis deals with, only two of them are bigger than 5 after 100 iterations, which means 1-2/76=97% of the GEH-indicators are smaller than 5. In a nutshell, as 97%>85%, the volumes which were virtually reconstituted after 100 iterations already consist in a precise and converged enough data set.

Then, Part 4 also aims to discuss the transferability of the model. First of all, it focuses on the case of countries which would be less developed than the –rather well industrialized- southern Brazil, and which are more likely to have to use manual count recording methods. These methods are very time consuming and labor intensive. Moreover, such countries are much more subject to rapid changes in terms of demography, transport habits, urbanization and land use in general. It makes the use of manual methods even less valuable, as these elements greatly shorten the duration of validity of the collected data. However, no counter argument, technically as well as theoretically, can be opposed to its application if such data were available and reliable.

Secondly, the case of industrialized countries is studied. The model should then be modified and get more complex, as some more data would be available and ready to use. Indeed, data about the current quality of the roads is for example likely to be available and to be usable as well, and not only the lengths between the origins and destinations. Moreover, the traffic allocation model should be gone over again because the 3-best-paths based model might appear a little obsolete. Indeed, the road network density of industrialized countries is incomparable with that of developing countries.

Finally, Part 4 also develops some other considerations about the model. Notably, it focuses on the accuracy of the choice of a Logit-model for the route choices, which does not consider the congestion of the links of the network. It also focuses on the fact that the model lacks a logistic approach.

The location of the counting stations is then the final issue that is discussed. Indeed, let us recall that the accuracy of the Origin/Destination matrix estimated will increase with the number of traffic counting locations. Of course, because of budget limitations, the increasing of the number of counting locations cannot be considered as the unique solution for the improvement of the accuracy. Instead, it is necessary to study at which locations the counting stations should be set so that they bring as much information as possible on the traffic distribution. An algorithm is proposed, and can be directly applied in the case of the German network.