TRANSPORTES
www.revistatransportes.org.br
anpet
associação nacional de
pesquisa e ensino em transportes
www.anpet.org.br

Disaggregated approach to urban trip distribution: a comparative analysis between artificial neural networks and discrete choice models

Abordagem desagregada de distribuição de viagens urbanas: uma análise comparativa entre redes neurais e artificiais e modelos de escolha discreta

Marina Urano de Carvalho Caldas[1], **Cira Souza Pitombo**[2], **Felipe Lobo Umbelino de Souza**[3], **Renan Favero**[4]

[1]University of São Paulo, São Paulo – Brazil, marinacaldas0@gmail.com
[2]University of São Paulo, São Paulo – Brazil, cirapitombo@usp.br
[3]University of São Paulo, São Paulo – Brazil, felipe.lobo@usp.br
[4]University of Florida, Florida – United States, renanfavero@ufl.edu

Recebido:
22 de agosto de 2021
Aceito para publicação:
5 de abril de 2022
Publicado:
26 de agosto de 2022
Editor de área:
Bruno Vieira Bertoncini

Keywords:
Destination choices.
Multinomial Logit.
Nested Logit.
Artificial Neural Networks.

Palavras-chave:
Escolhas de destinos.
Logit Multinomial.
Logit Aninhado.
Redes Neurais e Artificiais.

DOI:10.14295/transportes.v30i2.2686

ABSTRACT
Discrete choice models have been used over the years in disaggregated approaches to forecast destination choices. However, there are important constraints in some of these models that pose obstacles to using them, such as the Independence of Irrelevant Alternatives (IIA) property in the Multinomial Logit model, the need to assume specific structures and high calibration times, depending on the complexity of the case being evaluated. However, some of these mentioned constraints could be mitigated using Mixed Models or Nested Logit. Therefore, this paper proposes a comparative analysis between the Artificial Neural Network (ANNs), the Multinomial and Nested Logit models for disaggregated forecasting of urban trip distribution. A case study was conducted in a medium-sized Brazilian city, Santa Maria (RS), Brazil. The data used come from a household survey, prepared for the Urban Mobility Master Plan. For the sake of comparison, hit rates and frequency of trip distribution distances were analyzed, showing that ANNs can be as efficient as the Discrete Choice models for disaggregated forecasting of urban trip destination without, however, assuming some constraints. Finally, based on the results obtained, the efficiency of ANNs is observed for predicting alternatives with a low number of observations. They are important tools for obtaining Origin-Destination matrices from incomplete sample matrices or with a low number of observations. However, it is important to mention that discrete choice models can provide important information for the analyst, such as statistical significance of parameters, elasticities, subjective value of attributes, etc.

RESUMO
Este artigo propõe uma análise comparativa entre Redes Neurais Artificiais (RNAs), Logit Multinomial e Aninhado para uma previsão desagregada de distribuição de viagens urbanas. O estudo de caso foi a cidade de Santa Maria (RS). Os dados utilizados foram originados da pesquisa domiciliar, realizada para elaboração do Plano Diretor de Mobilidade Urbana. As comparações entre abordagens foram realizadas através de taxas de acertos e frequências de distâncias de viagens, mostrando que RNAs podem ser tão eficientes quanto os modelos de escolha discreta, sem assumir algumas restrições. Finalmente, com base nos resultados, pode-se afirmar que as RNAs são eficientes para previsão de alternativas com baixo número de observações. São importantes ferramentas para obtenção de matrizes O/D a partir de matrizes incompletas ou com baixos números de observações. Contudo, vale ressaltar que modelos de escolha discreta fornecem informações importantes, como significância estatística dos parâmetros estimados, elasticidades, valores subjetivos de atributos, etc.

1. INTRODUCTION

Over the years, many techniques and approaches have been used to forecast trip distributions, ranging from classic aggregate models (Casey,1955; Schneider, 1959; Evans 1970; Wilson, 1967; Evans and Kirby, 1974; Williams, 1976) to disaggregated individual behavior models (Fotheringham, 1983; Ben - Akiva and Lerman, 1985). Thus, the current literature presents important discussions related to the constraints of each method.

Among the aggregated classic models, the most relevant ones are the Fratar, Gravity and Intervening Opportunities ones. The Fratar model (Evans, 1970; Williams, 1976) is based on the growth factor and, although it is considered easy to implement and understand, its drawbacks include: the fact that it depends on the current matrix accuracy and that it does not consider, directly, changes in transport systems'configurations (Cascetta et al., 2007; Ortúzar and Willumsen, 2011).

In turn, the gravity model, developed by Casey (1955), is probably the most used model for the trip distribution stage. The consideration of impedance effects, such as time and cost, and the fact that it does not require a complete initial origin-destination matrix (Ortúzar and Willumsen, 2011) are the biggest advantages compared to the Fratar model.

Finally, the Intervening Opportunities model, conceived by Stouffer (1940), considers that the probability for a trip to have a specific destination is proportional to the amount of opportunities offered by it. Although it has a consistent theoretical basis (Wilson, 1967), this model is not usual due to the fact that it is difficult to understand, it has few advantages compared to the Gravity model and has a lack of appropriate applications (Ortúzar and Willumsen, 2011). Moreover, its applicability is indicated for studying trips in which the opportunities are decisive for the destination choice.

Despite having unique characteristics and properties, the aforementioned models have their aggregate nature in common, that is, the fact that they do not consider the individual and household characteristics that also guide the individual´s destination choice. This property of the classic models was decisive for discussing their efficiency in reflecting what happens in reality and in predicting future scenarios. In this context, the discrete choice models can be seen as an alternative to traditional aggregate models.

Although it is, in general, applied to the modal choice stage, the Multinomial Logit model has also been used, due to its disaggregated nature, for the trip distribution stage (Chow et al., 2005; Mishra et al., 2013), in which the destination options correspond to the different alternatives. Despite the good results obtained, the Multinomial Logit model has mathematical constraints, such as population distribution assumptions, multicollinearity problems and the fact that it considers IIA. This property assumes that the probability of choosing an alternative, rather than others, will not be affected by the inclusion or withdrawal of new alternatives (Luce and Suppes, 1965), which does not reflect the reality in the case of correlation between alternatives.

Although some of these limitations are absent in other models of discrete choice models, such as Mixed models and Nested Logit models, various studies have been published about applications of Artificial Intelligence (AI) tools for forecasting travel demand and their spatial interactions (Faghri and Sandeep, 1998; Tillema et al., 2006; Pitombo et al., 2009; Rasouli and Nikraz, 2013; Pitombo et al., 2017). This research concluded that AI tools can forecast spatial trip distribution accurately.

Regarding the Artificial Neural Networks (ANNs), Shmueli et al. (1996) emphasized the following characteristics: applicability for large databases and no requirements, as for other AI tools, to previously formalize associations between variables. Despite this, it is clear that there are many studies in the literature regarding the application of ANNs for travel mode choice, but there are few studies regarding its use for the trip distribution stage, especially with a disaggregated database. Studies, such as Black (1995), Subba Rao et al. (1998), Carvalho et al. (1998), Mozzolin et al. (2000), Cantarella and de Luca (2005), Rasouli and Nikraz (2013), and Tillema (2006) showed greater precision in using ANNs when compared to the Gravity, Multinomial Logit and Nested Logit models, even when the amount of data is scarce. Although Carvalho et al. (1998) used ANNs for trip distribution in a disaggregated approach, their research does not include socioeconomic characteristics of individuals, which could increase the model´s predictive power, and, in this case, it was applied to Stated Preference data.

Therefore, this paper aims to make a comparative analysis between ANNs, Multinomial and Nested Logit models for a disaggregated analysis of urban trip distribution. To do this, data from a household survey of a medium-sized Brazilian city was used, which was divided into Traffic Analysis Zones (TAZs), which are the possible destinations for each trip. In addition to the individuals' socioeconomic data, the models considered aggregated information about the origins and destinations TAZs and trip distances.

This article has 5 sections, in addition to this introduction. Section 2 briefly describes the tools used in this study: Artificial Neural Networks, the Nested and Multinomial Logit models. Section 3 provides a brief description about the city of Santa Maria (Rio Grande do Sul, Brazil), whose data were used in this study and which are also detailed in this section. In addition, the same section deals with the data processing carried out. Section 4 presents the proposed methodological procedure and the applications used. Then, the results corresponding to each stage of the method are presented and discussed in Section 5. Finally, the conclusions obtained from this study are discussed in Section 6.

2. TOOLS USED FOR ESTIMATING URBAN DESTINATION CHOICE

2.1. Artificial Neural Networks (ANNs)

One of the most promising Artificial Intelligence techniques is the Artificial Neural Networks (ANNs). This tool reproduces the behavior of mathematical functions, including non-linear ones (Smith, 1996) and works as a processor, consisting of processing units, called neurons, whose function is to store experimental knowledge and make it available for use. ANN's functioning is based on the structure of the human brain, since knowledge is acquired through a learning process. In the case of ANNs, the connection between neurons, known as synaptic weights, is used to store the knowledge acquired (Bishop, 1995).

In general, Neural Networks are trained to obtain output data from input data. Knowledge is then maintained in neurons, and this training process is called Machine Learning. Learning is a process that allows the network to gradually adjust and adapt synaptic weights and their connections to form an increasingly accurate model (Carvalho et al., 1998).

Regarding the advantages of the tool, Shmueli et al. (1996) enumerated the following characteristics of Artificial Neural Networks: (1) their ability to work with extensive databases and (2) no prior knowledge related to relationships between variables or population distributions. In contrast, Mozolin et al. (2000) highlight the inductive nature of ANNs by obtaining better results with traditional models for forecasting travel distribution.

An ANN consists of input layers (independent variables), hidden and output (estimated dependent variable) layers. The information saved in the input layer is transferred to the output layer through the hidden layers. In this study, we used the multilayer perceptron network according to the model in Figure 1. This network can extract more accurate data when compared to single layer networks. In addition, this model has greater precision, rapid convergence and allows a large amount of input data (Bishop, 1995).

Concerning architecture, when the network has a large number of hidden layers or is trained by many interactions, overtraining can occur. In this case, the network presents a good calibration for the trained database, but a low generalization power for untrained data. Due to this, a Neural Network must be sought that allows good precision, but with a smaller number of neurons and layers, so that the model is efficient (Dougherty, 1995). In this research, in order to maintain a maximum number of hidden layers, a limitation of a minimum of 1 and a maximum of 20 hidden layers was imposed.

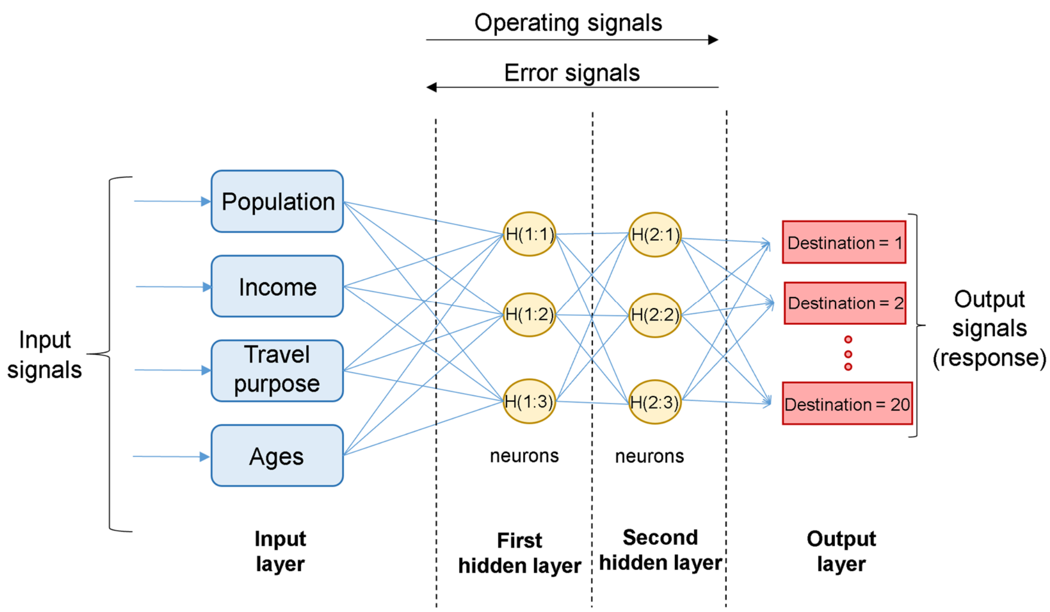

Figure 1. Representation of a Neural Network with two hidden layers. Adapted from Bishop (1995).

The results depend directly on the network's ability to extract information and replicate it. Therefore, according to Cantarella and De Luca (2005), an efficient architecture should preferably have a great reproductive power, measured by the error between simulated and calibrated observations; good generalization, measured by the error between simulated and validated observations; and low dependence on initial conditions, assessed by the error dispersion between observations and simulations for calibration. Thus, the architecture was selected in this study based directly on the IBM SPSS 24.0 software, which automatically chooses the best architecture.

Another important configuration for constructing the Neural Network is the applied training modality. According to IBM (2016), the type of training determines how the network processes records. In this study, batch, online and mini-batch training were tested. According to IBM (2016), batch training directly minimizes the total number of errors, but it may need many updates of the weights and, therefore, it needs many data transmissions.

The online training updates the synaptic weights after inserting each record into the training, commonly used in large databases. The mini-batch training, in turn, divides the observations into equally-sized groups and, for each batch, updates the weightings. In this step, the network with the highest number of hits was selected to make comparisons with the traditional parametric methods (in the case of this study, the Multinomial and Nested Logit models).

The interruption rules used for training were: a maximum of 10 steps with no error reduction, maximum training time of 15 minutes, automatic calculation of training periods, minimum relative change in the training error of 0.0001 and a maximum of 10,000 cases stored in the memory.

After training, the synaptic weights were saved and applied to the database with 30% of the records for validation. The results evaluated the number of hits and the accuracy of the obtained Neural Network.

2.2. Multinomial Logit

Discrete choice analysis was used to model preferences, based on the random utility theory (Mcfadden 1974; Ben-Akiva and Lerman, 1985). This theory assumes that every individual is a rational decision-maker, maximizing utility relative to their choices. The models adopted in this study comprise Multinomial Logit (MNL) and Nested Logit (NL). For instance, prior studies that compare logit models and Machine Learning in travel mode choice have an important limitation: the comparisons were usually made between the MNL model, (the simplest logit model in literature) and Machine Learning techniques (Zhou et al., 2020). Simpler structures were tested first, such as MNL models (Mcfadden, 1974), assuming that stochastic errors have an IID Gumbel distribution. The utilities are configured as follows:

$$V_{in} = a_i + b_{1i} \cdot x_{1\mathrm{i}} + c_{2i} \cdot x_{2\mathrm{n}} \tag{1}$$

V_{in}: the utility of alternative i for the individual n; $x_{1\mathrm{i}}$: the explanatory variable related to the i alternative; and $x_{2\mathrm{n}}$: the explanatory variable related to individual n; a_i, b_{1i} and c_{2i}: coefficients to be estimated.

After the utilities for each alternative have been defined, whose coefficients are estimated from the maximum likelihood, the probabilities of the alternatives to be chosen can be calculated, for each individual n, and they are defined by:

$$P_{in} = \frac{\mathrm{e}^{V_{in}}}{\sum_{j=1}^{z} \mathrm{e}^{V_{jn}}} \tag{2}$$

P_{in}: the probability of the alternative i to be chosen by individual n; z: number of alternatives.

The standard logit model exhibits independence from irrelevant alternatives (IIA), which implies proportional substitution across alternatives (Train, 2009). This assumption for the distribution of residuals is rather simplistic, as they depend on the hypothesis of independence and homoscedasticity of residues (Ben-Akiva et al., 2003).

2.3. Nested Logit

More complex logit models, such as the mixed logit and nested logit, can be derived similarly from different assumptions about the coefficients and error-term distribution. However, these models are more difficult to estimate. For instance, mixed logit models do not have closed-form solutions for the likelihood function and require the simulation of maximum likelihood for various parameter estimations (Zhao et al., 2020). This paper initially attempted to estimate mixed logit models, such as Mixed Logit with random coefficients to examine heterogeneity in behavior, and Mixed Logit with Error Components to analyze correlation among alternatives. Due to the simulation process in mixed logit models, it was observed that the processing time of the mixed model exceeded more than one hour, which made us exclude this model from the analyses presented in this study.

Therefore, Nested Logit (NL) models (Daly and Zachary, 1978; Williams, 1977) were estimated in order to include possible correlations between unobserved attributes of alternatives. NL is the model of the Generalized Extreme Value (GEV) family of models - best known and used. The main feature of GEV models is the correlation of the observed non-use of alternatives. In this case, non-independent alternatives can be grouped considering similarities. Alternatives contained in the same nest have the replacement pattern of independence probabilities of irrelevant alternatives (IAA) of the MNL model (Train, 2009; Ortúzar and Willumsen, 2011; Tavasszy and de Jong, 2014).

3. CASE STUDY APPLIED TO A MEDIUM-SIZED BRAZILIAN CITY

3.1. A description of the city of Santa Maria (Rio Grande do Sul, Brazil)

The information explored in this research is related to the TAZs of the city of Santa Maria (Rio Grande do Sul, Brazil). Information regarding individuals and trips were obtained from a household survey carried out to prepare the Urban Mobility Plan (IPLAN, 2013). The sociodemographic data of the origins TAZs are from the IBGE (Brazilian Institute of Geography and Statistics) Census (IBGE, 2010).

The household survey focused on the main urban district, in which 246,465 inhabitants live and work. According to data from the IBGE Census 2010, the district in question was formed by 41 neighborhoods (Figure 2). The city's structure is influenced by local physical characteristics, such as the railway line that divides downtown; this urban barrier makes accessibility difficult in some neighborhoods, especially for pedestrians, and isolates peripheral neighborhoods, as there are great distances between some of the 11 existing intersection points.

According to the IBGE survey in 2010, the city under study has a young population, in which residents between 15 and 29 years old predominate. Most men are between 1 and 14 years old, but among the population over 15 years old, the number of women exceeds the number of men. Urban population growth in Santa Maria compared to the rural population started in 1970 and the evolution of the population from 2002 to 2012 remained constant, showing an annual increase of 6%. The most relevant growth occurred in the most distant districts from downtown, such as Nova Santa Marta, Pinheiro Machado and Camobi (Figure 2). The highest concentration of residents is found in the neighborhoods Camobi, Downtown (Centro), Juscelino Kubitschek and Nova Santa Marta (Figure 2), which together represent about 30% of the population.

Regarding the inhabitant’s income profile, only 49% of people have an income above the minimum monthly wage (about US$290.00 in 2010). However, Santa Maria is considered the 28th city with a number of residents belonging to class A (with a family income equal to or greater than twenty minimum wages) and is one of the three cities in the state with the greatest consumption potential. Most of the wealth (represented by the sum of household monthly wages) is concentrated in the neighborhoods of Downtown (Centro), Nossa Senhora de Fátima and Camobi.

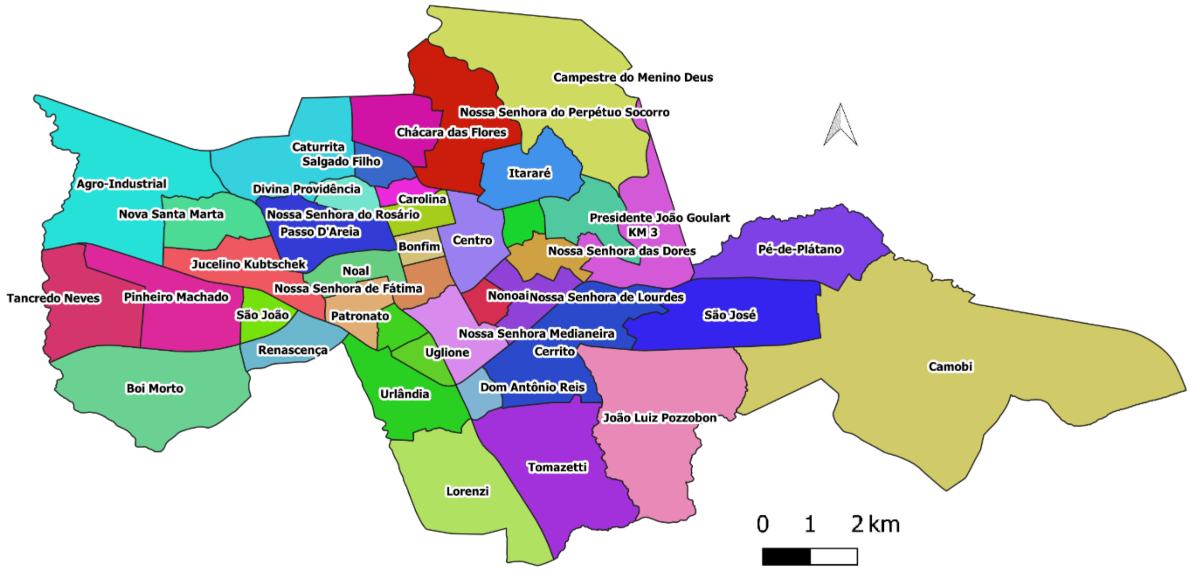

Figure 2. Layout of the 41 neighborhoods that comprise the main urban district of the city of Santa Maria.

The city is considered an attractive pole in the region because it is a reference for services in 36 cities in the central region of the state and influences more than one million people. Regarding the municipality's economy, the service sector stands out, representing about 81% of the generated income (IPLAN, 2013).

Santa Maria also stands out for having the status of “city of education”. The city has more than seven higher education institutions and more than 35,000 students, which is why it is considered the third city in Brazil with the highest number of masters and doctors per capita. Approximately 27% of the population has incomplete higher education and 13% complete higher education. These data point to a high level of formal education for the population as only 8.3% of the national population has finished higher education. Educational institutions represent major traffic-generating hubs, because together (primary, secondary and higher) they make a total of 95,784 students and 5,611 professors (27% of students and 30.7% of professors belong to the Federal University of Santa Maria (FUSM).

The main traffic generating centers, in the city, are health, educational, industrial and recreation facilities. The health facilities include 41 basic units, 5 emergency units and 11 emergency care units. The units have a capacity of 1,200 hospital beds and 1,100 medical professionals working in the city and the neighborhoods that most concentrate these units are Downtown (Centro) and Camobi. The educational facilities include 39 state schools, 80 municipal schools, 13 private schools and 2 military schools. Concerning universities, the biggest centers are the Federal University of Santa Maria, located in the Camobi neighborhood and where more than 30,000 people access it daily, and the Franciscan University, located in the Downtown (Centro) neighborhood. Industrial facilities account for a total of 572 industries, which formally employed 6,234 people in 2010. The main recreation places are located in the Nossa Senhora das Dores neighborhood, with the largest shopping and sports club in the city, Parque Itaimbé, located Downtown (Centro), and FUSM, in the Camobi neighborhood, which houses exhibitions, a community center and presentations. Readers can verify the spatial location of the mentioned neighborhood in Figure 2.

The largest survey of local data was developed as part of the development of the Urban Mobility Plan and its data was used as a basis for the development and understanding of this study.

3.2. Household survey

Important sources for travel demand forecasting and, consequently, for urban planning, household surveys seek to obtain characteristics about households, residents and their daily trips. In order to obtain the minimum number of interviews required for the city of Santa Maria, carried out in 2013, a 95% confidence level and a maximum sampling error of 10% were considered. By the end of the survey, there was a total of 3,758 records (IPLAN, 2013). By correcting any inconsistencies and lack of data, the sample was reduced to 3,136 observations.

3.3. Data processing

Due to the computational constraints related to calibrating the Multinomial and Nested Logit models, the 41 neighborhoods were aggregated into 20 TAZs, so that each of them would represent a possible destination. The purpose of this aggregation was to reduce the number of alternatives for Logit application. This aggregation was performed based on a k-means cluster analysis, in which the geographical coordinates of the centroids of the neighborhoods were considered as a criterion for the clusters. Figure 3 illustrates the layout of the 20 resulting TAZs and Table 1 identifies the neighborhoods that make up each TAZ at the end of the clustering. More details regarding the unit areas clustering can be found in the study carried out by Caldas et al. (2021).

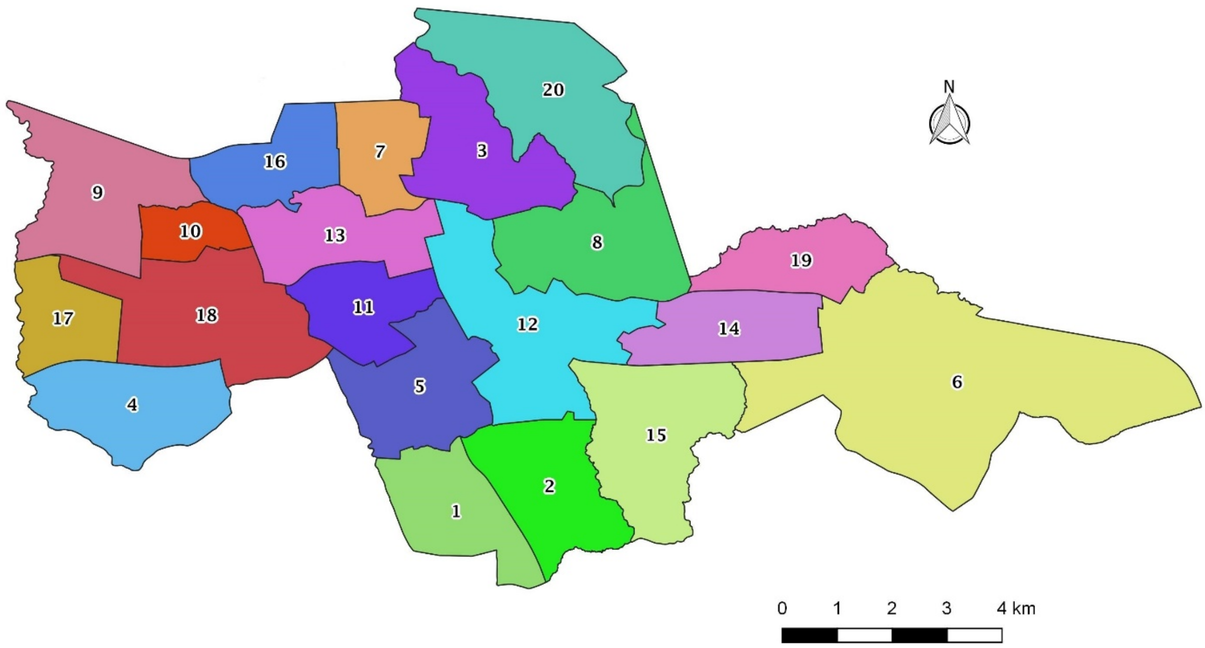

Figure 3. Division resulting from the neighborhood clustering in Santa Maria (RS) – Traffic Analysis Zones

In addition, the household survey database included information about the TAZs of origin and possible destinations and travel distances, which correspond to the Euclidean distances between the TAZs' centroids. In the case of trips within the same TAZ, to prevent systematic errors due to zero distances, a quarter of the distance between the centroid of the area unit in question and the nearest centroid neighbor area was considered (Thomas and Hugget, 1980). Readers who are interested in using more recent methods to calculate intrazonal trip distances can consult the study carried out by Plaza and Rodrigues da Silva (2021).

Table 1 - Characterization of the dependent variable

Zone	Neighborhood included
1	Lorenzi
2	Tomazetti
3	Itararé, Nossa Senhora do Perpétuo Socorro
4	Boi-Morto
5	Nossa Senhora Medianeira, Urlândia, Uglione; Dom Antônio Reis
6	Camobi
7	Carolina, Chácara das Flores, Salgado Filho
8	Km 3, Menino Jesus, Nossa Senhora das Dores, Presidente João Goulart
9	Agro-Industrial
10	Nova Santa Marta
11	Duque de Caxias, Noal, Nossa Senhora de Fátima, Patronato
12	Centro, Cerrito, Nonoai, Nossa Senhora de Lourdes
13	Bonfim, Divina Providência, Nossa Senhora do Rosário, Passo D'Areia
14	São José
15	João Luiz Pozzobon
16	Caturrita
17	Tancredo Neves
18	Juscelino Kubitschek, Pinheiro Machado, Renascença, São João
19	Pé-de-Platano
20	Campestre do Menino Deus

Table 2 characterizes the explanatory variables used in the ANNs, Multinomial and Nested Logit models. Figure 4 characterizes the dependent variable as to the percentage of choice for each TAZ (destinations choices), according to the final sample of 3,166 records, obtained after data treatment.

Table 2 - Variables used by the models

Variable	Description
School Frequency	(0) No; (1) Yes
Gender	(0) Male; (1) Female
Scholarity Level	(1) Illiterate; (2) Literate; (3) Incomplete Elementary School; (4) Complete Elementary School; (5) Incomplete Middle School; (6) Complete Middle School; (7) Incomplete High School; (8) Complete High School; (9) Incomplete Higher Education; (10) Complete Higher Education
Driver's License	(0) No; (1) Yes
Age	(1) Up to 17 years; (2) 18 to 28 years; (3) 29 to 39 years; (4) 40 to 51 years; (5) 52 to 65 years; (6) Over 66 years
Residents	Number of residents in the residence
Education Purpose	(0) No; (1) Yes
Integration Purpose	(0) No; (1) Yes
Recreation Purpose	(0) No; (1) Yes
Residence Purpose	(0) No; (1) Yes
Health Purpose	(0) No; (1) Yes
Work Purpose	(0) No; (1) Yes
Income*	(1) Up to 1 minimum wage*; (2) 1 to 2 minimum wages*; (3) 2 to 5 minimum wages*; (4) 5 to 10 minimum wages*; (5) Over 10 minimum wages*
Worker	(0) No; (1) Yes
Motor Vehicles	Number of motor vehicles in the residence
Origin Population	Number of residents in origin zone
Distance	Distance from the origin zone to possible destination zones
Bus Stops	Difference between the number of bus stops in the origin and destination zones

*Minimum wage reference value in 2013: US$314.00. Value in 2020: US$240.00

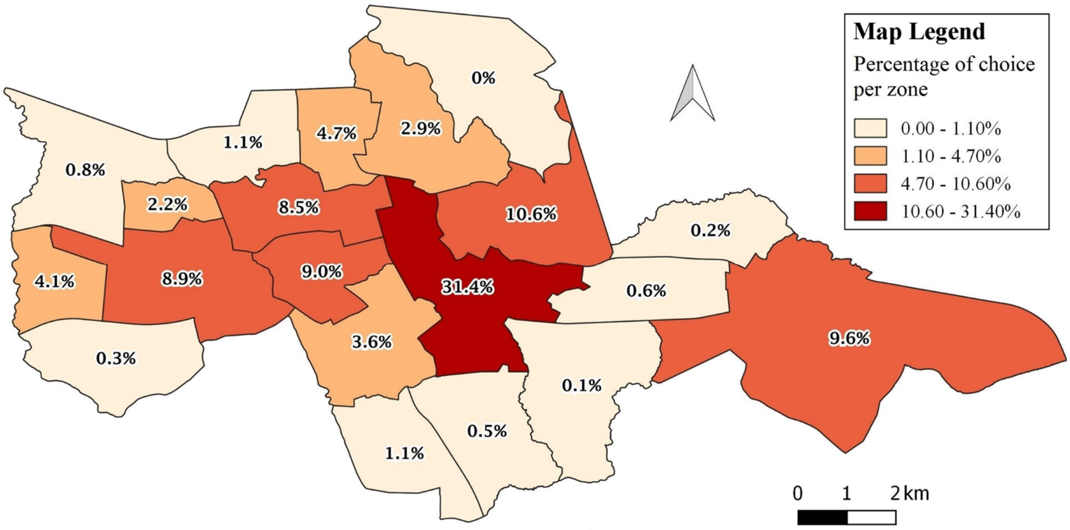

Figure 4. Characterization of the Dependent variable in relation to the percentage of choice per TAZ

Figure 4 shows that the database does not have trips with zone 20 as a destination (to the north of the city), comprised by the 'Campestre do Menino Deus' neighborhood. Thus, the analyses performed here do not include this destination in their estimates. Thus, the final choice set is formed by 19 alternatives.

4. METHODOLOGICAL PROCEDURE

The methodological procedure of this study, presented in Figure 5, begins with data processing, as described in Subsection 3.3. Then, Artificial Neural Networks, the Multinomial and Nested Logit models were applied to study and forecast destinations from the calibration sample, which corresponds to 70% of the total observations (calibration and training stage). Then, the calibrated models and the remaining 30% sample were used to forecast individual destinations (validation and test stage). Finally, comparisons and performance analyses were made between the three methods (Multinomial, Nested Logit and ANNs). To do this, the hit rate and the trip distribution distances were evaluated, calculated from the validation and test sample (30%). Some theoretical comparison regarding the three approaches was also made. The use of distance distribution is based on calibrating the Gravity distribution model based on adjusting the modeled and observed trip impedance distribution curves (Ortuzar and Willumsen, 2011).

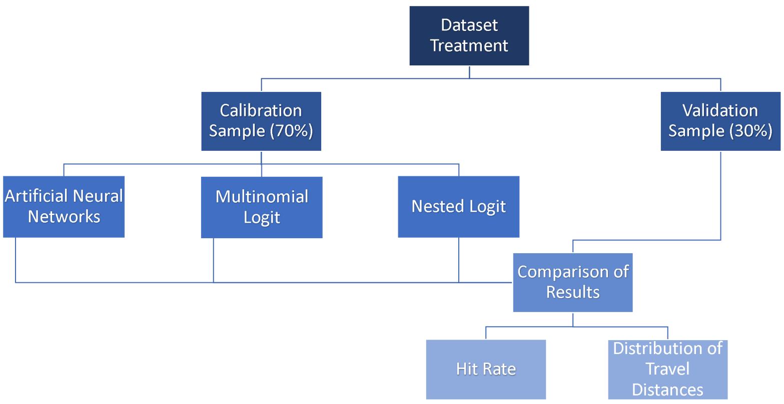

Figure 5: Method proposed for a comparative analysis between Artificial Neural Network, Multinomial and Nested Logit models for urban trip distribution in a disaggregated approach

4.1. Applications used

The analysis carried out using ANNs and the Cluster Analysis, for grouping the neighborhoods, were obtained through the IBM SPSS 24.0 application. The calibration of the discrete choice models (Multinomial and Nested Logit), in turn, was performed using R (R Core Team, 2020) and the Apollo package (Hess and Palma, 2019). The TAZs centroids and the distances between them were obtained using the QGIS software (version 3.6.3).

5. RESULTS

5.1. Artificial Neural Networks

The automatic selection of the ANN architecture of the IBM SPSS 24.0 software provided the most efficient network among the different trainings. Table 3 shows, for each type of ANN analyzed (batch, online and mini-batch), the architecture, errors and percentages of hit rates.

Table 3 - Architecture and results of the ANNs analyzed.

Training method	Batch	Online	Mini-batch
Hidden layers	1	1	1
Neurons in the hidden layer	16	18	18
Hidden layer activation function	Hyperbolic tangent	Hyperbolic tangent	Hyperbolic tangent
Units in the output layer	19	19	19
Output layer activation function	Softmax	Softmax	Softmax
Error function	Cross entropy	Cross entropy	Cross entropy
Correct preditions in training	46.50%	40.20%	37.60%
Stop rule used	Maximum number of seasons (100) exceeded	Relative change in training error (.0001) obtained	10 consecutive steps without any error reduction
Training time	0:00:6.77	0:00:5.16	0:00:02.86
Correct predictions in validation	36.38%	38.09%	35.21%

Although batch training is used for small databases, due to the greater computational time required, it can be concluded, from Table 3, that it proved to be the most efficient for the case study as its training time was only 6.77 seconds. This method provided the greatest number of hits in training, totaling 46.50%, showing a greater reproductive power, according to Cantarella and De Luca (2005). In the validation, the three ANNs exhibited similar correct prediction values, that is, the networks have the same generalization power. Based on these results, the data from the batch training network were chosen to be analyzed and compared to the Multinomial and Nested Logit models. It is worth mentioning that the training and test procedure used by ANNs enables the future projection of the variable of interest. Thus, it is possible to obtain future values of the dependent variable from future values of the independent variables and the model obtained by training stage.

As a result, the Neural Network also offers the normalized importance of the independent variables. The importance of an independent variable expresses how much it influences the selection of the predicted value. Its measurement is divided by the most important variable and the result is expressed as a percentage. Thus, the attributes “scholarity level”, “income” and “integration purpose” were considered the most important variables in the model, while the variable “bus stops” for zones 12, 13 and 10 were the least important.

5.2. Multinomial Logit

The calibration of the Multinomial Logit model followed the formulation described by Equation

(1) and (2), and the explanatory variables were those shown in Table 2. Several calibrations were performed, so that for each one, the parameters that were not significant in the previous calibration were excluded, considering a 95% significance level (p-value associated with the t-test less than 0.05). Therefore, this procedure was repeated until a calibration contained all statistically significant parameters. The final calibration resulted in the utility functions shown in Table 4, with ρ^2 having a value of 0.357 and ρ^2 adjusted with a value of 0.338.

Table 4 - Logit Multinomial - Utility functions for each destination Traffic Analysis Zone

$$U_1 = -6.26^* \cdot Distance_1 + 9.91^* \cdot BusStops_1$$
$$U_2 = -2.79^{**} \cdot Residents + 1.35^{**} \cdot RecreationPurpose - 3.16^* \cdot ResidencePurpose - 3.26^{**} \cdot Distance_2 + 14.2^* \cdot BusStops_2$$
$$U_3 = 5.73^* \cdot IntegrationPurpose + 4.42^* \cdot HealthPurpose - 6.03^* \cdot Distance_3 + 8.65^* \cdot BusStops_3$$
$$U_4 = -14.9^* + 4.86^* \cdot ScholarityLevel + 6.14^* \cdot Driver'sLicense - 4.46^* \cdot EducationPurpose - 9.44^* \cdot ResidencePurpose + 5.08^* \cdot Worker + 8.90^* \cdot BusStops_4$$
$$U_5 = -1.10^* \cdot Age - 3.01^* \cdot Residents - 1.95^* \cdot IntegrationPurpose - 3.78^* \cdot HealthPurpose - 5.06^* \cdot Distance_5 + 8.21^* \cdot BusStops_5$$
$$U_6 = 2.1^* \cdot OriginPopulation + 0.66^* \cdot Driver'sLicense - 2.52^* \cdot Residents + 1.74^* \cdot EducationPurpose - 3.72^* \cdot IntegrationPurpose - 2.46^* \cdot ResidencePurpose + 4.02^* \cdot HealthPurpose - 3.12^* \cdot Distance_6 + 9.01^* \cdot BusStops_6$$
$$U_7 = 2.56^* + 5.73^* \cdot IntegrationPurpose + 4.38^* \cdot HealthPurpose - 2.58^* \cdot Income - 5.25^* \cdot Distance_7 + 8.39^* \cdot BusStops_7$$
$$U_8 = 2.08^* - 1.94^* \cdot Residents + 5.25^* \cdot IntegrationPurpose - 2.16^* \cdot ResidencePurpose + 3.50^* \cdot HealthPurpose - 5.33^* \cdot Distance_8 + 8.22^* \cdot BusStops_8$$
$$U_9 = 1.99^* \cdot OriginPopulation - 1.60^{**} \cdot Gender - 6.81^* \cdot EducationPurpose - 3.72^* \cdot ResidencePurpose + 4.27^* \cdot HealthPurpose - 3.18^* \cdot Distance_9 + 9.60^* \cdot BusStops_9$$
$$U_{10} = -2.51^* \cdot SchoolFrequency - 3.55^* \cdot ScholarityLevel - 2.83^* \cdot Age - 2.13^{**} \cdot Residents + 6.80^* \cdot RecreationPurpose + 5.99^* \cdot ResidencePurpose + 11.10^* \cdot HealthPurpose + 5.71^* \cdot WorkPurpose - 3.11^* \cdot Distance_{10} + 5.51^* \cdot BusStops_{10}$$
$$U_{11} = 1.92^* \cdot ScholarityLevel - 1.68^* \cdot Residents + 5.49^* \cdot IntegrationPurpose - 1.47^* \cdot ResidencePurpose + 3.36^* \cdot HealthPurpose - 4.83^* \cdot Distânce_{11} + 3.61^* \cdot BusStops_{11}$$
$$U_{12} = 1.62^* \cdot ScholarityLevel - 2.26^* \cdot Residents + 6.07^* \cdot IntegrationPurpose - 2.66^* \cdot ResidencePurpose + 3.75^* \cdot HealthPurpose - 2.50^* \cdot Distance_{12} + 7.99^* \cdot BusStops_{12}$$
$$U_{13} = 2.51^* - 1.65^* \cdot Residents - 4.09^* \cdot IntegrationPurpose - 1.76^* \cdot ResidencePurpose + 2.77^* \cdot HealthPurpose - 6.74^* \cdot Distance_{13} + 7.15^* \cdot BusStops_{13}$$
$$U_{14} = 2.43^* \cdot EducationPurpose - 1.74^* \cdot HealthPurpose - 5.55^* \cdot WorkPurpose - 6.21^* \cdot Distance_{14} + 8.18^* \cdot BusStops_{14}$$
$$U_{15} = 6.05^* \cdot SchoolFrequency + 2.87^{**} \cdot Driver'sLicense - 11.5^* \cdot Age + 6.56^* \cdot RecreationPurpose - 6.07^* \cdot ResidencePurpose + 5.90^* \cdot WorkPurpose$$
$$U_{16} = -4.95^* \cdot OriginPopulation - 1.98^* \cdot IntegrationPurpose - 5.93^* \cdot RecreationPurpose - 1.26^* \cdot ResidencePurpose - 3.89^* \cdot HealthPurpose - 1.59^{**} \cdot Income - 14.30^* \cdot Distance_{16}$$
$$U_{17} = -2.85^* \cdot Residents - 2.27^* \cdot IntegrationPurpose + 3.10^* \cdot HealthPurpose - 4.19^* \cdot Distance_{17} + 7.18^* \cdot BusStops_{17}$$
$$U_{18} = -2.79^* \cdot Residents + 0.823^* \cdot EducationPurpose - 3.26^* \cdot IntegrationPurpose + 4.04^* \cdot HealthPurpose - 1.65^* \cdot Income - 4.14^* \cdot Distance_{18} + 5.56^* \cdot BusStops_{18}$$
$$U_{19} = -6.49^* \cdot SchoolFrequency - 9.06^* \cdot ResidencePurpose - 3.22^* \cdot HealthPurpose + 5.81^{**} \cdot BusStops_{19}$$

Significance level: P-value < 0.01 **; P-value < 0.05 *

By analyzing Table 4, the factors that influence the choice of each destination by individuals can be understood. In the equation that represents zone 6, for example, it is observed that Study and health trip purposes significantly increases the utility of this zone, and, consequently, the probability of the individual choosing it as a destination. The opposite happens if the trip is motivated by integration. This result is easily understood when observing in zone 6, comprising the Camobi neighborhood, the presence of the Federal University of Santa Maria (FUSM), an important attractive hub for students in the city, and of health facilities, as explained in Subsection 3.1. On the other hand, there are no importante integration hubs in this region, which would justify the negative effect of this motivation on the destination in question.

Regarding “age”, the negative influence of this variable can be seen in the choice of zone 10, comprising the “Nova Santa Marta” neighborhood. This can be explained by the fact that, according to the 2010 demographic census, this district is, among all 41 neighborhoods in the Santa Maria Headquarters district, the 41st in percentage of the elderly population (aged 60 or over) and the 1st in percentage of population in the minority (under 18 years old) (IBGE, 2010).

Another possible analysis concerns the “gender” variable, of which female individuals (represented by the number 1 in the database), for zone 9, represented by the “Agro-Industrial” neighborhood, are negatively related to utility of this location, which can be explained by the fact that this neighborhood, among all those that comprise the Headquarters district, is the one with the highest percentage of men (represented by the number 0 in the database), according to the 2010 population census (IBGE, 2010).

5.3. Nested Logit

The calibration of the Nested Logit model also followed the formulation described by Equation (1), and the explanatory variables were those shown in Table 2. As for the grouping of alternatives, different structures of NL models have been tested. This paper presents the best structure found in which we include three nests: (1) Nest 1: zones 1, 9, 10 and 15; (2) Nest 2: zones 2, 5, 17 and 18; (3) Nest 3: zones 6, 8 and 12.

Nest 1 consists of destinations with lower average household income (1.17 – 1.94 MW – Figure 6) and few observations – under 2.5% (Zone 1 – 1.1 %; Zone 9 – 0.8%; Zone 10 – 2.2%; Zone 15 – 0.1% - Figure 4). Nest 2, in turn, comprises low-middle-income destinations, but larger than those of Nest 1 (1.94 – 2.72 MW – Figure 6). Finally, Nest 3 is formed by destinations with high-average income (3.49 – 5.04 MW – Figure 6) and a large number of observations (Figure 4). Following that, Figure 6 represents thematic maps regarding average household income.

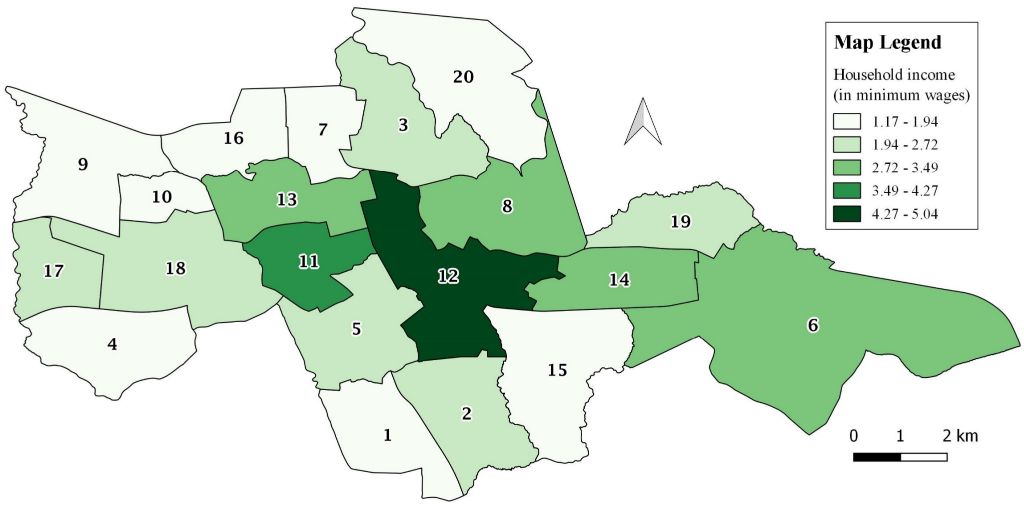

Figure 6. Characterization of TAZs - Average household income (Caldas et al., 2021).

In the implementation of the Nested Logit model, each nest must have a lambda parameter (λ) associated to it. For the model to be consistent with utility maximisation, the estimated value of the λ parameter of all nests should be between 0 and 1. When the values of λ falls between 0 and 1, the model is consistent with utility maximization for all possible values of explanatory variables (Train, 2009). The estimated NL model presented the following λ values: 0.7070 (Nest 1); 0.772 (Nest 2) and 0.6421 (Nest 3). The final calibration resulted in the utility functions shown in Table 5, with ρ^2 having a value of 0.357 and ρ^2 adjusted with a value of 0.339.

Table 5 - Nested Logit - Utility functions for each destination Traffic Analysis Zone

$$U_1 = -4.87^{**} \cdot Distance_1 + 10.56^{**} \cdot BusStops_1$$
$$U_2 = -2.30^{*} \cdot Residents - 2.77^{**} \cdot ResidencePurpose + 13.11^{**} \cdot BusStops_2$$
$$U_3 = 11.86^{**} \cdot IntegrationPurpose + 12.24^{**} \cdot HealthPurpose - 5.12^{**} \cdot Distance_3 + 9.16^{**} \cdot BusStops_3$$
$$U_4 = -30.07^{**} + 4.80^{**} \cdot ScholarityLevel + 13.29^{**} \cdot Driver'sLicense - 12.15^{**} \cdot EducationPurpose - 15.68^{**} \cdot ResidencePurpose + 13.01^{**} \cdot Worker + 9.95^{**} \cdot BusStops_4$$
$$U_5 = -0.92^{*} \cdot Age - 3.06^{**} \cdot Residents - 4.98^{**} \cdot IntegrationPurpose - 5.23^{**} \cdot Distance_5 + 9.32^{**} \cdot BusStops_5$$
$$U_6 = 2.05^{**} \cdot OriginPopulation + 0.48^{**} \cdot Driver'sLicense - 2.39^{**} \cdot Residents + 1.22^{**} \cdot EducationPurpose - 2.30^{**} \cdot ResidencePurpose + 11.61^{**} \cdot HealthPurpose - 2.06^{**} \cdot Distance_6 + 9.84^{**} \cdot BusStops_6$$
$$U_7 = 2.49^{**} + 12.06^{**} \cdot IntegrationPurpose + 12.05^{**} \cdot HealthPurpose - 2.56^{**} \cdot Income - 5.18^{**} \cdot Distance_7 + 9.25^{*} \cdot BusStops_7$$
$$U_8 = 2.28^{**} - 2.04^{**} \cdot Residents + 11.77^{**} \cdot IntegrationPurpose - 2.17^{**} \cdot ResidencePurpose + 11.25^{**} \cdot HealthPurpose - 3.75^{**} \cdot Distance_8 + 8.46^{**} \cdot BusStops_8$$
$$U_9 = 1.51^{**} \cdot OriginPopulation - 2.56^{**} \cdot ResidencePurpose + 12.05^{**} \cdot HealthPurpose - 3.04^{**} \cdot Distance_9 + 10.02^{**} \cdot BusStops_9$$
$$U_{10} = -2.02^{**} \cdot SchoolFrequency - 2.96^{**} \cdot ScholarityLevel - 2.35^{**} \cdot Age - 1.96^{**} \cdot Residents + 6.03^{**} \cdot RecreationPurpose + 5.46^{**} \cdot ResidencePurpose + 18.21^{**} \cdot HealthPurpose + 5.21^{*} \cdot WorkPurpose - 2.82^{**} \cdot Distance_{10} + 6.86^{**} \cdot BusStops_{10}$$
$$U_{11} = 1.61^{**} \cdot ScholarityLevel - 1.76^{**} \cdot Residents + 11.83^{**} \cdot IntegrationPurpose - 1.39^{**} \cdot ResidencePurpose + 11.03^{**} \cdot HealthPurpose - 5.05^{*} \cdot Dist\hat{a}nce_{11} + 7.94^{**} \cdot BusStops_{11}$$
$$U_{12} = 1.24^{**} \cdot ScholarityLevel - 2.19^{**} \cdot Residents + 12.33^{**} \cdot IntegrationPurpose - 2.47^{**} \cdot ResidencePurpose + 11.41^{**} \cdot HealthPurpose - 2.04^{**} \cdot Distance_{12} + 8.56^{**} \cdot BusStops_{12}$$
$$U_{13} = 2.22^{**} - 1.63^{**} \cdot Residents - 1.68^{**} \cdot ResidencePurpose + 10.47^{**} \cdot HealthPurpose - 6.82^{**} \cdot Distance_{13} + 8.05^{**} \cdot BusStops_{13}$$
$$U_{14} = 2.23^{**} \cdot EducationPurpose - 12.89^{**} \cdot WorkPurpose - 6.17^{**} \cdot Distance_{14} + 8.94^{**} \cdot BusStops_{14}$$
$$U_{15} = -22.96^{**} + 12.48^{**} \cdot SchoolFrequency + 2.39^{*} \cdot Driver'sLicense - 11.49^{**} \cdot Age + 13.59^{**} \cdot RecreationPurpose + 13.18^{**} \cdot WorkPurpose$$
$$U_{16} = 9.41^{**} - 5.27^{**} \cdot OriginPopulation - 4.48^{**} \cdot IntegrationPurpose - 13.86^{**} \cdot RecreationPurpose - 1.23^{**} \cdot ResidencePurpose - 1.62^{*} \cdot Income - 14.43^{**} \cdot Distance_{16}$$
$$U_{17} = 3.36^{**} - 2.78^{**} \cdot Residents - 5.22^{**} \cdot IntegrationPurpose + 10.99^{**} \cdot HealthPurpose - 4.01^{**} \cdot Distance_{17} + 7.55^{**} \cdot BusStops_{17}$$
$$U_{18} = 1.92^{**} - 2.75^{**} \cdot Residents + 0.69^{*} \cdot EducationPurpose + 11.57^{**} \cdot HealthPurpose - 1.27^{**} \cdot Income - 4.13^{**} \cdot Distance_{18} + 6.79^{**} \cdot BusStops_{18}$$
$$U_{19} = -13.06^{**} \cdot SchoolFrequency - 15.31^{**} \cdot ResidencePurpose - 5.96^{**} \cdot HealthPurpose + 6.51^{*} \cdot BusStops_{19}$$

Significance level: P-value < 0.01 **; P-value < 0.05 *

Observing Table 5, it can be stated that the same analyses performed on the utility functions of the Multinomial Logit model can be applied to the Nested Logit model, except for the analysis on the variable "gender" in the target function 9, which was discarded.

5.4. Comparison between approaches

The following subsections present different comparisons between the approaches. Initially, a comparison between the discrete choice models is carried out aiming to identify the best model regarding some traditional metrics. Following, we made a comparison between ANNs, MNL and NL models taking into account the trip impedance distribution and hit rates. Finally, a theoretical and brief discussion related to the advantages and disadvantages of ANNs and discrete choice models is presented.

5.4.1. Discrete choice models

According to Zhao et al. (2020), the selection of discrete choice models can be performed based on theoretical measures (R2, adjusted McFadden's pseudo R2, AIC, and/or BIC) in order to determine a best-fitting model. The adjusted McFadden's pseudo R2 is the most commonly used. However, AIC and BIC are often used to compare models with different numbers of variables.

It is noteworthy that to estimate the discrete choice models, we include only the significant parameters with at least 95% confidence. To estimate the NL model, 7 significant parameters in the MNL model had to be excluded because they were not significant in the NL model. Table 6 compares statistics from the MNL and NL models. Both models showed a good fit considering McFadden's pseudo R2. Values between 0.2 to 0.4 are generally considered as indicating the satisfactory model fit (McFadden, 1973). When comparing models, the model with the lowest AIC and BIC scores – measures that penalize the presence of several variables in the model – is preferred. For the case of this article, the NL model had better metrics.

Table 6 - Comparison between Multinomial and Nested Logit models

	1 - MNL	2 - NL
Sample size	2196	2196
Log likelihood (final)	-4161.292	-4156.52
R^2	0.3564	0.3569
Adj. R^2	0.3377	0.3392
AIC	8564.58	8544.87
BIC	9253.61	9194.03
N. Parameters	121	114
Time taken (mm:ss)	3'26''	24'07''

5.4.2. Trip distribution distances

The use of distance distribution is based on the calibration of the Gravity distribution model based on adjusting the observed and modeled trip impedance curves (Ortuzar and Willumsen, 2011). Thus, the frequency distribution of the observed and estimated travel distances for the validation and test sample were analyzed to identify the method that minimizes the difference between them. To do this, the distance values were normalized, which were distributed into 8 categories, as illustrated in Figure 7.

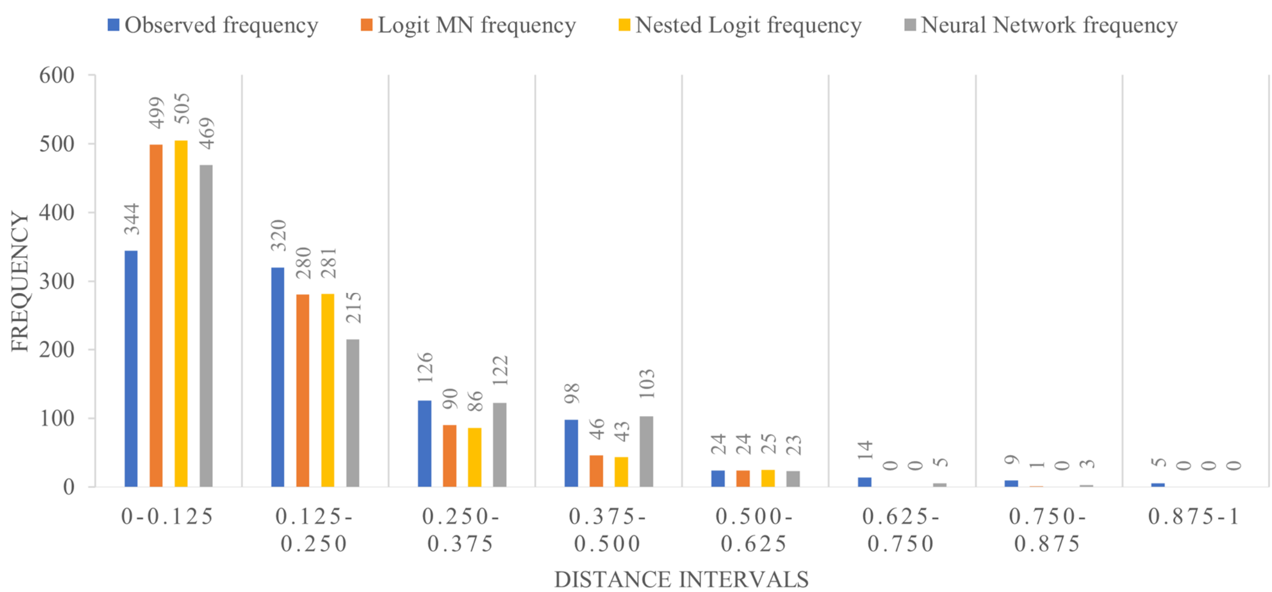

Figure 7. Histogram showing estimated and observed travel distances

It can be observed from the distance histogram that the ANNs performed better than the Multinomial Logit and Nested Logit models in almost all normalized distance intervals, except between 0.125 and 0.250. In addition to the histogram analysis, Mann-Whitney, Kolmogorov-Smirnov and Median non-parametric statistical tests were performed to assess similarities in population distributions and median values between the distances observed and estimated by the three approaches. Table 7 presents the findings regarding the non-parametric tests.

Table 7 - Comparison of Trip distances regarding distribution and median value

	Statistical Tests		
	Mann-Whitney*	**Kolmogorov-Smirnov***	**Median****
Artificial Neural Networks (ANNs)	Not rejected	Rejected	Not rejected
Multinomial Logit	Not rejected	Rejected	Not rejected
Nested Logit	Not rejected	Rejected	Not rejected

*Null Hypothesis: The observed and estimated trip distances have similar distributions
**Null Hypothesis: The observed and estimated trip distances are similar regarding median

Considering the null hypothesis as the equality of the observed and estimated trip distance frequencies regarding distributions (Man-Whitney and Kolmogorov sminorv) and central value (Median test), the Mann-Whitney test retained it for the three approaches, while the Kolmogorov-Smirnov test refuted the null hypothesis for all approaches. The Median test, in turn, retained the null hypothesis of similarity of medians of the observed and estimated trip distances for all the models.

5.4.3. Hit Rates

Applying the calibrated and trained models for all approaches, and for the validation and test samples, the estimated destination choices were compared with those observed, thus obtaining a hit rate of 39.15% for the Multinomial Logit model, 39.89% for the Nested Logit model and 36.38 % for ANNs. Although the rates are very close, when analyzing them for each destination zone, there is a considerable difference between their values for some alternatives, as shown in Table 8.

Table 8 - Hit rate for each alternative

Hit rate per destination (%)																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
MN Logit	9	0	8	0	0	49	27	17	0	5	5	83	6	0	0	40	18	29	0
Nested Logit	9	0	0	0	3	53	27	18	0	5	5	83	6	0	0	40	18	28	0
ANN	0	0	8	0	18	55	16	27	0	0	10	66	15	0	0	40	18	29	0

Table 8 shows that the ANN hit rate was equal or higher for most destinations, and was lower only for zones 1, 7 and 12, which represent 37.2% of all observations (see Figure 4). This result indicates that the ANN model does not need as many observations for each alternative as the Discrete choice models. Thus, ANNs could be a feasible alternative taking into account a sample Origin-Destination Matrix with many fewer observations, especially for the case of data collection problems.

5.4.4. Discrete Choice Models and ANNs: Theoretical discussion

Based on the literature related to Machine Learning and Discrete Choice models and considering the findings of this paper, we can verify the performance of the three approaches associated to different criteria. However, this article is not attempting to point out the best tool, as it depends on the study case and objectives. Taking into account the urban destination choice problem, which is a large choice set approach, we can list some advantages and disadvantages when comparing discrete choice models and Artificial Neural Networks, as shown in Table 9. Table 9 shows a summary of the main results for each criterion in addition to the time required for executing each technique.

Table 9 - Comparison between discrete choice models and ANNs

	Models		
Criteria	**MNL**	**NL**	**ANNs**
Discrete Choice Model Metrics		+	
Trip Impedance Distribution			+
Global Hit Rates	+	+	
Hit Rates for fewer observations			+
Advantages	Statistical significance of parameters, elasticities, subjective value of attributes	Allows grouping correlated alternatives, statistical significance of parameters, elasticities, subjective value of attributes	Adaptation to large data sets, -Accept different input variables, -Computationally less expensive, -Reduced runtime.
Disadvantages	Irrelevant Alternatives (IIA), - Assumptions for the parameters and distributions of the error term.	It may exhibit rigidity in their application, and they may have greater computational requirements (Guerrero et al., 2020).	- Non-parametric, - Impossibility of interpreting relationships between variables and alternatives.
Time taken	3’26’’	24’07’’	06’’

6. CONCLUSIONS

This paper presented an analysis of the performance of Artificial Neural Networks to estimate disaggregated urban trip distribution taking into account a comparative analysis regarding the calibration of the traditional disaggregated Nested and Multinomial Logit models.

Considering the results obtained for the three approaches, it can be concluded that the application of ANNs for the urban trip distribution stage, with Revealed Preference data, has an overall performance similar to that of the Nested and Multinomial Logit models. As advantages, we can mention that ANNs do not present any traditional restrictions, such as multicollinearity problems, IIA property (for MNL models), assumptions of specific structures and limitations for cases of large choice sets and a high number of parameters to be estimated.

In addition, the distance histograms indicate that the ANN tool has a better performance regarding the differences between observed and estimated trip distances. Concerning the number of hits for each alternative, in turn, the ANNs indicate less need for many observations for each alternative compared to discrete choice models, because it resulted, in general, to similar or superior predictive power for the destinations with smaller amounts of observations. It is also worth mentioning the advantage of ANNs in terms of operational efficiency, as the time required for the training performed, in this study, for the ANNs was about 7 seconds, while the calibration times for the Nested and Multinomial Logit models were of approximately 24 and 3 minutes, respectively.

It can be observed, however, that the aforementioned analysis has a purely quantitative content, as it is not possible to properly analyze relationships between variables using the ANNs, because it is a semiparametric technique. Thus, discrete choice models, based on the interpretation of the parameters of the calibrated utility functions, allow an understanding of the factors that influence the choice of destinations by individuals, helping to develop strategies to solve urban mobility problems, for example. However, it can be concluded that, for origin-destination matrices, the Artificial Neural Networks could be more efficient than the tested discrete choice models, even operationally. To assess the effect of explanatory variables on destination choices, in terms of estimated parameters, ANNs are not adequate.

ACKNOWLEDGEMENTS
This research was carried out with the financial support from the National Council for Scientific and Technological Development (CNPq 304345/2019-9) and CAPES (Coordination for the Improvement of Higher Education Personnel). The authors would also like to thank the Planning Institution of Santa Maria (Instituto de Planejamento de Santa Maria/RS – IPLAN) for providing the household survey data.

REFERENCES

Ben-Akiva, M., D. Bolduc, J. Walker (2003) Specification, Identification, and Estimation of the Logit Kernel (Or Continuous Mixed Logit) Model. Working Paper, *5th Invitational Choice Symposium*, Asilomar, California.

Ben-Akiva, M.; Lerman, S. R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, Cambridge, MA: MIT Press.

Bishop, Christopher M. (1995) *Neural networks for pattern recognition*. Oxford University Press, Oxford.

Black, W. R. (1995) Spatial interaction modeling using artificial neural networks. *Journal of Transport Geography*, v. 3, n. 3, p. 159–166. DOI: 10.1016/0966-6923(95)00013-S.

Caldas, M. U. C.; Pitombo, C. S.; Assirati, L. (2021) Strategy to reduce the number of parameters to be estimated in discrete choice models: an approach to large choice sets. *Travel Behaviour and Society*, v. 25, p. 1-17. DOI: 10.1016/j.tbs.2021.05.001.

Cantarella, G. E.; De Luca, S. (2005) Multilayer feed forward networks fortransportation mode choice analysis: An analysis and a comparison with random utility models. *Transportation Research Part C*, v. 13, p. 121-155. DOI: 10.1016/j.trc.2005.04.002.

Carvalho, M. C. M.; Dougherty, M. S.; Fowkes, A. S.; Wardman, M. R. (1998) Forecasting travel demand: A comparison of logit and artificial neural network methods, *Journal of the Operational Research Society*, v. 49, p. 717-772. DOI: 10.1057/palgrave.jors.2600590.

Cascetta, E.; Pagliara, F.; Papola, A. (2007) Alternative approaches to trip distribution modelling: A retrospective review and suggestions for combining different approaches, *Papers in Regional Science*, v. 86, n. 4, p. 597-620. DOI: 10.1111/j.1435-5957.2007.00135.x.

Casey, H. J. (1955) Applications to traffic engineering of the law of retail gravitation. *Traffic Quartely IX*, p. 23-35.

Chow, L., Zhao, F., Li, M., Li, S. (2005) Development and evaluation of aggregate destination choice models for trip distribution in Florida, *Transportation Research Record*, v. 1931, n. 1, p. 18-27. DOI: 10.3141/1931-03.

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.

Dougherty, Mark. (1995). A review of neural networks applied to transport. *Transportation Research Part C*, v. 3, n. 4, p. 247-260. DOI: 10.1016/0968-090X(95)00009-8.

Evans, A.W. (1970) Some Properties of Trip Distribution Methods. *Transportation Research*, v. 4, n. 1, p. 19-36. DOI: 10.1016/0041-1647(70)90072-9.

Evans, S.P., Kirby, H.R. (1974) A three-dimensional furness procedure for calibrating gravity models. *Transportation Research*, v. 8, n. 2, p. 105-122. DOI: 10.1016/0041-1647(74)90037-9.

Faghri, A., Hua, J. (1992) Evaluation of Artificial Neural Network Applications in Transportation Engineering, *Transportation Research Record*, v. 1358, p. 71-80.

Fotheringham, A.S. (1983) Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A: Economy and Space*, v. 15, n. 8, p. 1121-1132. DOI: 10.1068/a151121.

Guerrero, E., Laguna, R., Damián, M., Ortíz, J., Zárate, R. (2020). Basic considerations for the application of discrete choice experiments: a review. *Revista Mexicana de Ciencias Forestales*, v. 11, n. 59, p. 4-30. DOI: 10.29298/rmcf.v11i59.676.

Hess, S., D. Palma (2019) Apollo: A Flexible, Powerful and Customisable Freeware Package for Choice Model Estimation and Application. *Journal of Choice Modelling*, v. 32. DOI: 10.1016/j.jocm.2019.100170.

IBGE - Instituto Brasileiro de Geografia e Estatística, 2010. Censo Demográfico 2010.

IBM Corp. Released 2016. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

IPLAN (2013) *Plano Diretor de Mobilidade Urbana de Santa Maria* (1ª Edição). Santa Maria: Prefeitura Municipal de Santa Maria.

Luce, R. D., Suppes, P. (1965) Preference, utility, and subjective probability. In R. D. Lute, R. R. Bush, & E. Galanter (Fds.), *Handbook of Mathematical Psychology*, v. 111, p. 249-410.

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behaviour. *Frontiers in Econometrics*, p. 105-142.

Mcfadden, D. (1974) The Measurement of Urban Travel Demand. *Journal of Public Economics,* v. 3, n. 4, p. 303-328. DOI: 10.1016/0047-2727(74)90003-6.

Mishra, S., Yanli, W., Xiaoyu, Z., Rolf, M, Subrat, M. (2013) Comparison between Gravity and Destination Choice Models for Trip Distribution in Maryland, *Transportation Research Board 92nd Annual Meeting*, p. 1-22.

Mozolin, M. J. C., Thilland, E. L. (2000) Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. *Transportation Research Part B*, v. 34, n. 1, p. 53-73. DOI: 10.1016/S0191-2615(99)00014-4.

Ortúzar, J. D., Willumsen, L. G. (2011) *Modelling Transport*. Wiley, London.

Plaza, C.V., Rodrigues da Silva, A.N. (2021). Estimating intrazonal trip distances in a Brazilian medium-sized city. Case Studies on Transport Policy, v. 9, n. 2, p. 626-636. DOI: 10.1016/j.cstp.2021.03.001.

Pitombo, C. S., de Souza, A. D., Lindner, A. (2017) Comparing decision tree algorithms to estimate intercity trip distribution, *Transportation Research Part C*, v. 17, p. 16-32. DOI: 10.1016/j.trc.2017.01.009.

Pitombo, C. S.; Sousa, A. J.; Filipe, L. M. N. (2009) Classification and regression tree, principal components analysis and multiple linear regression to summarize data and understand travel behavior. *Transportation Letters*, v. 1, p. 295-308. DOI: 10.3328/TL.2009.01.04.295-308.

R Core Team (2020) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rasouli, M., Nikraz, H. (2013). Trip Distribution Modelling Using Neural Network. *Transport Research Forum*, Brisbane, Australia.

Schneider, M. (1959). Gravity model and trip distribution theory, *Papers in Regional Science*, v. 5, n. 1, p. 51-56. DOI: 10.1111/j.1435-5597.1959.tb01665.x.

Shmueli, D., Salomon, I., Shefer, D. (1996) Neural network analysis of travel behavior: Evaluating tools for prediction, *Transportation Research Part C*, v. 4, p. 151–166. DOI: 10.1016/S0968-090X(96)00007-1.

Smith, M. (1996) *Neural Networks for Statistical Modeling*. International Thomson Computer Press, Londres, Inglaterra.

Stouffer, S. A. (1940) Intervening Opportunities: A Theory Relating Mobility and Distance, *American Sociological Review*, v. 5, p. 845-867. DOI: 10.1016/S1003-6326(10)60620-6.

Subba Rao, P. V., Sikdar, P. K., Krishna Rao, K. V., Dhingra, S. L. (1998) Another insight into artificial neural networks through behavioural analysis of access mode choice. *Computers, Environment and Urban Systems*, v. 22, n. 5, p. 485-496. DOI: 10.1016/S0198-9715(98)00036-2.

Tavasszy, L.; De Jong, G. (2014) *Modelling Freight Transport.* 1ª. ed. London: Elsevier.

Thomas, R.W., Hugget, R.J. (1980) Modeling in Geography: A Mathematical Approach. Barnes and Noble, Totowa.

Tillema, F., Van Zuilekom, K. M., Van Maarseveen, M. F. (2006) Comparison of neural networks and gravity models in trip distribution. *Computer-Aided Civil and Infrastructure Engineering*, v. 21, n. 2, p. 104-119. DOI: 10.1111/j.1467-8667.2005.00421.x.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A: Economy and Space*, v. 9, n. 3, 285–344. DOI: 10.1068/a090285.

Williams, I. (1976) A comparison of some calibration techniques for doubly constrained models with an exponential cost function, *Transportation Research*, v. 10, n. 2, p. 91-104. DOI: 10.1016/0041-1647(76)90045-9.

Wilson, A. G. (1967) A statistical theory of spatial distribution models, *Transportation Research*, v. 1, n. 3, p. 253–269. DOI: 10.1016/0041-1647(67)90035-4.

Zhao X, Yan X, Yu A, Van Hentenryck P. (2020) Prediction and behavioral analysis of travelmode choice: A comparison of machine learning and logit models. Travel Behaviour and Society, v. 20, p. 22-35. DOI: 10.1016/j.tbs.2020.02.003.