ESTIMATION OF DWELL TIME USING FUZZY REGRESSION IN BRT SYSTEMS

Laio Oriel Seman
Universidade do Vale do Itajaí
Victor Hugo Rodrigues Machado
Universidade Tecnológica Federal do Paraná
Luiz Alberto Koehler
Universidade Regional de Blumenau
Eduardo Camponogara
Universidade Federal de Santa Catarina

ABSTRACT
This paper presents adjustments and validation of a model for the process of passenger boarding and alighting of BRT and express bus systems. Such models are of fundamental importance for the implementation of strategies to control the operation of those systems, which allow for quality and efficiency gains in the service. As a way to improve the model, real data were collected from the boarding/alighting times of passengers in the Troncal 10 system in the city of Blumenau - SC. These data were used in fuzzy regression to take into account the uncertainties associated with the dwell time of the passengers.

RESUMO
Este artigo trata do ajuste e validação de um modelo para representar o processo de embarque e desembarque de passageiros de sistemas BRT e ônibus expresso. Tais modelos são de fundamental importância para implementação de estratégias de controle da operação desses sistemas, o que permite ganhos de qualidade e eficiência no serviço. Como forma de aprimorar o modelo, foram coletados dados reais dos tempos de embarque/desembarque de passageiros na linha Troncal 10 da cidade de Blumenau – SC. Estes dados foram utilizados em uma regressão fuzzy que levou em conta as incertezas associadas aos tempos de embarque/desembarque dos passageiros.

1. INTRODUCTION

Mass public transport, and more generally urban mobility, has been a topic of constant discussion and mobilization of society, which is also verified in Brazil. Among the main topics discussed, besides the costs of transport tariffs, we can mention the quality and efficiency of the services. A negative factor observed in a large number of bus-based public transportation systems is the deviation of the schedule and/or interval between buses (tendency of bunching). These deviations increase the waiting time of users and the costs of operating the transportation system, contributing to the choice for individual transportation over public transportation.

A solution to the problem of operating public transport systems, in the case of buses, is the implementation of control strategies for their operation. Given the need for differentiated quality of express bus and BRT systems in terms of infrastructure and system operation, the application of real-time operation strategies in these systems is of paramount importance, particularly so the holding strategies at stops (Eberlein et al. 2001, Hickman 2001, Zolfaghari et al. 2004, Koehler et al. 2011, 2018, Zimmermann et al. 2016).

These real-time strategies make use of prediction models for boarding and alighting times of passengers at stops and/or terminals and stations. However, those commonly used models can be too simplistic, implying a representation that does not match the actual observed behavior or are too complex for real-time applications. In a first-world country scenario, with buses that do not operate crowded and with adequate road and station infrastructures, these simplifications typically used may not be significant. In the case of developing and underdeveloped countries, the typical reality of many Brazilian cities, the real behavior of the boarding and alighting process is significantly de-characterized.

The boarding and alighting process of passengers or "dwell time" is defined as the time spent at a bus stop (Abkowitz & Engelstein 1983). This time typically represents a considerable portion of the total travel time, which justifies the importance of a suitable model for representing this process. For example, Levinson (1983) found that in American cities from 1957 to 1980 the average dwell time represented about 20% of total travel time within urban areas and increased to 26% in central areas. From data collected in Sydney, considering on-board billing, Tirachini (2013*b*) concluded that the average "dwell time" is about 15% of total travel time.

As a way to improve the representation of the boarding/alighting process, this paper uses as a base the model developed and already presented by Koehler et al. (2018), which contemplates BRT characteristics typically present in several other public transport systems in Brazil and elsewhere. However, this model requires a more detailed study of the most appropriate value of some parameters, which reflect the behavior of the passengers during the boarding/alighting processes.

The development of the model presented in this paper makes use of real data collected from the passengers boarding/alighting process of the BRT line Troncal 10 located in the city of Blumenau, SC. These data were used in a fuzzy regression with the objective of developing a model that best represents the times associated with "dwell time", taking into account the number of onboard passengers.

The presented model uses a series of parameters that describe the behavior of the passengers, both in the boarding and alighting process, as well as parameters associated with the BRT system. Thus, suitable values for these parameters is of fundamental importance for the model to adequately represent the real behavior of the system. Another feature present in the model is the ease of integration in optimization algorithms for real-time application, without loss of modeling proprieties and efficiency.

Other proposed models make use of nonlinear functions (Guenthner & Sinha 1983, Lin & Wilson 1992), details associated with the passenger's location preference in the bus (Li et al. 2006), door width (Fernández et al. 2010), bus floor height (Dueker et al. 2004), ticketing technique (Dorbritz et al. 2008), passenger age (Tirachini 2013*a*), and grouping of people (Currie et al. 2013). Some of these particularities are present directly and indirectly in the model to be used here, others imply in a model too complex, infeasible for real-time applications. It is considered that the model used here has potential to improve the representation that describes the dwell time of the Troncal 10 BRT system of the city of Blumenau, with applicability in other urban public transport systems found in Brazilian cities.

Therefore, this paper is divided as follows. Section 2 presents the model used in the development of the work. Section 3 presents a review on fuzzy regression. The results are presented in Section 4. Finally, Section 5 presents the conclusion of the work.

2. BRT MODEL

The BRT and express bus model used for analysis and simulation is an update of the previously presented by Koehler et al. (2018). This model allows the real-time control of BRT-type systems, more specifically, the control of bus spacing through the holding of buses in stations. The model is deterministic and presented in the form of a frequently updated mathematical program with a rolling horizon. Before presenting the mathematical programming model, some preliminary definitions are necessary. First, the sets and indices used are:

- n_I: number of buses of the BRT system;

- n_K: number of bus stops of the BRT system;
- n_N: number of stops in the prediction horizon of the buses;
- $\mathcal{I}$: set of buses of the BRT system;
- $\mathcal{K}$: set of stops of the BRT system;
- i: bus index, $i \in \mathcal{I}$;
- k: stop index, $k \in \mathcal{K}$;
- m: index of the bus lap in the closed BRT circuit;
- $\mathcal{K}_i$: set of stops belonging to the prediction horizon of the bus i;
- $\mathcal{K}_i^0$: set of stops of bus i for which the initial values $d_{i,\omega}$ (departure times) and $l_{i,\omega}$ (onboard passengers) apply.

Considering a circular itinerary, the following lap arithmetic is used: $\omega = (k, m)$; $\langle i, \omega^- \rangle$ represents for bus i, the stop immediately preceding stop k, which can be stop $(k-1)$ on the same lap m, or else stop n_K of lap $(m-1)$ if the current stop is the first in the circuit $(k = 1)$; $\langle i^-, \omega \rangle$ corresponds to the previous bus that arrived at bus stop k before bus i in lap m, which can be a bus in the same lap or in the previous one.

System parameters and initial conditions that are assumed known and constant are the following:

- C_0: time for the start of passengers boarding and alighting process from the arrival of the bus to the stop [s];
- C_a: time spent for each passenger alighting the bus [s/passenger];
- C_b: time spent for each passenger boarding the bus [s/passenger];
- $C_{\max}$: bus capacity [pax];
- r_k: trip time from stop $k-1$ to stop k;
- q_k: fraction of passengers alighting at stop k;
- λ_k: passenger arrival rate at stop k [pax/s].

For describing the system dynamic in the BRT system, the following state and control variables are used in the formulation:

- $d_{i,\omega}$: departure time of bus i from stop k in lap m [s], with $\omega = (k, m)$;
- $h_{i,\omega}$: holding control time for bus i at stop k in lap m [s];
- $l_{i,\omega}$: onboard passengers of bus i when it departs from stop k in lap m.

To help express the dynamic model of the BRT system, the following auxiliary variables are defined:

- $a_{i,\omega}$: arrival time of bus i at stop k in lap m [s];
- $p_{i,\omega}$: number of passengers to board the buses at the stops [pax];
- $re_{i,\omega}$: passenger residue at stops [pax];
- $s_{i,\omega}$: dwell time of bus i at stop k in lap m [s];
- $t^a_{i,\omega}$: bus passenger alighting time [s];
- $t^b_{i,\omega}$: bus passenger boarding time [s];
- $v_{i,\omega}$: bus remaining capacity [pax].

The total passengers' waiting time for the horizon considered is given by:

$$f = \sum_{i \in \mathcal{I}} \sum_{\omega \in \mathcal{K}_i} \begin{bmatrix} \frac{\lambda_k}{2} \left(d_{i,\omega} - d_{\langle i^-,\omega \rangle} \right)^2 \\ + re_{\langle i^-,\omega \rangle} \left(d_{i,\omega} - d_{\langle i^-,\omega \rangle} \right) \\ + (1 - q_k) l_{\langle i,\omega^- \rangle} \left(h_{i,\omega} + s_{i,\omega} \right) \end{bmatrix}, \qquad (1)$$

subject to the following constraints:

$$d_{i,\omega} = d^0_{i,\omega}, \quad \forall \omega \in \mathcal{K}^0_i \tag{2}$$
$$l_{i,\omega} = l^0_{i,\omega}, \quad \forall \omega \in \mathcal{K}^0_i \tag{3}$$
$$re_{i,\omega} = re^0_{i,\omega}, \quad \forall \omega \in \mathcal{K}_i \cup \mathcal{K}^0_i \tag{4}$$
$$a_{i,\omega} = d_{\langle i,\omega^- \rangle} + r_k \tag{5}$$
$$v_{i,\omega} = C_{\max} - l_{\langle i,\omega^- \rangle} (1 - q_k) \tag{6}$$
$$p_{i,\omega} = re_{\langle i^-,\omega \rangle} + \left(d_{i,\omega} - d_{\langle i^-,\omega \rangle}\right) \lambda_k \tag{7}$$
$$d_{i,\omega} = a_{i,\omega} + s_{i,\omega} + h_{i,\omega} \tag{8}$$
$$d_{i,\omega} \geq a_{\langle i^-,\omega \rangle} \tag{9}$$
$$l_{i,\omega} = \min\{(1 - q_k)\, l_{\langle i,\omega^- \rangle} + p_{i,\omega}, C_{\max}\} \tag{10}$$
$$re_{i,\omega} = \max\{0, p_{i,\omega} - v_{i,\omega}\}. \tag{11}$$

Initial conditions of the bus departure time from a stop is given by (2), bus on-board passengers when departing from a stop by (3), and initial residue of passengers at a stop by (4). The constraint (5) represents the bus arrival time to the stop k; (6) the bus remaining capacity and (7) the number of passengers to board the bus. Departure time is given by (8), and (9) guarantees that bus overtaking is not allowed. The number of on-board passengers after bus departure is given by (10), and the residue of passengers by (11).

The boarding and alighting processes are defined as follows:

$$t^a_{i,\omega} = C_a q_k l_{\langle i,\omega^- \rangle} \tag{12}$$
$$t^b_{i,\omega} = \min\{C_b v_{i,\omega}; C_b\,(p_{i,\omega} - h_{i,\omega} \lambda_k)\} \tag{13}$$
$$s_{i,\omega} = C_0 + t^a_{i,\omega} + t^b_{i,\omega} \tag{14}$$

The bus passengers alighting time is given by (12). The boarding time is given by the time spent to board the passengers who occupy the remaining capacity of the bus or the time spent for boarding the accumulated passenger queue at the stop. This is given by means of a minimum operator as (13). Considering terminals and stations as bus stops of traditional BRT systems, the passenger alighting process occurs first. The boarding process, with pre-billing, proceeds through the same bus rear doors used for alighting. Thus, dwell time is calculated as the sum of the boarding and alighting processes as (14), Fig. 1 illustrates this proccess.

As can be seen in (12), (13) and (14), boarding and alighting times depend on the parameters C_a and C_b which are typically considered constant. However, what is observed in real systems is that both parameters, besides the typical variation around the average value, vary as a function of other parameters of the BRT system such as bus capacity. In this context and given the characteristics and peculiarities of urban transport systems by buses in Brazil, the following is a proposal for a model of boarding and alighting of passengers that takes into account these nuances, with characteristics compatible with real-time applications.

In the new proposed boarding/alighting model, the time spent for each passenger boarding the bus (C_b) and the time spent for each passenger alighting the bus (C_a) can vary depending on the number of onboard passengers. Those times are given by means of a fuzzy regression of real collected data. Consider the following set containing the limits of the categorical variable

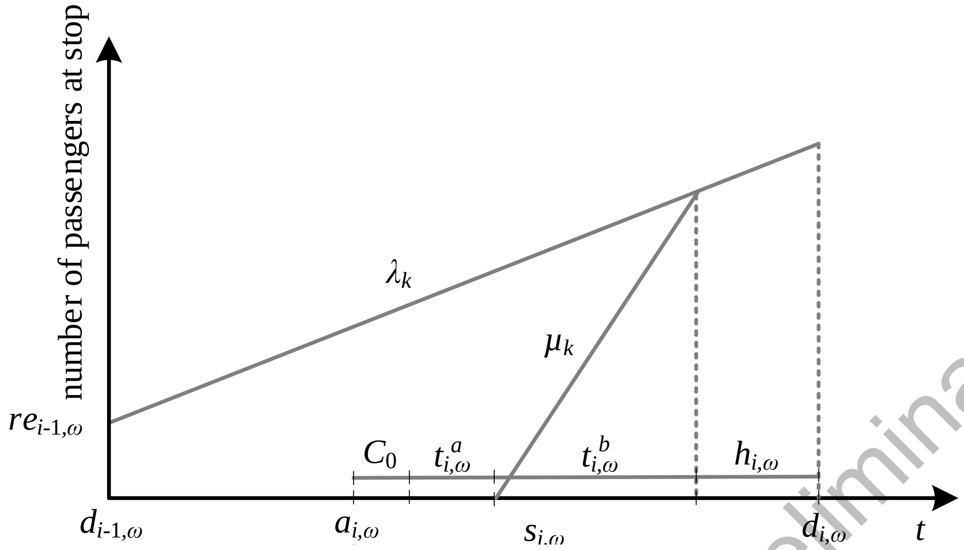


Figure 1: Boarding and alighting processes for terminal and stations with the number of boarding passengers smaller than the bus remaining capacity. The queue boarding rate is μ_k.

"quantity of onboard passengers":

$$\mathcal{U} = \{\text{low, medium, large}\} \quad (15)$$
$$\mathcal{U} = \{(0,40),(41,50),(51,60)\}. \quad (16)$$

The decision between the categorical variables for the boarding/alighting time can be represented as a Generalized Disjunctive Program (GDP):

$$\bigvee_{(\underline{u},\overline{u})\in\mathcal{U}} \begin{bmatrix} g_{i,\omega,(\underline{u},\overline{u})} \\ \underline{u} \le l_{\langle i,\omega^-\rangle} \le \overline{u} \\ t^a_{i,\omega} = C_{a,(\underline{u},\overline{u})} q_k l_{\langle i,\omega^-\rangle} \\ t^{b0}_{i,\omega} = C_{b,(\underline{u},\overline{u})} v_{i,\omega} \\ t^{b1}_{i,\omega} = C_{b,(\underline{u},\overline{u})} \left(p_{i,\omega} - h_{i,\omega}\lambda_k\right) \end{bmatrix} \quad (17)$$

$$g_{i,\omega,(\underline{u},\overline{u})} \in \{True, False\}, \forall (\underline{u},\overline{u}) \in \mathcal{U} \quad (18)$$

Big-M and convex hull reformulations are the most common methodologies for transforming a GDP into a MILP (Castro 2015).

2.1. Testing setup

The approach of this paper uses the fuzzy regression coefficients of real collected data as deterministic values in the control model, namely C_a and C_b. To do so, the measurements were made considering the number of passengers to board/alight, the total "dwell time", and the number of passengers already boarded (defined as a categorical variable: low, medium, high). During the control coefficients (holding) calculation stage, the equations (17) and (18) are responsible for identifying which categorical variable to be used in the process.

For the stochastic model used in the simulation (transit model), boarding/alighting times are given randomly. Thus, fuzzy regression data were used to collect samples from a triangular distribution over the interval. The triangular distribution is a continuous probability distribution with lower limit left, peak at mode, and upper limit right. Unlike the other distributions, these parameters directly define the shape of the probability density function (pdf). The pdf for the triangular distribution is:

$$P(x;l,m,r) = \begin{cases} \frac{2(x-l)}{(r-l)(m-l)} & \text{for } l \leq x \leq m, \\ \frac{2(r-x)}{(r-l)(r-m)} & \text{for } m \leq x \leq r, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

For this strategy, the setup presented in Fig. 2 will be used. Further details of the simulation parameters can be found in (Koehler et al. 2018).

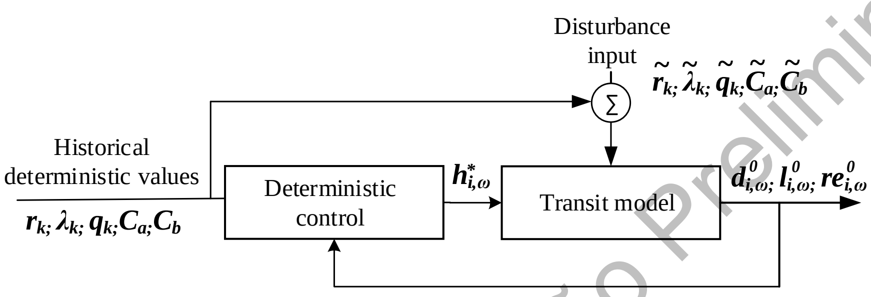


Figure 2: Test setup.

3. FUZZY REGRESSION

Fuzzy regression was proposed in the literature as a way to consider the uncertainties of the coefficients of a linear regression, mostly used when the model is indefinite, the relationships between model parameters are vague, sample size is slow or the data is hierarchically structured (Hideo Tanaka et al. 1982, Tanaka et al. 1989, Tanaka & Lee 1998, Martinkova & Skrabanek 2018). This kind of analysis was initially proposed by Hideo Tanaka et al. (1982) based on the concepts of fuzzy functions presented by Zadeh (1973) that aimed to take into account the natural uncertainties present in a certain phenomenon.

While for traditional linear regression the deviation between the observed and estimated values are assumed to be due to observation errors, Hideo Tanaka et al. (1982) assumed that these differences occur because of the vagueness of the system structure. The fuzzy parameters of a linear system measured by the model correspond to the system's possibility distribution, being responsible for measuring its uncertainty (fuzziness).

3.1. Possibilistic Regression

By generating parameters corresponding to the possibility distribution of a system, the model presented by Hideo Tanaka et al. (1982) is known in the literature as possibilistic regression and has as a general form given by the equation below:

$$\tilde{y}(k) = \tilde{A}_0 + \tilde{A}_1 x_1(k) + \cdots + \tilde{A}_N x_N(k) \tag{20}$$

with $k = 1, 2, \ldots, K$ and where $x = [x_0, x_1, \ldots, x_N]^T$ is a vector of independent variables, $\tilde{A} = [\tilde{A}_0, \tilde{A}_1, ..., \tilde{A}_N]^T$ a vector of fuzzy coefficients denoted in the form of a triangle by $\tilde{A}_j = (a^c_j, a^w_j)$, having its pertinence function described by:

$$\mu_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \frac{|a^c_j - a_j|}{a^w_j}, & a^c_j - a^w_j \leq a_j \leq a^c_j + a^w_j \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

where a_j^c and a_j^w are respectively the center point and the maximum dispersion of the function, as shown in Fig. 3. Thus, the regression presented in (20) can be rewritten as:

$$\tilde{y}(k) = (a_0^c, a_0^w) + \sum_{n=1}^{N} (a_n^c, a_n^w) x_n(k) \tag{22}$$

Differently from the ordinary least squares regression, the deviation between the data and the

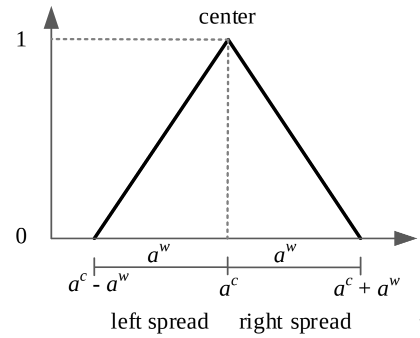


Figure 3: Symmetric triangular membership function.

estimated model depends on the imprecision of the parameters and not on the measurement errors. In this case the model of Hideo Tanaka et al. (1982) proposes that, in order to minimize the uncertainty of the estimated model, one can minimize the total spread of the system's fuzzy coefficients (Romano 2006).

This spread may also depend on a degree of pertinence known as the "h factor", defined by the operator, which specifies the degree of feasibility of the system conditions (Shapiro 2005). The higher the degree of viability, the higher the spreading of the system, i.e., the h factor expands the confidence interval of the model.

Fig. 4 illustrates the behavior of the spreading of the system with the increase of the h factor, both to the right and to the left of the central part of the system. Its increase is directly proportional to the expansion of the limits, allowing the inclusion of more or less data between the limits of the model. While the typical values are contained within the limits even with a reduced h factor, the outliers are included with the factor expansion. From these considerations, Hideo Tanaka et al. (1982) show that the possibilistic regression is reduced to a linear programming problem whose objective is to minimize the spread of the system, or its uncertainties, as presented in (23) subject to the constraints present in (24).

$$\min_{a^c, a^w} J = \sum_{n=1}^{N} \sum_{k=1}^{K} a_n^w x_n(k) \tag{23}$$

$$\begin{aligned} s.t. \quad & \left(\sum_{n=1}^{N} a_n^c x_n(k)\right) - (1-h)\left(\sum_{n=1}^{N} a_n^w x_n(k)\right) \leq y(k), \forall k \in \mathcal{K} \\ & \left(\sum_{n=1}^{N} a_n^c x_n(k)\right) + (1-h)\left(\sum_{n=1}^{N} a_n^w x_n(k)\right) \geq y(k), \forall k \in \mathcal{K} \\ & a^w \geq 0 \end{aligned} \tag{24}$$

where $a^c = [a_1^c, \ldots, a_n^c]$, $a^w = [a_1^w, \ldots, a_n^w]$, y is the dependent variable to be predicted by the regression, and $\mathcal{K} = \{1, \ldots, K\}$ is the set of points.

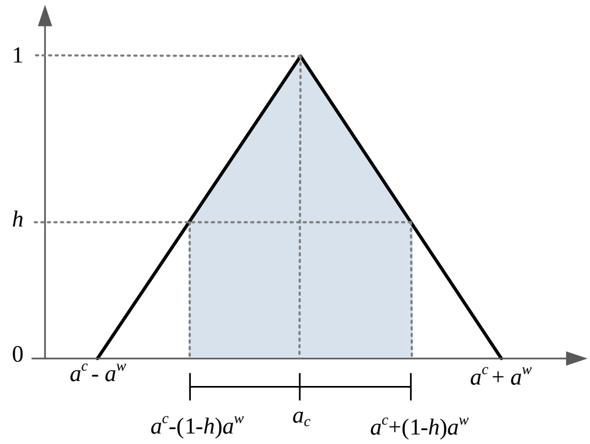


Figure 4: Triangular membership function with h factor interval.

3.2. Hybrid Regression

Criticism of the initial model presented by Hideo Tanaka et al. (1982), including the uncertainty of how it relates to the traditional model of least squares regression (Diamond 1988) has led to the creation of hybrid models.

One of these hybrid models seeks to maintain the central tendency of the model, Savic & Pedrycz (1991) introduced a two-stage fuzzy regression considered important by its results and low computational cost (Chan & Wuz 2007). In the two-stage fuzzy regression, the central trends of the model are defined, typically by ordinary least squares, which are then used as known variables in the optimization.

Let a^c be a central trend vector previously defined, the two-stage fuzzy regression can determine the spreads as in (25) subject to the conditions presented in (26). Notice that in order to allow an optimization capable of differentiating between the possibility of left and right propagation, one can rewrite the quadratic possibilistic regression decomposing a^w in a^{wL} and a^{wR}.

$$\min_{a^{wL},a^{wR}} J = \sum_{n=1}^{N} \sum_{k=1}^{K} (a_n^{wL} + a_n^{wR}) x_n(k) \tag{25}$$

$$\begin{aligned} s.t. \quad & \left(\sum_{n=1}^{N} a_n^c x_n(k)\right) - (1-h)\left(\sum_{n=1}^{N} a_n^{wL} x_n(k)\right) \leq y(k), \forall k \in \mathcal{K} \\ & \left(\sum_{n=1}^{N} a_n^c x_n(k)\right) + (1-h)\left(\sum_{n=1}^{N} a_n^{wR} x_n(k)\right) \geq y(k), \forall k \in \mathcal{K} \\ & a_n^{wL} \geq 0 \\ & a_n^{wR} \geq 0 \end{aligned} \tag{26}$$

This model can be expanded to use quadratic programming in order to increase the diffusion of the coefficients. Rewriting (25) in its quadratic form it is possible to arrive at (27), very similar to the model of Tanaka et al. (1989). Since the vector a^c is already defined in the first stage of the regression and does not need to be optimized, the function is naturally strictly convex.

$$J = \sum_{n=1}^{N} (a_n^{wL} + a_n^{wR})^2 x_n^T x_n \tag{27}$$

With two-stage fuzzy regression it is possible to maintain the original central tendency of the model, which can be replaced by other types of estimation (ridge, minimax), and minimize spreads based on these values to include the human uncertainties in the model.

3.3. Index of Confidence (IC)

As a way of interpreting the results of a fuzzy regression and collaborating for the interpretation regarding the regression interval, Wang & Tsaur (2000) have proposed an index of confidence (IC) that measures the variation between the independent upper and lower limits of the variable, which is similar to the goodness-of-fitness R^2 in statistical regression. Thus, we consider that $y^L(k)$ and $y^R(k)$ are respectively the lower and upper bounds of a fuzzy estimate $\tilde{y}(k)$, which has as a central estimation $y^c(k)$.

The index of confidence (IC) is calculated as (28), and the larger the value, the greater the system representation by the regression performed.

$$IC = \frac{SSR}{SST} \tag{28}$$

The SST (Total Sum of Squares) measures the total variation of $y(k)$ to the left spread ($y^L(k)$) and to the right spread ($y^R(k)$), calculated as:

$$SST = \sum_{k=1}^{K} \left(y(k) - y^L(k)\right)^2 + \sum_{k=1}^{K} \left(y^R(k) - y(k)\right)^2 \tag{29}$$

The SSR (Regression Sum of Squares) measures the variation of the center estimate ($y^c(k)$) to both spreads ($y^L(k)$ and $y^R(k)$), calculated as:

$$SSR = \sum_{k=1}^{K} \left(y^c(k) - y^L(k)\right)^2 + \sum_{k=1}^{K} \left(y^R(k) - y^c(k)\right)^2 \tag{30}$$

The h factor is directly related to the index of confidence, being directly proportional to its increase (Wang & Tsaur 2000).

4. RESULTS

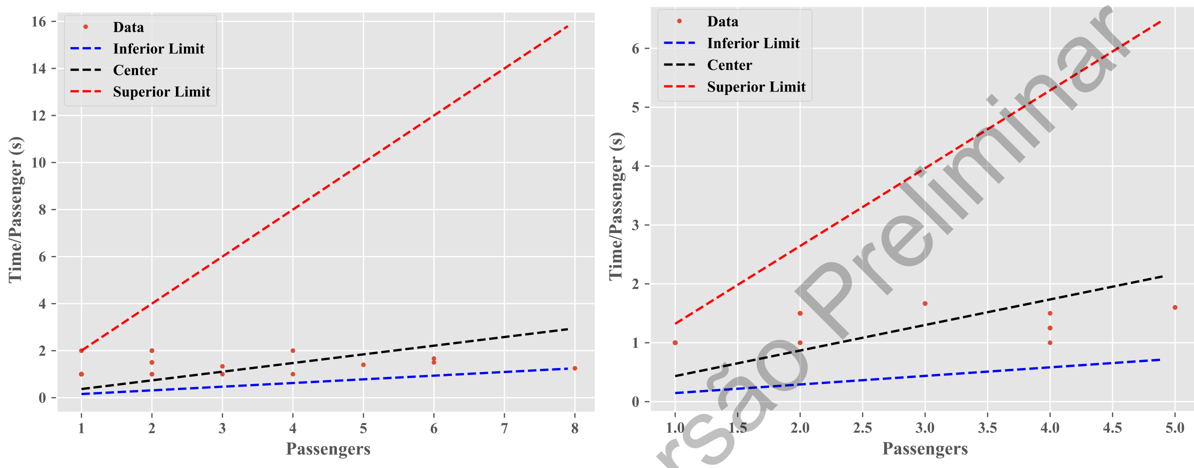
During the data collection, filming was carried out at bus stops in the city of Blumenau/SC for later determination of boarding/alighting times, taking into account the number of passengers to be processed, as well as the number of passengers already onboard. With these data, a hybrid fuzzy regression was performed for each of the categorical variables (related to the number of passengers standing on the buses), considering the collected data, which are presented in Table 1 as [a^{wL}; a^c; a^{wR}]. The h factor was defined as 0 for all the preliminary tests.

To illustrate the results of the fuzzy regression, Fig. 5(a) shows the regression obtained for C_a^{lo} (alighting time from a bus with a low number of passengers), while Fig. 5(b) shows the results obtained for C_b^{lo} (boarding time into a bus with a low number of passengers).

As a next step, the setup of Fig. 2 was used to test each of the fuzzy regression boundaries (lower limit, central, and upper limit) as control values. Considering each case, the simulation using the test setup was performed, whose results of the passenger waiting time appear in Table 2. The average headway distribution is also shown for each possibility in Fig. 6.

Table 1: Fuzzy regression results

Item	Number of onboard passengers	Fuzzy regression model	IC
C_a^{lo}	low	[0.156; 0.368; 2]	0.944
C_a^{me}	medium	[0.07; 0.367; 2.848]	0.937
C_a^{hi}	high	[0.134; 0.388; 2.275]	0.921
C_b^{lo}	low	[0.145; 0.433; 1.32]	0.916
C_b^{me}	medium	[0.129; 0.489; 1.532]	0.770
C_b^{hi}	high	[0.140; 0.565; 2]	0.836



(a) C_a^{lo}: Alighting. (b) C_b^{lo}: Boarding.

Figure 5: Fuzzy regression for alighting time for passengers onboard of a bus with low occupancy.

Table 2: Passengers waiting time after 5 simulation steps.

Deterministic Coefficient	Passenger's waiting time
Inferior	161,519.65
Central	161,996.71
Superior	167,886.60

From the table results it can be noticed that the use of the lower limit of the fuzzy regression reduces the waiting time of the passengers. While the use of the upper limit, taking into account outliers of the captured data, generates an overestimation of the boarding/alighting times of the passengers, which, consequently, increases the total waiting time in the transit model. When analyzing the trajectories of a single bus for the three distinct cases, it can be seen that the holding times are smaller when using the lower coefficient, as in Fig. 7.

5. CONCLUSION

This paper presented a methodology using fuzzy regression to estimate the time of boarding/alighting of passengers to be used in control of BRT systems. In order to do so, data from the real Troncal 10 scenario was collected in the city of Blumenau/SC, measuring the number of passengers to be processed, the time of boarding/alighting by passengers and the number of passengers already boarded.

The use of fuzzy regression results in the stochastic simulation model help to better represent the random behavior present in the transit system. In this way, the methodology presented here can be used to test other forms of control in a most accurate scenario.

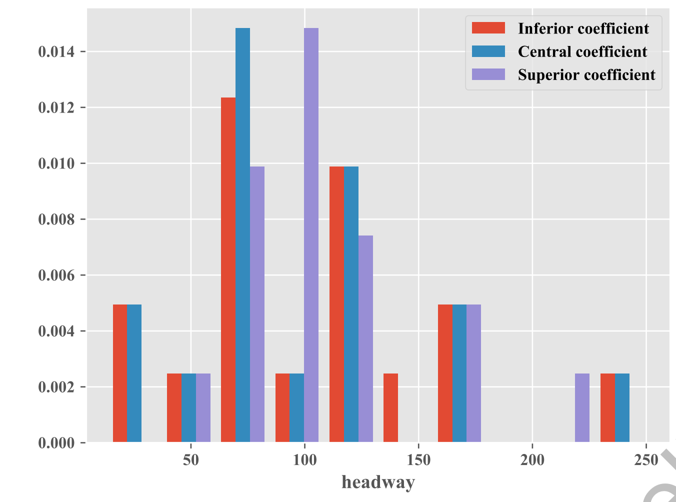


Figure 6: Headway distribution with each fuzzy coefficient.

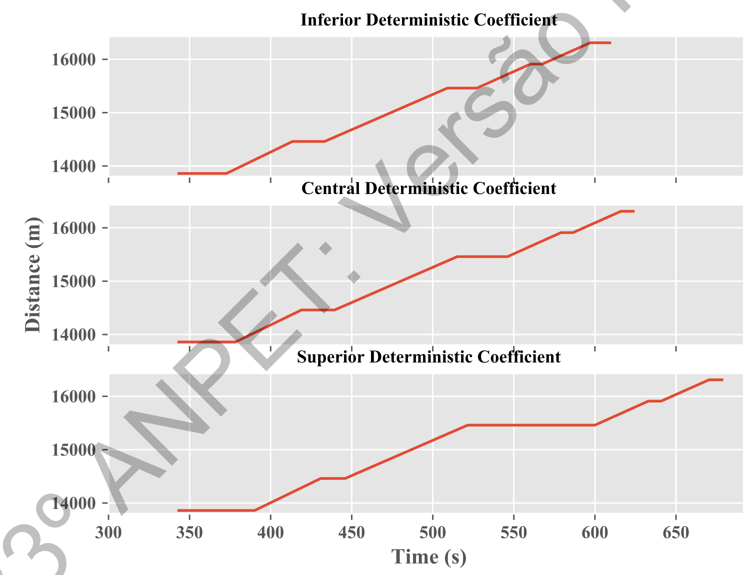


Figure 7: Bus trajectories after 5 simulation steps for the three different possible coefficients.

Simulation results suggests that, for the studied scenario, the use of the lower limit of the fuzzy regression as deterministic control values for the parameters C_a and C_b implies in a lower total waiting time of the passengers of the BRT system. While the use of the upper limit, taking into account outliers of the captured data, generates an overestimation of the boarding/alighting times of the passengers, which, consequently, increases the total waiting time.

For future work it is suggested the expansion of the methodology used in the fuzzy regression, taking into account the optimization of the factor h to be used, and metrics such as "average fuzzy spread" and "mean fuzzy credibility".

REFERENCES

Abkowitz, M. D. & Engelstein, I. (1983), 'Factors affecting running time on transit routes', *Transportation Research Part A: General* **17**(2), 107–113.

Castro, P. M. (2015), 'Tightening piecewise McCormick relaxations for bilinear problems', *Computers and Chemical Engineering* **72**, 300–311.

Chan, L. & Wuz, M. (2007), 'Fuzzy Regression; Methods and Evaluations', *Journal of Macau University of Science and Technology* .

Currie, G., Delbosc, A., Harrison, S. & Sarvi, M. (2013), 'Impact of Crowding on Streetcar Dwell Time', *Transportation Research Record: Journal of the Transportation Research Board* **2353**(1), 100–106.

Diamond, P. (1988), 'Fuzzy least squares', *Information Sciences* **46**(3), 141–157.

Dorbritz, R., Luthi, M., Nash, A. & Weidmann, U. (2008), 'Effects of on-board ticket sales on public transport reliability', *Institue for Transport Planning and Systems (IVT) Swiss Federal Institute of Technology, ETH Zurich, Switzerland* .

Dueker, K., Kimpel, T., Strathman, J. & Callas, S. (2004), 'Determinants of Bus Dwell Time', *Journal of Public Transportation* **7**(1), 21–40.

Eberlein, X. J., Wilson, N. H. M. & Bernstein, D. (2001), 'The holding problem with real-time information available', *Transportation Science* **35**(1), 1–18.

Fernández, R., Zegers, P., Weber, G. & Tyler, N. (2010), 'Influence of Platform Height, Door Width, and Fare Collection on Bus Dwell Time', *Transportation Research Record: Journal of the Transportation Research Board* **2143**(1), 59–66.

Guenthner, R. P. & Sinha, K. C. (1983), 'Modeling bus delay due to passenger boardings and alightings', *Transportation Research Record 915* **7**(13).

Hickman, M. (2001), 'An analytic stochastic model for the transit vehicle holding problem.', **35**(3), 215–237.

Hideo Tanaka, Uejima, S. & Asai, K. (1982), 'Linear Regression Analysis with Fuzzy Model', *IEEE Transactions on Systems, Man, and Cybernetics* **12**(6), 903–907.

Koehler, L. A., Kraus, W. & Camponogara, E. (2011), 'Iterative quadratic optimization for the bus holding control problem', *IEEE Transactions on Intelligent Transportation Systems* **12**(4), 1568–1575.

Koehler, L. A., Seman, L. O., Kraus, W. & Camponogara, E. (2018), 'Real-time integrated holding and priority control strategy for transit systems', *IEEE Transactions on Intelligent Transportation Systems* pp. 1–11.

Levinson, H. S. (1983), 'Analyzing transit travel time performance', *Transportation Research Record: Journal of the Transportation Research Board* **915**, 1–6.

Li, M. T., Zhao, F., Chow, L.-F., Zhang, H. & Li, S.-C. (2006), 'Simulation Model for Estimating Bus Dwell Time by Simultaneously considering Numbers of Disembarking and Boarding Passengers', *Transportation Research Record: Journal of the Transportation Research Board* **1971**(1), 59–65.

Lin, T.-m. & Wilson, N. H. M. (1992), 'Dwell Time Relationships for Light Rail Systems', *Transportation Research Record* (1361), 287–295.

Martinkova, N. & Skrabanek, P. (2018), fuzzyreg: An r package for fuzzy linear regression, *in* P. Cech & D. Svozil, eds, 'ENBIK2018 Conference Proceedings', University of Chemistry and Technology, Prague, p. 7.

Romano, R. (2006), Fuzzy Regression and PLS Path Modeling: a combined two-stage approach for multi-block analysis, PhD thesis.

Savic, D. A. & Pedrycz, W. (1991), 'Evaluation of fuzzy linear regression models', *Fuzzy Sets and Systems* **39**(1), 51–63.

Shapiro, A. (2005), 'Fuzzy regression models', *Article of Penn State University* pp. 1–17.

Tanaka, H., Hayashi, I. & Watada, J. (1989), 'Possibilistic linear regression analysis for fuzzy data', *European Journal of Operational Research* **40**(3), 389–396.

Tanaka, H. & Lee, H. (1998), 'Interval regression analysis by quadratic programming approach', *IEEE Transactions on Fuzzy Systems* **6**(4), 473–481.

Tirachini, A. (2013*a*), 'Bus dwell time: The effect of different fare collection systems, bus floor level and age of passengers', *Transportmetrica A: Transport Science* **9**(1), 28–49.

Tirachini, A. (2013*b*), 'Estimation of travel time and the benefits of upgrading the fare payment technology in urban bus services', *Transportation Research Part C: Emerging Technologies* **30**, 239–256.

Wang, H. F. & Tsaur, R. C. (2000), 'Insight of a fuzzy regression model', *Fuzzy Sets and Systems* **112**(3), 355–369.

Zadeh, L. A. (1973), 'Outline of a New Approach to the Analysis of Complex Systems and Decision Processes', *IEEE Transactions on Systems, Man and Cybernetics* **SMC-3**(1), 28–44.

Zimmermann, L., Kraus, W., Koehler, L. A. & Camponogara, E. (2016), 'Holding Control of Bus Bunching without Explicit Service Headways', *IFAC-PapersOnLine* **49**(3), 209–214.

Zolfaghari, S., Azizi, N. & Jaber, M. Y. (2004), 'A model for holding strategy in public transit systems with real-time information', *International Journal of Transport Management* **2**, 99–110.