An integrated method to optimise the infrastructure costs of bus rapid transit systems

Bernardo Caicedo, Manuel Ocampo, Mauricio Sanchez-Silva

*Universidad de los Andes, Civil and Environmental Engineering, Bogota, 1 Colombia

First published on: 20 July 2010
An integrated method to optimise the infrastructure costs of bus rapid transit systems

Bernardo Caicedo*, Manuel Ocampo and Mauricio Sanchez-Silva

Universidad de los Andes, Civil and Environmental Engineering, Cr 1 #18A-70, Bogota, 1 Colombia

(Received 13 July 2009; final version received 24 March 2010; accepted 30 May 2010)

A significant effort has been directed in many countries to develop economically feasible transportation solutions which include, in many cases, buses that run on express lanes or that are used as part of a feeder system; these systems are identified as ‘bus rapid transit systems’ (BRT). The purpose of this paper is to present a reliability cost-based optimisation model of these systems. The model couples the transportation requirements and the mechanical performance of asphalt pavement structures. Also it includes critical transportation parameters such as bus sizes, frequency of the routes and passenger demands. Regarding the pavement, the model takes into account the life-cycle of existing or new pavements, which involves a stochastic model of the pavement’s mechanical properties and its overall performance. The model is applied to the actual case of Bogota’s rapid transportation system, i.e. Transmilenio, showing the importance of this integrated approach to build efficient BRT systems.

Keywords: infrastructure planning; life-cycle cost; optimisation; pavement design; reliability; transport planning

1. Introduction

There is renewed interest in many developing and developed countries in finding ways of providing efficient and effective public transport that does not come with a high price tag. Although light rail has often been promoted as a popular ‘solution’, there has been progressively emerging an attractive alternative in the form of bus rapid transit (BRT). Hensher and Golob (2008). As an example, Wright and Hook (2007) describe the best practice bus rapid transit (BRT) systems in South America such as Curitiba in Brazil and Transmilenio in Bogota, Colombia, as follows: bus rapid transit is ‘...a high-quality bus based transit system that delivers fast, comfortable, and cost-effective urban mobility through the provision of segregated right-of-way infrastructure, rapid and frequent operations, and excellence in marketing and customer service. BRT essentially emulates the performance and amenity characteristics of a modern rail-based transit system but at a fraction of the cost. A BRT system will typically cost four to 20 times less than a light rail transit (LRT) system and 10 to 100 times less than a metro system.’

Although a significant effort has been directed in optimising BRT systems, these efforts consider separately the optimisation of the transport system and the optimisation of the infrastructure. From a broader perspective, BRT transportation systems include not only the physical infrastructure but also the operation. Therefore, a transportation system model should be extended to include not only the vehicles (e.g. buses) and their operation requirements (service frequency) but also the design and maintenance of the infrastructure (i.e. pavement). However, there is a lack of coupled models that integrates the costs of both transportation service and infrastructure operation within a single framework. As BRT systems are characterised by dedicated infrastructure and homogeneous bus size, the choice of the size of buses is an element that has a great influence on the cost of the BRT system and is a clear example for which a coupled optimisation cause significant improvements in the costs.

The purpose of this paper is to present a reliability cost-based optimisation model of BRT transportation systems that couples the operation of the transportation service and the mechanical performance of asphalt pavement structures. The model takes into account the uncertainty in the pavement mechanical controlling parameters and the life-cycle of existing and new pavements. It assumes a pavement maintenance programme which follows a renewal model based on the expected deterioration of the pavement so that the programmed investments can guarantee their financial feasibility and optimum operation.

The objectives of this paper are:

- Present a life-cycle cost model of transportation systems that integrates both physical (e.g. pavement structure) and operational (e.g. bus services) aspects.
Formulate a transportation cost-based optimisation problem to determine optimum design and operation parameters of the transportation system (e.g. optimum pavement structure, bus size and frequency).

Illustrate the proposed approach with an actual case: Bogota’s rapid transit system, i.e. Transmilenio.

This paper is organised as follows: it first formulates an integrated life-cycle model of transportation systems. This section includes a cost-based formulation and the description and derivation of costs. Afterwards, the paper presents the numerical derivations of the failure rates for both the transportation service and the pavement performance; they are used as input for the life-cycle analysis. Finally, the proposed approach is applied to the case of the feeder route of the Bogota’s BRT transportation system.

2. Coupled operation and infrastructure design

Decisions related to resource allocation in transportation systems should take into account their long term performance and the cost for society. In an integrated approach to transportation systems, the service and physical infrastructure should be optimal with respect to the investment during its lifetime. The life-cycle cost (LCC) of a project is defined as the total cost that is incurred, or may be incurred, in all stages of the project life cycle. LCC analysis provides a framework to support decisions about resource allocation related to the design, construction and operation of transportation systems at minimum life cycle cost (Sánchez-Silva et al. 2009). Then, the decision becomes an optimisation problem with the following objective function,

\[ Z(p) = B(p) - C_0(p) - C_T(p) - R(p) \]  

where \( B(p) \) is the benefit from the existence of the system, \( C_0(p) \) represent the initial investment of the system, \( C_T(p) \) the operational costs, and \( R(p) \) the expected value of losses due to forced interventions or pavement failure. The vector parameter \( p \) accounts for all design parameters necessary to make the decision: the frequency and the vehicle size for the transportation system, and the material properties and the thickness for the pavement’s layers.

The benefits and costs included in Equation (1) account for the whole system: transportation fleet as follows:

- The major benefit of the transportation system, \( B(p) \), is the reduction in the transportation time of the users of the system, all other benefits are considered independent of the variables included in the optimisation process. There are two possibilities for including the transportation time in the objective function: (i) including the reduction in the transportation time explicitly as a benefit in Equation (1), or (ii) including the waiting and travel time as a cost in the objective function. This second approach is used in this research.

- The initial cost, \( C_0(p) \), represent the initial investment in both components of the system: investment in the transportation fleet, \( C_{0f}(p) \), and investment in the infrastructure, \( C_{0i}(p) \).

- The operational cost, \( C_T(p) \), includes the cost necessary to operate the fleet, \( C_{op}(p) \), and the user cost that is related to the time that the users spend in the system, \( C_{us}(p) \), therefore the total operational cost is: \( C_T(p) = C_{op}(p) + C_{us}(p) \).

- The replacement costs of the system, \( R(p) \), includes the cost to replace the transport vehicles, \( R_V(p) \), and the cost necessary to repair and reconstruct the pavement, \( R_P(p) \), therefore the total replacement cost is: \( R(p) = R_V(p) + R_P(p) \).

As in a LCC analysis the decision about the project has to be made at time \( t = 0 \), all costs have to be discounted, for instance, by using a continuous function \( \delta(t) = e^{-\gamma t} \); where \( \gamma \) is the interest rate and \( t \) is the time in suitable units (Rackwitz 2000). As described, the benefits related to the transport time of the users are included in the transportation cost, and all the other benefits are taken as constant, i.e. independent of the optimisation variables. As a result, only the costs are included in the objective function and therefore the optimisation is the minimisation of the LCC costs, \( C_{LCC} \).

Two cases are considered:

1. The operation of buses on existing roads.
2. The operation of buses on new roads.

In the first case the variables to optimise are the size of the vehicles, \( K \), and the frequency, \( f \), the objective function is then:

\[ \min_{K, f} C_{LCC}(K, f) \]

Subject to \( R_0(f) \leq 1 \)  

(2)

In the second case the size of the structure of the pavement, \( H \), is included in the optimisation, then the objective function is:

\[ \min_{K, f, H} C_{LCC}(K, f, H) \]

Subject to \( R_0(f) \leq 1 \).  

(3)
In these equations \( R_0(f) \) represents the relative occupancy of the busses that is defined later on.

The optimum solution obtained by solving Equation (3) leads to a combination, or a value of \( H, f \) and/or \( K \) that minimises the total cost. However, in decision theory, optimum decisions should be based on expected values; therefore, the optimisation should be expressed as:

\[
\min_{\Phi} E[CLCC(\Phi)]
\]

Subject to \( R_0(f) \leq 1 \)

where \( \Phi \) is a vector parameter containing the decision variables (e.g. \( H, f, K \)). In order to compute the expected value with respect to the decision variables, the function \( CLCC \) should be integrated with respect to the joint density function of the variables considered. In what follows, the terms used to compute \( CLCC \) for each case will be defined and derived.

2.1. Benefit and initial cost

As described, in this paper the most important benefit is the reduction in the transportation time, this aspect is considered in the transportation cost as the cost of the time that the users spend in the system.

On the other hand, the initial investment on infrastructure and fleet are considered only at time \( t = 0 \), therefore the LCC cost, \( C_0^\infty(p) \) is:

\[
C_0^\infty(p) = C_0(p) = C_{0F}(p) + C_{0V}(p).
\]

2.2. Operational costs

The main link between the transportation service and the performance of the infrastructure is the vehicle size, which affects strongly the maintenance and replacement programs. Thus, operational costs are directly related to the frequency of vehicles and the number of passengers transported. In order to evaluate operational costs, Jara-Díaz and Gschwender (2009) considered a corridor served by one circular bus line \( L \) kilometres long, operating at a frequency \( f \) with a fleet of \( F \) vehicles. This circular route is used by a total of \( y \) passengers per hour uniformly distributed along the corridor. It is also assumed that each passenger travels a distance \( l \). Then, the cycle time \( t_c \) can be computed as:

\[
t_c = T_m + t_b y / f
\]

where \( T_m \) is the time in motion of a vehicle within a cycle, and \( t_b \) is average boarding and alighting time per passenger. The frequency \( f \) is given by the ratio between fleet size \( (F) \), and cycle time \( (f = F/t_c) \). Then the fleet size required to provide the service is:

\[
F = fT_m + t_b y.
\]

If traffic congestion is taken into consideration, a reduction in vehicle speed and an increment in interactions are expected, augmenting the time in motion of a vehicle on the route. The time in motion of a vehicle can be computed based on the speed-flow model proposed by Hoban et al. (1994) (Figure 1). In Figure 1, \( PCSE \) are the passenger car space equivalents; \( Q_0 \) is the flow level below which traffic interactions are negligible in \( PCSE/\text{h} \); \( Q_{nom} \) is the nominal capacity of the road in \( PCSE/\text{h} \); \( Q_{ult} \) is the ultimate capacity of the road for stable flow in \( PCSE/\text{h} \); \( S_{nom} \) is the speed at the nominal capacity in km/h; \( S_{ult} \) is the speed at the ultimate capacity in km/h; and \( S_{free} \) is the free flow speeds of different vehicle types in km/h.

The model predicts that below a certain traffic volume there are no traffic interactions and all vehicles travel at their free speed. When traffic interactions occur, the speed of individual vehicles decreases until the so called nominal capacity, which is the condition for which all vehicles will be travelling at the speed of the slowest vehicle class. The speed may eventually decrease until it reaches the ultimate capacity of the system. The flow level, i.e. \( Q \), of the circular route considered in this paper depends on (1) the initial passenger car in the route, i.e. \( Q_0 \); (2) the frequency of the bus system \( f \); and (3) the vehicle size \( K \):

\[
Q(f) = Q_0 + f k_s(K)
\]

where \( k_s(K) \) is a function relating the size of the vehicle (e.g. bus) \( K \) with its passenger car space equivalent (PCSE). Note that this model considers a fixed travel demand of private vehicles. However, it is possible that as a result of the improvement of the bus service level a

![Figure 1. Hoban’s speed-flow model.](image-url)
modal shift from car to bus might happen; further development of the model could consider this effect. The time in motion of a vehicle within a cycle considering the traffic congestion is:

\[ T_m(f) = \frac{L}{S(f)} \]  

(9)

where \( S(f) \) is the speed of the vehicles obtained after calculating the flow \( Q(f) \) (Equation (8)) and applying the Hoban’s model. According to Jara-Diaz and Gschwender (2009), the total costs per hour of the transportation service (i.e. \( C_T \)) is the sum of the operational cost \( (C_{op}) \), and the user cost \( (C_{us}) \), i.e. \( C_T = C_{op} + C_{us} \). The operator cost \( (C_{op}) \) is assumed to be a linear function of the vehicle size (i.e. \( K \)):

\[ C_{op} = F(c_0 + c_1 K) \]  

(10)

where \( c_0 \) and \( c_1 \) are constants. The user’s costs \( (C_{us}) \) depend on the values of waiting and in-vehicle time, respectively, \( P_w \) and \( P_v \):

\[ C_{us} = P_w \frac{p}{2} + P_v \frac{l}{L} (c_0 + c_1) \]  

(11)

Then, the total costs per hour as a function of the frequency becomes:

\[ C_T = C_{op} + C_{us} = (f T_m(f) + t_b y)(c_0 + c_1 K) \]  

\[ + P_w \frac{1}{2f} y + P_v \frac{l}{L} \left( T_m(f) + \frac{t_b y}{f} \right). \]  

(12)

If \( N_{hY} \) is the total amount of hours operated in a year (i.e. sum of the operation of buses all days), the annual operation cost for all buses operating the circular route becomes:

\[ C_{T_y} = N_{hY} C_T. \]  

(13)

Finally, the discounted operation cost required as input to Equation (1), is computed as:

\[ C_T(t) = \int_0^t C_{T_y}(t)e^{-\gamma t} dt \]  

(14)

which is the same as:

\[ C_T(t) = N_{hY} [(f T_m(f) + t_b y)(c_0 + c_1 K) \]  

\[ + P_w \frac{1}{2f} y + P_v \frac{l}{L} \left( T_m(f) + \frac{t_b y}{f} \right)] e^{-\gamma t} - 1 \]  

(15)

with the following asymptotic solution \( (t \to \infty) \),

\[ C_T^\infty = N_{hY} \left[ (f T_m(f) + t_b y)(c_0 + c_1 K) \right. \]  

\[ + P_w \frac{1}{2f} y + P_v \frac{l}{L} \left( T_m(f) + \frac{t_b y}{f} \right) \]  

\[ \left. \frac{1}{\gamma} \right] \]  

(16)

2.3. System failure costs

The total cost of losses can be computed as: \( H_{T}(p) = C(p) + H(p) \). In this expression \( C(p) \) is the total reconstruction or replacement cost and \( H(p) \) are all additional losses in which the owner incurs as a result of the intervention (e.g. opportunity cost).

The total cost of losses of the entire system includes costs associated with the pavement structure and the operation cost of the transportation fleet. The life-cycle performance of both physical infrastructure and transportation service can be modelled as a renewal process. This implies that once the system is put in service, it deteriorates until it reaches a pre-specified threshold. This threshold is a minimum operation standard prescribed by the authorities or by specific maintenance requirements. Then, the system is replaced and taken to its original standards. This process of successive failure and replacement is called a renewal process, which extends to infinity. Under these assumptions and based on the work by Rackwitz (2000), the total system failure costs can be rewritten as,

\[ R^\infty = \sum_{i=p,v} \left( C_i(p) + H_i(p) \right) \frac{1}{\gamma t_i} \]  

(17)

where \( 1/t_i \) is the failure rate; the terms \( p \) and \( v \) represent the cases for pavement and bus service (e.g. bus operation). The model proposed is based on the assumption that costs associated to the operation of the fleet and the investments on the pavement constitute a renewal process. This means that every time there is a failure (e.g. bus replacement, pavement intervention) the system (or the component) is repaired and the operation is restated. For this renewal process with events (failures) occurring at random times, there is an analytical solution proposed by Rackwitz (2000), which depends only on the failure rate of the system (or system components). This failure rate depends on the design parameters of the system and is computed in Equation (17) as \( 1/t_i \).

The failure rate is computed in Section 3 for both pavement and transportation failure modes (Equations (26) and (28) for pavements and Equation (31) for bus operation).
3. System’s time to failure

3.1. Model of pavement failure

Pavements are multilayer structures consisting of the sub-grade, the base layer(s) and a top layer consisting of all bitumen-bound layers (Figure 2). Despite the simplicity of the pavement structure, the material performance and the interaction between constituent layers is a rather complex phenomenon. In pavement design, the target of the analysis focuses on controlling:

1. Strains and stresses induced by traffic in the bituminous layer and in the sub-grade (radial and vertical strain, Figure 2).
2. The damage caused by fatigue in the bituminous layer.
3. Permanent deformation on the sub-grade.

The model of pavement failure used in this paper will focus mainly on the following performance (damage) criteria: (1) fatigue of the bituminous layer; and (2) permanent deformations on sub-grade.

3.1.1. Fatigue damage model of bituminous layers

Structural fatigue is the result of repeated cycles of stress/strain, which cause decay in the stress capacity and ultimately the structural failure (i.e. crack initiation). Fatigue is usually computed in terms of stress/strain amplitudes or stress/strain ranges of a cycle. Fatigue behaviour is described by an $S–N$ diagram or stress-cycle diagram; often the values of $N$ are plotted on a logarithmic scale since they are generally quite large. The variable $N$ describes the fatigue life (cycles to failure) and $S$ the cycle strain amplitude to reach the failure. Several models have been proposed to describe the $S–N$ relationship. Nevertheless, a widely accepted model is the Basquin equation, which is valid only for the high cycle range (i.e. $N > 10^4$ cycles, which is the normal case for pavements). This is,

$$NS^m = C$$  \hspace{1cm} (18)

where $S$ is the stress amplitude (or stress range) and $m$ and $C$ are empirical constants. For bituminous materials, the *French Design Manual for Pavement Structures* (LCPC 1997) recommends the following expression for fatigue life.

$$N\left(\frac{\varepsilon}{\varepsilon_0}\right)^m = 10^6$$  \hspace{1cm} (19)

where $\varepsilon$ is the strain amplitude induced on the material by the loading, $\varepsilon_0$ is the strain amplitude for which the failure is obtained at $10^6$ cycles, and $m$ is an experimental constant. This paper will focus only on the fatigue process under constant temperature and climate conditions; these and other relevant factors might be easily included in the model.

A global measure of fatigue damage is important to assess the overall state of the pavement and to define appropriate intervention measures. If a structure is subjected to a constant amplitude and frequency stress/strain, the extent of fatigue damage can be calculated by computing the rate between the number of loading cycles and the cycles required to reach the failure. However, under real fatigue circumstances, the frequency and amplitude of external loading varies from cycle to cycle; therefore, it is necessary to quantify the damage in every loading cycle and compute the accumulated damage based on the loading history. The effect of cyclic loads on any structure is normally

Figure 2. Design criteria for single wheel and dual wheel loading types.
accounted for with cumulative damage rules. These rules try to relate fatigue behaviour under a complex loading history to the known behaviour under constant amplitude loading. The Palmgren-Miner hypothesis is the most common linear damage accumulation rule and it is commonly known as Palmgren-Miner’s rule. Under variable amplitude loading, the Palmgren-Miner’s rule predicts the damage at time \( t \) as,

\[
D(t) = \sum_{i=1}^{k} \frac{n_i(t)}{N(S_i)}
\]  

(20)

where \( n_i(t) \) is the number of harmonic stress/strain cycles, at amplitude \( S_i \) at time \( t \); in other words, it is the incremental damage due to the \( i \)th block of constant stress/strain range \( S_i \); \( N(S) \) is the number of cycles to failure under the stress/strain \( S \) and \( k \) the total number of stress/strain groups. Despite some criticism about the validity of this index, it is widely accepted that it is a reasonable and convenient way of measuring global fatigue damage.

3.1.2. Fatigue damage rate for pavements

In order to compute the time to failure of a pavement subject to cyclic loading it is necessary to determine the time at which the damage index (Equation (20)) exceeds a pre-specified threshold \( \xi \) with \( \xi \leq 1 \) (Figure 3).

Extending Equation (20) to the continuous case and conditioning out the strain, the expected value of fatigue damage, \( D_F \), for a time, \( t \), can be expressed as (Sanchez-Silva et al. 2005):

\[
D_F(p, t) = \int_0^t \left[ \int_0^\infty \frac{N_e(\tau)}{N(S)} f_\varepsilon(p, S) dS \right] d\tau
\]  

(21)

where \( f_\varepsilon(p, S) \) is the strain probability density function, \( N_e(t) \) is the expected traffic at time \( t \), and \( N(S) \) is the number of loading cycles of strain \( S \) required to reach the failure (\( S-N \) model). Note that the integral inside the square brackets corresponds to the total number of loading cycles at a given time. The integral over time (i.e. outer integral) considers the time lapse from \( t = 0 \) to the evaluation time \( t \). Therefore,

\[
D_F(p, t) = \int_0^t \frac{N_e(\tau)}{N(S)} \left( \int_0^\infty f_\varepsilon(p, S) dS \right) d\tau
\]  

(22)

where \( N(S) = [(\varepsilon(p)/\varepsilon_0)^m] \). Therefore,

\[
D_F(p, t) = \int_0^t \frac{N_e(\tau)}{10^6} \left( \int_0^\infty f_\varepsilon(p, S) dS \right) d\tau
\]

\[
= \int_0^t \frac{N_e(\tau)}{10^6 \varepsilon_0^m} E[\varepsilon(p)^m] d\tau
\]  

(23)

Consequently, the crossing rate can be calculated as function of the damage threshold value, \( \xi \) (Figure 3) as: \( D_F(p, t_{PF}) = \xi \) where \( t_{PF} \) corresponds to the average time for the asphalt fatigue damage index to cross the threshold, \( \xi \). For a steady state with constant traffic, the cumulated traffic at \( t_{PF} \) is computed as:

\[
\int_0^{t_{PF}} N_T d\tau = N_Y t_{PF}.
\]  

(24)

Therefore, the fatigue damage at the time \( t_{PF} \) becomes:

\[
D_F(p, t_{PF}) = \frac{E[\varepsilon(p)^m]}{10^6 \varepsilon_0^m} N_Y t_{PF} = \xi
\]  

(25)

and the rate at which the system reaches a threshold \( \xi \) is:

\[
\frac{1}{t_{PF}} = \frac{N_Y E[\varepsilon(p)^m]}{10^6 \varepsilon_0^m \xi}.
\]  

(26)

The cumulated number of loadings due to traffic in one year, \( N_Y \), is obtained by the product of the hourly frequency \( f \) and the total number of operational hours \( N_{hY} \) in a year, i.e. \( N_Y = f N_{hY} \).
3.1.3. Damage model for permanent deformations on sub-grade

In addition to fatigue, the second main failure mode of pavements is due to excessive deformations on the sub-grade. The resistance of granular layers and the sub-grade is commonly measured in terms of permanent deformation, which is a time-dependent variable. Currently, there is not a well established mechanical model for describing this process. Empirical models have been proposed based on cyclic loading tri-axial test results. The scope of these results is limited by factors such as the loading history and the material physical characteristics. Fundamental methods, which take into account the mechanism of accumulation of permanent deformation, are still under construction, and there is not a widely accepted solution. In this paper, the empirical model proposed by Sweere (1990) will be used.

The limiting strain, \( \varepsilon_R \), at a given time \( t \) is defined as (LCPC 1997),

\[
\varepsilon_R = aN(t)^b = 0.012[N_e(t)]^{-1/4.5}.
\]

The Cost 333 (1999) action report recommends that design criteria for permanent deformations must be transformed into an incremental format (e.g. Miners law). Then, the damage is defined as a linear function of the number of loading cycles modelled also by Palmgren-Miner rule. Then, the crossing rate for the accumulated permanent deformation on the sub-grade can be computed as,

\[
\frac{1}{t_{PR}} = \frac{E[\varepsilon_R(p)^{4.5}]N_Y}{\psi0.012^{4.5}}.
\]

Equation (28) corresponds to the average number of loading cycles needed to reach the threshold value \( \psi \), which is a pre-specified limit for damage due to permanent deformations. The value of \( \psi \) is an input value for the optimisation; for further details see Sánchez-Silva et al. (2005).

3.1.4. Combination of pavement failure modes

As mentioned before, the failure of the pavement is governed by different failure modes: fatigue and permanent deformations on the sub-grade; therefore, if they are assumed to be independent, Equation (17) can be written as,

\[
R(p)^\infty = \frac{1}{\gamma} \left[ \frac{C_{Asph}(p) + H_{Asph}(p)}{t_{PF}} + \frac{C_{sg}(p) + H_{sg}(p)}{t_{PR}} \right]
\]

where \( C_{Asph} \) and \( C_{sg} \) are the costs of the asphalt layer, and of the entire pavement structure; note that its assumed complete replacement as a result of sub-grade failure, \( H_{Asph} \) and \( H_{sg} \) are the opportunity costs in which the owner incurs due to the damage of the bituminous layer, and the sub-grade respectively.

3.1.5. Other variables affecting the performance of the pavement

The mechanical performance of the pavement is affected by other variables in addition to the wheel loads, i.e. the environmental inputs (temperature, rain, etc.), which alter the performance of the layers of the pavement. As a result, the environmental conditions affect the number of loadings on the pavement to achieve the rutting or fatigue failure. For the cases considered in this paper, (busses operating on existing or new roads), the effect of the environment is considered independent of the optimisation variables. This is clear for existing roads since the environment affects the pavement in the same way for all the sizes of the vehicle and frequencies. In the case of new roads the size of the pavement is a variable included in the optimisation; in that case the environment could affect the pavement in different ways depending on the size of the structure. However, to avoid including more complexities in the model, the optimisation in this case is considered independent of the environment conditions too. Further developments could include this aspect.

3.2. Renewal model for the transportation fleet

In addition to pavement damage, losses may also occur in reference with the operation of the transportation fleet. It will be assumed in this paper that the fleet is replaced continually once each bus travels a fixed length, \( K_{FR} \). The total length that all the fleet travels in a year (\( T_L \)) is obtained by the product of the total cycles in one year (\( N_{hY} / t_c \)), multiplied by length of the route (\( L \)) and the size of the fleet (\( F \)):

\[
T_L = \frac{N_{hY}}{t_c} F L.
\]

Then, the mean amount of buses to be replaced each year can be estimated as the relationship between the total length travelled by the fleet, \( T_L \), and \( K_{FR} \): \( T_L / K_{FR} \). As a result, the rate of replacement of one bus, i.e. \( 1/t_Y \) is computed as the relationship between the number of buses that are replaced in a year divided by the size of the fleet (\( F \)):

\[
\frac{1}{t_Y} = \frac{N_{hY} L}{t_c K_{FR}}.
\]
Consequently, the cost associated to the renewal of the fleet is:

\[
R_V^\infty = \frac{1}{\gamma} \left( T_m(f) + t_h \frac{x}{C_{16}} \right) K_{FR} F C_V(K) \tag{32}
\]

where \( C_V(K) \) correspond to the cost of an individual vehicle with passenger capacity \( K \).

4. Characteristics of Transmilenio transport system

Before presenting the application of the proposed model to Bogota’s rapid bus transport system, some essential information about the system is provided in this section. Three fundamental aspects of Transmilenio will be presented and discussed in this section:

- Characteristics of the transport service (buses).
- Mechanical properties and performance of existing pavements.
- Definition of the pavement damage thresholds.

4.1. Characterisation of bus sizes and loading

In a previous study carried out at Los Andes University (2008), the actual axle loads of Transmilenio feeder buses were measured with the results shown in Figure 4. It can be observed that there is a linear relationship between bus capacity and axle loads. However, it should be noted that the relationship depends on the position of the engine in the bus. For example, when the engine is in the rear, the rear axle load increases and the loads applied by the front axle decrease. In this paper, only two axle buses with engines located at the front will be considered; nevertheless, the results can be easily extended to study other cases.

The actual axle load depends on the occupancy, \( K(f) \), which can be estimated in terms of bus frequency as (Jara-Díaz and Gschwender 2009):

\[
K(f) = \frac{y \ell}{\gamma L}. \tag{33}
\]

Similarly, the occupancy \( R_0(f) \) can be defined as the relationship between the actual occupancy and the bus capacity:

\[
R_0(f) = \frac{K(f)}{K} = \frac{y \ell}{\gamma L K}. \tag{34}
\]

The actual axle load, \( W_a \), depends on:

1. the axle load at maximum bus capacity, \( W_{\text{max}} \);
2. the relative occupancy; and
3. the relationship between axle loads for empty and a full bus occupancy \( W_m/W_T \), then:

\[
W_a = W_{\text{max}} \frac{W_m}{W_T} + \left( 1 - \frac{W_m}{W_T} \right) R_0(f). \tag{35}
\]

The relationship \( W_m/W_T \) depends of each bus as it is shown in Figure 5 for measurements of Transmilenio in Bogotá. As a first approximation, the relationship \( W_m/W_T \) is chosen as a constant for each axle and for each type of bus (two axles or articulated buses).

4.2. Description of Bogota’s pavement performance

The pavement structure of existing roads in Bogotá is highly heterogeneous, thus, the strains on the pavement as a result of the loading are not easy to compute. Considering an elastic behaviour of the pavement structure, the actual strain \( S_a \) is obtained by the

![Figure 4. Axle loads of two axle buses and articulated buses used in Bogotá.](image-url)
relationship between the strain at full occupancy of the bus ($S_{\text{max}}$), and the actual and maximum load of the bus: $S_o = W_a/W_{\text{max}}S_{\text{max}}$. Then, considering all the pavement structures, the effect of traffic loads is modelled as the summation of the effect of loads on discrete structures, and the equivalent strain, $S_e$:

$$E[S_e^m] = \sum_i sp_i \left( \frac{W_a}{W_{\text{max}}} S_{\text{max}} \right)^m$$

(36)

where $sp_i$ is the frequency of occurrence of a type of pavement structure $i$.

Los Andes University (2008) identified 10 predominant pavement structural profiles, (Figure 6) conforming the main feeder routes of Transmilenio, which were selected for this study.

For each structure (Figure 6) and each bus size (Figure 4), the tensile strain in the asphalt layer, and the vertical strain in the sub-grade were calculated using the software Alise$^\text{\textregistered}$ (LCPC software, Balay et al. 1997). One of the studied cases is shown in Figure 7 in order to illustrate the results of the analysis. A summary of the results of the mean tensile strain on the asphalt layer and the mean vertical strain on the sub-grade, for all 10 pavement structures considered (Figure 6), is shown in Figure 8.

4.3. Determination of threshold values of fatigue and rutting

The pavement repair cost depends on the time to reach a predetermined damage threshold for fatigue and rutting. The selection of those values was made by carrying out accelerated loading tests were performed at Los Andes University using the test track showed in Figure 9. Two axle loads, both with double tyre configuration corresponding to a rear axle of heavy and light buses, were applied, i.e. 100 kN and 50 kN.

The evolution of the density of fissures and the rut depth as a result of the number of loading cycles were recorded. Figure 10 shows the result of the density of fissures due to fatigue for two loading cases. The rut depth for the same two cases is depicted in Figure 11. Based on the results from the test track, a value of $\xi = 0.5$ was chosen as a threshold for the fatigue and a value of $\psi = 0.8$ as a threshold for rutting.

4.4. Results of the optimisation

During the design and construction of the Transmilenio, the transportation service characteristics, such as bus sizes and operational frequencies, were selected based mainly on passengers demand and leaving aside the infrastructure characteristics. The Transmilenio feeder’s system operates on small secondary roads that were not designed to withstand the loading. This decision led to a non-optimal solution since the required investments on maintenance and repair of the pavement infrastructure has proven to be extremely high. In fact, optimal solutions can only be found by integrating both operational decisions and investment in physical infrastructure; i.e. pavements. To illustrate the benefits of an integrated optimisation model, the model described in previous sections will be applied to two cases in Bogotá:

1. Feeder’s buses circulating on existing roads.
2. Feeder’s buses circulation on new pavement structures.

In the first case the optimisation process is intended to find the optimum bus size and bus frequency. In the second case, the optimisation will also include finding of the optimum pavement structure.
Case 1: Optimisation of frequency and bus size of buses operating on existing roads

As described, two basic assumptions were made for this case of optimisation:

1. The benefit was taken as constant, i.e. independent of the size of vehicles and the frequency.
2. The construction cost of the pavement $C_0$ is neglected since the buses run on existing roads.

Then the optimal frequency and bus size is obtained solving the following problem:

$$\min \mathbb{E}[C_{LCCA}(K,f)]$$

$$= \min \mathbb{E}[C_T(K,f) + C_{0F}(K,f) + R^\infty_f(K,f) + R^{>}_f(K,f)].$$  \hspace{1cm} (37)$$

Subject to $R_0(f) \leq 1$ (Equation 34).

The values of the parameters and values of the mechanical properties of materials, the bus operation characteristics and costs used to obtain the results shown here are summarised in Appendix 1. In reference to Equation (37) it is important to stress that, as stated by Jara-Díaz and Gschwender (2009), vehicle capacity constraints will always be active, since the objective function $C_{LCCA}$ does not improve when $K$ increases.

The calculated cost functions for the case of 3000 passengers/hour are shown in Figure 12. Figure 12 (a) corresponds to the case when all the transportation costs are considered and Figure 12 (b) correspond to the case where users costs (Equation (11)) are excluded. It can be observed that transportation costs grow with bus capacity but the total cost of losses for the infrastructure grow exponentially. As a result, the optimum bus size should be found decreasing the bus capacity and increasing the frequency.

Figure 6. Pavement structure on existing feeders routes in Bogotá. In this figure CA is asphalt concrete, GB is granular base, GSB is granular sub base and S is sub-grade.
The optimum bus capacity and operation frequency are shown in Figure 13. It shows the optimum bus sizes within a range from $K = 12$ (250 passengers/hour) to $K = 40$ (12,000 passengers/hour). For the case where the user's costs are not considered the optimum bus size is almost independent of the number of passenger per hour, i.e. $K_{op} = 42$ (250 passengers/hour) and $K_{op} = 40$ (12,000 passengers/hour). The trend for constant optimum bus size is similar to the trend found by Jara-Díaz and Gschwender (2009); however, the result of the optimum bus size is decidedly different since Jara-Díaz and Gschwender (2009) found an optimum of $K_{op} = 130$ while the present study shows an optimum value of $K_{op} \approx 40$. 

Figure 7. Tension strain and vertical strain in asphalt layer on sub-grade for structure E3 and maximum bus capacity (articulated bus and two axle bus $K = 92$).

Figure 8. Mean tension and vertical strain in asphalt and sub-grade layer considering all pavement structures.
The reason for this difference is the influence of the renewal considerations for modelling the pavement. Then a more reasonable solution would be optimising the vehicle size to preserve the infrastructure.

4.4.2 Case 2: Optimisation of frequency and bus size of feeders systems operating on new roads

As described, in this case the optimisation process deals with three variables:

1. Optimum bus size.
2. Optimum frequency.
3. Optimum pavement structure, \( H \).

Therefore, the construction cost of the pavement is included in the objective function, which becomes:

\[
C_{LCCA}(K, f, H) = C_T(K, f, H) + C_{0F}(K, f, H) + C_{0P}(K, f, H) + R_T^V(K, f, H) + R_P^V(K, f, H).
\] (38)

It is important to stress that although there are many parameters that may influence the construction cost of new pavements, most models found in the literature assume that these costs are proportional to the volume of material plus some fix cost of construction; this approach was used in the paper but the model can be easily extended to include more accurate construction cost models.

Based on Equation (37), the optimal frequency and bus size is obtained solving the following problem:

\[
\min_{K, f, H} E[C_{LCCA}(K, F, H)]
\] (39)

Subject to \( R_0(f) \leq 1 \).
There are infinite possible solutions for new pavement structures as a result of combining material properties and layer thicknesses. For the purpose of this analysis, this problem was approximated by using the eight pavement structures shown in Figure 14 for which the tensile strain on the bituminous layer and the vertical strain on sub-grade were calculated. Strains were computed for each bus size and the results were represented as a function of the equivalent thickness of the structure \( H_e \), using the equation suggested by Odemark (1949) and Kirk (1961):

\[
H_e = 0.8 \sum_{i=1}^{n-1} h_i \frac{E_i}{E_n} \quad (40)
\]

where \( n \) is the number of layers including sub-grade (layer \( n \)), \( h_i \) is the thickness of each layer and \( E_i \) is its modulus. It was assumed that all these structures rest on the same sub-grade having a Young modulus of \( E = 30 \) kPa.

Figure 15 (a) shows the tensile strain in the bituminous layer as a function of the vehicle size for different equivalent thickness of pavement; note that a linear relationship has a good agreement with the calculated strains. Figure 15 (b) shows the slope of this linear relationship as a function of the equivalent thickness, in this case a potential equation has a good agreement with the results of the calculations. Vertical strain on sub-grade follows the same trends than tensile strain. Then tensile and vertical strains for two axle buses can be represented by the following equation:

\[
e_{t,z} = a_{t,z} \left( \frac{1}{H_e} \right)^{n_{t,z}} K. \quad (41)
\]
For articulated buses only the bus having 160 passenger’s capacity is considered. As a result the strains are independent on the size $K$, and then Equation (41) becomes:

$$\varepsilon_{t,z} = \varepsilon_{t,z}^{\text{art}} \left( \frac{1}{H_e} \right)^{n_{t,z}},$$  \hspace{1cm} (42)$$

The optimisation process considers bus sizes and frequencies in a discrete set of values, and the effect of the thickness as a continuous form using Equations (41) and (42). Figure 16 shows the results of a route having 3000 passengers/hour with and without considering user’s cost. This figure shows that for each set of bus size, $K$, and frequency, $f$, there is an equivalent thickness having the minimum cost.

Figures 17 and 18 show the results of the optimisation process: optimum thickness, optimum bus size, and optimum frequency for different demands when new pavements are considered. These figures show that the optimum bus size for a small demand of passengers per hour is smaller than the optimum bus size for the case of existing pavement structures. This result is more evident mainly when the user’s costs are not considered, in fact, in this case the optimisation leads to a very thin pavement structure and lighter buses. On the other hand, for higher demands, e.g. 12,000 passengers/hour, the optimisation leads to stronger pavement structures and therefore larger buses ($K = 100$ considering users costs and $K = 160$ without considering users costs).

4.4.3. Some remarks concerning the costs and the results of the optimisation

The costs presented in the previous examples correspond to a particular case of study, for this reason the absolute value of the costs does not have a major interest for the analysis. However, a relative comparison of the costs illustrates the relevance of formulating a coupled analysis including the costs of the transport system and the infrastructure in the decision process of urban transportation systems. In fact, the decision based exclusively in the transport variables (passengers per hour) has a tendency to choose larger capacity buses, (buses having a capacity of 80 to 90 for the feeder’s routes in Bogotá). Note that for the particular
case illustrated in Figure 12, the total cost corresponding to buses of 90 passengers capacity are two to four times higher than the cost corresponding to the optimum bus size, depending on considering or not the users cost. Similar results can be found in the case of buses operating on new roads.
5. Conclusions
The design of BRT transportation systems is usually carried out following two separate stages:

1. Design the transportation system taking into account the demand.
2. Design the infrastructure taking in to account bus weight and traffic frequency.

This paper presents a model that integrates the operation measured through bus size and frequency, with the characteristics of the pavement. This model considers both the transportation costs (i.e. operation and user’s costs) as proposed by Jara Díaz and Gschwender (2009); and the infrastructure costs (i.e. pavement construction and renewal) as proposed by Sanchez-Silva et al. (2005).

The model has been applied using actual data from Bogota’s BRT transportation system. In particular, the optimum operation conditions were identified for the cases of existing and new pavement structures. The results of the integrated model show that the optimum bus size is considerably reduced to preserve the infrastructure. Although the vehicle size on the infrastructure costs. The results obtained considering existing pavement structures show that the optimum size of the buses is considerably reduced to preserve the infrastructure. Although when a new pavement structure is considered in the optimisation process, the size of the buses grows for high demand mainly if the user’s cost is not considered in the optimisation process. Furthermore the optimisation shows that the strong link between infrastructure and bus size remains active when new pavements structures are considered.

This new model could be an essential tool to establish new policies for designing the BRT systems.

Acknowledgements
This research was partially funded by the Instituto de Desarrollo Urbano IDU of Bogotá.

References


Los Andes University, 2008. Estudio de costos de mantenimiento de acuerdo con las configuraciones de ejes equivalentes de rutas alimentadoras del sistema Transmilenio. Bogotá, October, Internal Report Universidad de los Andes.


Appendix 1

Relationship between the sizes of the vehicle \( K \) with its passenger car space equivalent, PCSE:

\[
k_K(K) = 1 + 0.125 K \text{(PCSE)}.
\]

Cost of the vehicles as a function of its size:

\[
C_V(K) = 20864.10^{0.0112k} \text{(US$)}.
\]
Table 1. Values of the parameters for the transportation component of the model, Jara Díaz and Gschwender (2009).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$c_0$ (US$/h)$</th>
<th>$c_1$ (US$/h)$</th>
<th>$t_b$ (s)</th>
<th>$l$ (km)</th>
<th>$L$ (km)</th>
<th>$P_w$ (US$/h)$</th>
<th>$P_v$ (US$/h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>8</td>
<td>0.15</td>
<td>5</td>
<td>10</td>
<td>60</td>
<td>2.5</td>
<td>1.275</td>
</tr>
</tbody>
</table>

Table 2. Values of the parameters for the renewal model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$e_0$ (μstrain)</th>
<th>$m$</th>
<th>$ξ$</th>
<th>$ψ$</th>
<th>$N_{h,Y}$</th>
<th>$K_{FR}$ (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>160 $10^{-6}$</td>
<td>5</td>
<td>0.5</td>
<td>0.8</td>
<td>6935</td>
<td>600.000</td>
</tr>
</tbody>
</table>

Table 3. Values of the relationship of axle loads for empty and full occupancy buses.

<table>
<thead>
<tr>
<th></th>
<th>2 axle buses</th>
<th>Articulated buses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Front axle</td>
<td>Rear axle</td>
</tr>
<tr>
<td>$W_m/W_T$</td>
<td>0.75</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table 4. Values of the parameters of the Hoban’s speed-flow model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$S_{free}$ (km/h)</th>
<th>$S_{ult}$ (km/h)</th>
<th>$Q_0 = Q_{norn}$ (PCSE/h)</th>
<th>$Q_{ult}$ (PCSE/h)</th>
<th>$Q_{in}$ (PCSE/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>40</td>
<td>7</td>
<td>0</td>
<td>400</td>
<td>200</td>
</tr>
</tbody>
</table>