School location and capacity modification considering the existence of externalities in students school choice ☆

Iván Castillo-López [a], Héctor A. López-Ospina [b],*

[a] Departamento de Ingeniería Industrial, Universidad de Chile, República 701, Santiago, Chile
[b] Industrial Engineering Department, Pontificia Universidad Javeriana, Cra 7 #40-62, Ed. Jose Gabriel Maldonado P.3, Bogotá, Colombia

ABSTRACT

Geographic and socioeconomic characteristics of rural zones in Chile have made schools located in these areas present inefficiencies such long travel times and multi-degree courses[1] that affect the academic performance of their students. In this paper, a model of location and modification of school capacity is presented as an alternative to reduce these inequalities. In Chile a student school choice is a process that depends not only on the time and income constraints but also on the decisions made by other students (segregation). This behavior is modeled using a microeconomic approach; thereby a constrained multinomial logit discrete choice model is derived. By incorporating the student's school choice in an optimization model, it becomes nonlinear. A Tabu Search metaheuristic is proposed, which unlike other implementations requires solving a fixed point system of equations to evaluate each solution. A computing experience for instances of 10 and 45 zones is developed; in the first the quality of the solution is evaluated compared to the optimum obtained by enumeration and in the second different scenarios are analyzed.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Geographic and socioeconomic characteristics in rural zones in Chile and many Latin American countries have made that rural education present inefficiency. The schools located in these zones have a very low teacher-student ratio and many of their courses are multi-degree. According with SIGER (2009) Report, in Chile in 2007, 59.26% of rural establishments had less than 4 teachers and 69% of the courses were multigrade.

Another important aspect is that students are affected by long travel times to attend their schools. In SIGER (2009) it is shown that about 36% of the high school students (segundo medio) using public transportation indicated that they travel more than 60 min. With respect to the students' socioeconomic status (SES), more than a half belong to the poorest group, so that the financial resources they have for education are quite low. The above inefficiencies definitely affect school performance. 12% of rural schools have a poor performance on math state tests.

Furthermore, since 1981 the Chilean educational system is open for private providers to enter the market of education. This allowed

that all families can freely choose the school where their children will attend, only restricted by the time and income they have. This deepened inefficiencies of rural education. Cordova (2006) indicates that between 1992 and 2006 enrollment in rural zones fell by 16%, which shows that many students are willing to sacrifice themselves doing long travels and attending urban schools in search of a better education.

An alternative to reduce the rural education inequalities could be the location and modification of schools. A new configuration that improves the school's structural variables that affect the students' performance could lead to a higher quality education.

Thus, this paper proposes a location and modification school capacity model. In Chile as well as in other countries, a student school choice is a process that depends not only on the time and income constraints but also on the decisions made by other students (segregation). This behavior is modeled using a microeconomic approach; thereby a constrained multinomial logit discrete choice model is derived. By incorporating the student's school choice in an optimization model, it becomes nonlinear. A Tabu Search metaheuristic is proposed, which unlike other implementations requires solving a fixed point system of equations to evaluate each solution.

The other part of this paper is organized as follows: In Section 2 a literature review of schools location works is presented and it is established the context in which this research is conducted. In Section 3 the behavior of students and the optimization model is

---

described. Section 4 is devoted to the heuristic of solution. In Section 5 computational and evidence is reported, applied to a test zone. Finally, Section 6 describes the conclusions and final remarks.

2. Literary review

The models used in the location of schools have been developed in parallel with those that locate other facilities such as hospitals, households, firms, etc. Two contexts in which these models have been developed are: mathematical programming or optimization and the urban economy.

Optimization models seek the most efficient location with respect to a profit measure as distance, time, cost, etc. subject to different constraints such as security, resources availability, and time. The location of schools in this context was studied initially as a problem of students' allocation and known in the literature as *the school (re)districting problem*. The two features that distinguish the problem are that students in a geographic zone must be allocated to a school according to some criterion, for example distance, and that this allocation should not exceed the available capacity. In this sense, in the work of Caro, Shirabe, Weintraub, and Guignard (2004), the properties of a good allocation are established, they formulate an MIP with the aim of reducing the distance traveled by students and they make an application using a GIS, it is proved that imposing maximum impedance to travel generates more compact allocations.

Later, models of location and allocation were developed. While locating schools, students are allocated to them, with the aim of optimizing the distance. Early works only considered the opening and closing of schools. Tewary and Jena (1987) use a maximum coverage model to locate a fixed number of schools and maximize the population covered within a maximum travel distance. These models do not explicitly consider the capacity and only locate schools of the same size. Pizzolato and Silva (1997) use a p-median model to make a population clustering, each cluster is analyzed by comparing the educational supply and demand, so that, the opening or closing of schools in areas of shortage or oversupply is considered.

The works described previously do not consider the capacity explicitly and therefore they only locate schools of a same size. For that reason, Teixeira and Antunes (2008) use a p-median model with capacities, so that the allocation is performed to the nearest school with available capacity. In those works, the authors present a discrete hierarchical location model for the public facilities planning (schools). The main features of the model are: accessibility maximizing, several levels of demand and facilities, and capacity constraints. However, the authors assume that the system has the capacity to meet the demand. Pizzolato, Broseghini, and Nogueira (2001) also use a p-median model with insert capacities in a GIS to re-locate schools of different sizes in zones with shortages and oversupply of these ones, however, the authors assume that the system has the capacity to meet the demand.

The works described above do not take into account that it is financially appropriate to amend the current schools than opening new ones. Thus, Cohen, Martinez, Donoso, and Aguirre (2003) and SIGER (2009) develop models that locate new schools and modify the existing ones. The latter uses an MIP within a GIS for determining which schools should be opened, closed or modified. The model allocates students to the nearest school with available capacity to minimize transportation cost, operation and investment costs.

Antunes and Peeters (2000, 2001) describe a dynamic optimization model to formulate planning proposals for the school networks development based on an extension of the capacitated p-median model. The model allows the facility closure or downsizing, as well as the facility opening and its size expansion. The costs of the facilities are divided into a fixed component and two variable components, which depend, respectively, on the capacity and attendance. In Antunes, Berman, Bigotte, and Krass (2009) a model that seeks to maximize the total accessibility of the population to all different kinds of facilities is presented, considering that the location decisions influence the spatial distribution of the population growth.

Delmelle, Thill, Peeters, and Thomas (2014) develop a multi-period capacitated p-median model for the facility location planning that minimizes transportation costs, whereas the functional and operational costs of the education system are subject to a budget constraint. Furthermore, the allocations which are considered impractical because of the distance are penalized with a parameter associated with travel time.

The works described above assume that the students attend the nearest school. However, today's competitive markets allow a student to choose freely the service provider. Thus, a line of research that incorporates user behavior through discrete choice models in location models emerged.

The application of this methodology in other industries for example the location of airline hubs proposed by Eiselt and Marianov (2009), where the user's choice is defined by a gravitational function that considers travel time and fares; the problem is solved using a heuristic concentration method.

Marianov, Ríos, and Icaza (2008) propose a model for locating facilities, so that the market capture is maximized under the assumption that customers choose the facility where they want to be assisted according to the travel time and the waiting time, such choice is represented by a multinomial logit model. The authors demonstrate that under certain conditions there is an equilibrium demand and the problem is solved through a metaheuristic.

Colomé and Serra (2001) define several ways to incorporate the user's behavior in the context of location models and coverage. Such paper analyzes the optimal location from a competitive viewpoint including consumer's behavior aspects such as distance and transportation costs. To solve their formulations metaheuristics based on GRASP and tabu search are used.

With respect to the schools location, Gac, Martinez, and Weintraub (2009) develop a linear optimization model in which the students' preferences are introduced through a utility function whose variables are the characteristics of each school, travel and school costs. The bidders' utility is defined as the difference between revenue minus operating costs. The model seeks to maximize the profit of the bidders and applicants in the education system. However, the solution found does not involve a balance in prices, so this could get away from optimality.

A recent work is Hasse and Müller (2013), a model of location planning of the school network is proposed, seeking to maximize the expected utility of all students taking into account capacity constraints and a given budget. The utility value of each student is derived from a random utility model thus obtaining an endogenous demand. However the existences of externalities are not considered neither the impact of endogenous constraints in the students' utility.

Another context in which location models are developed is the urban economy. These models mainly seek to balance in the location of facilities (supply) and customer allocation (demand), the type of solutions generated are of a macro level and are mainly used for urban planning. School location methodologies were also developed in this context. Martínez, Tamblay, and Weintraub (2011) develop two models, one of equilibrium and another of optimization. The first one considers that households are agents who choose to attend school in accordance with each school characteristics, distance and price, and supply acts as an agent that maximizes its profit. The behavior of all the agents is modeled by logit constrained models. The second model seeks an optimal

solution, understood as the equilibrium which maximizes the whole system utility.

2.1. Variables that affect academic performance and school choice

In the Chilean context, Zuñiga (2009) researched the variables that influence the academic performance of students in rural zones. Starting from 21 variables (from school, family, and teachers) and using econometric models with different segmentations, the author found the influence of these variables on school performance. Although the models found present low adjustments, which according to the author, were mainly due to the presence of not-easy-to-measure subjective variables, those variables that are significant in the models are identified, from which, we conclude that reasonable travel times and large schools without multigrade classes will positively influence school performance. One of the conclusions of the previous research is that the school is 18% responsible for school performance and low SES students are the most affected by therein. Therefore, taking into account that in rural zones a majority of students belong to the lowest SES, modifications in the schools structure and location would be very beneficial to improve the quality of education.

Zuñiga (2009) also conducted a study on the variables affecting the schools choice in rural zones of Chile. The research was based on the revealed preferences of the parents. She concludes that the structural variables affecting parent's choices are: the SIMCE school score, the presence of secondary education, the enrollment, the number of combined grades and the socioeconomic status of students. These variables can be classified into two types: endogenous and exogenous. The first, also called externalities consist of those ones that depend on the decision made by other students.

Thus, this work contributes to the location modeling as follows: First, a school location model close to reality is developed, since it considers that a student not only chooses his/her school but also the mode transport, action that was not included in any school location model before. Second, several facility location models that incorporate a discrete choice model in its formulation, since the beginning, include a utility function with the variables that the authors consider relevant and ignore the presence of externalities in the choice process. In this paper, similarly to Martínez et al. (2011), a justification for the existence of externalities in the choice process is provided and a new scheme for its introduction into an optimal location model is proposed.

3. Formulation and development of the model

Due to the complexity of the problem, the analytical treatment is developed in two stages: the first one develops a microeconomic model that describes the student's behavior, and solutions which are characterized by the lack of incentive to change choice unilaterally are obtained (Nash Equilibrium); in the second, an optimization model to improve those structural variables of schools that affect the academic performance of students is developed.

3.1. Students' behavior modeling

Just as in the works by Martínez et al. (2011) it will be supposed that the students are rational agents who choose their school by maximizing a utility function under income and time constraints. A logit constrained model will be used to modeling this behavior, which is based on a microeconomic approach of discrete choices (Martinez, Aguila, & Hurtubia, 2009).

Based on the microeconomic approach by Jara-Diaz (2007) and defined in the Chilean school choice context (Zuñiga, 2009), the following variables will be considered in the utility function of a student: continuous consumption goods, travel time from home, and hedonic attributes like school academic performance and externalities associated with the quantity and socioeconomic status of students attending the school.

Therefore, a utility function is postulated for a student belonging to the SES h located in the zone i is

$$U_{hi}(\bar{x}, \bar{z}_{jk}, t_{ijm}) \tag{1}$$

where $\bar{x}$ is the vector of goods, $\bar{z}_{jk}$ the vector of hedonic attributes of the school type k located in j, and t_{ijm} is the travel time from the zone of residence i to the zone where the school j is located in the transport mode m. It is assumed that each student will maximize his/her utility (1) under budget and time constraints, and will know the transportation costs and the price of attending school.

Using the discrete choice approach, utility is maximized in two stages, in the first the utility function is maximized, conditional on an alternative (j, k, m) (school type k, located in j, which is accessible by the transport mode m) thus the conditional indirect utility function in the chosen alternative V_{jkm}. In the second stage, the student simply selects the choice (j, k, m) that generates the highest utility among all possible discrete alternatives. The problem is formulated analytically as:

$$\underset{j,k,m}{Max}[\underset{\bar{x}}{Max}\, U_{hi}(\bar{x}, \bar{z}_{jk}, t_{ijm})] \tag{2}$$

$s.t.$

$$\sum_l p_l x_l + r_{jk} + c_{ijm} \leqslant I_h \quad \forall h, i, j, k, m \tag{3}$$

$$t_{ijm} \leqslant \tau_m \quad \forall i, j, m \tag{4}$$

$x_l \geqslant 0 \quad \forall l$

h: Index for socio-economic cluster to which the student belongs
i: zone index for the student's home location
j: zone index for school location
k: index to shool type
m: index to mode transport
$\bar{z}_{jk}$: vector of hedonic attributes of type school k located in zone j,
t_{ijm}: travel time from home located at the zone i to school located in zone j in transport m
τ_m: maximum time that a student would be willing to ride the transport mode m
p_l, x_l: goods consumption and their respective prices
r_{kj}: Tariff of school type k located in the zone j
c_{ijm}: cost of transport mode m from origin zone i to destination zone j
I_h: average income of a student h

The income constraint (3) states that the expense of a student in a period consists of three items: consumption of goods (clothing, school supplies, etc.), school fees (tariff) and transportation costs. The total cost should be less than average revenue per student in the same period. The time constraint (4) states that a school which is infeasible due to travel time cannot be chosen.

Thus, assuming quasi-linear direct utility function, then, conditional indirect utility function is obtained

$$V_{hijkm} = \sum_n \beta_h^n z_{jk}^n + \gamma_h \cdot t_{ijm} + \theta_h(I_h - q_{ijkm}) \tag{5}$$

where β_h^n is the parameter that indicates the student's assessment for the n-th attribute of the school, γ_h is the attribute of the travel time, θ_h parameter indicating the impact of money in the utility level achieved and q_{ijkm} is the sum of the cost of education and transportation.

According to the random utility theory, McFadden (1978), it is assumed that the utility has a stochastic component that specifies the idiosyncratic user's behavior, namely:

$$V_{hijkm} = \overline{V}_{hijkm} + \varepsilon_{hijkm}$$

where $\overline{V}_{hijkm}$ is the deterministic part, established by (5) and ε_{hijkm} is the error stochastic term associated with the level of ignorance, which in this model it is assumed equally and identically Gumbel distributed, with scale factors μ. Therefore, given a school type k located in a zone j and which is accessible in the mode m, an SES student h will attend it and in that mode, if this option generates a greater utility over the other alternatives. Therefore, the choice probability for an SES student h located in i for a school (j, k, m) will be:

$$P_{jkm/hi} = \text{Prob}(V_{hijkm} > V_{hij'k'm'}; \ \forall h, i, j', k', m')$$

Or equivalently

$$P_{jkm/hi} = \frac{\exp\left[\mu(\overline{V}_{hijkm})\right]}{\sum_{j'k'm' \in \Omega} \exp\left[\mu(\overline{V}_{hij'k'm'})\right]} \quad \forall h, i \tag{6}$$

This choice (j, k, m) offers the maximum utility compared to other alternatives where Ω is the set of alternatives (school-zone-mode) available for an SES student h located in the zone i

The following sections describe how to include constraints and externalities within the stochastic discrete choice model (6).

3.1.1. Choice alternatives set constraint

A student does not consider all the schools of the system as choice alternatives; therefore, it is needed to delimit this set. The first condition to consider a school as an alternative is that if must exist. Thus, the probability of choice is:

$$P_{jkm/hi} = \frac{x_{jk} \exp\left[\mu(\overline{V}_{hijkm})\right]}{\sum_{j'k'm' \in \Omega} x_{j'k'} \exp\left[\mu(\overline{V}_{hij'k'm'})\right]} \quad \forall h, i \tag{7}$$

where the variable x_{jk} is obtained from an external model of the discrete choice associated with an optimization model and takes the value 1 if in the zone j there is a school type k or zero otherwise.

This optimization model will be described in Section 3.2. Strictly speaking, Eq. (7) corresponds to a multinomial logit model with supply correction.

3.1.2. Travel time and income constraints

In addition a student cannot make long travel times nor spend on transport and tariff more than his/her income allows. Hence, a school must meet these time (4) and income (3) constraints to become a choice alternative. These constraints are considered using the constrained logit model (CMNL) by Martinez et al. (2009). In this model, individuals impose thresholds to the attributes directly in the utility function, by penalties specifications in the binomial logit form, in order to greatly reduce the utility of those alternatives that do not meet any restrictions. In Castro, Martinez, and Munizaga (2013) analyzed the parameters estimation of CMNL model using real data. The authors conclude that the model CMNL is more suitable in some applications, having better fit than the Multinomial Logit model. Besides, there are significant differences in the value of time and elasticity values, showing that these differences increase as the attribute thresholds are triggered.

Thus, the following terms are joint to the indirect utility function

$$\frac{1}{\mu} \ln w_{ijm} \tag{8}$$

$$\frac{1}{\mu} \ln u_{hijm} \tag{9}$$

$$w_{ijm} = \frac{1}{1 + \exp\left[\omega_w(t_{ijm} - \tau_m + \varphi_w)\right]} = \begin{cases} \to 1 & \text{if} \quad t_{ijm} < \tau_m \\ \eta_w & \text{if} \quad t_{ijm} = \tau_m \end{cases}$$

$$u_{hijm} = \frac{1}{1 + \exp\left[\omega_u(q_{hijm} - I_h + \varphi_u)\right]} = \begin{cases} \to 1 & \text{if} \quad q_{hijm} < I_h \\ \eta_u & \text{if} \quad q_{hijm} = I_h \end{cases}$$

η_w, η_u are parameters that indicate the population proportion that do not meet the constraints of time and income respectively, and in this research it is assumed that their values are equal to 0.05. As well as, ω_w, ω_u are binomial scale parameters and determine how fast w_{ijm} or u_{hijm} take extreme values 1 and η. It is known that very large values of ω_w, ω_u will make w_{ijm}, u_{hijm} behave in a deterministic way, wherefore it will be assumed they have the value 1. With respect to φ_w, φ_u, these are adjustment parameters and are defined as:

$$\varphi_w = \frac{1}{\omega_w} \ln\left(\frac{1 - \eta_w}{\eta_w}\right)$$

$$\varphi_u = \frac{1}{\omega_u} \ln\left(\frac{1 - \eta_u}{\eta_u}\right)$$

Thus, the utility associated with schools that do not comply with a particular constraint will be diminished, making the alternative show a less choice probability.

Thus the utility function would be described as:

$$V_{hijkm} = \sum_n \beta_h^n z_{jk}^n + \gamma_h \cdot t_{ijm} + \theta_h(I_h - q_{ijkm}) + \frac{1}{\mu} \ln w_{ijm} + \frac{1}{\mu} \ln u_{hijm}$$

And the choice probability with the time and income constraints is formulated as follows:

$$P_{jkm/hi} = \frac{w_{ijm} \cdot u_{hijm} \cdot x_{jk} \exp\left[\mu(\overline{V}_{hijkm})\right]}{\sum_{j'k'm' \in \Omega} w_{ij'm'} \cdot u_{hij'm'} \cdot x_{j'k'} \exp\left[\mu(\overline{V}_{hij'k'm'})\right]} \quad \forall h, i$$

3.1.3. Capacity constraint

Finally, there is a capacity constraint which states that the number of students attending a school must not exceed its capacity. Therefore, when the capacity use of a school is 100%, it will cease to be a choice alternative for other students. This constraint is also included through a constrained logit model (CMNL), in which the same previous values can be assumed for the parameters η_ϕ and ω_ϕ. However it should be noted that in this case, the number of students attending a school (enrollment) is an externality of choice.

$$\frac{1}{\mu} \ln \phi_{jk} \tag{10}$$

$$\phi_{jk} = \frac{1}{1 + \exp\left[w_\phi(mat_{jk} - cap_{jk} + \varphi_\phi)\right]} = \begin{cases} \to 1 & \text{if} \quad mat_{jk} < cap_{jk} \\ \eta_\phi & \text{if} \quad mat_{jk} = cap_{jk} \end{cases}$$

$$mat_{jk} = \sum_{him} H_{hi} \cdot P_{jkm/hi}$$

mat_{jk}: Number of students who attend school type k, located in zone j.
cap_{jk}: Capacity of the school type k, located in zone j.

Thus incorporating the capacity restriction in the utility function, we have

$$V_{hijkm} = \sum_n \beta_h^n z_{jk}^n + \gamma_h \cdot t_{ijm} + \theta_h(I_h - q_{ijkm}) + \frac{1}{\mu} \ln w_{ijm} + \frac{1}{\mu} \ln u_{hijm}$$
$$+ \frac{1}{\mu} \ln \phi_{jk}$$

And the choice probability with these constraints is formulated as follows:

$$P_{jkm/hi} = \frac{\phi_{jk} \cdot w_{ijm} \cdot u_{hijm} \cdot x_{jk} \exp\left[\mu(\overline{V}_{hijkm})\right]}{\sum_{j'k'm' \in \Omega} \phi_{j'k'} \cdot w_{ij'm'} \cdot u_{hij'm'} \cdot x_{j'k'} \exp\left[\mu(\overline{V}_{hij'k'm'})\right]} \quad \forall h, i \tag{11}$$

where the system of equations described by (11) represents a fixed point system of equations, given that the capacity of each school generates an externality associated with the competition for some options. That is, the school choice decision is conditioned to the fact that other students choose it too.

3.1.4. Socioeconomic and size externalities

With regard to the choice externalities, it follows that the vector $\bar{z}_{jk}$ collects the school hedonic attributes. This implies that there are endogenous and exogenous attributes within the choice process.

The number of enrolled students and the school SES are endogenous attributes and constitute choice externalities. The way to introduce them into the students' behavior modeling is as attributes of the indirect utility function, because depending on the value they take they can make an alternative more or less attractive.

The number of students who enroll in a school is an attribute valued by the students when choosing the school. However, it must be considered that schools will not necessarily be at 100% capacity. Hence, schools will also be classified in $\hat{k} \in K$ types because of the number of students enrolled. Thus, the following term is added to the utility function:

$$\sum_{\hat{k}} \eta_{h\hat{k}} \cdot tam_{jk}(P_{jkm/hi}; \forall m, h, i) \tag{12}$$

$\eta_{h\hat{k}}$: parameter indicating the assessment of an SES student h by a school type $\hat{k}$
$tam_{jk}(P_{jkm/hi}; \forall m, h, i)$: 1 if the school located in j is type $\hat{k}$ and 0 otherwise.

Another important assumption is that each student values, when choosing a school, the attribute associated with the other students' SES who attend this school. The following expression is introduced into the utility function when considering this externality:

$$\sum_{h'} \lambda_{hh'} \cdot ns_{jh'}(P_{jkm/hi}; \forall m, i) \tag{13}$$

$\lambda_{hh'}$: parameter indicating the assessment of an SES student h, for a school where most students are SES h'
$ns_{jh'}(P_{jkm/hi}; \forall m, i)$: 1 if in the school located in j the majority of students are SES h' and 0 otherwise.

Then, the utility function with all the considerations previously described is of the form

$$V_{hijkm} = \sum_{n} \beta_h^n z_{jk}^n + \gamma_h \cdot t_{ijm} + \theta_h(I_h - q_{ijkm}) + \frac{1}{\mu}\ln w_{ijm} + \frac{1}{\mu}\ln u_{hijm}$$
$$+ \frac{1}{\mu}\ln \phi_{jk} + \sum_{k'} \eta_{hk'} \cdot tam_{jk'} + \sum_{h'} \lambda_{hh'} \cdot ns_{jh} \tag{14}$$

Thus, we have that the new probability of attending to a school type k, located in j, in the mode m is given by:

$$P_{jkm/hi} = \frac{\phi_{jk} \cdot w_{ijm} \cdot u_{hijm} \cdot x_{jk}\exp\left[\mu \cdot \bar{V}_{hijkm}\right]}{\sum_{j'k'm'\in\Omega} \phi_{j'k'} \cdot w_{ij'm'} \cdot u_{hij'm'} \cdot x_{j'k'}\exp\left[\mu \cdot \bar{V}_{hij'k'm'}\right]} \quad \forall h,i \tag{15}$$

Note that ϕ_{jk} as well as $\bar{V}_{hijkm}$ depend on $P_{jkm/hi}$ then (15) generates a fixed point system of equations. The solution of this system is a condition for the existence of a static equilibrium. In Martinez et al. (2009) its convergence has been demonstrated for similar choice probabilities

3.2. Optimization model

After the students' behavior modeling, the constrained logit model (15) must be introduced into an optimization model that seeks to find a configuration (location, size) of schools with lower average travel times, improving the average number of students per school, and the number of schools with multigrade classes. However, considering only such kind of objectives can lead to financially infeasible solutions; for this reason the objective of minimizing the investment and operational costs is included.

Due to the fact that a problem can consider different types of schools, then to explain the problem modeling it is assumed that $k, k' \in \{0 = \text{closed}; 1 = \text{multigrade}; 2 = \text{small}; 3 \text{ medium}; 4 = \text{large}\}$

$$\text{Min } \alpha\left(\frac{CT_1 - CT_0}{CT_0}\right) - (1-\alpha)$$
$$\times \left[v_1\left(\frac{Tv_0 - Tv_1}{Tv_1}\right) + v_2\left(\frac{Tcol_1 - Tcol_0}{Tcol_0}\right) + v_3\left(\frac{Nm_0 - Nm_1}{Nm_1}\right)\right] \tag{16}$$

$$H_{hijkm} = H_{hi} \cdot P_{jkm/hi} \quad \forall h,i,j,k,m \tag{17}$$

$$\sum_k x_{jk} cap_k = \sum_k y_{jk} cap_k + a_{jk} - s_{jk} \quad \forall j \tag{18}$$

$$a_{jk} \leqslant M \cdot add_{jk} \quad \forall j,k \tag{19}$$

$$a_{jk} \geqslant add_{jk} \quad \forall j,k \tag{20}$$

$$s_{jk} \leqslant M \cdot sub_{jk} \quad \forall j,k \tag{21}$$

$$s_{jk} \geqslant sub_{jk} \quad \forall j,k \tag{22}$$

$$\sum_k x_{jk} = 1 \quad \forall j \tag{23}$$

$$add_{jk} + sub_{jk} \leqslant 1 \quad \forall j,k \tag{24}$$

$$sub_{j0} = 0 \quad \forall j \tag{25}$$

$$add_{j4} = 0 \quad \forall j \tag{26}$$

$$CT_0 = \sum_{j,k} C1_{jk} y_{jk} \tag{27}$$

$$CT_1 = \sum_{j,k,k'}\left(C1_{jk} \cdot x_{jk} + C2_{jkk'} \cdot y_{jk} \cdot x_{jk'}\right) \tag{28}$$

$$Tv_1 = \frac{\sum_{h,i,j,k,m} H_{hijkm} \cdot t_{ijm}}{\sum_{hi} H_{hi}} \tag{29}$$

$$Tcol_1 = \frac{\sum_{hi} H_{hi}}{\sum_{j,k\neq 0} x_{jk}} \tag{30}$$

$$Nm_1 = \sum_{j,k=1} x_{jk} \tag{31}$$

$$v_1 + v_2 + v_3 = 1 \tag{32}$$

$$x_{jk}, add_{jk}, sub_{jk} \in \{0,1\}; \quad a_{jk}, s_{jk} \geqslant 0; \quad P_{jkm/hi} \in [0,1] \tag{33}$$

The parameters are

y_{jk}	1 if currently there is a school type k in the zone j
H_{hi}	number of students in the group h in the zone i
cap_k	school type k capacity
$C1_{jk}$	annual operating cost of school type k located in the zone j [\$]; which considers only the teachers and principals' salaries
$C2_{jkk'}$	annualized modifying cost of a school located in the zone j, from the type k to the type k' [\$]; this parameter includes (when it corresponds) the costs of increasing or decreasing the capacity or opening a new one

Decision variables

x_{jk}	1 if a school type k is installed in the zone j, 0 otherwise
add_{jk}	1 if to the school type k in the zone j the capacity is increased, 0 otherwise

sub_{jk}	1 if to the school type k in the zone j the capacity is decreased, 0 otherwise
a_{jk}	size in which the school type k capacity increases in the zone j (students)
s_{jk}	size in which the school type k capacity decreases in the zone j (students)
$P_{jkm/hi}$	probability that a student located in a zone i, and belonging to a group h, attends a school type k located in the area j in the transport mode m. Note that this variable depends on x_{jk} as explained in Section 3.1
H_{hijkm}	Number of students in the group h, located in the zone i attending the school type k located in zone j in the transport mode m

Auxiliary variables

CT_1, CT_0	total cost of schools operation with and without intervention plus subsidy cost as appropriate [$]
Tv_1, Tv_0	average travel time per student with and without intervention (min)
$Tcol_1, Tcol_0$	Average number of students enrolled per school with and without intervention (students)
Nm_1, Nm_0	Number of multigrade schools with and without intervention (schools)

The model (16)–(33) is multi-objective optimization since it seeks to improve the student's performance and reduce system costs. The objectives are not similar or comparable each other, so that these are converted to indexes. Thus, the percent variation of each objective in the new configuration is optimized regarding the one that was initially had.

The objective function (16) indicates operating costs and investment in the education system are minimized simultaneously, and in turn the average travel time is minimized too, maximizing the average amount of enrolled students per school, and minimizing the number of schools with multigrade classes. The planner must establish the importance of each objective by modifying the parameters α and υ.

Eq. (17) shows the relationship between the number of students attending a particular school with the choice probability.

With respect to the constraints, Eq. (18) establishes a balance between the final situation of a school size and the current one, such that the capacity of each school must increase and decrease when it corresponds. Constraints (19)–(22) provide that when a school capacity is increased or decreased there will be a binary variable respectively that should take a value of one, and zero otherwise. These constraints consider a value M, which corresponds to a very large value that can be the maximum capacity of the schools. Eq. (23) states that there must be only one school per zone; this implies that the zoning was done considering this assumption. Eq. (24) shows that a school can be increased o reduced in capacity, but not both simultaneously. The constraints (25) and (26) indicate that capacity cannot be reduced in a closed school and in a large school cannot be increased respectively. The constraints (27)–(32) are equations that determine the value of auxiliary variables used in the objective function. Eq. (33) indicates the nature of the decision variables.

4. Solution method

Due to the combinatorial nature of the problem, the nonlinearity of the model and the need to solve fixed point systems of equations, there is no exact method for its solution, except for the exhaustive enumeration which for combinatorial optimization problems is very inefficient. For this reason the use of a metaheuristic to find a good solution is proposed.

Heuristics used in location problems are *Genetic Algorithms (GA)* (Jia, Ordoñez, & Dessouky, 2007; Song, Morrison, & Ko, 2013; Toro-Díaz, Mayorga, Chanta, & McLay, 2013), *Tabu Search (TS)* (Cura, 2010; Diaz, Ferland, Ribeiro, Vera, & Weintraub, 2007), *Simulated Annealing (SA)* (Karaoglan, Altiparmak, Kara, & Dengiz, 2012; Taheri & Zomaya, 2007; Yu, Lin, Lee, & Ting, 2010), *Heuristic Concentration (HC)* (Eiselt & Marianov, 2009), *Greedy Randomized Adaptive Search Procedure (GRASP)* (Bautista & Pereira, 2006; Marianov et al., 2008), *particle swarm optimization* (Samarghandi, Taabayan, & Jahantigh, 2010). Unlike other problems, in this a fixed point problem should be solved every time the objective function is evaluated, therefore a heuristic that is efficient in this sense and evaluates the objective the least amount of times is needed while a good solution is being found simultaneously. In this paper, the use of Tabu search is proposed (Glover, 1989, 1990, 1994).

Tabu Search is a meta-heuristic developed by Glover (1990). It is an iterative process starting with an initial solution in search of a better one. In each iteration a neighborhood is defined, which is composed of solutions that are accessible from a movement from the current one, then the best of these solutions is selected even if this does not improve the current solution. To avoid returning to the same solution the procedure maintains a list of constrained items, called tabu list. Likewise, the procedure applies criteria to leave local optima and make more intensive searches.

The form in which the use of meta-heuristics is proposed in the problem is as follows: it considers that in each candidate area $j \in J$ a school type $k \in K$ can be found, where $k \in \{0 = closed, 1 = multigrade, 2 = small, 3 = medium, 4 = large\}$. Therefore, a feasible solution is of the form $x = \{x_1, x_2, \ldots, x_{|J|}\}$. From which a local search is made, it is a procedure that permits to move to a neighboring solution x'. The elements (solutions) of the neighborhood are obtained by making modifications $m \in M$ to the current solution. A neighbor can be denoted as:

$$x' = x \oplus m \quad m \in M$$

Therefore, a neighbor is obtained when making a modification $m \in M$ to the solution x, hence the neighborhood $N(x)$ can be specified as follows:

$$N(x) = \{x' : x' = x \oplus m / m \in M\}$$

In this problem a modification is defined as follows:

1-opt: Set of modifications where in any area $j \in J$ the size of the school increases or decreases. Therefore, let $x' = x \oplus m$ for $m \in 1 - opt$ and any area $j \in J$, then: if $x_j = k$ $x'_j = k^* \in K/$ ($k^* = k + 1 \cup k^* = k - 1$).

This means if for example a school is of medium size, in a modification that school can only be large or small.

The number of neighbors (neighborhood) that are caused by a movement *1-opt* is sized $|J| \times 2$, considering that for each one a complex fixed point system of equations must be solved a strategy to reduce the neighborhood is needed to be used. The proposed strategy is as follows: not to consider movements that a priori will not improve the solution, so only if a school is with some use of its capacity its size will increase or decrease. For example if the students attending a school are less than half the capacity of the school, it would not be advisable to increase its size. Therefore, if a current solution is $x_j = k^*$, the neighboring solution for that school will be given by: $x'_j = k^* + 1 \iff enrollment_j \geqslant T_3 \cdot Cap_k$ or $x'_j = k^* - 1 \iff enrollment_j \leqslant T_4 \cdot Cap_k$. Where T_3 and T_4 are parameters to be specified. With this strategy, the number of neighbors decreases to a maximum $|J|$.

In each iteration, the best solution x' is selected, it could happen that in some iterations the best solution x' is worse than the current solution. However this movement is permitted because it

is a strategy to avoid local optima. In this application, the elements of the tabu list (TL) include the zones where schools were opened or where interventions were performed to the existing schools, that is to say if $x' = x \oplus m$ for $m \in 1 - opt$ in any area $j \in J$, then $j \in LT$. A school will be on the list for a maximum number of T_1 iterations, also called *tabu ternure*. The stopping criterion of the heuristic is a maximum number of iterations without improvement (*ITER_MAX*).

In addition to the basic tabu procedure described, it is drawn on aspiration, intensification and diversification strategies that use the short and long term memory structures of heuristic. In this way, more intensive searches in promising regions are done and searches in unvisited regions.

Every time the objective function is evaluated, it must be calculated $P_{jkm/hi}$ (probability that a student of SES h located in the area i attend a school of size k, located in j in mode m), for which it is required to solve a complex fixed point system of equations, with $|H| \times |I| \times |J| \times |K| \times |M|$ equations.

When a school is resized, the choice probabilities of the whole system should be re-calculated, which requires too much computational effort. But this change only affects some zones, because a student may only attend schools that are within a maximum travel time (Tmax), therefore the following postulate is formulated:

Postulate: Whenever a school that is located in the area j product of a 1-opt move is modified, only the choice probabilities of the zones that are within a time Rmax of the zone j (region of influence Z_j) will change, whereas the rest of probabilities will remain unalterable.

To understand the postulate, the following figure must be considered (Fig. 1): the circles correspond to zones where students are located and the squares are schools. For example, if school A is modified, a region of influence Z_A is generated. Then, according to the postulate only the choice probabilities of the students from zones belonging to the region of influence (1, 2, 3, 4 and 5) would change, whereas the rest of probabilities of other zones will remain unalterable.

The above is justified because when the school A is modified, those directly affected are the students of the zones that can access it at a maximum travel time Tmax (2 and 4) but due to the presence of externalities students from other zones would also be affected (1, 3, 5, 8, etc.). However, it has been verified that the enrollment of the schools closest to A are the ones that suffer the most changes, and as schools are further away from A, their enrollment varies less. Hence, the choice probabilities variations will also be decreasing according the zones move away from A. Therefore, since the modification made to school A in each iteration is controlled and as product of this modification the variation of the

Table 1
Comparison of results with and without consideration of postulate.

Type of modification	Postulate	F.O.	Time (seg.)	Num. iter.
Open small school	Yes	−0.169	13	58
	Not	−0.169	4	30
Close small school	Yes	−0.181	11	44
	Not	−0.181	3	24

other schools enrollment gradually decreases according the zones move away from A; and furthermore an Rmax greater than a Tmax is used, it can be ensured the postulate accomplishment. In conclusion, when a modification in a school located in the area j is made, only the probabilities $P_{jkm/hi}$ $\forall i \in Z_j$ will change whereas the rest of the probabilities will remain unalterable.

In Table 1, the results of comparing the postulate with different modifications made in an instance are shown (see Section 5). The first column indicates the modification description made to a school; the second column indicates whether or not the postulate is considered; the third column shows if there are differences in the solutions from the objective function value of the optimization model; the last two columns show the calculation time and the number of iterations with and without the postulate for a specific modification type. It can be seen that, after modifying a school, the results obtained with and without the postulate are the same. The objective function value obtained from each solution indicates that the average travel time and the average number of students per school are the same. In turn, the calculation times and the number of iterations to find the solutions change. Therefore, if the heuristic considers the postulate, the same solution would be obtained as that obtained without it, but instead a saving in calculation time would be achieved. This result is very important and will help resolve larger problems.

With respect to the fixed point problem (15), it can be shown that it converges and has a unique solution, therefore it becomes an unconstrained nonlinear programming problem and solved by commercial software.

5. Application and computational experience

The application will be made to locate secondary schools. The students at this level are characterized by greater freedom for longer trips; on the other hand as the quality the students receive at this level is critical to achieve a good result in the PSU[2] they are willing to sacrifice more resources to attend good quality schools.

The application is performed in two networks of 10 and 45 zones, in which four scenarios are analyzed: The first consists of locating and modifying schools without any constraint, in the second it is constrained the fact that there must be at least a school 30 min away, the third scenario assumes a grant to school transport for the lowest SES students and the latter constrains the number of schools that can be closed.

The application seeks to locate schools in the rural area, which is modeled as a network whose nodes correspond to zones $i \in \{1, 2, \ldots, I\}$ and the arcs represent the existing road network. The students in each area differ in homogeneous groups by their families' SES $h \in= \{1, 2, 3\}$, differentiated mainly by their income level I_h. It is assumed that each area is made up by the same percentage of students from each SES. Furthermore it is considered that depending on the SES, each household allocates a percentage of their income to their children education. Since each family has an average number of children, the necessary annual income for education that a student belonging to a particular SES is obtained.
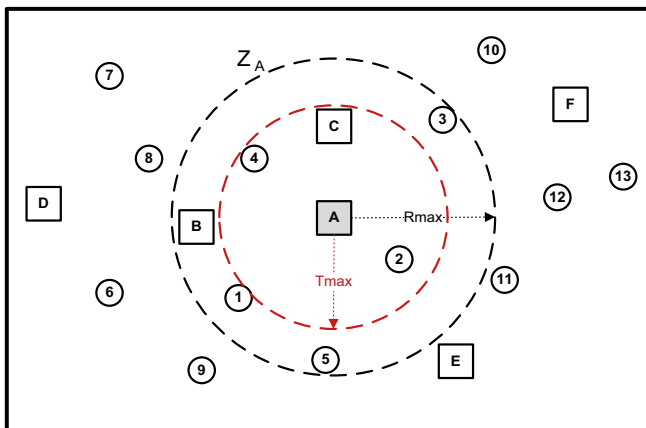

Fig. 1. Influence region for school A.

Table 2
Solution quality and time resolution (Instance of 10 zones).

Sce	Cost parameter (α)	Quality parameter ($1-\alpha$)	Constraint of maximum time	Transport subsidy	Constraint of schools closed	F.O. Enumeration	F.O. heuristic	GAP (%)	Time resolution for enumeration (hh:mm:ss)	Time resolution for heuristic (hh:mm:ss)
1	0.5	0.5	No	No	No	−0.0202	−0.0202	0.0%	20:45:12	00:10:18
2	0.5	0.5	Yes (30 min)	No	No	−0.0981	−0.0981	0.0%	12:35:18	00:00:59
3	0.5	0.5	No	yes	No	0.0825	0.0825	0.0%	32:12:33	00:18:20
4	0.50	0.50	No	No	Yes (1 school.)	−0.0202	−0.0202	0.0%	15:23:54	00:10:05

Table 3
Solution of location and modification of schools (instance 45 zones).

Esc.	Size[d]				Closed schools[a]	Opened schools	Modified schools[d]	Average travel time[b] (min)	% utilization[c]	Socioeconomic status			Performance SIMCE			Costs (million Chilean pesos)[d]
	Small	Med	Large	Total						Low	Medium	High	Def.	Good	Outstan-ding	
Initial	29 (48%)	4 (0%)	1 (0%)	34	–	–	–	23.5 (14.1%)	77% (108)	30	4	0	14 (40%)	14 (40%)	6 (20%)	0
Asig	32 (34%)	1 (0%)	0 (0%)	33	5 (60%)	4	0	23.8 (14.7%)	88% (111)	31	2	0	11 (33%)	17 (51%)	5 (15%)	−26
1	13 (23%)	11 (36%)	2 (0%)	26	8 (88%)	0	11	23.2 (12.2%)	80% (141)	22	4	0	7 (27%)	13 (50%)	6 (23%)	−45.2
2	27 (48%)	6 (17%)	1 (0%)	34	0	0	6	23.5 (14.0%)	75% (108)	30	4	0	14 (40%)	14 (40%)	6 (20%)	80
3	1 (0%)	12 (42%)	4 (0%)	17	17 (53%)	0	17	24.7 (10%)	81% (216)	16	1	0	5 (30%)	6 (35%)	6 (35%)	−77.5
5	21 (33%)	9 (44%)	1 (0%)	31	3 (100%)	0	9	23.3 (14%)	76% (119)	28	3	0	11 (31%)	14 (45%)	6 (19%)	48.8

[a] The values in parentheses correspond to the percentage of schools with low performance in mathematics test: SIMCE.
[b] The values in parentheses correspond to the percentage of students who exceed the threshold of 45 min.
[c] The values in parentheses are the average number of students per school.
[d] Difference between the cost of the system once the changes have been made and the system's initial configuration (1 US dollar is approximately 500 Chilean pesos).

Table 4
Variation in the costs and quality of the system compared with the initial solution.

Scenarios	Costs (%)	Quality			Global (%)[a]
		Travel time (%)	Schools size (%)	Total quality (%)	
Assigns	−1.54	1.09	−42.19	−21.64	−10.05
1	−2.68	−1.29	53.13	27.21	14.94
2	4.73	0.04	12.50	6.23	0.75
3	−4.55	4.86	89.06	42.10	23.32
5	2.89	−0.86	26.56	13.71	5.41

[a] Value corresponding to the objective function of the respective scenario.

The schools can be located in each of the zones $j \in \{1, 2, \ldots, J\}$ assuming that in each course there can be at most 30 students, schools are classified into 4 types $k \in \{1, 2, 3, 4\}$, closed, small, medium and large respectively, all of municipal dependence. These are located using a maximum coverage model (*Location Set Covering Problem, see for example* El-Darzi & Mitra, 1990; Farahani, Asgari, Heidari, Hosseininia, & Goh, 2012) so that all students are covered by at least one school within a maximum travel time. A score will be randomly allocated to the existing schools, so that 40% have a poor performance, 40% good and 20% excellent. The new settled schools will only be municipal and it will be assumed that they will have an outstanding performance. With regard to the costs, 3 types are only considered: operating, investment and closure costs.

Regarding the mode of transportation, there are 4 alternatives: on foot, by bicycle, public and private transport $m \in \{1, 2, 3, 4\}$ respectively, the travel time is estimated from the average speed in each of the modes. The cost of a walking trip is zero, the cycling cost has an annualized cost because of the bike purchase, and the cost in private and public transport varies depending on the traveled stretches.

The multinomial logit function parameters are built such that they reasonably reflect the students' choice, ensure consistent results and facilitate the calculation of the fixed-point systems. The heuristic was programmed in MATLAB 7.6.0, the fixed-point

systems are solved with GAMS/ MINOS 3.5, all in a notebook with AMD Turion 64 × 2 Dual Core, 2 GB RAM, OS Windows Vista.

The optimal solution of the instance of 10 zones is obtained by complete enumeration.[3] However because of the combinatorial nature of the problem, this method cannot be used to obtain the optimal solution for the instance of 45 zones. Thus, the first instance will evaluate the quality of the solutions, whereas the second will be an analysis of the solutions.

In Table 2 the scenarios are described and the values of the objective function are compared as well as the computation times in the scenarios of the instances of 10 zones. In this table, the results obtained by numbering are the same obtained by using the heuristic; therefore the GAP in each of the scenarios is 0%. The times used for obtaining the optimal solution by numbering become between 12 and 32 h, whereas the heuristic finds them in minutes. In this way the efficiency of the heuristic used is verified.

In Table 3 different indicators obtained from the found solutions for the instance of 45 zones are shown. The first column identifies the scenario. The second column shows the size of the schools, specifying the number of small, medium and large schools.

---

[3] The number of solutions for a problem is j^k where j is the number of zones and k is the school type.

The third, fourth and fifth columns indicate the number of closed, open and modified schools respectively. The sixth column shows the average travel time and the percentage of students traveling above the permitted threshold. The seventh column shows the percentage of the schools capacity use and the average number of students per school. The eighth and ninth columns show the number of schools by SES and SIMCE performance respectively. Finally the last column shows the difference between the system cost with the modifications and the initial situation.

Table 4 shows the variation between the costs and the quality variables compared to the initial situation. The first column identifies the scenario, the second column shows how much the system total costs vary including as appropriate the grant, the third column shows the variation in the quality of the system, specifying the percentage change in each quality variable, the last column indicates the global system variation.

The initial solution of the system, obtained from a maximum coverage model, has 34 schools (29 small, 4 medium and 1 large). The average travel time is 23.5 min and 14.1% of students travel above the threshold. The percentage of the schools capacity use is 77%, with an average of 108 students per school. The performance of each school is randomly obtained. Fig. 2 shows the initial situation.

If the solution of a model that allocates students to the closest school such a way to minimize the travel time is used, it is evidenced with respect to the initial situation, the average travel time increases 1.1% (23.8 min) and the percentage of students who travel above the trip threshold increases from 14% to 14.7%. 97% of schools are small and schools with an outstanding performance are closed. Therefore, in this application the solution of a model that allocates students to the nearest school does not improve the educational system quality and generally it worsens it at 21.6%. Fig. 3 shows the hypothetical allocations and Fig. 4 what would happen in reality.

Scenario 1, which does not consider any constraint on the modification and location of schools, has a solution, with respect to the initial configuration the number of schools is reduced to 26 (13 small, 11 medium and 2 large). The average travel time decreases a 1.3% and the percentage of students exceeding the threshold is reduced from 14% to 12.2%. There are on average 141 students per school. The number of schools with poor performance halves and the system annual costs decrease a 2.7%. Clearly this solution improves the quality of education; the quality variables show an improvement of 27.2%. The number of small schools decreases and medium and large schools are increased, demonstrating a consolidation of institutions. There are also a smaller number of
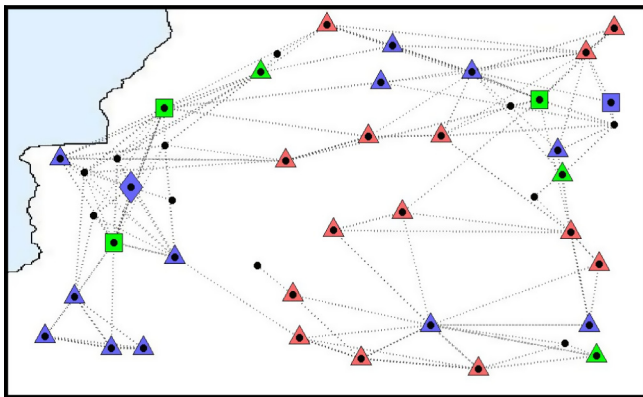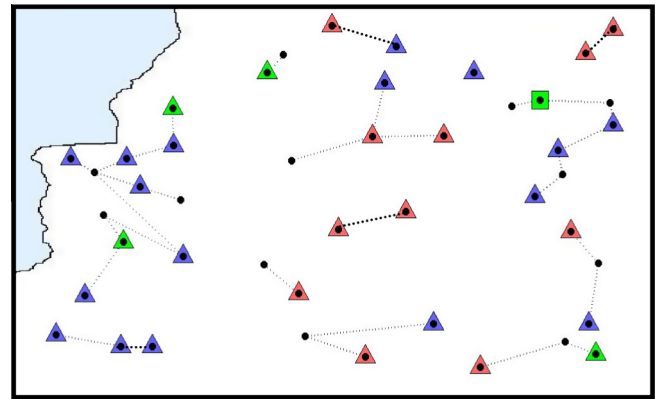

Fig. 3. Situation assuming that students attend to the nearest school. △ Small school, ☐ Medium school, ◆ Large school, ■ Poor performance, ■ Good performance, ■ Outstanding Performance, • Zone, ——- If at least 2 students traveling from one zone to a school.
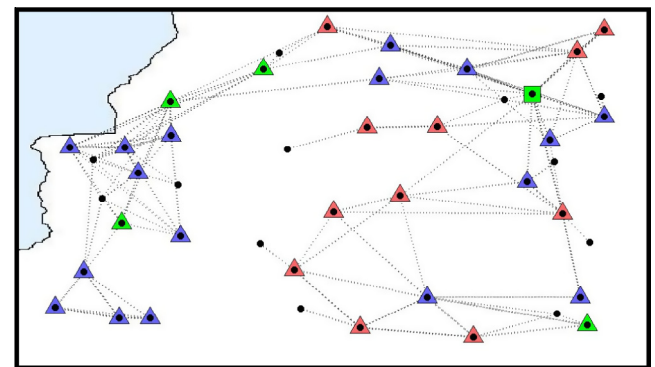

Fig. 4. Real situation with the hypothetical solution that students attend to the nearest school. △ Small school, ☐ Medium school, ◆ Large school, ■ Poor performance, ■ Good performance, ■ Outstanding Performance, • Zone, ——- If at least 2 students traveling from one zone to a school.

schools with poor performance, because 88% of the closed schools are of this type. Depending on the objective function value this scenario achieves an overall improvement of 15% compared to the initial solution and with respect to the rest of scenarios is the second best solution. Fig. 5 shows this Scenario.

The solution of scenario 2, which requires the existence of at least one school 30 min away, shows that schools are kept at 34, but with different sizes (27 small, 6 medium and 1 large). The average travel time and the percentage of students who exceed the threshold are almost the same. This solution increases the annual cost of the educational system in 4.7% and the quality variables improve a 6.2%.

Scenario 3, which simulates a possible 100% grant for the transport of low SES students, has as solution that the number of schools is reduced to 17 (1 small, 12 medium and 4 large). This new configuration significantly reduces the number of small schools and increases the amount of medium and large schools. As a result of the grant, the average travel time increases a 4.86%, but the percentage of students exceeding the threshold is reduced from 14% to 10%. The use of the installed capacity also increases to 81% having on average 216 students per school. The number of schools with poor performance decreases by more than a half. For this, it is necessary a grant that reaches $ 190 million per year and benefits the 45% of the student population. The system annual costs (including the grant) decrease a 4.5%. This
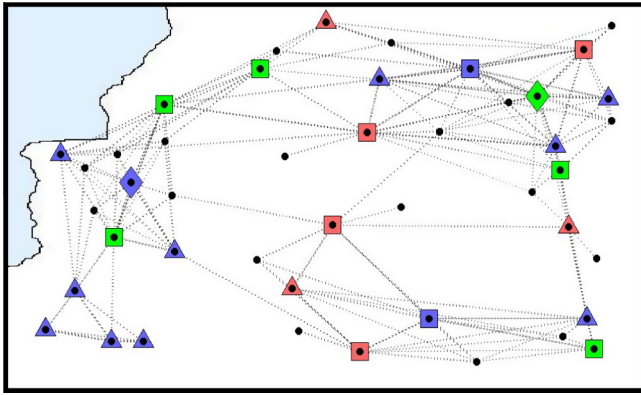

Fig. 2. Initial situation. △ Small school, ☐ Medium school, ◆ Large school, ■ Poor performance, ■ Good performance, ■ Outstanding Performance, • Zone, ——- If at least 2 students traveling from one zone to a school.

Fig. 5. optimization scenario 1. △ Small school, □ Medium school, ◆ Large school, ■ Poor performance, ■ Good performance, ■ Outstanding Performance, ● Zone, ——– If at least 2 students traveling from one zone to a school.

Fig. 7. Optimization scenario 4. △ Small school, □ Medium school, ◆ Large school, ■ Poor performance, ■ Good performance, ■ Outstanding Performance, ● Zone, ——– If at least 2 students traveling from one zone to a school.

solution greatly improves the quality of education. The quality variables show a 42.1% improvement. In addition the poor performance schools decrease from 14 to 5. The objective function value indicates that the system is globally improved a 23.3%. Fig. 6 shows this solution.

Scenario 4, which constrains the number of schools that can be close to 3, has as solution: a configuration with respect to the initial solution the number of schools is reduced to 31 (21 small, 9 medium and 1 large). The average travel time is reduced a 0.86%. The capacity use decreases to 76% having on average 119 students per school. The three closed schools have a poor performance. And the system annual cost increases a 2.9%. This solution only improves the quality a 13.7%. And globally the system a 5.4%. Fig. 7 shows this solution.

The scenario that gets greater reduction in the system costs is 3. This solution achieves an annual reduction (including the transportation grant) of 4.5%. Also, this scenario is the one that achieves better results in terms of quality, the solution improves the quality variables a 42.1%. This solution is characterized by closing almost all small schools and greatly increasing the number of medium and large schools. Due to the possibility of doing longer trips, the average travel time increases a 4.9%, but the percentage of students who exceed the travel threshold decreases from 14% to 10%. The number of poor performance schools is reduced to more than a half, having a greater number of outstanding than deficient schools.
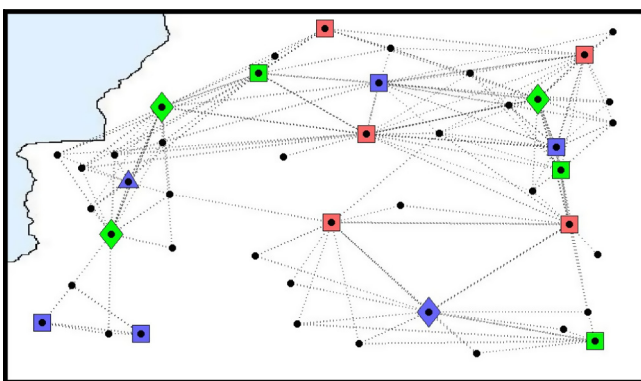
The scenarios that impose constraints on the modifications are the ones which get fewer improvements to the system. Scenario 2, which constraints the solution to the existence of at least one school 30 min away, is the one with the worst results. In almost all scenarios the opening of schools is not considered, although it is assumed that the new schools will have a good performance. This is because the initial solution installs enough schools, thus increasing the amount increases the system costs. There are not any high SES schools either, because the percentage of the population belonging to this SES is very small and scattered as to form a school of this level.

6. Conclusions

This research presents a mathematical model to determine the location and size of new schools and the intervention type that must be made in the current ones, so that the investment and operating costs and the schools structural variables that influence on academic performance are optimized. The model incorporates two modeling approaches: the discrete choice theory by which the students' behavior is modeled and the mathematical programming which optimizes the variables that influence the school performance. The need to incorporate the time in the modeling is achieved by introducing the modal choice problem. Also using a suitable microeconomic model, the choice variables and the externalities presence within the election process are justified. Because a student has constraints that limit the set of choice alternatives, a constrained multinomial logit model (CMNL) arises.

By incorporating the constrained logit model within an optimization model, it becomes an endogenous nonlinear problem and solving it requires using a Tabu Search heuristic procedure. Even so the problem is different from typical applications of the heuristic since each time the objective function is to evaluate it is needed to solve a fixed point system of equations. Given this complexity, the heuristic incorporates reduction strategies of the neighborhood size and efficient probabilities calculation. Precisely because of these strategies, this methodology can be applied to solve larger problems.

The model is applied to the location of secondary schools in instances of 10 and 45 zones, in the first the solution quality is evaluated by comparing it with the optimal solution obtained by enumeration. The best heuristic solutions for 10 zones are obtained in a few minutes and compared with those obtained by enumeration it is verified that they are optimal, which demonstrates the efficiency of the heuristic used. In the instance of 45 zones, the optimal solution cannot be obtained by enumeration due to

Fig. 6. Optimization scenario 3. △ Small school, □ Medium school, ◆ Large school, ■ Poor performance, ■ Good performance, ■ Outstanding Performance, ● Zone, ——– If at least 2 students traveling from one zone to a school.

the combinatorial nature, but these demonstrate significant improvements in the educational system compared to the initial situation. The solution obtained from a model that allocates students to the nearest school shows that with respect to the initial solution the costs decrease a 2.7%, but the quality also decreases a 21.6%. Therefore, doing interventions in the schools, obtained from an allocation model, would not improve the quality of education. The scenario that achieves better results in terms of costs and quality is the one that grants transportation for low SES students. This solution achieves an annual reduction in costs of 4.5% (including transportation grant) and improves the quality variables a 42.1%.

In general the methodology is very flexible and can be used to analyze different scenarios, depending on the policies expected to be implemented. In addition it can be used to locate other facilities in which the choice depends not only on the location but also on the cost, time, or the choices made by others (externalities).

As future research it is expected to test the model in real instances, finding the characteristic parameters of the discrete choice problem as well as adding or deleting some of the postulates used in this research or using other heuristics to solve the same problem.

References

Antunes, A., & Peeters, D. (2000). A dynamic optimization model for school network planning. *Socio-Economic Planning Sciences, 34*(2), 101–120.

Antunes, A., & Peeters, D. (2001). On solving complex multi-period location models using simulated annealing. *European Journal of Operational Research, 130*(1), 190–201.

Antunes, A., Berman, O., Bigotte, J., & Krass, D. (2009). A location model for urban hierarchy planning with population dynamics. *Environment and Planning A, 41*(4), 996–1016.

Bautista, J., & Pereira, J. (2006). Modeling the problem of locating collection areas for urban waste management. *An application to the metropolitan area of Barcelona. Omega, 34*(6), 617–629.

Caro, F., Shirabe, T., Weintraub, A., & Guignard, M. (2004). School redistricting: embedding GIS tools with integer programming. *Journal of the Operational Research Society, 55*(8), 836–849.

Castro, M., Martinez, F., & Munizaga, M. (2013). Estimation of a constrained multinomial logit model. *Transportation, 40*(3), 563–581.

Cohen, E., Martinez, R., Donoso, P., & Aguirre, F. (2003). Localización de infraestructura educativa para localidades urbanas de la Provincia de Buenos Aires. In CEPAL, Santiago de Chile, Serie Políticas Sociales (Vol. 79).

Colomé, R., & Serra, D. (2001). Consumer choice and optimal locations models: Formulations and heuristics. *Regional Science* (80), 439–464.

Cordova, C. (2006). Elección de escuela en sectores pobres de Chile. Simposio: Políticas educacionales y gestión de sistemas escolares. Universidad de Complutense de Madrid.

Cura, T. (2010). A parallel local search approach to solving the uncapacitated warehouse location problem. *Computers & Industrial Engineering, 59*(4), 1000–1009.

Delmelle, E., Thill, J-C., Peeters, D., & Thomas, I. (2014). A multi-period capacitated school location problem with modular equipment and closest assignment considerations. *Journal of Geographical Systems, 16*(3), 263–286.

Diaz, A., Ferland, J., Ribeiro, C., Vera, J., & Weintraub, A. (2007). A tabu search approach for solving a difficult forest harvesting machine location problem. *European Journal of Operational Research, 179*(3), 788–805.

Eiselt, H., & Marianov, M. (2009). A conditional p-hub location problem with attraction functions. *Computers & Operation Research, 36*(12).

El-Darzi, E., & Mitra, G. (1990). Set covering and set partitioning: A collection of test problems. *Omega, 18*(2), 195–201.

Farahani, R., Asgari, N., Heidari, N., Hosseininia, M., & Goh, M. (2012). Covering problems in facility location: A review. *Computers & Industrial Engineering, 62*(1), 368–407.

Gac, I., Martinez, F., & Weintraub, A. (2009). A deterministic linear optimization model for allocating school to zones. *Journal of the Operational Research Society, 60*, 895–905.

Glover, F. (1989). Tabu-search part I, ORSA. *Journal on Computing, 1*, 190–206.

Glover, F. (1990). Tabu search: A tutorial. *Interfaces, 20*(4), 74–94.

Glover, F. (1994). *Tabu search fundamentals and uses.* University of Colorado at Boulder.

Hasse, K., & Müller, S. (2013). Management of school locations allowing for free school choice. *Omega, 41*(5), 847–855.

Jara-Diaz, S. (2007). *Transport economy theory.* Elsevier.

Jia, H., Ordoñez, F., & Dessouky, M. (2007). Solution approaches for facility location of medical supplies. *Computer & Industrial Engineering, 52*, 257–276.

Karaoglan, I., Altiparmak, F., Kara, I., & Dengiz, B. (2012). The location-routing problem with simultaneous pickup and delivery: Formulations and a heuristic approach. *Omega, 40*(4), 465–477.

Marianov, V., Ríos, M., & Icaza, M. (2008). Facility location for market capture when users rank facilities by shorter travel and waiting times. *European Journal of Operational Research, 191*(16).

Martinez, F., Aguila, F., & Hurtubia, R. (2009). The constrained multinomial logit: A semi-compensatory choice model. *Transportation Research Part B*, 102–122.

Martínez, F. J., Tamblay, L., & Weintraub, A. (2011). School locations and vacancies: A constrained logit equilibrium model. *Environment and Planning A, 43*(8), 1853–1874.

McFadden, D. (1978). Modeling the choice of residential location. In *Spatial interaction theory and planning models* (pp. 75–96). Amsterdam: North-Holland.

Pizzolato, N., Broseghini, F., & Nogueira, L. (2001). School location methodology in urban zones of developing countries. *International Transactions in Operational Research* (11), 667–681.

Pizzolato, N. D., & Silva, H. B. (1997). The location of public schools: Evaluation of practical experiences. *International Transactions in Operations Research, 4*(1), 13–22.

Samarghandi, H., Taabayan, P., & Jahantigh, F. (2010). A particle swarm optimization for the single row facility layout problem. *Computers & Industrial Engineering, 58*(4), 529–534.

SIGER (2009). Diseño y Aplicación de un Sistema de Información Geográfica para la Educación Rural. Universidad de Chile. Santiago de Chile.

Song, B., Morrison, J., & Ko, Y. (2013). Efficient location and allocation strategies for undesirable facilities considering their fundamental properties. *Computers & Industrial Engineering, 65*(3), 475–484.

Taheri, J., & Zomaya, A. (2007). A simulated annealing approach for mobile location management. *Computer Communications, 30*, 714–730.

Tewary, V. K., & Jena, S. (1987). High school location decision making in rural India and location–allocation models. In *Spatial analysis and location–allocation models.* New York: Van Nostrand Reinhold.

Teixeira, J., & Antunes, A. (2008). A hierarchical location model for public facility planning. *European Journal of Operation Research, 185*(1), 92–104.

Toro-Díaz, H., Mayorga, M., Chanta, S., & McLay, L. (2013). Joint location and dispatching decisions for Emergency Medical Services. *Computers & Industrial Engineering, 64*(4), 917–928.

Yu, V., Lin, S., Lee, W., & Ting, C. (2010). A simulated annealing heuristic for the capacitated location routing problem. *Computers & Industrial Engineering, 58*(2), 288–299.

Zuñiga, C. (2009). Estudio de Variables que Inciden en la Calidad y en la Elección de establecimientos Rurales para el Desarrollo de un Sistema de Localización. Universidad de Chile, Santiago de Chile, Memoria para optar al título de Ingeniaría Civil Industrial.