COUPLING GIS AND OPTIMIZATION SOFTWARE IN PUBLIC FACILITY PLANNING

João Teixeira[1] and António Antunes [2]

ABSTRACT

In the last ten years, GIS became a fundamental tool for planning activities with a geographic component. During the same period, optimization software went through enormous improvements and became useful for many planning purposes. In this article we show how we coupled one of the GIS more widely available on the market (ArcView GIS) with one of the most powerful discrete optimization programs (XPRESS-MP) in a public facility planning process – the preparation of a school network development plan for the municipality of Coimbra, Portugal. The plan was aimed at defining the optimum configuration for the municipality's primary and secondary school networks in 2015. The GIS was applied in four types of operations: to define the demand centers included in each local community; to determine the shortest path matrix for all pairs of demand centers and facility sites (i. e., sites where schools were or could be located); to compute data for a special type of assignment constraints included in the optimization model; and to generate visual representations of model solutions. The coupling of GIS and optimization software was crucial to simplify activities that otherwise would be much more difficult to perform.

KEY WORDS

GIS, optimization software, public facility planning, school network, Coimbra.

INTRODUCTION

In this article we show how we coupled one of the Geographical Information Systems (GIS) more widely available on the market (ArcView GIS) with one of the most powerful discrete optimization programs (XPRESS-MP) in a public facility planning process developed in 2004-2005 – the preparation of a school network development plan for Coimbra, a medium size municipality of 320 km^2 and 150,000 inhabitants located in the center-littoral region of Portugal. The plan aims at defining the configuration for the municipality's primary and secondary school networks in 2015, in terms of the location, type, size and catchment areas of schools. The plan addresses two major challenges. First, current aggregate school capacity is excessive because of the strong decline of school-age population in the last two decades.

---

[1] Researcher, Civil Engineering Department, University of Coimbra, 3030 – 290 Coimbra, Portugal, Phone (+351) 239 797131, FAX (+351) 239 797142, jtex@dec.uc.pt

[2] Civil Engineering Department, University of Coimbra, 3030 – 290 Coimbra, Portugal, Phone (+351) 239 797139, FAX (+351) 239 797142, antunes@dec.uc.pt

Second, school typology needs to be changed according to a recent reorganization of the Portuguese educational system.

To support the planning process, we developed a discrete facility location model, considering decisions of opening or closing schools, and of assigning students to schools. The objective of the model is to maximize accessibility, that is, to minimize total student travel distance, while satisfying a set of constraints on school capacity and student-to-school assignments. The plan is to be approved by a council composed of the municipal administration, school administrations, parents' organizations and teacher unions. The multiple interests of all parties involved, often conflicting, make this decision-making process complex. The merit of the approach presented here is to provide rational solutions as a basis for discussion.

In the last ten years, GIS became a fundamental tool for the input and output operations involved in all planning activities with a geographic component (Brail and Klosterman, 2001; Longley *et al.*, 2001). In addition to being used for storing information, the GIS was applied in four types of operations: to define the demand centers included in each local community; to determine the shortest path matrix for all pairs of demand centers and facility sites; to compute data for a special type of path assignment constraints included in the optimization model; and to generate visual representations of the model solutions obtained through the optimization software.

During the same period, discrete optimization software went through enormous improvements and became useful for many planning purposes. Basic facility location models are now relatively easy to solve. However, more complex model variants, e.g. considering facility capacities, are much harder. Several specialized techniques are available for solving them, either exact (provably giving an optimal solution) or heuristic (giving good solutions relatively quickly but with no guarantee of optimality). The progress of discrete optimization software systems in the last two decades has been remarkable and they are now capable of solving large-scale problem instances arising in practical applications (Atamturk and Savelsbergh, 2005; Bixby et al., 2000).

Facility planning has a natural geographic component that is perfectly matched with the capabilities of a GIS for data manipulation and visualization. Furthermore, software like ArcView 3.x (ESRI, 1996b) and ArcInfo 8.x (ESRI, 2001) already includes algorithms for solving some discrete optimization models, such as shortest paths, routing and basic facility location. However, more complex location models require either the interface with dedicated modeling and solving software or the development of specialized algorithms. Church (2002) provides an excellent review on the integration of location models and GIS systems.

This article is organized as follows. First, we briefly describe the school network planning problem in Coimbra. Then, we present the optimization model addressing primary schools (for the sake of brevity, a second model, addressing secondary schools, is omitted). In the third section, we describe the interface between GIS and optimizer, emphasizing how the GIS was used to prepare model data and to generate visual representations of model solutions. Finally, we offer some conclusions on our experience with the application of discrete location models supported by a GIS.

PRESENTING THE SCHOOL NETWORK PLANNING PROBLEM

The planning problem to be solved consists of defining the location, type and size of the schools composing Coimbra's secondary and primary school networks in 2015, given the existing school network and the typology conversion required by a reorganization of the Portuguese education system. Specifically, primary education will be composed of two education levels, offered at two types of school – EB1 (for the first level only) and EB12 (for both levels). Secondary education will be composed of two levels, offered at a single type of school – ES12.

The problem includes decisions of closing existing schools and, possibly, building new schools. Indeed, despite the existing excess capacity, it can be advantageous to build new schools either to adjust the location of schools to recent housing developments or to replace existing small schools by larger ones, provided with better equipment (laboratories, libraries, sports buildings, etc.). A solution to the problem should meet a set of constraints prescribed by the guidelines of the Portuguese Ministry of Education (MinEdu, 2000) for redeploying the school network. These constraints include maximum travel distance of students to schools and maximum and minimum numbers of students per classroom and classrooms per school.

Three objectives are pursued by education authorities. First, all the population should be covered by either public schools or subsidized private schools (that is, schools located in areas not covered by public schools, where students do not pay tuition fees). Second, the accessibility of students to schools should be maximized. Third, the changes to the existing network should be minimized, either because of scarce public budgets to build new schools or to avoid public reactions against school closure (particularly from parents and teachers).

In Coimbra, existing public schools will be converted to the new typology according to the following rules. Current secondary schools are converted to ES12 (offering six instead of the current three years of education); current large primary schools (24 classrooms or more) are converted to EB12 (offering six instead of the current five years); current primary schools EB1 remain dedicated to the first level of primary education, comprising four years of education, if they have four or more classrooms. Smaller EB1 schools (about two thirds of existing schools, concentrating 40% of total EB1 capacity), which do not offer adequate pedagogic conditions, will either be closed or converted to kindergartens, to expand coverage of pre-school education.

Coimbra's existing school network is composed of 106 public schools and 19 private schools. Accompanying a national trend, the number of students in Coimbra decreased in the last decade. For instance, in the period 1998/99-2003/04, enrollments in primary and secondary education have decreased by 17% and 32%, respectively. As a consequence, current aggregate occupation is less than 80% for primary schools, and less than 60% for secondary schools. Even though the proportion of 0-4 years old population is now stabilizing, the number of students in primary and secondary education in the year 2015 is still expected to decrease by 11% and 18% relative to the values of 2003/04 (Teixeira et al., 2005). As a consequence, the aggregate occupation in 2015 of the current schools converted to the new typology will be around 70% (Table 1). That is, existing capacity will remain excessive in the future, even though some schools are not considered in this analysis (small EB1 schools and private schools where students pay tuition fees).

Table 1:Occupation of Existing Schools in 2015

Type of education	Type of school	Number of schools			Capacity (students)	Number of students (2015)	Occupation rate
		Public	Private	Total			
Primary	EB1	28	0	28	13875	9446	68%
	EB12	9	5	14			
Secondary	ES12	7	5	12	12750	9271	73%

MODELING AND SOLVING WITH THE OPTIMIZER

In this section, we present the discrete facility location model developed for the school planning problem. For the sake of brevity, in this article we will focus on primary schools (for more detailed expositions, see Teixeira and Antunes, 2005, and Teixeira et al., 2005, respectively focusing on primary and secondary schools).

Applications of facility location models are numerous, both in the private sector (location of factories or warehouses in production-distribution systems, location of concentrators on telecommunication networks, etc.) and in the public sector (location of hospitals, schools, solid waste disposal sites, etc.). The book by Daskin (1995) provides an introduction to discrete location models. Current *et al.* (2002) give a recent review of facility planning models. Representative examples of school network planning are reported by Antunes and Peeters (2001) and Pizzolato *et al.* (2004).

The usual setting of discrete facility location models is the following. Demand for the services provided by the facilities is measured in number of users (e.g., students) and is assumed to be concentrated in points named centers, which may represent regions, municipalities, towns or neighborhoods. Supply of facilities (e.g. schools) is assumed to be possible at specified points, named sites, which represent either one of the above geographical entities or specific plots of land. Centers and sites are connected by a transportation network.

The main ingredients of the model are an accessibility maximization objective, multiple demand and facility levels, maximum and minimum facility capacities, and center-to-facility assignment constraints. Multiple levels are considered in a nested hierarchy: a facility level can serve lower or equal level demands, e.g. level-1 facilities are the EB1 schools (serving the first level of primary education) and level-2 facilities are the EB12 schools (serving both levels of primary education). The purpose of minimum capacities serves both to ensure economically feasible operation of the schools that remain open and to prevent significant capacity unbalances across schools.

The purpose of assignment constraints is to prevent solutions with undesirable spatial patterns, such as splitting the demand of a given center among several facilities or assigning a centre to a facility much further than the closest (Teixeira and Antunes, 2005). Specifically, they are of three types: single assignment (all users in each center must be assigned to the same facility); closest assignment (each center must be assigned to the closest open facility); path assignment (all centers crossed in the travel path from a given center to a given facility must also be assigned to that facility).

For formulating the model, consider the following notation. (i) Sets: I is the set of centers, J is the set of sites, and S is the set of demand (and facility) levels. (ii) Decision variables: x_{ijs} is the fraction of the level-s demand of centre i satisfied by a facility at site j; $y_{js} = 1$ if a level-s facility is installed at j, and equals zero otherwise; z_{jst} is the capacity occupied with demand level s of a level-t facility located at j. The formulation is:

(H):

$$\text{Minimize} \quad \sum_{i \in I} \sum_{j \in J} \sum_{s \in S} d_{ij} u_{is} x_{ijs} \tag{1}$$

$$\text{Subject to:} \quad \sum_{j \in J} x_{ijs} = 1, \ \forall i \in I, s \in S \tag{2}$$

$$x_{ijs} \leq \sum_{t \in S | t \geq s} y_{jt}, \ \forall i \in I, j \in J, s \in S \tag{3}$$

$$\sum_{t \in S, t \geq s} z_{jst} = \sum_{i \in I} u_{is} x_{ijs}, \ \forall j \in J, s \in S \tag{4}$$

$$\sum_{s \in S, s \leq t} z_{jst} \geq b_{jt} y_{jt}, \ \forall j \in J, t \in S \tag{5}$$

$$\sum_{s \in S, s \leq t} z_{jst} \leq B_{jt} y_{jt}, \ \forall j \in J, t \in S \tag{6}$$

$$\sum_{k \in J | d_{ik} \leq d_{ij}} x_{iks} \geq y_{jt}, \ \forall i \in I, j \in J, s \in S, \ \forall t \in S | t \geq s \tag{7}$$

$$\sum_{k \in P_{ij}} x_{iks} \geq \left| P_{ij} \right| \cdot x_{ijs}, \ \forall i \in I, j \in J, s \in S \tag{8}$$

$$\sum_{j \in J \setminus J_s^0} y_{js} \leq p_s, \ \forall s \in S \tag{9}$$

$$\sum_{j \in J_s^0} y_{js} \geq \left| J_s^0 \right| - q_s, \ \forall s \in S \tag{10}$$

$$x_{ijs} = 0, \ \forall i \in I, j \in J, s \in S \,|\, d_{ij} > D_s \tag{11}$$

$$x_{ijs} \in \{0,1\}, \ y_{js} \in \{0,1\}, \ z_{jst} \geq 0, \ \forall i \in I, j \in J, s \in S, t \in S \tag{12}$$

The objective (1) of this mixed-integer optimization model is to minimize the total travel distance, where u_{is} is the level-s demand of centre i and d_{ij} is the distance between center i and site j. Constraints (2) state that all demand must be satisfied. Constraints (3) define the facility hierarchy rules: a demand level can only be satisfied by a facility of equal or higher level. Constraints (4) define the capacity variables z_{jst}. Expressions (5) and (6) are the capacity constraints, where B_{js} and b_{js} are the maximum and minimum capacities of a level-s facility at site j. Note that capacity is shared by all the demand levels served from a facility.

In addition, note that the co-location of facilities of different levels at the same site is allowed by constraints (3) and (4), which may be advantageous to satisfy maximum capacity constraints.

Expressions (7) are closest assignment constraints, stated separately per demand level, that is, each demand level of each center must be assigned to the closest facility of equal or higher level. Expressions (8) are path assignment constraints, stated separately per demand level, and work as follows. Given P_{ij}, the set of centers crossed by the travel path from center i to site j, if i is assigned to j, then all centers in P_{ij} must also be assigned to j. In addition, single assignment is imposed in constraints (12), by defining variables x_{ijs} as binary. Note that, in this formulation, closest and path assignment constraints are present simultaneously. However, they can be used as alternatives, as adding these constraints comes at the cost of reducing solution alternatives. Indeed, with the data set of Coimbra's existing primary school network, closest assignment constraints (7) were too restrictive and no feasible solutions could be found. Thus, they were removed and only path assignment constraints (8) were used. These are less "rigid" while still eliminating undesirable assignment patterns.

Constraints (9) and (10) limit the number of new facilities to open and existing facilities to close, where J_s^0 is the set of sites with existing level-s facilities, p_s is the maximum number of new level-s facilities to open, and q_s is the maximum number of existing level-s facilities to close. Constraints (11) limit the user-to-facility travel distance for each demand level s to a maximum of D_s. Finally, constraints (12) define the decision variables.

The model was implemented in the language XPRESS Mosel 1.4 (Dash, 2004) and solved with XPRESS MIP Optimizer version 15 (Dash, 2005), running under Windows XP on a computer with a Pentium-M 1.3 GHz CPU and 512 MB of memory. The instance used for Coimbra's primary school network had 68 centers and sites (coincident) and two facility levels (EB1 and EB12). The resulting size of model (H), excluding the closest assignment constraints (7), is 9588 variables (of which 9384 are binary), and 19044 constraints (not counting the simple constraints (11)). The model was easily solved in under 3 minutes, as, in spite of its complexity, relatively few changes were allowed to the current school network (controlled by parameters p and q).

APPLYING THE MODEL WITH THE GIS

The study reported in this article relied heavily on a Geographic Information System (GIS) for storing information, preparing data and analyzing results. Specifically, the GIS was applied in four types of operations: (i) to define the demand centers included in each local community; (ii) to compute the shortest paths matrix for all pairs of demand centers and facility sites (i.e., sites where schools were or could be located); (iii) to compute data for path assignment constraints; (iv) to generate visual representations of the model solutions.

For this purpose, we built a prototype system based on Arcview GIS version 3.2 (ESRI, 1996a). Operations (ii)-(iv) were accomplished through scripts written in Avenue, Arcview's programming language. We used this version of Arcview as these scripts require the Network

Analyst extension (ESRI, 1996b), which became available for the more recent ArcGIS 9.1 only in June 2005 (ESRI, 2005). The steps for applying the model in practice using the GIS and the optimizer are shown in Figure 1. Typically, the model is solved using different model parameters (e.g. number of schools to open), and possibly different data (e.g. school capacities), for analyzing different scenarios. In the remainder of this section, we describe the individual operations carried out with the GIS.

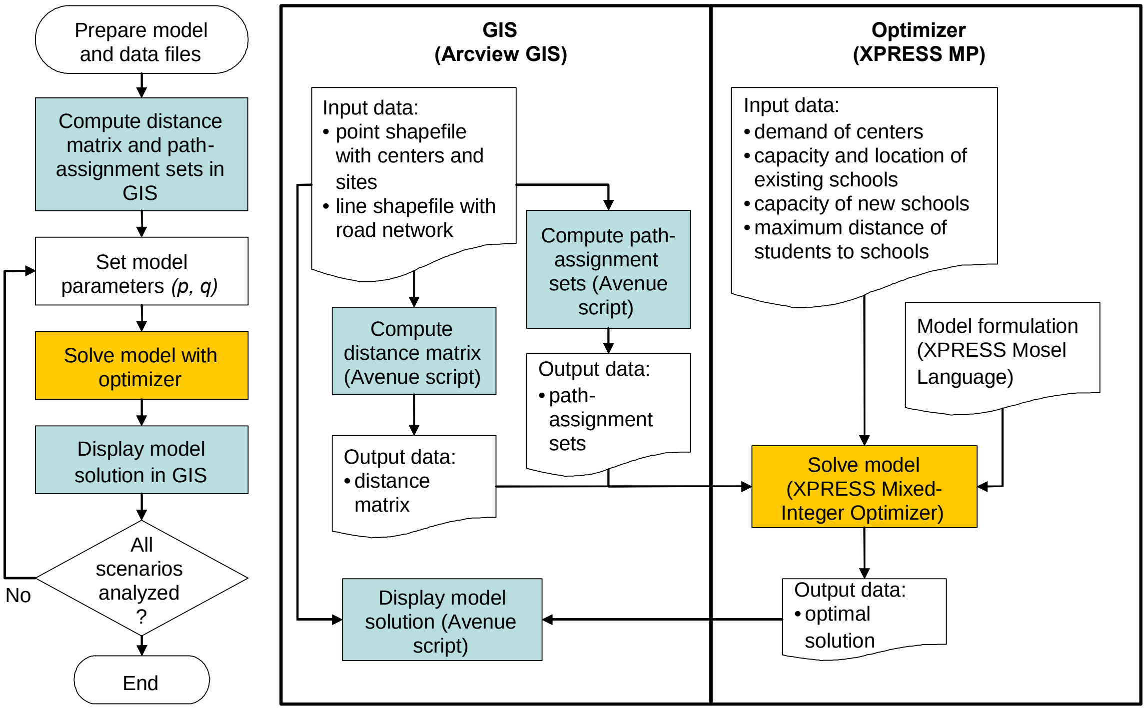


Figure 1: Applying the Model with the GIS and the Optimizer

Geographic data available for the municipality of Coimbra included community boundaries and population centers, census tracts, the road network, and existing school locations. Demand centers were defined through appropriate aggregation of census tracts taking into account geographic issues (e.g. topography), zoning regulations and existing school locations. An aggregation in 68 centers was used for primary schools (Figure 2), and sites were assumed to be coincident with centers.

The distance matrix consisting of all shortest path lengths between centers and sites was computed over the main road network projected for 2015 (Figure 2), comprising around 1,400 links and 1,000 intersections. The computation makes use of the Network Analyst extension, specifically the “FindClosestFac” function was used to compute distances between all centers and all nodes with a single call. Although computing all shortest paths on a graph is an easily solvable problem, for which there are efficient algorithms, using a GIS is still convenient and saves burdensome work. For instance, it is not necessary to export the road network, which must include nodes for all road intersections, and to track the correspondence between network nodes and centers or sites. For the purposes of our study, given the relatively large aggregation level used and the long time horizon of the planning problem, travel was assumed to be made only through main roads at a constant average speed. However, the Network Analyst of Arcview can model networks in a more detailed manner, including one-way streets, prohibited turns, variable link speeds, etc. This functionality has been further expanded in ArcGIS 9.1.

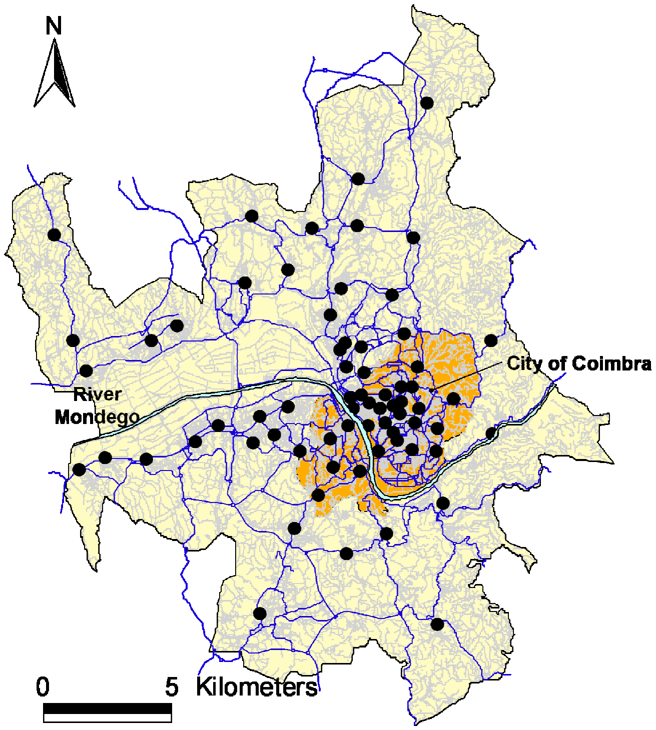


Figure 2: Municipality of Coimbra: Demand Centers and Main Road Network

The computation of data for path assignment constraints is carried out in two main steps (recall that the path-assignment set P_{ij} for a given center *i* and site *j* contains all centers "near" the travel path between *i* and *j*). First, buffers around each center were created with a radius of half the distance to the nearest neighbor, truncated to a maximum of 1 km, measured along the road network. This operation was carried out with the "FindServiceArea" function of Network Analyst, returning polygons with an irregular shape (Figure 3). Second, all centers whose buffer is intersected by the shortest path from *i* to *j*, using the road network, are added to set P_{ij}. In the example of Figure 3, $P_{38,8} = \{38,36,8\}$, which means that if center 38 is assigned to school 8, then center 36 must be assigned to the same school.

Finally, mapping model solutions largely facilitates the diagnosis of model errors and the interpretation of model results. Within the prototype system, solutions are exported from the optimizer as text files and read by a script in the GIS, where they are displayed as school locations and center assignments. In addition, all solution data (such as school occupation) can easily be displayed and inspected in tabular form (Figure 4).

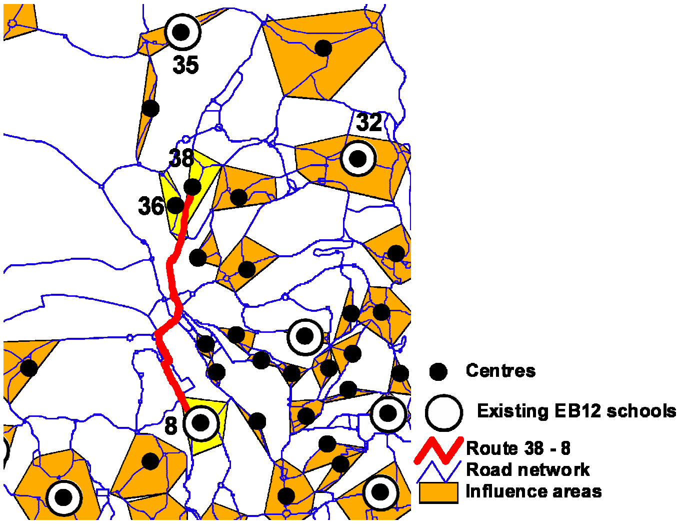
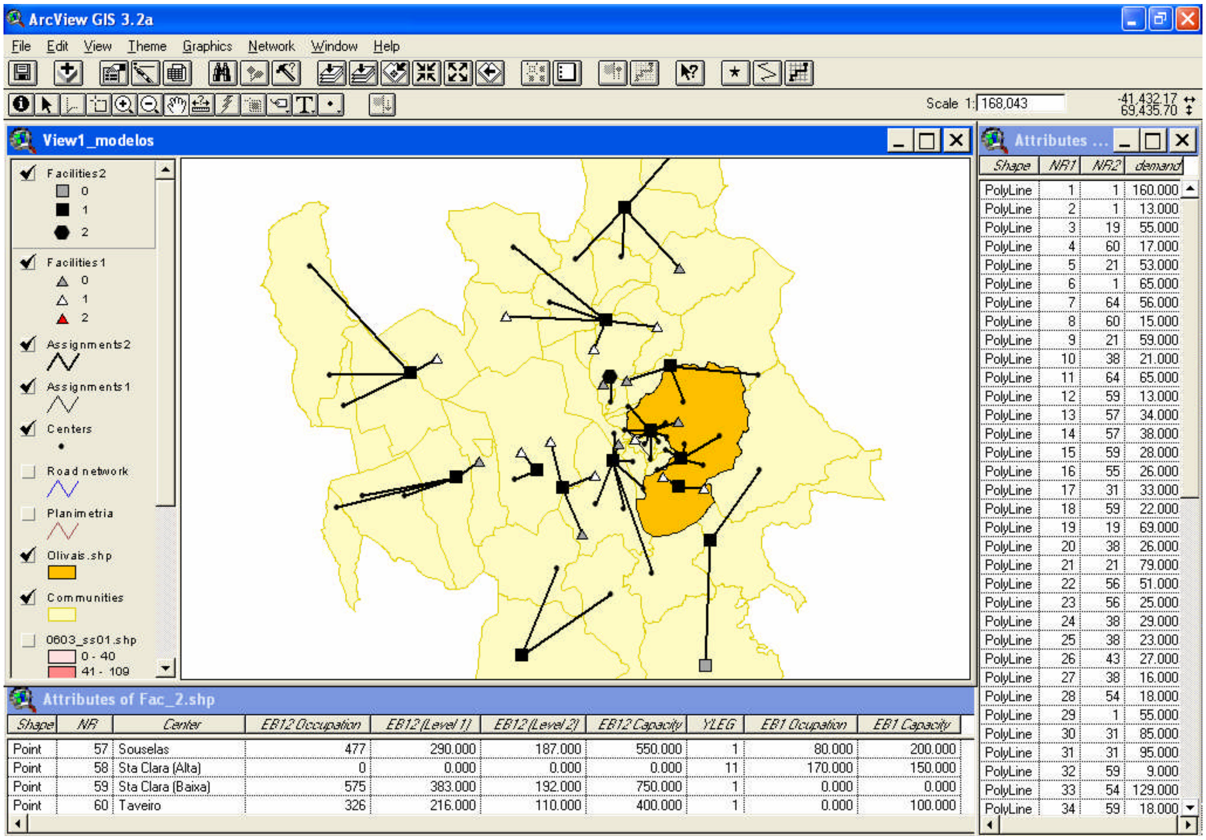


Figure 3: Computation of Path Assignment Data



Shape	NR	Center	EB12 Occupation	EB12 (Level 1)	EB12 (Level 2)	EB12 Capacity	YLEG	EB1 Ocupation	EB1 Capacity
Point	57	Souselas	477	290.000	187.000	550.000	1	80.000	200.000
Point	58	Sta Clara (Alta)	0	0.000	0.000	0.000	11	170.000	150.000
Point	59	Sta Clara (Baixa)	575	383.000	192.000	750.000	1	0.000	0.000
Point	60	Taveiro	326	216.000	110.000	400.000	1	0.000	100.000

Shape	NR1	NR2	demand
PolyLine	1	1	160.000
PolyLine	2	1	13.000
PolyLine	3	19	55.000
PolyLine	4	60	17.000
PolyLine	5	21	53.000
PolyLine	6	1	65.000
PolyLine	7	64	56.000
PolyLine	8	60	15.000
PolyLine	9	21	59.000
PolyLine	10	38	21.000
PolyLine	11	64	65.000
PolyLine	12	59	13.000
PolyLine	13	57	34.000
PolyLine	14	57	38.000
PolyLine	15	59	28.000
PolyLine	16	55	26.000
PolyLine	17	31	33.000
PolyLine	18	59	22.000
PolyLine	19	19	69.000
PolyLine	20	38	26.000
PolyLine	21	21	79.000
PolyLine	22	56	51.000
PolyLine	23	56	25.000
PolyLine	24	38	29.000
PolyLine	25	38	23.000
PolyLine	26	43	27.000
PolyLine	27	38	16.000
PolyLine	28	54	18.000
PolyLine	29	1	55.000
PolyLine	30	31	85.000
PolyLine	31	31	95.000
PolyLine	32	59	9.000
PolyLine	33	54	129.000
PolyLine	34	59	18.000

Figure 4: Visual Representation of a Solution

CONCLUSION

In this article we showed how we coupled a GIS and a discrete optimization program to solve a school network planning problem. Modern optimizers are capable of solving relatively large, complex models suitable for real-world applications that could only be solved with heuristics ten years ago. The main ingredients of the model used to represent the problem are an accessibility maximization objective, maximum and minimum school capacities and centre-to-facility assignment constraints. The latter include single sourcing, closest assignment and path assignment, and their purpose is to prevent solutions with undesirable spatial patterns.

The use of the GIS in conjunction with the optimizer was extremely valuable. In addition to simplify the analysis of results, it was used for computing shortest paths and data for path assignment constraints. The prototype system we built for this purpose could be developed into a stand-alone application, for instance by embedding the model and optimizer in the user interface of the GIS, by using the XPRESS-MP Builder Component Library (Dash, 2005b).

ACKNOWLEDGMENTS

The authors thankfully acknowledge support received from Fundação para a Ciência e a Tecnologia through grants POCTI/AUR/32599/2000 and SFRH/BD/12672/2003.

REFERENCES

Antunes, A. and Peeters, D. (2001). "On Solving Complex Multi-Period Location Models Using Simulated Annealing". *European Journal of Operational Research* 130 (1) 190-201.

Atamturk, A. and Savelsbergh, M.W.P. (2005). "Integer-Programming Software Systems." *Annals of Operations Research* 140 (1) 67-124.

Bixby, R. E., Fenelon, M., Gu, Z., Rothberg, E., and Wunderling, R. (2000). "MIP: Theory and practice - closing the gap." In Powell M. J. D. and Scholtes, S. (eds). *System Modelling and Optimization: Methods, Theory, and Applications*. Kluwer Academic Publishers, 19-49.

Brail. R. K., Klosterman, R. E. (Ed.) (2001). *Planning Support Systems: Integrating Geographic Information Systems, Models, and Visualization Tools.* ESRI Press, Redlands, CA.

Church, R. (2002). "Geographical information systems and location science". *Computers & Operations Research* 29 (6) 541-562.

Current, J., Daskin, M., and Schilling, D. (2002). "Discrete network location models." In Drezner, Z. and Hamacher, H. (eds.). *Facility Location: Applications and Theory*. Springer-Verlag, Berlin, Germany, 81-118.

Dash Optimization, Ltd (2004). *XPRESS-Mosel Language Reference Manual - Release 1.4*. Blisworth, UK.

Dash Optimization, Ltd (2005a). *XPRESS-Optimizer Reference Manual - Release 15*. Blisworth, UK.

Dash Optimization, Ltd (2005b). *Xpress-BCL Builder Component Library - Release 2.6.* Blisworth, UK.

Daskin, M. S. (1995). *Network and discrete location: Models, algorithms, and applications*. Wiley Interscience, New York, USA.

ESRI – Environmental Systems Research Institute, Inc. (1996a). *Using ArcView GIS*. Redlands, CA.

ESRI – Environmental Systems Research Institute, Inc. (1996b). *Using ArcView Network Analyst*. Redlands, CA.

ESRI – Environmental Systems Research Institute, Inc. (2001). *ArcInfo Version 8.1* [Computer Software].

ESRI – Environmental Systems Research Institute, Inc. (2005). *ESRI Software 9.1 Library Tutorials - ArcGIS Desktop Extensions* [CD].

Longley, P. A, Goodchild, M. F., Maguire, D. J. and Rhind D. W. (2001). *Geographic Information Systems and Science*. Wiley, Chichester, UK.

MinEdu – Ministério da Educação (2000). *Critérios de Reordenamento da Rede Educativa (Criteria for the Redeployment of School Networks)*. Lisbon, Portugal.

Pizzolato, N. D., Barcelos, F. B. and Lorena, L. A. N. (2004). “School location methodology in urban areas of developing countries.” *International Transactions in Operational Research* 11, 667-681.

Teixeira, J. and Antunes, A. (2005). “An hierarchical location model for school network planning”. Working paper, Department of Civil Engineering, University of Coimbra, Portugal.

Teixeira, J., Antunes, A., and Peeters, D. (2005). “An optimization-based study on the redeployment of a secondary school network”. Working paper, Department of Civil Engineering, University of Coimbra, Portugal.