Revista Brasileira de Cartografia (2013) N⁰ 65/6: 1189-1198
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto
ISSN: 1808-0936

EVALUATION OF THE CHANGES IN THE SPATIAL DISTRIBUTION OF COMPANIES IN OSASCO, (SAO PAULO - BRAZIL) BY A DATA ENVELOPMENT ANALYSIS (DEA)

Avaliação das Mudanças na Distribuição Especial de Empresas em Osasco (são Paulo-Brasil) Utilizando a Técnica de "Data Envelopment Analysis (DEA)"

George Eduardo Serra Almeida[1]; Linda Lee Ho[2];
Mariana Abrantes Giannotti[3] & José Alberto Quintanilha[3]

[1]GEOSERT - Consultorias em Geotecnologias e Geomarketing
Rua Euclides Miragaia 620 Centro Cep:12245820 São José dos Campos - SP, Brasil.
grgserra@gmail.com

[2]Escola Politécnica da Universidade de São Paulo – EPUSP
Departamento de Engenharia de Produção
Avenida Professor Luciano Gualberto, no. 530, Butantã - São Paulo - SP, Brasil. 05508-010
lindaleeho@usp.br

[3]Escola Politécnica da Universidade de São Paulo - EPUSP
Departamento de Engenharia de Transportes – Laboratório de Geoprocessamento
Av. Prof. Almeida Prado, Travessa 2, no. 83 Butantã - São Paulo - SP, Brasil. 05508-070
{mariana.giannotti, jaquinta}@usp.br

Recebido em 11 de Janeiro, 2013/ Aceito em 03 de Maio, 2013
Received on January 11, 2013/ Accepted on May 03, 2013

ABSTRACT

This paper presents the results of GIS application related to the spatial distribution of companies in a Brazilian city of a metropolitan region. To compose urban scenarios in different time periods, cadastral and socioeconomic data are collected and integrated in Data Envelopment Analyze (DEA). An installation index was obtained using the input and output variables. The input variables were: the number of companies in 1998 and 2003 and the average distance from companies to the highway accesses (which can be viewed as an advantage of the companies' location). The output variables were: population and average income (from the national census in 1991 and 2000) and average price of property (from the municipality files in 1994 and 2005). Results were displayed by a geographical information system (GIS). Comparative analysis reveals that the companies have moved from the central area to sites near access to highways and the land use of the study region is different from what was stated by regulated laws.

Keywords: GIS, DEA, Urban Planning, Installation Index, Spatial Analysis.

RESUMO

Este artigo apresenta uma aplicação de sistemas de informações geográficas relacionada à distribuição espacial de empresas, na região metropolitana de uma cidade brasileira. Dados cadastrais e econômicos foram usados e integrados via um modelo de Data Envelopment Analyze (DEA), para compor cenários urbanos em duas épocas diferentes.

Um índice de instalação foi criado utilizando variáveis de entrada e de saída no DEA. As variáveis de entrada foram: o número de empresas instaladas no município em 1998 e em 2003 (antes e depois da construção do trecho oeste do Rodoanel de São Paulo) e a distância media dessas empresas ao acesso ao Rodoanel (visto como uma vantagem para as empresas se localizarem). As variáveis de saída foram: a população e renda média (obtidas dos censos nacionais de 1991 e 2000, respectivamente) e o preço médio das propriedades (a partir de arquivos municipais de 1994 e 2005). Os resultados da aplicação do modelo são mostrados via um sistema de informações geográficas (SIG). Análises comparativas revelam que companhias se mudaram da área central do município para o entorno do Rodoanel e que a ocupação da região difere daquela prevista na legislação.

Palavras chaves: SIG, DEA, Planejamento Urbano, Índice de Instalação, Análise Espacial.

1. INTRODUCTION

The Metropolitan Region of São Paulo (Brazil) is composed of the Capital of São Paulo state (also named São Paulo) and others 38 municipal districts. It presents a dynamism in population growth and industrial and urban development. This fact has started in the 40's and 50's and nowadays this region represents the Brazilian financial, economical and cultural center (see Grostein, 2001).

The city named Osasco is one of the municipal components of this metropolitan region. It is the fifth largest city of the state and ninth municipal district in Gross Domestic Product - GDP *per capita*. According to IBGE – Instituto Nacional de Geografia e Estatística, the Brazilian Bureau of Census and Statistics (http://www.ibge.gov.br/english/presidencia/noticias), in 2004, "*Osasco (SP) and Fortaleza (CE) were the municipalities with the biggest gain in terms of participation in services; the activity of public administration accounted for over 70% of the GDP in 39 municipalities*". Its urban and economic development was favored by the proximity to São Paulo's Municipal district and by a good existing transportation infrastructure.

The objective of this study is to present an analysis of the dynamic urban management of Osasco municipality using a Geographical Information System (GIS), through exploratory spatial analysis, and an index that measures the settlement of the companies (associated with the urban zones) in this municipality.

The employment of spatial exploratory analysis tools from a GIS with Data Envelopment Analysis (DEA) allowed a comparison of the quantities of companies installed in this municipality before and after the construction of the western section of the highway named Rodoanel "Mário Covas" (known as Rodoanel, that means "beltway"). The construction of this section started in 1998 and finished in 2002. These changes were identified and marked by aerial photos and satellite images (from 1991 to 1998, before the construction and from 2002 to 2006, after the conclusion). These analyses permitted modeling the existing potential market and will consequently support the urban planning and infrastructure of this region.

In previous works, GIS and DEA were used associated with accessibility indices as in Chang and Lee (2008) and Martín and Reggiani (2007). In the first case they dealt with an accessibility analysis of Korean high-speed rail, a systemized accessibility analysis with a case study of the Seoul metropolitan area was made. The authors used a reduced form of a Hansen-type accessibility measure, an analysis of variance (ANOVA) test and a GIS-based mapping audit were used as tools for the assessment. At the end, zones of opportunity that could yield the greatest demand increase of high-speed rail are identified and some metropolitan railway expansions for improving region accessibility were recommended.

In Martín and Reggiani (2007), the authors conducted a research where they analyzed and compared synthetic indices of accessibility in order to measure the impact of high-speed trains on European cities. In particular, it considers DEA and PCA (principal component analysis), as methodologies, yield a consistent ranking for global accessibility of European cities for three different scenarios: 1996, 2005 and 2015.

They referred to the data and results concerning two previous articles by Gutiérrez et al. (1996) and Martín et al. (2004 and 2007) that analyzed the accessibility gains produced by the construction of Trans-European Transport Networks (TEN-Ts). A geographic information system (ARC/INFO) was used to calculate three different partial accessibility indicators: location, potential market and daily accessibility, for the three different scenarios mentioned.

Martín et al. (2004) applied DEA methodology to analyze gains in the accessibility of Spanish territory produced by the new high-speed train (HST) Madrid–Barcelona–French border. A geographic information system was used to calculate four complementary partial accessibility indicators: an index of location, economic potential, relative efficiency of the network and daily accessibility. Based on DEA, the authors obtain a composite accessibility index that produces a better understanding of the overall accessibility changes produced by the new infrastructure. The results of the model will be used to extract some policy considerations with respect to the (polarizing/balancing) effects of the line within the Spanish territory.

Ashby and Longley(2005), compare Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA), to generate robust efficiency scores as a tool in which police forces might use geodemographics analysis to better deploy resources at a variety of spatial scales in England and Wales.

Klimberg and Ratick (2008), develop and test location modeling formulations that utilize data envelopment analysis (DEA) efficiency measures to find optimal and efficient facility location/allocation patterns. They believe that solving for the DEA efficiency measure, simultaneously with other location modeling objectives, provides a promising rich approach to multiobjective location problems. This is accomplished by first formulating the simultaneous DEA linear program, and then combining that formulation, in a multiobjective framework, with both the uncapacitated and capacitated fixed charge facility location problem. In this way we have been able to provide to decision makers a way to measure the interactions between different location patterns and the effects that the site attributes of those chosen locations may have on facility performance. The results obtained using the multiobjective formulations demonstrate that the trade-offs between site and facility efficiencies may in some cases be substantial. The ability to use location models to test these trade-offs provides a promising new rich approach for multiobjective location analysis.

Cullinane et al. (2006) compare DEA against Stochastic Frontier Analysis (SFA) measuring the efficiency of the container port industry for the world's largest container ports and compares the results obtained. A high degree of correlation is found between the efficiency estimates derived from all the models applied, suggesting that results are relatively robust to the DEA models applied or the distributional assumptions under SFA. High levels of technical efficiency are associated with scale, greater private-sector participation and with transhipment as opposed to gateway ports. In analysing the implications of the results for management and policy makers, a number of shortcomings of applying a cross-sectional approach to an industry characterized by significant, lumpy and risky investments are identified and the potential benefits of a dynamic analysis, based on panel data, are enumerated.

The content of this paper is organized as follows: the methodology is described in section 2. In this section the reader will find the description of the region, the database, how it is organized since data of this region are collected from different sources and stored in different format and extensions. A brief description of DEA is also in this section. The results are the subject of the next section and the paper closes with a conclusion section.

2. METHODOLOGY

Three subsections with different subjects are the components of this section. In the first subsection, the region study is briefly described. In the second subsection the components of database used to make the comparative study is presented. In the third section a short introduction of DEA is made and a general vision of methodology is presented in order to facilitate future replication of this work.

2.1 Region Study

The municipality Osasco is crossed by important highways as Castello Branco, Raposo Tavares and Anhanguera (Figure 1). It is located in a privileged region with an extension of 70 Km^2 of urban area when compared with other neighbor municipalities. It is the fifth largest city with a population of 686,872 inhabitants (a demographic growth rate of 1.29% per year in 1996-2000 according SEADE, 2000); and it presents the ninth municipal district in GDP per capita (SEADE, 2003). These rates made this municipality one of the cities with higher development index of the São Paulo State, Brazil.

Figure 1 shows the localization of the municipality and the section of the highway Rodoanel.

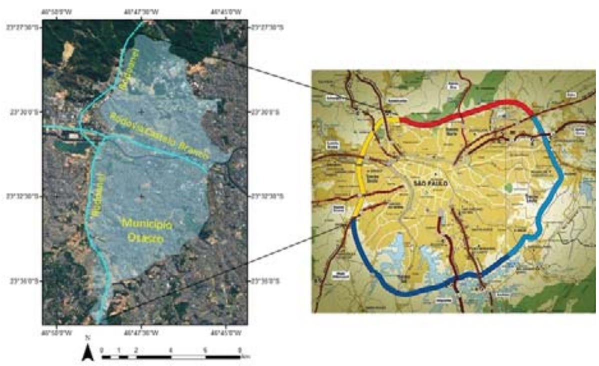


Fig. 1: Study region: Osasco municipal district and its location related to highway Rodoanel Mário Covas – SP/Brazil.

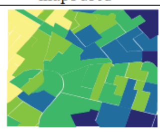
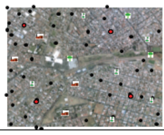
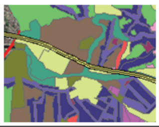
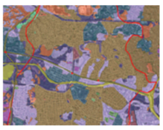
2.2 Database

To make comparison of the quantities of companies installed before and after the construction of the highway Rodoanel, sets of geographic and social-economical information were collected in two distinct periods (1993 and 2006) with the aim to understand the possible business transformation in this region due to the construction of the highway Rodoanel.

The initial stage consisted of the organization of the geographical database and turned compatible the different formats of spatial data and tables. The information of several public and private organizations was stored in different software formats and extensions. For example, as the vector data were CAD (Autodesk), Shapefile and Coverage (ESRI); in tables were Access and Excel (Microsoft); in the image formats were IMG (ERDAS/Leica), Grid (ESRI) and JPeg, moreover data in analogical format. These data had additionally different cartographic projection systems with also different scales and precisions.

The format of the software ArcGIS/ESRI denominated Geodatabase was chosen as a reference file, which allowed the storage of geographical information and tables in an RDBMS, as well as the use of ArcGIS/ArcInfo for the creation of structures for base maintenance and spatial integrity, like: spatial relationships, tables junction, geocoding and topology. The list of the variables used in this study is summarized in Table 1.

Table 1: Considered variables.

Data	Ilustrative figures of the maps used	Variables
Census blocks - polygon 1991, 1996 and 2000 Source: IBGE -		• Number of domiciles • Verticalization index • Population • Average income • population density
Map of real estate values - point 1994 and 2004 Source: 's Municipal District		• Terrain average price values geocoded by blocks
Geocoded companies - point 1998 and 2003 Source: RAIS-MTE		• Number of companies (in 1993 and 2003); • Distance from companies to the highway accesses
Map of urban zoning - polygon 1993 and 2006 Source: 's Municipal District		• Mapping and geographical areas of the zones in Km²
Map of Land use 2003 - polygon Source: Satellite Spot Image		• Mapping and geographical areas of land use in Km²

2.3 Data envelopment analysis (DEA)

Data Envelopment Analysis (DEA) is largely used in the operational research to quantify and compare the efficiency of production process. According to Charnes et al. (1978), the DEA models are based on aggregations of input and output variables, new virtual variables called *efficiencies* are created as a result of a linear combination of the original variables. The weights used in this combination are calculated by linear programming, so that each decision unit or DMU (decision making unit) benefits from the best combination of weights, thus maximizing their efficiency.

The present analysis evaluates the companies efficiency grade of changes location of the *n* gravity centers of each municipal zone unit in 1993 and 2006, considered as the DMUs, supposed as the economic activity centres, using the distance to this gravity centers to the nearest assess of the highway Rodoanel.

To apply DEA model in this study, the variables listed in Table 1 have to be classified as input or output variables. The inputs variables $X_{ij}\ (i = 1,...,m)$ considered are:

- the number of companies in 1998 and 2003
- the average distance from companies to the highways access (it can be viewed as an advantage of the companies' location).

and the output variables $Y_{rj}\ \ (j = 1,...,s)$ are:

- Population and Average income (from the national census in 1991 and 2000);
- Average price of terrains (from the municipality files in 1994 and 2005).

The input variable X_i and output variable Y_r are weighed respectively by (v_i, u_r) to produce a new variable h_0 denominated as efficiency. *DMU* with $h_0 = 1$ is a unit with high efficiency otherwise ($h_0 < 1$) is low efficiency or inefficiency. In this paper the new variable will be named as *installation index*.

For this study, DEA's Model named CCR (Constant Returns to Scale) was used, in which variations in the inputs produce proportional changes in the outputs (Charnes et Al., 1978). In this mathematical formulation it is considered that DMU_j, $j = 1,...,n$, is a unit of production that uses *m* inputs X_{ij}, $i = 1,\ldots,m$, to produce *s* outputs Y_{rj}, $j = 1,\ldots,s$. The model maximizes the ratio of the linear combination of outputs and the linear combination of inputs, with the restriction that for any DMU the ratio cannot be greater than 1.

Therefore, the CCR model can be described as follows:

$$\max h_0 = \sum_{r=1}^{s} u_r y_{r0}$$

$$s.a. \begin{cases} \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 \\ \sum_{i=1}^{m} v_i x_{i0} = 1; \\ -u_r \le \xi; \\ -v_i \le \xi \end{cases}$$

where:

n DMUs, each DMU_j (j=1,…,n) characterized by:

m *inputs* – vector x_{ij}

s *outputs* – vector y_{rj}

(v_i, u_r) – Weights of the input and output variables

$\sum_{r=1}^{s} u_r y_{rj}$ - Amount of virtual output from DMU_j

$\sum_{r=1}^{m} v_i x_{ij}$ - Amount of virtual input from DMU_j

$h_0 = 1$ – high efficiency

$h_0 < 1$ – low efficiency or inefficiency

The linear programming problem is solved for every municipal zone unit considered as an economic activity center in order to obtain the installation index by DEA. The installation index obtained is calculated as the inverse of the maximum proportional output installation (population, average income and average price of terrains) that be obtained for the indicated installation inputs can (the number of companies in 1998 and 2003 and the average distance from companies to the highways access).

To summarize a general vision of the methodology is illustrated in Figure 2. Basically, the map of geocoded companies was used to calculate the number of companies in each municipal zone and also the average distance to the highways access (used as input variables in DEA model). Census data and the map of real state value points were considered to compute the output variables, more precisely the population and average income, as well as average price of terrain, respectively. All these spatial data were organized and processed in the GIS database in order to make it proper for DEA modeling. The result of DEA modeling was the installation index, as explained previously, that was then compared, by spatial analysis to the land use/ cover map and the urban zone map in order to allow better interpretation of the results.

3. ANALYSIS OF RESULTS

The results can be visualized in the zone maps for 1993 and 2006 (see Figure 3), classified by the intensity of installation index. For visualization and analysis the ArcGIS spatial tools are used.

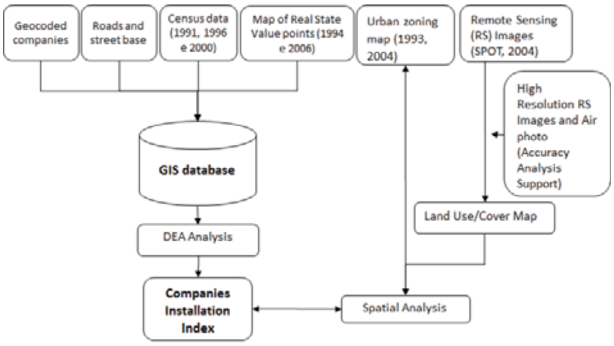


Fig. 2 - Methodology flowchart.

Dark colors represent zones with high efficiency (that is installation index equal to one) and clear colors show areas with smaller efficiency or "inefficiency" in installation of companies. The zones marked by codes 1, 2 and 3 on the map of 1993 (left-hand of Figure 3), present significant values for the index, denoting a strong effectiveness in the companies' installation for these regions. For both periods, the index represents a strong efficiency concentration in installation of companies in the middle of municipal district of Osasco, although in 1993 this efficiency was represented by a horizontal imaginary axis addressed to the proximity with São Paulo city limits.

A strong concentration of business from commerce and services sector at the centre of the city was observed for 2006, but also a larger dispersion in comparison with the previous period. A change was noted: shifts from south region and the proximities of São Paulo to the main access highways of the region, configuring thus a vertical imaginary axis of business density.

The results obtained by the installation index allowed an analysis of spatial distribution of the business by zone and its relationship with the others variables. Zones that presented larger values of installation index were related to commercial and industrial areas, although some high values have been associated with residential zones, since it refers to businesses distributed in commerce activities, services and industries. These areas can be characterized as zones of multiple activities in a municipal district urban land use plan.

Low values of the index were related to regions denominated as *green areas*. Those ones are inserted in the areas protected by laws where

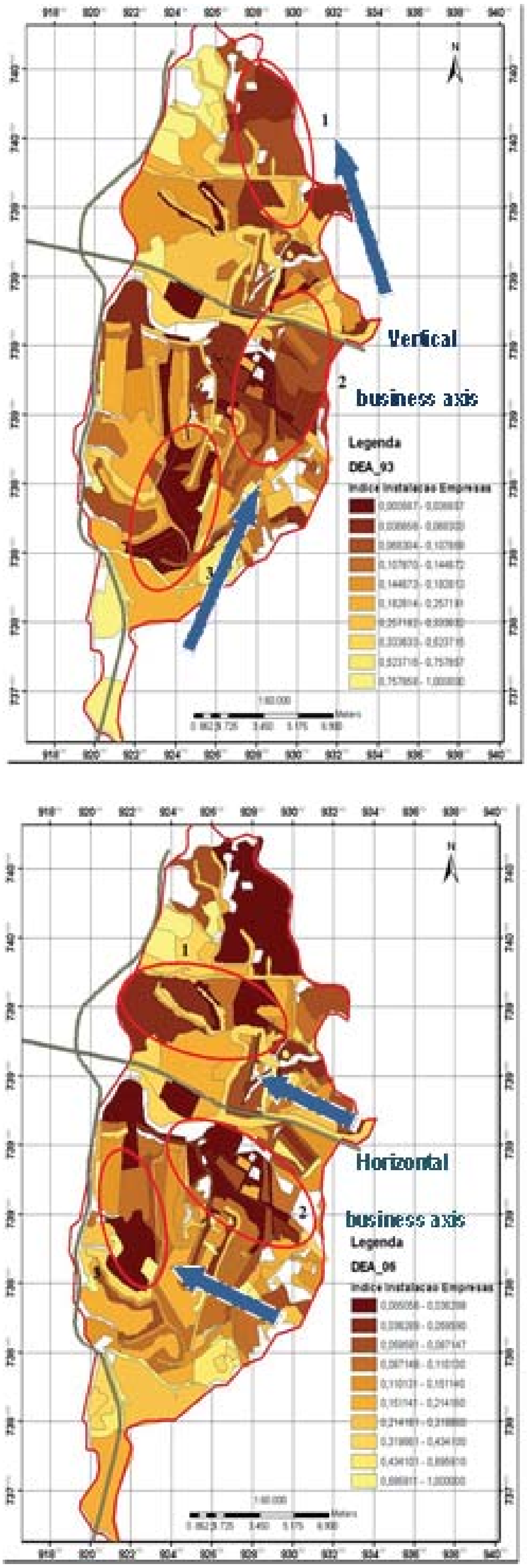


Fig. 3 - Comparison of installation index - 1993 versus 2006

installment of companies in them requires special license. So low concentration of companies in them and they are reserved for recreational use or environmental protection.

To illustrate these results, Figure 4 presents the results for 1993. The zone areas in Figure 4(a) are those classified as a high efficiency one, their profiles are: residential zones with high demographic density and near of commercial and services areas. An amount of approximately four thousand micro or small companies is installed in this area; it distances averagely of 3.5 km from the main access of highways. The zone areas in Figure 4(b) are those classified as low efficiency; it is a residential zone with five thousand domiciles and a population of 24 thousand in an area of one km^2) but it distances averagely of 800 meters from the main access. Less company (around of 30 in diversified activities as commercial and services) is installed in this region.

Similar analysis on the installation index of 2006 reveals interesting observations. It is noted changes in the values of efficiency and a rearrangement of the spatial distribution of the business in this municipality (see Figure 3). To exemplify, some regions with high and low values of efficiency and different social-economical profiles are chosen. Figure 4 illustrates these cases. Figure 5(a) is an

Fig. 4(a) - High efficiency

Fig. 4(b) - Low efficiency

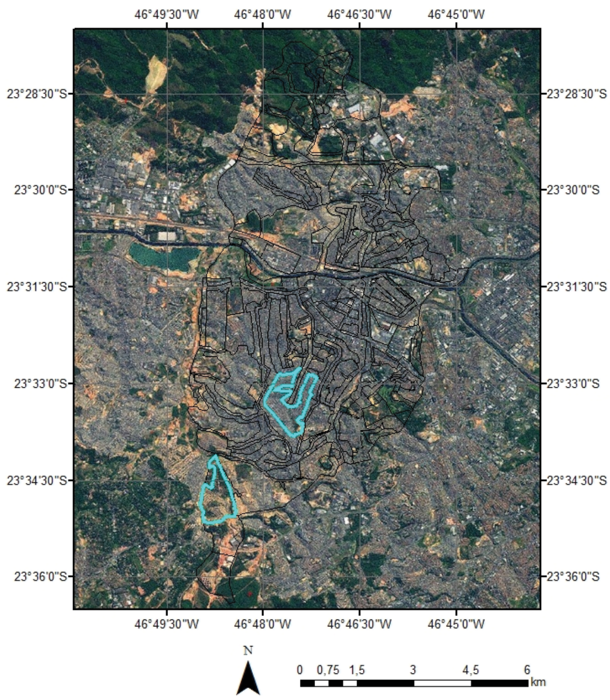


Fig. 4(c) - Total area

Fig. 4 - Areas of high (a) and low (b) efficiency, and total area (c) in 1993 according to the installation index.

example of low efficiency. It is characterized by residential zones with low demographic density and average area of 350 m^2.

The price of terrain is low, no installation of the companies, few domiciles. These areas are far from the center or the access of the highways; few infrastructure of access with some sparse streets and low degree of urbanization. Figure 5(b) illustrates a high efficiency case. It is a residential zone; a population of 35 thousand of inhabitant with an average income of $ 800.00; 9 thousand of domiciles in an area of 1.5 Km^2.

More than 2 hundred of companies are installed in this region (most them in commercial activities) and it distances averagely 3 km from the main access of the highways.

Fig. 5(a) - Low efficiency

Fig. 5(b) - High efficiency

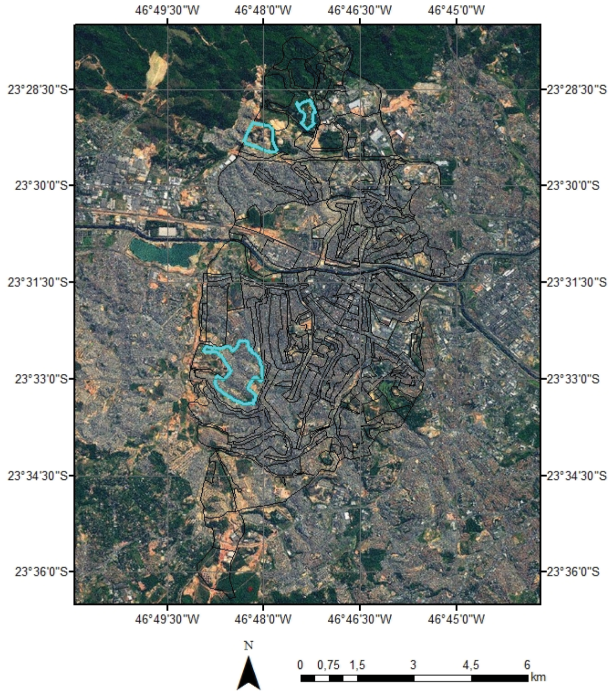


Fig. 5 - Areas of high (b) and low (a) efficiency and total area (c) in 2006 according to the installation index.

The study area concentrates strong urban and commercial activities. Some areas in this region classified as ZAV (zones with environment protection and greenery areas) deserve special attention. In Figure 6, the distribution of ZAV areas according to its land use category, obtained via satellite image classification, can be visualized. These areas (the ZAV's) have specific laws created for land use, however, they have been received multiple activity use rather than the determined and regulated by laws. So reformulation of planning actions to mitigate the occupation process, to avoid damages to the environment and the population must be taken.

4. CONCLUSIONS AND FINAL REMARKS

The aim of this study was to analyze the dynamism of the spatial distribution of the companies (industries, commerce and services) in urban land of the city of Osasco in two periods: 1993 and 2006.

For that SIG and DEA were jointly employed and the comparative analysis allowed verifying different patterns of settlement of companies. Some of them have moved from the central region to locals near of the main highways. This result is coherent of what Biderman (2001) stated. According to Biderman

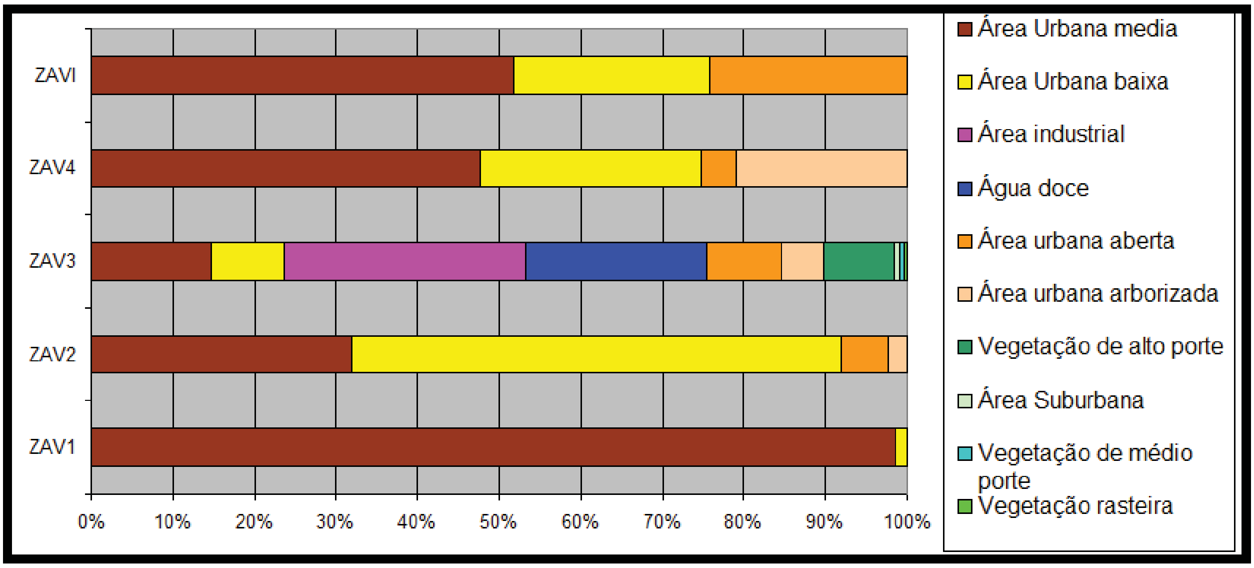


Fig. 6 - Distribution of the land use of ZAV's.

(2001) the transport cost is a reason to install (the company located in dense urban regions may save costs) or uninstall companies (high density may increase the traffic jam generating urban problems).

Moreover the map of land use reveals a reality of the land use different from was determined by the laws. So reformulation of planning actions to mitigate the occupation process, to avoid damages in environment must be taken.

ACKNOWLEDGEMENTS

The authors would like to tacknowledge to Escola Politécnica da Universidade de São Paulo - EPUSP and Conselho Nacional de Pesquisa – CNPq to support this research; to Imagem – Soluções de Inteligência Geográfica for the disposal of SPOT images; to Prefeitura de Osasco (Osasco municipality) by the files disposal; and to Dr. Anderson Laécio Galindo Trindade for his help in Matlab routines.

BIBLIOGRAPHIC REFERENCES

ASHBY, D. I.; LONGLEY, P. A. Geocomputation, Geodemographics and Resource Allocation for Local Policing. **Transactions in GIS**, 2005, 9(1): 53–72.

BIDERMAN, C. Forças de Atração e Repulsão na Grande São Paulo. (**Phd. Thesis**). 210 p. Fundação Getúlio Vargas, São Paulo, 2001 (in portuguese).

CHANG, J. S.; LEE, JANG-HO Accessibility Analysis of Korean High-speed Rail: A Case Study of the Seoul Metropolitan Area. **Transport Reviews**, Vol. 28, No. 1, 87–103, January 2008.

CHARNES, A.; COOPER, W. W.; LEWIN, A. Y.; SEIFORD, L. M. **Data Envelopment Analysis**: Theory, Methodology, and Application, Kluwer Academic Publishers, Boston, 1994.

CULLINANE, K.; WANG, T.-F.; JI, D.-W. S. P. The technical efficiency of container ports: Comparing data envelopment analysis and stochastic frontier analysis. **Transportation Research Part A: Policy and Practice**, Vol. 40, Issue 4, May 2006, Pages 354-374.

DERSA – Informações gerais sobre o Rodoanel. 2005. Disponível em: <http://www.dersa.sp.gov.br/> (in portuguese).

GROSTEIN, M. D. **Metrópole e expansão urbana:** a persistência de processos "insustentáveis". São Paulo, v.15, n.1, p.13-19, jan/mar. 2001 (in portuguese).

Fundação Sistema Estadual de Análise de Dados – SEADE. **Pesquisas Econômicas**. Disponível em: <http://www.seade.gov.br/> Acesso em: 25 Jul. 2006 (in portuguese).

GUTIÉRREZ, J., GONZÁLEZ, R. AND GÓMEZ, G. (1996) The European high-speed train network: predicted effects on accessibility patterns, **Journal of Transport Geography**, 4(4), pp. 227–238.

KLIMBERG, R. K.; RATICK, S. J. Modeling data envelopment analysis (DEA) efficient location/allocation decisions. **Computers & Operations Research** Vol. 35, Issue 2, February 2008, p. 457-474. Part Special Issue: Location Modeling Dedicated to the memory of Charles S. ReVelle.

MARTÍN, JC.; REGGIANI, A. Recent Methodological Developments to Measure Spatial Interaction: Synthetic Accessibility Indices Applied to High-speed Train Investments. **Transport Reviews**, Vol. 27, No. 5, 551–571, September 2007.

MARTÍN, J. C., GUTIÉRREZ, J. AND ROMÁN, C. (2004) Data Envelopment Analysis (DEA) index to measure the accessibility impacts of new infrastructure investments: the case of the high-speed train corridor Madrid–Barcelona–French border, **Regional Studies**, 38(6), pp. 697–712.

MARTÍN, J. C., GUTIÉRREZ, J. AND ROMÁN, C. (2007) Accessibility impacts of European TENs railway network, in: M. van Geenhuizen, A. Reggiani and P. Rietveld (Eds) **Policy Analysis for Transport Networks**, pp. 189–213 (Aldershot: Ashgate).