Big Data and Systemic Geo-Intelligence:
Innovation in optimizing scenarios for utility companies

ABSTRACT

What will utility operations be like in 3 years' time? And what about 30 years from now? How many operational bases will there be and in which locations? This research studies the influence of geographic variables and optimization models on operational KPIs of AES Eletropaulo, an energy distribution company in Brazil, in its emergency response service. Different information perspectives were used (traffic, microclimate, trees, real estate, occurrences and crew movements) on an innovative spatial discretization for computational purposes. Four algorithms (Brute Force, Hillclimbing, Genetic, K-Means) were developed to simulate base composition scenarios. The results enable assessments to be carried out that support better decisions for utility companies. This is, therefore, a Big Data platform with geographical optimization that is unprecedented in the Brazilian electricity sector.

Keywords: Big data; Geo-intelligence; Utility companies.

1. INTRODUCTION

This work, the result of ANEEL R&D Project 0390-1076/2014 – "A Study for Proposing and Analyzing Composition Scenarios for Operational Bases by way of a Systemic Geo-Intelligence Platform" looks at the influence of geographic variables on the key performance indicators (KPIs) of the operating area of AES Eletropaulo when answering emergency calls in the region covered by its energy distribution concession in the São Paulo Metropolitan Region. The focus of the study is to identify the influence of the location of the operational bases in the emergency response operation and to check the consequences that possible changes of address of these bases have on their performance indicators.

It is important to stress that the current location of the bases is a natural consequence of the historical evolution of the provision of energy distribution services in the city of São Paulo, and that obviously it did not take into account the city's current geographic situation; the opportunities to acquire available real estate (plots of land) were cross-checked against the sensitivity of the planning area of the operation. AES Eletropaulo serves more than 6 million consumer units, representing approximately 16.6 million inhabitants and approximately 25% of the Brazilian GDP.

The analyses and conclusions of this work took into account the extensive database made available by AES Eletropaulo and relied on the vision of the concessionaire's professionals as far as concerned their advice with regard to the *modus operandi* of the company. Statistical prediction, evaluation, analysis and validation models were created, which served as the basis for the construction of software that will be available to AES Eletropaulo for future consultation and for supporting its strategic decisions. The innovation lies in the fact that since this software was made available, the concessionaire's planning area has had an unprecedented geospatial view of the relevant variables in the concession area, precisely in order to check the consequences of urban changes on the performance of its teams.

The most surprising conclusions of the work are the decisive influence of the traffic variables on operational performance and the relationship between the geographical location of the bases and the road network. In particular, by simply redefining the coverage area of each current base, possible immediate gains are noted for each team by way of a significant improvement in the performance of the service time for the first occurrence of the day, which is the occurrence that is impacted by the location of the base.

In the long term, as the software is used by the planning teams, the tool that has been developed will allow for a drastic adjustment in the number of operational bases, thereby freeing up an important company fixed asset. It may also modify the logistical procedures used for positioning company vehicles and equipment.
As the urban dynamics of cities involve evolutionary phenomena – the occupation profile of neighborhoods, the volume of new real estate developments, the increase in the number of vehicles, the opening of new roads, among others - the software developed is capable of incorporating into its calculations other quantities that represent the urban complexity and can, therefore, over time be used to check whether such changes are capable of influencing the flow of traffic, thereby generating unforeseen difficulties for meeting service response times.
The results achieved also allow the company to check whether the number of call centers and their position in the urban network are the most appropriate at each moment in the evolution of the urban characteristics of the São Paulo Metropolitan Region. This equation is sufficiently complex to suggest that the conclusions are exposed not only by the simple identification of the best addresses for positioning the service bases, but also by the generation of service time estimates, in order to provide elements for decision making, and to decide at what moment the concessionaire should review the positions due to possible changes.
For this to be possible, the work of analyzing variables and developing evaluation and results exposure criteria uses a digital map platform, which introduces the concepts of geographic intelligence and spatial statistics into the decision process (LEE; KANG, 2015; FRANCISCO *et al.*, 2006; LLOYD, 2010). Digital maps help analysts understand the evolution of changes in operating conditions in urban areas better and offer the possibility of quickly performing complex spatial calculations, which consider street and road situations with different traffic capacities and journey speeds at various times of the day and days of the week. Combined with climate change, local tree cover density and the real estate value of the regions, the software uses geographical algorithms that indicate the best locations for positioning the operational bases in each situation, both today and in the future.
The software that was developed allows for assessments to be carried out and decisions made about repositioning the operational bases immediately and routinely, provided that access to the databases used for monitoring traffic, the weather, the characteristics of the tree cover and real estate values used in constructing the geographic-digital analysis model is maintained.

2. RESEARCH DEVELOPMENT

2.1 Project Objectives and Product

In general terms, the project objective and product are shown in Figure 1.
The inclusion of the existing KPIs that are traditionally monitored by AES Eletropaulo, and the adaptation and/or possible creation of new indicators linked to real estate performance and its idiosyncrasies are an intrinsic part of this work. This platform will be used to analyze the KPI Predictive Models by way of a flexible model into which variables can be dynamically and continuously inserted for monitoring growth in the demand for assets and the need for operational bases that satisfy the different conditions as defined in the energy distribution rules.

Objective	To develop a methodology based on statistical models, operational research and geo-referenced data analysis that allows for an optimized configuration of AES's assets, without interfering with the operational performance of the technical and commercial services in the field.
Project Product	An analytical platform that has a tool for exploring the various scenarios resulting from the possible structuring of the operational bases (number/location/size) and the supply of services in the field, which will enable the concessionaire to replicate the analyses carried out, by systematically optimizing the geographic performance of the location of their operational bases.

Figure 1. Project Objectives and Product

2.2 Methodology

Figure 2 shows the macro-stages involved in developing the methodology. The sub-topics describe each stage in greater detail.

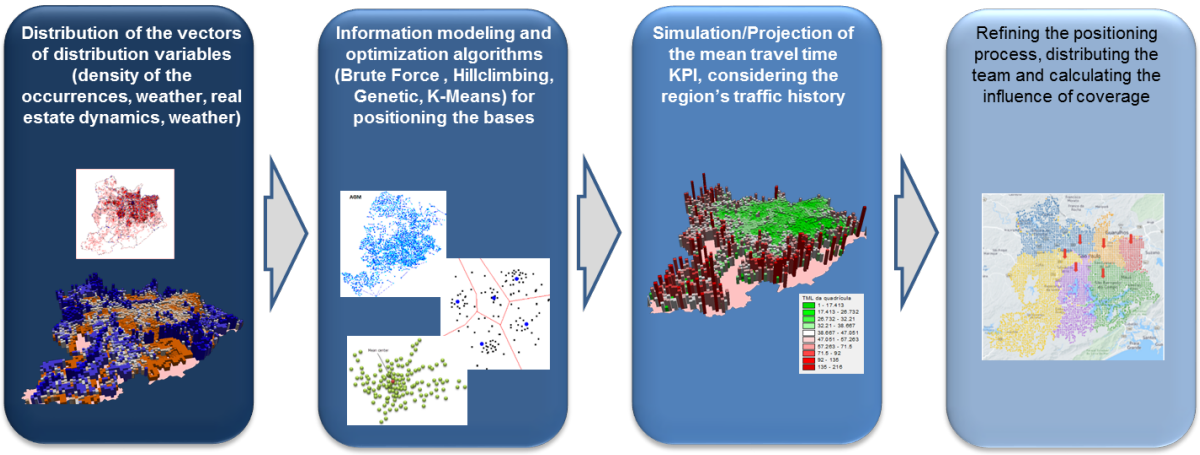


Figure 2. Stages in the development of Big Data and Systemic Geo-intelligence methodology

2.2.1 Survey of the variables, analysis and using the data

The development of all the ideas and related solutions started with AES Eletropaulo making of an operational database available that had more than 1.2 million records of emergency response calls handled by Eletropaulo's teams in the region it serves between December 2012 and November 2015. The emergency response data were supplied to the R&D team in data-tables containing a total of 12 million pieces of data (variables and observations) to be analyzed and including the following details: location of the emergency; the time the request for assistance was registered; the time the team left; the time the team arrived at the place of the emergency; the team responsible; the equipment; and the type of occurrence.

These data were incorporated into several other service processes, in an initial phase called "Proof of Concept", in which other services, especially Meter Reading and Bill Delivery, Fraud Inspection and Technical Notes were analyzed.

The recorded field occurrence data, especially the registration, departure and arrival times, were compared with consistent traffic data from the GisBI/Here external database. The consistency of the data collected/annotated by the field teams was then checked by reading the journey times

of vehicles (light and heavy) in the urban area, which were consolidated by way of São Paulo Metropolitan Region's traffic monitoring systems. The work used as its reference data the information that had been processed and worked on by GisBI, and based on the data of Here North America, LLC, taken from its service called "Traffic Patterns". This originated from the systematic monitoring of vehicles equipped with GPS and other types of journey control, whose license to use during the R&D development period was included in the contractual amounts.
As input for identifying (and updating) the typical traffic time for each road (to and from), the traffic base used in this project has a minimum history of 3 years that is being constantly updated (every 15 minutes), with hourly averages for each typical day of the week. This represents 13,171 different patterns of average journey times per day of the week and category of functioning roads.
For preliminary comparison purposes, journeys between the operational bases and the first occurrences registered for each team were considered for a period of one year (from December 2012 to November 2013), for total success in 111,912 occurrences, with two criteria being calculated for each comparison (shortest route and fastest route), which are available in the GisBI/Here database.

2.2.2 Enriching the variables and conceptual modelling

Based on the conclusions of the debates and methodological definitions of the previous stage, and to make it possible to add the different possible variables to the model that was to be proposed for locating the operational bases, the team started detailing and enriching the databases. This work required special dedication so that all the bases to be included in the methodology retained a spatial geographic coherence with the original data and, at the same time, allowed for the mathematical integration of different amounts, which were conveniently described in the actual development stage of the geospatial analysis computer model.
Previous studies (FRANCISCO, 2010; KIRSCHNER, 1971) showed the feasibility of concentrating a large volume of data into a grid as a possible simplification of the geographic distribution of different size databases with different meanings.
Despite this on-screen simplification possibility, the researchers nevertheless considered that the problem of locating AES Eletropaulo's operational bases in the defined concession region involved not only the question of the distribution of variables across the territory (and its geography), but the observation of journey times along routes (streets and avenues), as characterized by the region's traffic data coming from the GisBI/Here base. In other words, it required that each (and every) square on the grid be characterized as the origin and destination of journeys from all the other squares.
This conjugation of information and necessary iteration of the data of different origins and measures, it was noticed, turned the solution of the problem into an example of "NP-Complete", which in the theory of computer complexity means a special level of difficulty for solving problems in polynomial time.
This motivated the group to take the initiative of seeking aggregate representations and computable algorithms for solving the search for the optimal positioning of the bases within the grid, in order to generate conditions for AES Eletropaulo's planning teams to take decisions under operational conditions.

2.2.3 Development of the Spatial Analysis Algorithm

The title adopted for this R&D proposes using the expression "systemic geo-intelligence" (A Study for Proposing and Analyzing Composition Scenarios for Operational Bases by way of a

Systemic Geo-Intelligence Platform) as an innovation of the analysis and operational planning process within a power distribution concessionaire (LACERDA, 2000).
The aim of the innovation is not only to modify and/or increase the theoretical and practical instruments used by the professionals involved in these tasks within energy distribution companies by way of yet another piece of geographical technology. The idea is to transform all the complexity of spatial analyses into an instrument that is simple enough to be adopted in the day-to-day practice of operationally planning this type of public service, by incorporating the analytical possibilities of geographic platforms, which were made accessible very recently for this type of public service thanks to the rapid development of digital maps and their dissemination for use by everybody, notably by way of applications linked to the planning of traffic routes in cities, among other aspects (BALLOU, 2006; COELIS, 2006; BARROS, 2005; LACERDA, 2000).
Companies that already used maps and graphic resources in their most sophisticated strategic analyses now have the chance to develop new planning routines based on the intensive use of spatial reasoning on simplified platforms that do not require specific GIS software training. The assumption is that from this work it becomes possible to spread the geospatial culture among analysis and operational planning professionals, by way of interfaces that can be easily manipulated (FRANCISCO, 2011; WANKE, 2001).

2.2.4 Urban conditioners

The initial hypothesis was that some urban characteristics would need to be considered in formulating a mathematical-statistical analysis algorithm that was capable of recognizing, at any time, the influence of changes in the concession area, in particular, under conditions involving:

- Traffic on routes used by emergency response vehicles;
- Identification of and access to occurrence times;
- Local tree-cover patterns (interference with overhead cabling)
- Intensity of atmospheric phenomena (rain, wind and electrical discharge)
- Real estate vectors (land value of the operational bases and urban occupation trends)

In terms of traffic it was necessary to pay special attention to the times noted down by the field teams for their vehicle journeys to the place of occurrences. This work was carried out in the preliminary phase of this study, when a small number of inconsistencies was detected in a sample of 111,912 team journeys, which undoubtedly derived from the operational forms being badly completed, which is fully understandable, although undesirable when checking performance, given the emergency conditions and sometimes the urgency of these services.

2.2.5 Volume of Data versus Processing Capacity

After this initial cleaning up of the database of real journey times between the base address and the occurrence address, the challenge was to check the journey times on these same routes, given by the platform GisBI/Here, whose data make traffic patterns available that have an impact on journey times and that, in theory, could be measured every 15 minutes, 7 days a week.
This comparison work of both the journey times for each specific emergency response call-out was carried out in the preliminary phase and indicated a strong correlation, which allowed the GisBI/Here databases to be adopted more readily and confidently as a reference for producing the final algorithm for supporting the decision and for planning better locations for the operational bases.

One of the major challenges identified by the R&D team was to find a way to simplify the geolocation issue of eventual occurrences in such a way that the volume of information to be considered in the calculation of journey times from the operational base to the occurrences could simultaneously respect the strength of this correlation, without, however, overloading the algorithm to the point where it could not be processed by ordinary computers and in the worst of hypotheses would require a supercomputer.

2.2.6 Address representation squares

The solution adopted was based on the premise that it was necessary to reduce the number of geographic variables in the matrix of occurrence and base address possibilities in order to make a computer solution possible that was accessible to the concessionaire's team of professionals. In other words, it was necessary to check the hypothesis of representing a series of real addresses in the city from a random division of the concession area territory into squares with 1000-meter long sides, and converging all the addresses from that square on the address closest to its centroid (FRANCISCO, 2010; WANGENHEIM, 2006; WANKE, 2001).

In doing so an undefined (and certainly huge) number of possibilities for addressing occurrences started being controlled in a smaller matrix of squares, each representing a portion of the concession area territory. The centroid addresses then became the origin or destination of the journeys, the times of which would be given by the GisBI/Here traffic routes, which are made available in the concession area and standardized every 15 minutes of the day and night, 7 days a week.

This decision implied checking other proposition possibilities involving squares with sides less than 1000 meters in length to make the calculation even more precise and representative. But simulation of the number of mathematical interactions in each of the hypotheses indicated that a square with sides 1000 meters long was a good choice, because of the size of the concession area for making calculations feasible by the algorithm, since all addresses made available (from any centroid) should be checked in terms of journey times to all other addresses (other centroids) (LLOYD, 2010).



Figure 3. Relationship between the size of the square and processing time for the computer solution adopted by the study

Figure 3 shows the estimated processing time, which is a function of the number of iterations between the various addresses (closest to the centroids of the squares), according to the number of squares. The graph makes obvious the physical limitation of processing a data matrix of the "all addresses for all addresses" type. It shows that the simplification of 1000-meter-sided

squares with representative addresses closest to the centroids of the squares is at the limit of computer feasibility for the AES Eletropaulo concession region.

2.2.7 Cost versus Time Matrix

After simplifying the address matrix, it was necessary to check the behavior of the data available in the various databases chosen for undertaking the work. Each of the squares, represented by the address closest to its centroid, received all the geolocated information belonging to that particular square.

The databases of occurrences (AES Eletropaulo), real estate value, weather and tree cover (information taken from AES Eletropaulo's own description of the occurrences, because the city administration does not have territorial coverage of this information) were transferred to the squares, as of course were the journey times between the squares, two by two (traffic by GisBI/Here database), concentrating them in the address near the centroid.

Any type of information that can be expressed by its geographical location characteristics could be placed in the squares in the same way, which gives a universal character to the research carried out by the R&D team. It is certainly capable of being reproduced in other analyses, whether logistical or not, for any territory in any city in Brazil, or the world.

In this work, the combination of the various factors (databases) inserted in the grid, weighted according to the interest of the analysis, was called the "cost matrix" of the system analyzed, whose values can be assessed from the reference KPIs.

Because of the *ANEEL* performance requirements, in the analyses presented below this cost matrix will always offer as an output the journey times between each particular square and those squares that represent the location of the best placed operational bases in the territory, from the table of occurrences weighted by the variables of the system's cost matrix.

Interpretation of the territory as squares allows for mathematical computer modeling and therefore considers:

- The adjustment of the positioning of the AES operational bases depends to a certain extent and among other factors on the journey time between the bases and the occurrences.
- The road network is dealt with as a graph comprising nodes and the vertices that link them, as shown in Figure 4 below.
- To make computation of the results feasible on everyday commercial equipment some simplifications were introduced, such as dividing the area into 1 km-sided squares.

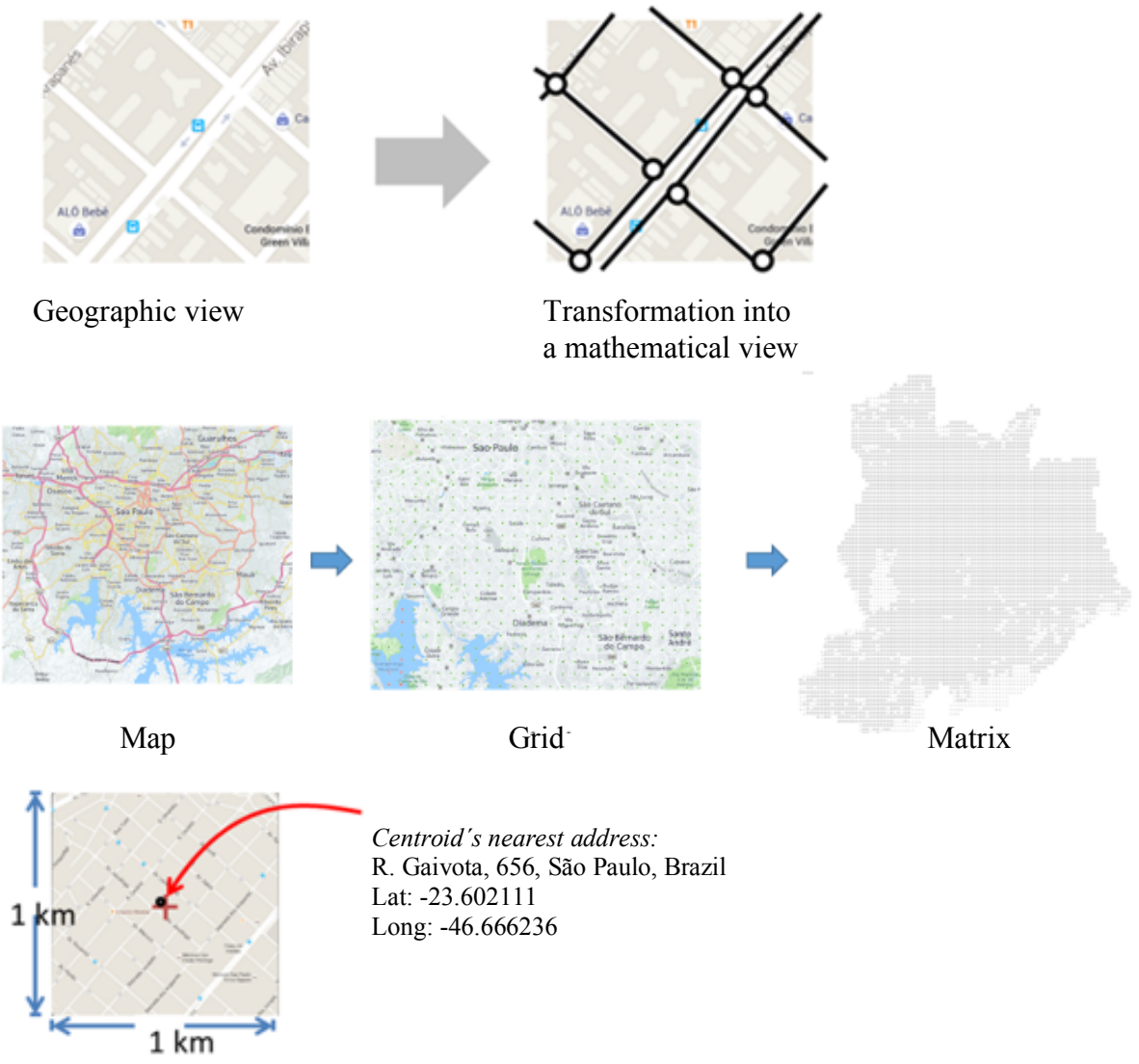


Figure 4. Transforming the geographic view into a mathematical view (matrix representation) for computer enablement of the algorithms developed in the project

2.2.8 Calculation of the best location for the bases

The simplification introduced by the squares allowed for an understanding of the calculation processes, but did not modify the nature of the computing problem to be solved. In other words, to find the best location for the bases would require mathematical calculations that are computationally intractable because of the volume of interactions between all the squares.

To overcome this difficulty, modeling had to incorporate different calculation algorithms by approximation, known for the solution of NP-complete problems. These algorithms were used, therefore, to define the best place for positioning the bases, when the volume of calculations became computationally intractable. For these calculations by approximation, Brute Force (an attempt to solve the NP-complete problem), Hill Climbing, Genetic and K-means algorithms, which are explained below, were applied.

For the traffic data, Here's API was used, which adopts a traffic pattern that is the result of studies that take into account the historical data of several years, and supplies the time and distance between two points on the road network in the concession region, in this case, the addresses closest to the centroids of each square.

As collection and preparation of the traffic data takes several days to process, traffic data for 8:30 a.m. on December 15, 2015 (Tuesday) were extracted for the simulations shown below.

2.2.9 The geographic problem transformed into a matrix problem

The geographic problem of determining the best position for the bases was then converted into the mathematical problem of determining which positions give the lowest weighted location cost of the bases in a matrix in which each cell represents a square, as shown in Figure 2, above. For calculating the best location, a cost matrix was considered that determines the journey cost (time) between each origin and each possible destination, and a weighting of the square that determines the weight of this square as a journey destination.
The algorithm can therefore be used to determine the best position for the bases according to any criterion. For example, it is possible to adopt the journey time between the squares as the cost matrix and the number of occurrences for each square as weighting.
In this way, it was possible to determine the locations of the operational bases that allow the smallest average journey time to occurrences, although this mathematical/matrix approach allows the position to be calculated according to any other criterion that it is wished to use.

2.3 Development of the optimization algorithms

As already mentioned, the work analyzed four iterative calculation methodologies, the characteristics of which will be presented below (Items II.3.1 and II.3.2). The methodologies for producing iterations were introduced into the algorithm using a C ++ program (C ++ 11 standard) for processing on an ordinary computer.
The basic calculation of the concession area under these conditions is developed considering the journey costs between all the squares, two by two, with the square destined to receive an operational base being the one that has the lowest cost in the cost matrix. The configuration cost is given by the sum of the multiplications of the weighting of each square and the journey cost from the operational base to that square. The average cost per occurrence in each operational base location situation is the total cost divided by the weighting, as shown in Figure 5.

$$Custo_{configuracao} = \sum_{i=1}^{N_{bases}} \sum_{j=1}^{M_{quadriculasDaBase_i}} Custo_{ij} * Ponderacao_j$$

$$Custo_{medio} = \frac{\sum_{i=1}^{N_{bases}} \sum_{j=1}^{M_{quadriculasDaBase_i}} Custo_{ij} * Ponderacao_j}{\sum_{k=1}^{P_{quadriculas}} Ponderacao_k}$$

Figure 5. Formal definition of the average cost per occurrence

2.3.1 The "Brute Force" method – Implementation (not feasible with more than 3 bases) of the NP-Complete problem

This is the most intuitive method: it tests journey costs from the operational bases placed on each one of the squares to all the other squares.
The operational base is placed on the square that has the lowest total cost.

- The method is not feasible for configurations with a lot of squares;
- That is why, for the purposes of this work, the maximum number of iterations was fixed at 10 million, in other words, just over one hour of processing time on the normal computer with a good performance that was used in this project;
- This time corresponds to calculating the position of two bases, if the data of the AES concession area are taken into account;
- It can be used to measure the results of the other methods;
- It can be used for areas that are smaller than the concession area.

It provides a precise result, although massive processing makes it unfeasible in the case being studied, which has a greater number of operational bases. To check the feasibility of iterations (and processing) according to the number of bases, the processing time was estimated from the number of iterations necessary, considering from one to ten simultaneous operational bases. Processing time becomes computationally unfeasible (years of machine processing) starting from the optimization of 4 operational bases.

2.3.2 The Hill Climbing, Genetic and K-means methods

The table in Figure 6 highlights the concepts, characteristics and limitations of each optimization method developed in the project.

2.4 Implementation and Performance Tests

Tests were performed considering from 1 to 2 bases with simulation of the 4 algorithms and verification of the location of the optimal bases in comparison with the Brute Force (ideal solution) method. For 3 or more bases, the tests consisted in evaluating the performance of the solutions.

	Description of the algorithm – Characteristics and Limitations
Hillclimbing	This method comprises calculation iterations considering the following stages: An initial configuration is chosen by draw. One base at a time is moved to each one of the squares adjacent to it. The cost of each one of the nine possible configurations is calculated and the base is moved to the position on which the configuration has the lowest cost. The procedure is repeated for all the bases successively until no movement of any base results in a cost reduction. The result depends on the initial draw, in other words, the "initial seed" for the draw of the initial configurations. Sometimes the initial position is such that a final position is arrived at which is a "local minimum", not representing the best possible situation. The calculation module allows the number of initial draws of the method to be determined, which is very fast and allows thousands of attempts to be calculated. The number of draws that gives the best result is unknown, because it depends on the heterogeneity of the cost matrix and the weightings. Generally, the bigger the number the better the result.
Genetic	This comprises the following stages: 1. An initial population of configurations is chosen by draw. 2. The cost of each configuration is calculated. A number of the lowest cost survivors is chosen. 3. Each survivor generates a number of descendants and equal to the number of bases in the following way: each configuration base is substituted one at a time by another drawn base. The descendant configuration is compared to a list containing all the configurations tested in order to generate more repetitions. 4. The cost of the new population along with the survivors of the previous generation and the new survivors are chosen. 5. The process is repeated until the desired number of generations is reached. It is not limited to the minimum number of locations. The quality of the solution depends on the number of generations. There is no guarantee that the solution will be optimal, or the best of all solutions.
K-Means	1. An initial configuration is chosen by draw. 2. The coverage of each base on the configuration is calculated. 3. In each area of cover the best position for the base is calculated. 4. The two previous steps are repeated (in other words, no new draw is carried out) until the positions for the bases are established. It depends on the initial draw. The quality of the solutions depends on the number of repetitions. There is no reason to believe that stability is reached at the point of lowest cost.

Figure 6. Description of the algorithms implemented in the project

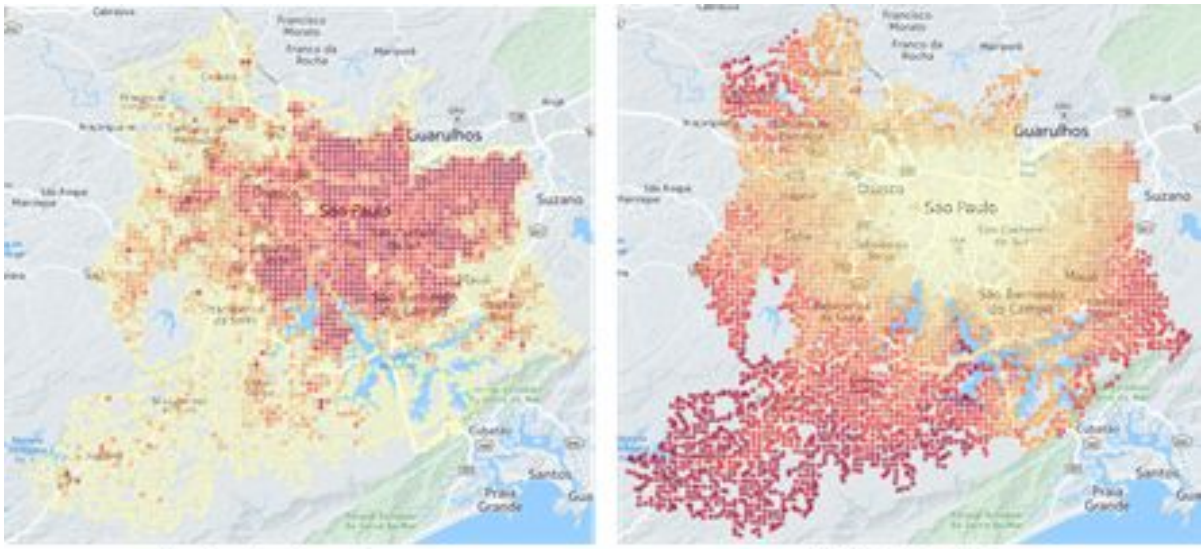
3. RESULTS

Among the main findings are that traffic has a great and significant influence on the positioning of the bases.

In the conditions observed, the Hill Climbing method converges faster to the best solution, but the intermediate results do not consistently approximate to the best result. The Genetic method requires more iterations, but consistently approximates to the best result: the more generations there are the better.

These results influenced the decision to make a fourth genetic method available with refinement, which we call "GENETIC + HILL CLIMBING". This method uses the genetic method, as intended, and takes the best solution to apply the Hill Climbing method to it just once. The idea is to take the best possible generation and optimize it in a way that the Genetic method is unable to do in itself. The performance needs to be taken into account with all the data in order to assess the best uses for the algorithm in each case. The K-means method apparently converges on a point that is not lowest cost.

Figure 7 shows the arrival and leaving costs of each square in AES Eletropaulo's concession area (as described in 2.3). Figure 8 shows an example of the final configuration for the 5-base simulation. General maintenance of the current service indicators is estimated (in particular, average service and journey times) as are cost reductions resulting from a reduction in the total constructed area of the bases, which reach tens of millions of *Reais*.



Arrrival cost Departure cost

Figure 7. Costs of reaching and leaving squares in the AES Eletropaulo concession area

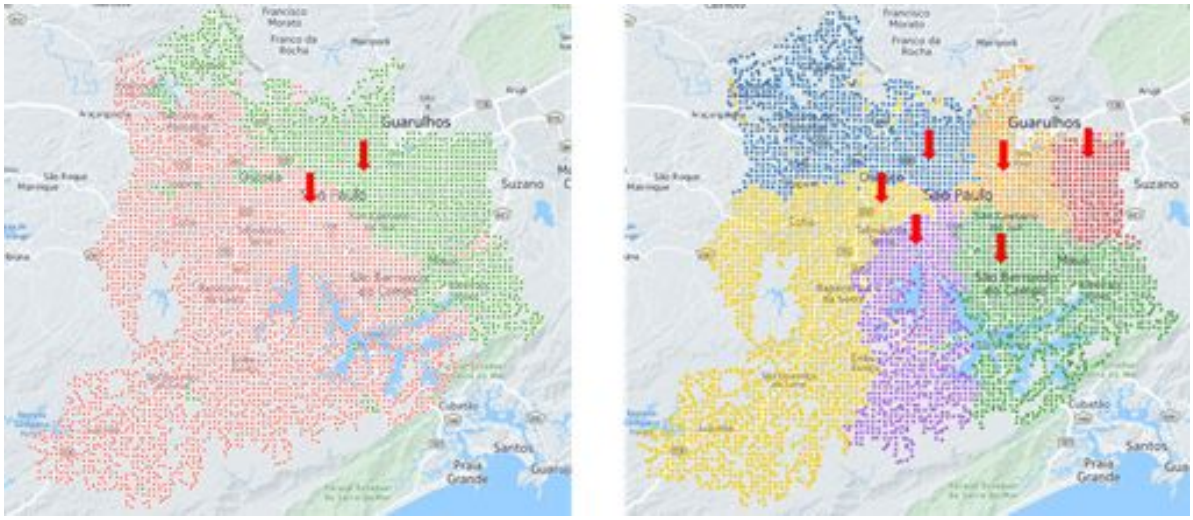


Figure 8. Example of optimization in the composition of 2 (left) and 5 (right) operational bases

In completing the implementation, development of the system platform (software) was completely web-based to allow AES Eletropaulo analysts to access it via a browser in order to define the simulation variables (traffic, real estate dynamics, weather and tree data) and the weights of these variables, to select the number of bases, the optimization method and the specific parameters and to simulate the KPI scenarios and their measurement.

4. CONCLUSIONS

The introduction of the concepts and possibilities of geotechnology and geointelligence into the operational activities of utility concessionaires can be considered an inexorable trend in methodological advances and business practice, notably of the company that dedicates itself to activities that are distributed throughout the territory, in both urban and rural areas.

The R&D project that was developed shows that the concepts, considerations and methodologies of geostatistical calculations can be introduced into the day-to-day business routine without the professional planning and operational management professionals having to learn or incorporate all the complexity and nuances of GIS (Geographic Information Systems) software that is available in the market.

The system developed (software) aims to offer an automated implementation instrument for the geographical calculation methodology for reallocating the operational bases in AES Eletropaulo, and is the main result of the work that was carried out. The system considers simple objectives for the introduction of the variables to be considered in the analyses and reasonable computing times with the possibility of carrying out studies in multiple change scenarios involving the urban and weather conditions of the concession region.

The same system may be used in other regions, provided local variables are made available and introduced in combination with the operational data of any power distribution company in the country, or even in other countries. The digital map resources are integrated into the tool that was developed and, therefore, the results of this R&D are not limited to the geographical area of the AES Eletropaulo concession, and can also serve other types of public service concessionaire.

For the concessionaires' planning and operational management teams, the results of this R&D offer the possibility of immediate incorporating geotechnologies into their professional activity, without requiring an in-depth knowledge of spatial statistics. This possibility may induce a new level of operational performance, and prove that significant improvements may be introduced

in the provision of public energy distribution services, thus constituting great methodological success for those who adopt the computer system that was developed for reallocating the operational bases.
This is, therefore, a Big Data project with geographical optimization that is unprecedented in the Brazilian electricity sector.

BIBLIOGRAPHY

BALLOU, R. H. Gerenciamento da Cadeia de Suprimentos/Logística Empresarial. 5a Edição. Porto Alegre: Bookman,2006.
BARROS, M. C. Warehouse Management System (WMS): Conceitos teóricos e Implementação em um Centro de Distribuição. Rio de Janeiro: 2005. 132p. Dissertação de Mestrado - Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.
COELIS, E. L. Logística Empresarial. Belo Horizonte: 2006. 10p. Instituto de Educação Tecnológica. On-line. Disponível em <http://www.techoje.com.br/site/techoje/categoria/detalhe_artigo/507>
FRANCISCO, E. R: Indicadores de Renda Baseados em Consumo De Energia Elétrica: Abordagens Domiciliar e Regional na Perspectiva da Estatística Espacial. Tese de Doutorado - Fundação Getúlio Vargas - EAESP. São Paulo: 2010.
FRANCISCO, E. R. . RAE-eletrônica: Exploração do Acervo à Luz da Bibliometria, Geoanálise e Redes Sociais. ERA (Impresso), v. 51, p. 280-306, 2011.
FRANCISCO, E. R., ARANHA, F., ZAMBALDI, F., GOLDSZMIDT, R. Electricity Consumption as a Predictor of Household Income: a Spatial Statistics Approach. Advances in Geoinformatics. pg. 267-282. GEOINFO 2006. Campos de Jordão, Brazil, 2006.
KIRSCHNER, P: A study of vision as a measurement of decision-making behavior: 1971. Thesis (Ph.D.) - Dept. of Industrial Engineering, Stanford University.
LACERDA, L. Armazenagem Estratégica: Analisando Novos Conceitos. Mar. 2000. On-line. Disponível em <http://www.ilos.com.br>
LEE, J-G, KANG, M. Geospatial Big Data: Challenges and Opportunities. Big Data Research, 2: 74-81. http://dx.doi.org/10.1016/j.bdr.2015.01.003. 2015
LLOYD, C. Spatial Data Analysis, Oxford University Press, 2010.
WANGENHEIM, A. Análise de Agrupamentos 2006. On-line. Disponível em <http://www.inf.ufsc.br/~patrec/agrupamentos.html>
WANKE, P . Aspectos Fundamentais do Problema de Localização de Instalações em Redes Logísticas. 2001. On-line. Disponível em http://www.centrodelogistica.org/new/fr-aspec.htm