Spatiotemporal Data Mining for Distribution Load Expansion

Hector Gustavo ARANGO[1], Germano LAMBERT-TORRES[2],
Carlos Henrique Valério de MORAES[3], Luiz Eduardo BORGES DA SILVA[3]
[1]*South of Minas College for Economic Sciences, FACESM, Itajuba, MG, 37504-066, Brazil*
[2]*Gnarus Institute, Itajuba, MG, 37500-052, Brazil*
[3]*Itajuba Federal University, UNIFEI, Itajuba, MG, 37500-903, Brazil*
germanoltorres@gmail.com

***Abstract*—The load spatial forecasting is fundamental for the electric energy distribution systems planning. Several methods using different conceptions have been proposed to determine the future configuration of the electric markets. This paper proposes a dynamic model of load expansion, based on concepts of local analysis using ideas and applications from urban poles theory. Thus, the load expansion is simulated in a dynamic way, maintaining a continuous change in the conditions for localization of a new load unit. An algorithm generating a snapshot that represents the distribution system configuration at that instant determines the geometry of the market in a given instant. The proposed dynamic model, based on the urban poles theory, has the capacity for summing up the information from economic variables sets, expressed in terms of interchange flow laws, which are modeled by distance and transportation functions. This supplies the model with the capacity for being used even though the number of available explanatory variables is reduced.**

***Index Terms*—data mining, intelligent systems, load modeling, power distribution, urban areas.**

I. Introduction

An adequate planning of the electric distribution systems comprises economic and strategic factors, whose importance increases in proportion to the increase in supply restrictions. This is a natural consequence of demand growth and lower availability of energy sources. It occurs mainly in cases where the demand growth is generated by increasing economy process and income redistribution. Also, it makes the network system to operate near its operational limit [1]. In the short-range, new electric energy generating and transmission units are not expected to alleviate the load demand [2].

In the case of distribution system, the equipment dimensions will depend upon the expansion prevision for load density [3]. In other words, as it is basically a radial system, simply estimating the total load is not enough; it is also necessary to determine the space distribution of the load [4]. There are several methods for the prevision of the space load. The first methodologies are from the sixty-decade and the number of that such studies have been increasing by the first half of the seventy-decade [5]. In the second half of the eighty-decade, the available computer resources allowed the increase of the number or programs that combine the prevision methods in graphic environment with some complexity and precision [6]. In the beginning of this century, elements from geographic information systems (GISs) have started to be used in this spatial forecasting [7]. Also, spatiotemporal data mining has been used in different area of knowledge, such as climate change [8], car mobility [9], city expansion [10], cyber-security [11], reservoir system operation [12], and cell-phone network [13].

This paper presents a data mining model for the simulation of the expansion of electric energy distribution system, based on the theories of urban growth and regional economy. Nowadays, a spatiotemporal load expansion, forecasting new loads in space and in time, are growing up its importance due to the smart-grid structures [14], in special for the planning activity when it occurs in several plans minimizing the distributed energy resources and maximizing their benefits [15].

II. Urban Economy Applied to Load Forecasting

A. Historic Evolution of Urban Economic Models

The initial models concerning urban and regional growth did not use the space dimension explicitly. In [16, 17], Fujita developed studies of space agglomeration in dynamic environments. The majority of the studies are based on Solow's growth model, which does not consider the economic growth optimum rate. In [18], Kanemoto proposed an optimum growth model, and derives the space-state ideal dimension of the city, maximizing the sum of the difference between the supplied utility by one instantaneous utility of the household and the space-state ideal level. The proposed agglomerate dynamic models explain the development with the help of production geographic differentiation, and the capital reproduction by the migration of factors between urban centers.

The agglomeration of manufacturing industries and consequently, the needs of the workers (homes, stores, and public services) are studied in [19]. This study takes into consideration the availability and variety of consumption goods in this area. According to Von Thünen's model, this area, named the "isolated state", has a single-centered economy and a balanced spatial configuration. Into this reality, a model of monopolistic competition is made from the isolated state, suggesting a mathematical formulation that allows the analysis of the parameters defining the spatial agglomeration rhythm. In addition, a balanced spatial configuration in a single-centered economy is defined, making variations in the parameters used for the analysis of spatial growth.

This work was supported in part by the CNPq – Brazilian National Science Foundation under Grant 306230/2010-0.

Digital Object Identifier 10.4316/AECE.2016.03010

The concepts of regional and urban economy allow the creation of models for controlling urban growth [20], expansion of urban infrastructure [21] and balance in space systems [22].

B. *Urban Economy Models in Load Forecasting*

The theoretical basis of the methods that use urban poles comes from the location analysis model, which is a branch of urban economy. This makes the analysis of the consumption and production a part of the microeconomic environment. Another research line investigates the reasons for the urban agglomeration, suggesting a model that has the accumulation of human knowledge as the main explicative factor. In this line, the topics discussed are: (i) the reasons for spatial agglomeration in the process of economic development; (ii) the reasons for the growth of some cities while others get smaller; (iii) the factors that generate the size of the city and the growth rates of regional incomes; (iv) the possibility of a town to achieve a socially ideal allocation in a free environment and (v) the government actions able to promote one efficient timely allocation in the local economy.

To attend to these questions, a dynamic model of general balance with spatial iteration is made in which human capital is the outward factor that represents the agglomeration centripetal force. Other agglomerate factors referred to in the literature would be availability of public resources, inside scale economy, outside scale economy, outwards combined in consumption and production, and gains from transactions with heterogeneous agents.

III. LOAD EXPANSION DYNAMIC MODEL BASED ON URBAN POLES

A. *Conceptual Presentation of the Proposed Model*

The proposed model uses the concepts and applications of the urban poles theory. However, the load expansion is simulated in a dynamic way, changing the conditions for the localization of a new load unit in each considered period of time. Thus, the market geometry in a given instant is determined by an algorithm that actuates in a continuous way, generating a sequence of space-state that represents the configuration of the distribution system in that instant. The formation of the spatial geometry that represents the load density, carried out by one algorithm that repeats itself indefinitely, has a process similar to the formation of a fractal. Models that consider the geometry of the chaos for explaining the population dynamic in space are also used in the urban economy [23]. Furthermore, in the proposed model, there is a component that actuates in the opposite direction from the urban pole, and that promotes the load redistribution from dense locals to spread-out locals, copying the urban migration phenomenon [24]. So, it may be considered as a load expansion dynamic model based on urban pole theory.

But there is a difference with respect to polycentric models since each load unit actuates as an urban pole, attracting new consumers in direct proportion of its demand. This characteristic of difference is important to guarantee the evolution process dynamic into one quasi-continuous space.

The proposed model describes the load growth for each unit through the union of three types of load: residential, commercial and industrial. Each one of these types is already divided and adjusted according to the load evolution and the nature of business [25]. For example, in case of commercial load, there are businesses that attract each other (those that work with complementary goods), and others that tend to keep apart (working with substitute goods) because of competition. The public load, for the purpose of this model, is considered as the residential load.

B. *Residential Load Model*

The algorithm for the residential load growth is implemented starting from a function that aggregates two components: deterministic and random. The deterministic component is concerned with the attraction produced by an individual load, and may be taken as an independent pole. The random component represents the probability for this pole not to exercise its attraction power. Hence, the model acts intrinsically with multiple poles, having as many poles as there are load units.

The function that unites the functions of attraction, f_A, and the random function, f_R, duly normalized, is named space disorder generating function, or composed attraction function. The modeling of this function in $\Re^1$ is made

$$f_{AR} = f_A + \rho \cdot f_R$$

where $0 < \rho < 1$ is the weight of the random component in the generating function of set space disorder and the vector indicates the function value in each point in space.

Thus, the gauging of ρ determines the higher or lower load space dependence.

The construction of vector f_A represents the normalized attraction function values, made by adding the individual influence of each load in space, based on the traditional factors:

$$f_A = f(M, \delta, d)$$

where M is field maximum intensity; d is distance between a new possible point and an existing one; and δ is load influence intensity variation rate ($\delta = \partial A / \partial d$).

If many points, located in different places, are considered in this space, there will be several attraction functions corresponding to the places where the points are located. Then, it is possible to have the attraction function value (A) for that particular point, which defines the set attraction function.

The modeling of the random component f_R is carried out by assuming one uniform function of weight ρ associated with the attraction function. In the simulation process, the introduction of each new load increase, whether in empty local or not, depends on the conditions immediately preceding its vision.

Thus, once the loads are located, the initial conditions for the localization of a new set of load points will be changed as a function of the previous set position. Thus, the process feedback happens until the maximum iteration number is reached.

The migration modeling is made through the named light attraction function (or the light attractor). This function is based on the idea of permanence probability. The permanence probability evaluates the chance for one standard to keep its position in space. In principle, this probability is high (above 0.9) and varies according to the density in the surrounding region to the element, so

$$D = n / u$$

where: D is density, n is number of standards, and u is load dimension unit .

As the number changes of the standard in each iteration, the density and, consequently, the permanence probability of each standard changes in each iteration. Hence, the permanence probability of standard i in the iteration I may be given as

$$P_i(I) = 1 - \gamma \, . \, D_i(I)$$

where γ is a setting parameter of the staying probability function.

The density may be obtained from the attraction function in the previous iteration. In this case, D is substituted by the value of the normalized attraction function in the point, f'_A. The probability of each standard to not stay in the place, or migrate, then results in the function

$$f_m = m - m \, . \, n \, . \, f'_A$$

where, m represents the migration tax and n is the number of considered points. It is important to notice that though the function value in each point depends on the density, the expected value of the migration function is equal to m.

The load migration follows the rule:

- rand $> f_m$ do not migrates
- rand $< f_m$ migrates

C. Commercial Load Model

The central problem of the commercial growth model is the determination of the laws that govern the location of the commercial loads in space. The approach to this problem may take into consideration two hypothesis: (i) the existence of a plan that set up the commercial expansion conditions, such as the growth rate, location areas, etc.; (ii) the commercial development as generated by the market behavior.

In the first hypothesis, restrictions of temporal and spatial nature are imposed upon commerce by a central form development. In this case, the aim is to reach an ideal social benefit, like the maximum satisfaction of the producer-consumer set. At these conditions, the spatial dynamics of commerce development may be explained inside the theory of land use.

In the second hypothesis, the commercial development is defined fundamentally by the set of interests of the producer. In this case, the objectives are in relation with the individual pay, the stability and survival of each commercial unit. The necessary behavior to achieve the set of objectives is limited by the existent economic conditions. Thus, the restrictions imposed upon the model are basically in relation with the predominant spatial and market structures. In the proposed commercial growth model, the following restrictions were considered: homogeneous space, consumer equivalent income, non-elastic demand, equivalent consumption trend, constant transaction number per time unit, and constant profit per dealt unit.

The criterion for localization of the commercial center is to have the highest income once the profit per unit sold is a constant.

Furthermore, the consumers have the same trend and the same budget toward consumption. However, the consumer uselessness in the acquisition of each unit of good depends on the unit cost of good and on the transportation cost up to the place where the consumer is located. Therefore, the distance up to the commercial center works like a restraint to consumption, reducing the commercial center income.

In Fig. 1, C represents the consumer number i, whose localization is Xi. P is the commercial center (localization Xp), K the production cost of the good sold and d, the distance between the consumer i and the supply center.

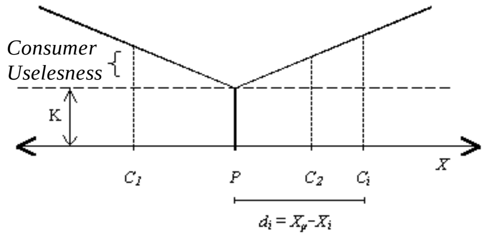


Figure 1. Distribution on $\Re^1$ of consumers supplied by one producer

In order to guarantee consumers the same purchase stimulus, the commercial center takes the decision of assuming the product transport up to the customer. With the assumed hypothesis, the ideal center localization will be in the place that guarantees lower operational cost, understood as the summation of each sold product and transported up to the consumer. Thus, Cti = f (K , di). Accepting that the transportation cost changes with the square of the distance, the ideal localization of the production unit occurs for

$$X^*_P = \overline{X}$$

Thus, the commercial load localization model starts by locating the first commercial consumer into the spatial gravitation center of resident consumers. However, once the first commercial consumer is located, it is assumed that another commercial consumer, that trades the same type of goods, will provide competition to the existing one.

Thus, once the first commercial unit is located, the second one decides upon its place of installation regarding the most advantageous localization, taking into consideration the effects of the competition and the residential consumer market. The modeling used for adding these effects considered the same laws for all the commercial centers with respect to the consumer market; namely, higher residential density increases the stimulus for localization (market function). With respect to competition, however, distance was considered a linear function of restriction regarding commerce of the same nature. This function, called repulsion function, gets activated at a certain fixed distance, saturation beginning from this same distance. The Fig. 2 shows this effect in the $\Re^2$.

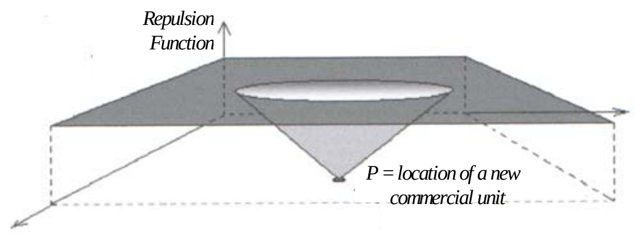


Figure 2. Repulsion function effect over the location of new commercial units

After the establishment of the first commercial unit, the location of subsequent units is decided based on a function of composed probability by the market functions (residential consumers) and the repulsion function that is modeled by the existent trading. For measuring the market function, the residential attraction function defined previously was used. Finally, the probability function for the commercial location is considered. Thus:

probability function of commercial location =
= attraction function – repulsion function

D. Industrial Load Model

The industrial load modeling is made through a parameter set that actuates whether in the temporal domain or in the space domain.

In a manner similar to commercial load modeling, the determination of localization time of industrial loads is given by a function that relates the number of residential loads and the number of industrial loads. Thus, one industrial load would emerge after the accumulation of a definite number of residential loads. This relation is linear in the temporal domain.

$$f_{ind} = 1 / n_{res}$$

where f_{ind} is the frequency of industrial load sprouts and n_{res} is the residential load number.

Once the industrial load localization rhythm is defined, two parameters determine when and where the load will be located. They are: the implantation coefficient and the localization coefficient.

The implantation coefficient is related with the efforts (fiscal, infrastructure, politic) of government officials in attracting the productive units into their influence area. The adoption of industrial development policies, through taking direct or indirect actions, works out as a temporal one-attractive agent, favoring the localization of industries in direct proportion to the intensity of those actions. The implantation coefficient varies between the interval $0 < \tau < 1$, indicating a total lack of interest and stimulus from the public in promoting the installation of industries in the lower limit, and the sprout of an industrial load for each n_{res} residential load in the upper limit.

The localization coefficient is a reference to the public power planning regarding the urban development, in particular, regarding the designation of specific areas for set up of production units, or industrial districts. The localization coefficient varies between the interval $0 < \lambda < 1$, where higher values of λ indicate higher probability for an industrial load set up into an area assigned only for industries. For $\lambda = 0$, the production units would follow the same localization standards as the residential loads (they would be ruled by the same residential localization function topology).

E. The Influence of the Industrial Load

The localization of previous residential and commercial loads influences the localization of subsequent loads of the same nature (either by induction or repulsion). Therefore, these types of loads produce attraction in the space domain. In the model being proposed, the industrial loads do not have this characteristic; they produce attraction in the temporal domain. This means that the rhythm of entry of residential/commercial loads is affected by the setting up of one industry, taking a step in the direction of temporal localization of this type of load. This step would be explained by the migration that comes from the increase in job opportunities in the region. This leads to the introduction of new exogenous loads to the model, thereby changing the forecast regarding the total regional load.

IV. Experimental Results with Real Data

A. Spatial Forecasting Error Computation

Error computation needs to be discussed before the presentation of the experimental results. Usually, in common load forecasting, only the numerical value of the error is important, i.e., just a number can express whether the forecasting is good or not. However, in spatial load forecasting, not just the numerical value but also the location and how this error is spread by the analyzed region are important. Fig. 3 presents two cases of possible error output results using a region divided into 16 areas.

In Fig. 3(a), two possible error distributions are presented. In the first one, the areas to the left of the analyzed region show a concentration of +1 while a concentration of -1 occurs in the areas to the right. In this case, a total unbalanced load occurs. In the second error distribution, a mixture of -1 and +1 occurs throughout the analyzed region. It means that the errors are homogeneously distributed. Although the total error is the same (zero), the second forecasting is must better than the first one because surplus power in an area can help a power deficiency in an adjacent area. In this case, area (1,2) can help area (1,3) by using, for example, a helping switch (normally open), but area (1,1) cannot help area (1,4).

In Fig. 3(b), the total error in the first distribution is +16, while in the second one, it is +13. If only the numerical value is considered, the second case appears to be better, but in reality, the first one is better than the second one. This is because in the second case, the error is not spread throughout the region and there may be a huge problem in distributing the load of area (1,4).

+1	+1	-1	-1
+1	+1	-1	-1
+1	+1	-1	-1
+1	+1	-1	-1

+1	-1	+1	-1
-1	+1	-1	+1
+1	-1	+1	-1
-1	+1	-1	+1

(a)

+1	+1	+1	+1
+1	+1	+1	+1
+1	+1	+1	+1
+1	+1	+1	+1

0	0	+1	+9
0	0	+1	+2
0	0	0	0
0	0	0	0

(b)

Figure 3. Examples of spatial forecasting error results

B. Estimation of Basic Parameters for the Simulation Model of Electric Market Expansion

In this section, a procedure is proposed for making adjustments to some parameters from this model to a space-actual temporal series, based on data from the city of Itajuba, Brazil. The aim is to test the potential of the laws that govern the model, and check whether it has the required flexibility for the production or reproduction of the universe from the possible states of an actual case, in both temporal and spatial dimensions as well as in combined form. This aspect is important once the spatial or temporal adjustments analyzed separately and then combined; do not produce the same results as the one that is obtained through simultaneous treatment of both these dimensions.

For this purpose, three data sets are composed. Each set contains scenarios ("snapshots") that correspond to the spatial configurations of the Itajuba city during a certain set of years and provides forecasting for another set of years, according the Table I

The objective is to try to approach the generated result by applying the simulation algorithm of market evolution (or of spatial disorder simulation) to the actual sets of space-temporal load. For the construction of these scenarios, data from Minas Gerais State Power System Co. (CEMIG), Brazilian Institute for Geography and Statistics (IBGE), Institute for Economic Research (INPEC) from South of Minas College for Economic Sciences (FACESM) and Itajuba City Hall, and Google Earth were used.

During the tests, the simulation algorithm of the market evolution simulation has had some of its parameters fixed, such as: the load growth rhythm, the scale, etc. The parameters taken as principals, however, were left free, so as to make possible to manipulate them in such a way that, they make the set that generates the higher possible proximity between the actual data and the simulated ones, as per one evaluation function that takes into consideration not only the average of a simulation set, but also its variability.

The process used for the optimization of the disorder program parameter was the Simulated Annealing (SA) [26], once the functions that generate the space disorder are not known. The tests aim to obtain sensitivity regarding the parameters used in simulated annealing.

C. Function of error rating (FER) from Simulated Annealing program (SA)

Let the generic set ph represent all the possible parameter sets. Feeding the spatial disorder program (SD) with this set will result in spatial disorder, as indicated in Fig. 4.

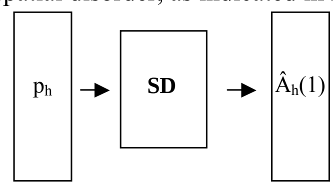


Figure 4. Space disorder generated by parameters set h

Here, $\hat{A}_h$ (1) is the vector (output set) that represents the generated spatial disorder by the set of parameters p_h in the first simulation. For N simulations, we have $\hat{A}_h$ (1), $\hat{A}_h$ (2), …, $\hat{A}_h$ (N), which is a vector set that may be grouped in a matrix I x N. We note that the columns of this matrix are the result of space-states generated by the N tests made.

The error in each one of the N tests made in relation with the vector column A, represents the observed (actual) series space-time. It may be rated in squared form by calculating the square of the difference between each element of matrix A and the corresponding elements in the matrix columns that aggregate the N vectors Â(h).

We also have the matrix E (I x N) that has all the squared errors that came from each test and in each point of space.

This matrix is converted into a scalar that is the representative parameter of the error from the N test sets related to true space-state. The scalar is calculated in the following way:

1. The summation of all the elements of each matrix column E results in one vector-line with N elements.

$$\varepsilon = \sum_{i=1}^{I} E_i$$

2. The order 1 moment for the vector elements described in 1 is calculated.

$$m1 = \sum_{j=1}^{N} \varepsilon_j \Big/ \mathrm{N}$$

The first order momentum, thus calculated, may be called the average squared error from the parameters set p_h, related to the observation set A.

The second order momentum may be calculated as follows:

$$m2 = \sum_{j=1}^{N} (\varepsilon_j - m1)^2 \Big/ \mathrm{N}-1$$

One composition out of the several moments described above may be used to compose the evaluation function of the simulated annealing program. An evaluation, up to the second order, would include sensitivity with regard to variability from results with the same parameter set; up to the third order, it would be possible to identify error concentrations to the left or right from the average error ($m1$). The evaluation function to be used is:

$$FER = \alpha_1 \,.\, m1 + \alpha_2 \,.\, m2$$

More details of this implementation and evaluation can be seen in [23].

D. Test of parameters estimated by SA

This long-term project contains three sets of input data (Table I). Each set of data contains detailed files from Itajuba County (each area of the grid is 100 m x 100 m). The first case (input data between 1980 and 1991 and forecasting from 1992 to 1996) was computed using simulated annealing algorithm for parameter adjustments. Its purpose was to verify the quality of the proposed methodology, and also to collect parameters in a period where the economic situation in Brazil was not good. These parameters can be used to forecast the load in similar economic situation. In this case, for the year 1996, the evaluation function value was 498.67.

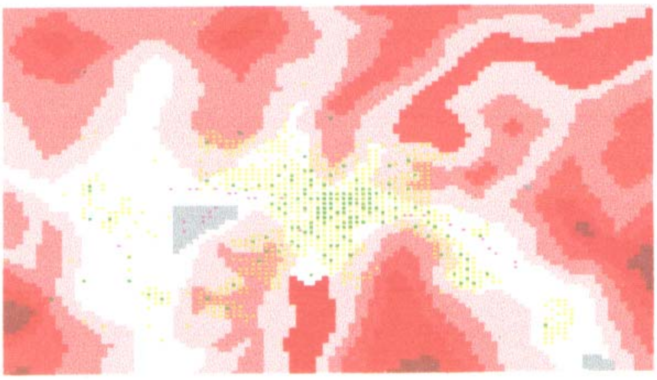
TABLE I. DATA SETS – OBSERVED YEARS AND FORECASTING YEARS

Case	Observed Years	Forecasting Years
1	1980-1991	1992-1996
2	1985-1998	1999-2004
3	1985-2005	2006-2014

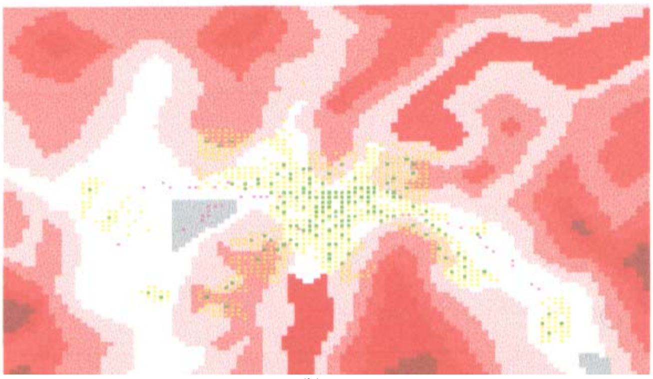
This value, as explained, corresponds to the average error composition and the variability, with equal weights, of simulated settings with respect to the actual setting of 1996. It should not be confused with the results obtained by the time of the adjustment process. A smaller detailing was also used (grid 50 m x 50 m) once in this case. A simulated setting for 1996, using these parameters, may be seen in Fig. 5(a). The actual setting from 1996 was included in Fig. 5(b) allowing visual comparison with this work.

The first test verified that the quality of data between 1980 and 1984 was not up to the mark and this part of data set was erased form the database for other simulations.

In the second case (input data between 1985 and 1997 and forecasting between 1998 and 2005), tests with these sets of data require external adaptations to compute the parameter adjustments. All adaptations have been provided automatically by the proposed methodology. The forecasting results of the tests were good, but these set of data contained a significant problem for load evolution. Itajuba had an enormous water flood in 2000, producing a break in load evolution in the years 2000 and 2001. The forecasting for 2005 has an evaluation function value equal to 386.95.



(a)



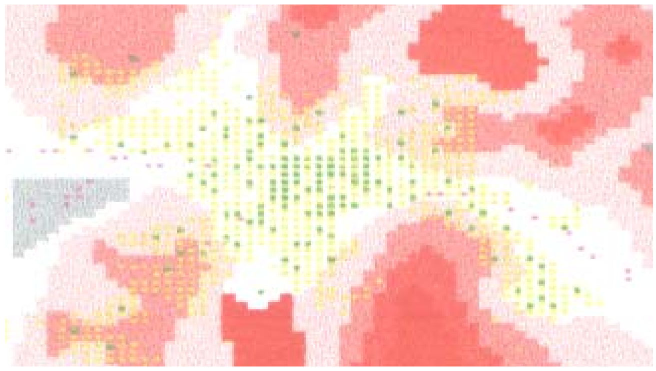
(b)

Figure 5. Snapshots for Itajuba County in 1996: (a) forecasting values; (b) real electric market

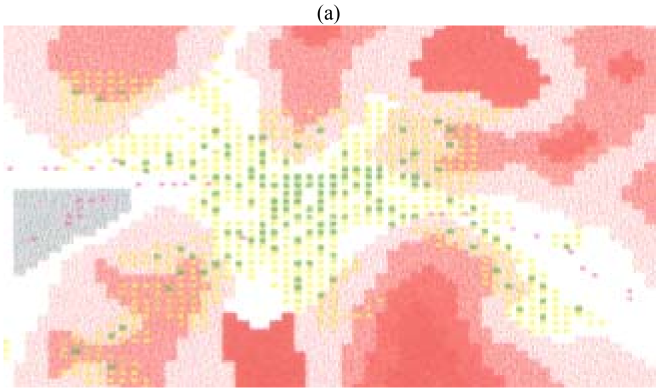
Finally, in the third case (input data between 1985 and 2005 and forecasting between 2006 and 2014), for the reason explained above, the data of the years 2000, 2001, and 2002 were not used in this forecasting. The forecasting for 2014 presents an evaluation function value equal to 213.85. Fig. 6 presents the spatial load forecasting and the actual load for 2014, obtained through the Google Maps [27].

Fig. 7 presents a partial vision of the load in the developed software for the Square 1 during the forecasting process. This scenario represents the forecasted load to 2008. The appendix shows an illustrative example of this software.

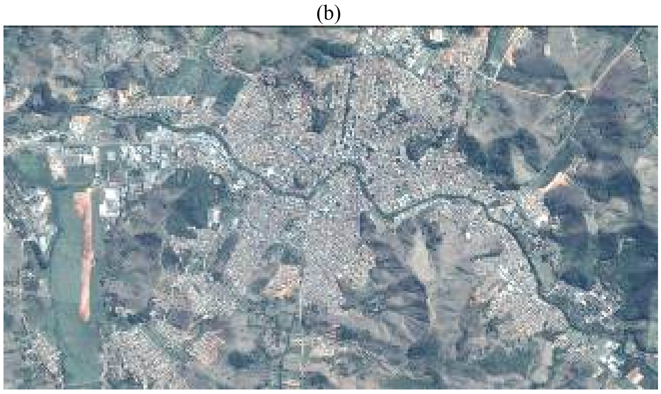
Fig. 8 shows the spatial distribution of forecasted errors for the residential electric market from Itajuba County in 2014, for the Square 2 shown in Fig. 6(a). The numbers inside the squares show the differences between the actual result and the result from one of the settings estimated by the simulator. Positive numbers indicate that the simulator has estimated consumers in excess, while negative numbers show that the simulator has under-evaluated the market growth. The areas with number zero indicate that the simulator had hit the exact load growth in these places. Please, notice that, sometimes the number zero is hit because the represented area is a forbidden area to new constructions and this type of hit is not computed in the final results.



(a)



(b)



(c)

Figure 6. Snapshots for central area Itajuba County in 2014: (a) forecasting values; (b) real electric market; and (c) Google Earth aerial photo

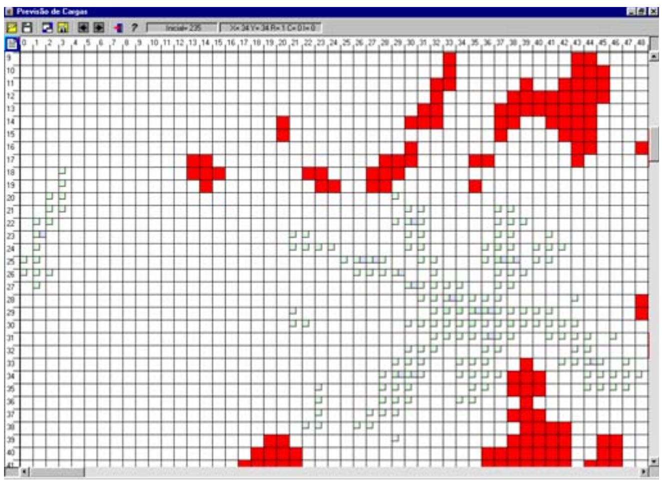


Figure 7. Scenario of forecasted load to 2008 (software window)

The areas with 'x' are places with restrictions on locating loads (as river, parks etc.). It is important to note how the growth in the number of residential consumers has been projected by the observed tax in the 1985-2005 time interval. An over-evaluation was made regarding the total number of consumers; hence the prevailing positive numbers in the space errors matrix. It occurs because Brazil had an impressive economic development between 2005 and 2008, before the 2009 world crisis. The major error, presented in Fig. 8 in the cell with the value +7, occurred because due to the mentioned economic development a first building appeared in this area. Finally, a new set of parameters were obtained training the system with the data from 2005 to 2008. This new set of parameters could be useful to load forecasting during good economic periods and/or to establish future scenarios.

With all information, it is possible to declare the spatial forecasting obtained using the proposed model to the Case 3 of Table I (presented in Figure 6) has an average error of 4.7% with a standard deviation of 2.1%; while the maximum error registered in a single square is +16.2% and only 2.3% of the squares has an error bigger than 10% in a set of more than 90000 squares.

+1	+1	+1	+1	0	0	x	x	0	0
+1	+1	0	0	0	0	0	0	x	0
0	0	-1	-2	+2	0	0	+1	+1	x
+2	-2	+2	-1	+5	+7	+1	0	0	0
0	-2	+3	+2	0	+1	+1	+2	0	0
0	+2	+2	0	+5	+1	+2	0	+1	0
0	+1	+1	-3	+1	+1	0	0	+1	+1
0	+2	+2	+1	0	0	0	0	-2	0
+1	0	0	0	0	+1	+1	+1	0	0

Figure 8. Matrix of spatial distribution error (forecasted vs. actual values for 2008, for residential load)

V. CONCLUSION

The existent methodologies for the space load forecast are based on simple adjustments of time series on land-use, on economic factors or demographic growth, on economic flows and interactions between economic development poles, and on others of lesser importance.

The proposed load evolution model differs from the existent models for two reasons. First, it is a model that makes the basic statement explaining the electric demand evolution process in time and space. Thus, it is looking to understand how the phenomenon works before emulating its behavior beyond the limits of the existent data. Thus, at least by construction, the model is more interpolator than extrapolator.

Second, it is a dynamic model that allows constant interaction of the explaining variables during the whole course of the consumer market evolution process. More clearly, the model may assume totally unpredictable behavior during the process that may certainly have some influence on the direction of the resultant series.

Thus, it is considered that the dynamic model of load evolution simulation is a different way for reproducing the distributor system space-state.

The proposed model divides the load in three main groups: residential, commercial and industrial. These loads provide different types of attractions among them. In our study, only the commercial load is divided in two types: attractive loads and repulsive loads; which play important roles to define new location of residential loads or other commercial loads. It is possible to define other levels of attractiveness for commercial loads, such as very attractive, attractive, indifferent, repulsive and very repulsive loads, just defining new repulsion functions.

Also in this study, only a single type of residential load and industrial load are used for each type of load. However, a similar concept of the commercial load can be used for these two other types of loads. For instance, a new industrial load can be attractive or repulsive in relation to the existing industrial loads; while the residential load can use information of the amount of electric demand, where for example a building can attract another building.

From the forecast point of view, the model corresponds to a focus different from classic forecasting. It is noticed that the adjustment result of the model parameters does not lead up to a point forecast or to system space-state interval, but to an algorithm composed by parameters which are able to generate infinite results similar to the evolution dynamic of a particular market. In other words, it is possible to generate infinite future scenarios for planning and to test how the diversity of these results can affect the decisions taken due to the different configurations that characterize these scenarios. In this way, the sensibility analyses of planning programs will be made based on sets of tests that make alternatives with which better solutions are found.

The simulation model from the proposed electric power consumer market takes into consideration the urban pole theory as the land-use. The resulting dynamic model allows the load growth estimation on time and space by three types of consumers: residential, commercial and industrial. The model also incorporates a set of contour conditions that allows the use of human knowledge with the simulations and knowledge incorporation of urban and regional economy.

Extracting knowledge from actual load series can make the obtained simulations from the proposed model. The adjustment was made using a simulated annealing approach, which used an evaluation function sensitive to either the spatial average error or to the result variability supplied by the parameters used in the simulation.

Many cases have been studied during this long-term project. Many typical situations in different cities provide relevant information to compose the best model for each type of consumer. In this paper, the example of Itajuba, a Brazilian city, was chosen to present a complete study of this city because it has several important features such as three rivers across the city, many mountains around the city, the development project of the city (new locations for industries, graves, state prison, city hall), and so on. Just to have an idea of the evolution of the city during the studied years, the population went up by 25%, eight bridges were built, and the electric market went up by 27%.

APPENDIX

Fig. 9 shows an illustrative example of the construction of a start scenario. Fig. 10 shows the evolution of this scenario for many years. In this example, the adjustment was made between 1993 and 1996, and forecasted to other years ahead.

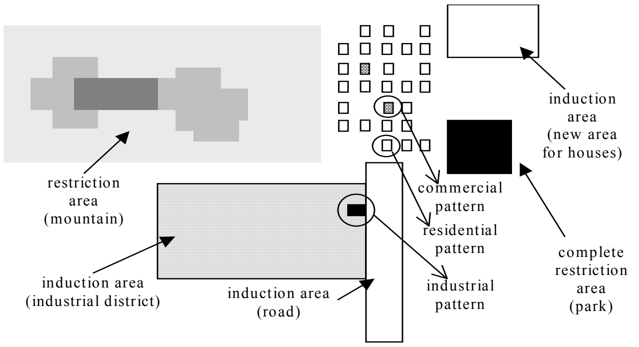


Figure 9. Construction of a new scenario

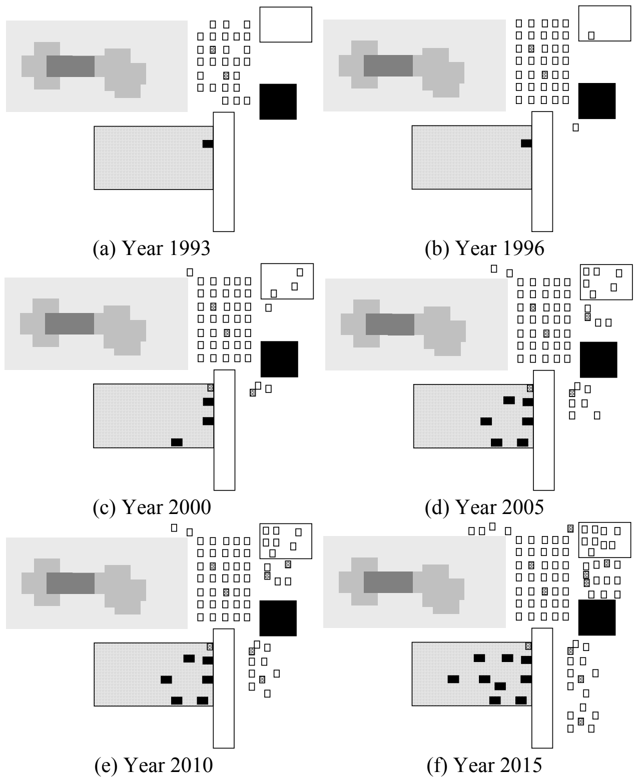


Figure 10. Evolution of the spatial load

ACKNOWLEDGMENT

The authors would like to thank the following Brazilian Research Agencies: CAPES, CNPq, FAPEMIG, and ANEEL R&D for the financial support of this project.

REFERENCES

[1] H. E. Gelani, F. Dastgeer, “Efficiency Analyses of a DC Residential Power Distribution System for the Modern Home,” Advances in Electrical and Computer Engineering, vol.15, no.1, pp.135-142, 2015. doi: 10.4316/AECE.2015.01018.

[2] A. A. A. Esmin, G. Lambert-Torres, “Application of Particle Swarm Optimization to Optimal Power Systems,” International Journal of Innovative Computing, Information and Control, vol.8, no.3A, pp.1705-1716, 2012, Available: http://www.ijicic.org/ijicic-10-12075.pdf.

[3] V.C. do Nacimento, G. Lambert-Torres, C.I.A. Costa, L.E. Borges da Silva, “Control Model for Distributed Generation and Network Automation for Microgrids Operation,” Electric Power Systems Research, vol.127, pp.151-159, 2015. doi: 10.1016/j.epsr.2015.05.025.

[4] H. G. Arango, A. C. Z. Souza, G. Lambert-Torres, A. P. Alves da Silva, “Difference between Regular and Deterministic Chaos Processes Based on Time Analysis of Load: An Example using CEMIG Data,” Electric Power Systems Research, vol.56, no.1, pp.35-41, 2000. doi: 10.1016/S0378-7796(00)00095-X.

[5] H. L. Willis, Spatial Electric Load Forecasting, CRC Press, New York, ISBN 9780824708405, 2002.

[6] Y. Kanemoto, “Evaluating benefits of transportation in models of new economic geography,” Economics of Transportation, vol.2, no.2-3, pp.53–63, 2013. doi: 10.1016/j.ecotra.2012.11.003.

[7] M. Y. Chow, H. Tram, “Application of fuzzy logic technology for spatial load forecasting,” IEEE Transactions on Power Systems, vol.12, no.3, pp. 1360-1366, 1997. doi: 10.1109/59.630482.

[8] A. R Ganguly, K. Steinhaeuser, “Data Mining for Climate Change and Impacts,” Proc. IEEE Int. Conf. Data Mining Workshop, ICDMW’08, IEEE Press, Pisa, pp.385-394, 2008. doi: 10.1109/ICDMW.2008.30.

[9] R. Trasarti, F. Pinelli, M. Nanni, “Mining Mobility User Profiles for Car Pooling,” Proc. 17th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, KDD’ 11, ACM Press, San Diego, CA, pp. 1190-1198, 2011. doi: 10.1145/2020408.2020591.

[10] H. S. Moghadama, M. Helbich, “Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains – cellular automata urban growth model,” Applied Geography, vol.40, pp.240-249, 2013. doi: 10.1016/j.apgeog.2013.01.009.

[11] M. P. Coutinho, G. Lambert-Torres, L. E. Borges da Silva, H. Lazarek, “A Rough Set Classification Algorithm for Detecting Attacks on Electric Power Systems and Other Critical Structures,” The International Journal of Forensic Computer Science, vol.3, no.1, pp.25-32, 2008. doi: 10.5769/J200801002.

[12] A. Mohan, A New Spatio-Temporal Data Mining Method and its Application to Reservoir System Operation, Master Thesis, University of Nebraska, 2014.

[13] M. R. Vieira, V. Frias-Martinez, N. Oliver, E. Frías-Martínez, “Characterizing Dense Urban Areas from Mobile Phone-Call Data: Discovery and Social Dynamics,” Proc. 2010 IEEE 2nd Int. Conf. Social Computing, IEEE SocialCom 2010, IEEE Press, Minneapolis, MN, 241-248, 2010. doi: 10.1109/SocialCom.2010.41.

[14] R. B. Gonzatti, S. C. Ferreira, C. H. da Silva, L. E. Borges da Silva, R. R. Pereira, G. Lambert-Torres, “Smart Impedance: A New Way to Look at Hybrid Filters,” IEEE Transactions on Smart Grid, vol.7, no.2, pp.837-846, 2016. doi: 10.1109/TSG.2015.2466613.

[15] J. Smith, M. Rylander, L. Rogers, R. Dugan, “It's All in the Plans: Maximizing the Benefits and Minimizing the Impacts of DERs in an Integrated Grid,” IEEE Power and Energy Magazine, vol.13, no.2, pp.20-29. doi: 10.1109/MPE.2014.2379855, 2015.

[16] M. Fujita, P. Krugman, “When is the economy monocentric?: von Thünen and Chamberlin unified,” Regional Science and Urban Economics, vol.25, no.4, pp.505-528, 1995. doi: 10.1016/0166-0462(95)02098-F.

[17] M. Berliant, M. Fujita, “The Dynamics of Knowledge Diversity and Economic Growth,” Southern Economic Journal, vol.77, no.4, pp.856-884, 2011, Available: http://www.jstor.org/stable/23057315.

[18] Y. Kanemoto, “Second-best cost–benefit analysis in monopolistic competition models of urban agglomeration,” Journal of Urban Economics, vol.76, pp.83-92, 2013. doi: 10.1016/j.jue.2013.03.006.

[19] M. Fujita, “The Evolution Of Spatial Economics: From Thünen To The New Economic Geography,” The Japanese Economic Review, vol.61, no.1, pp.1-32, 2010. doi: 10.1111/j.1468-5876.2009.00504.x.

[20] G. Duranton, M. A. Turner, “Urban Growth and Transportation,” Review of Economic Studies, vol.79, no.4, pp.1407–1440, 2012. doi: 10.1093/restud/rds010.

[21] E. L. Glaeser, S. P. Kerr, W. R. Kerr, “Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines,” Review of Economics and Statistics, vol.97, no.2, pp.498-520, 2015. doi: 10.1162/REST_a_00456.

[22] F. Kung, P. Wang, “A spatial network approach to urban configurations,” Canadian Journal of Economics, vol.45, no.1, pp314–344, 2012. doi: 10.1111/j.1540-5982.2011.01697.x.

[23] H. G. Arango, G. Lambert-Torres, A.P. Alves da Silva, “Load Evolution for Power System Distribution Planning Studies: An Approach Using Spatial Disorder,” Proc. IEEE Int. Conf. Systems, Man and Cybernetics, IEEE SMC-1996, IEEE Press, Beijing, pp.1910-1915, 1996. doi: 10.1109/ICSMC.1996.565403.

[24] M. Berliant, P. Wang, “Dynamic Urban Models: Agglomeration and Growth,” Contributions to Economic Analysis, vol.266, pp.531-581, 2004. doi: 10.1016/S0573-8555(04)66017-2.

[25] G. Grigoras, G. Cartina, E. C. Bobric, "Strategies for Power/Energy Saving in Distribution Networks," Advances in Electrical and Computer Engineering, vol.10, no.2, pp.61-64, 2010. doi: 10.4316/AECE.2010.02010

[26] L. Ingber, “Very fast simulated re-annealing,” Mathematical and Computer Modelling, vol.12, no.8, pp.967-973, 1989. doi: 10.1016/0895-7177(89)90202-1.

[27] Google Earth Pro - Available: http:// www.google.com.br/intl/pt-BR/earth/, accessed in March 3rd, 2015, typing Itajubá-MG.