FUZZY INFERENCE IN SPATIAL LOAD FORECASTING

Vladimiro Miranda Cláudio Monteiro
vmiranda@inescn.pt cmonteiro@inescn.pt

INESC Porto, Instituto de Engenharia de Sistemas e Computadores
and DEEC/FEUP, Faculdade de Engenharia da Universidade do Porto, Portugal
http://power.inescn.pt/

Abstract – Forecasting electric demand and its geographical distribution is a prerequisite to generate expansion planning scenarios for distribution planning. This paper presents a comprehensive methodology that uses a fuzzy inference model over a GIS support, to capture the behavior of influence factors on load growth patterns and map the potential for development. The load growth is spread over maps with cellular automata. The interaction with a scenario generator inputs data into a graph generator, which will serve as a basis for more classic network planning tools.

Keywords – Spatial Load Forecasting, Geographic Information Systems, Fuzzy Inference, Cellular Automata.

1. INTRODUCTION

In distribution planning, research has been concentrated for many years in developing algorithms to optimize network design. In the last years, however, the concept of optimization has been challenged.

Two things contributed for this: the development of multiple criteria models and the emergence of the risk analysis paradigm, in the context of multiple future possible scenarios. Also evolutionary computing techniques have taken over other methods and given opportunity for the development of really comprehensive truly dynamic in time.

The first author was deeply involved in this movement of change of point of view. The work in [1] reported the first Genetic Algorithm model for Distribution Planning and established the bases for future evolutionary computing approaches, including multiple criteria e a tree of futures. This work already presented a decision choice procedure based on the minimization of decision risks. But the clear distinction between risk analysis and other decision paradigms was to be made clear in other papers such as in [2].

It is now clear that we have witnessed the completion of a phase in solving the Distribution Planning problem. It must be recognized, however, that the departing point is still the same as 20 years ago: data must include a definition of nodes (for possible injections and loads) and a graph (for possible lines to be built thus forming and expanding the network).

We know how to solve the network design problem in the space of graphs. However, how does one define such graph to begin with?

The key factor in distribution planning is load forecast - but in distribution planning we deal with vast geographical regions, not with a few nodal loads. Utilities around the world are increasingly adopting Geographic Information Systems (GIS) to serve as the background digital representation of regions and networks.

Land-use spatial load forecast (SLF) simulation methods have been used to model the process of the load growth in order to predict load evolution in a spatial and temporal basis [3]. This methodology is particularly suited to high spatial resolution for long range forecasting and multi-scenario planning [4].

The challenge that researchers face presently, in the distribution planning problem, is to derive a consistent way of establishing an acceptable correspondence between the geographical space and the space of graphs, so that the geographical model may generate an adequate input to the network design models. The reverse projection of the results of such models into the geo-referenced space is quite straightforward.

In this paper is we describe an approach to generate adequate spatial load forecasts in regions and to generate graphs (with nodes and branches) representing possible investments for system development, which may serve as input for an evolutionary computing algorithm to proceed with actual system design proposals. The aim is to allow an automated (as much as possible) generation of robust solutions, that may be good regardless of the future or acceptable in a majority of credible future scenarios.

The model is based:

1. on a fuzzy inference engine (built after neuro-fuzzy concepts) which extracts knowledge from past data and produces a set of rules that condition demand growth in a region
2. on a cellular automata engine that helps in spreading to a whole map the effects of rules
3. on a scenario generator that allows building a tree of futures
4. on a grid generator that prepares graphs with potential branches and nodes to be fed into network design and optimization tools.

This allows one to generate maps representing potential for growth, from which one may build, for a succession of time stages, maps of forecasted load. From these, one may finally

derive a graph of nodal loads and of potential system branches - the departing point for final network design.

2. SPATIAL FORECASTING

Spatial Load Forecasting (SLF) refers to models used to predict load growth in a region based on the influence of several control factors, defined as "influence factors". Examples of those factors are, for instance, the influence of a radial distance to an urban center or of the distance to a waste treatment center.

In recent years some works have enhanced the land-use methods applied to urban redevelopment, using fuzzy logic, GIS, multi-objective programming[5, 6, 7].

In urban planning, a large work is being done to model land-use conversion. This is done under the assumption that a simulation approach under the self-organization paradigm is appropriate for addressing the process of land development [8, 9]. To simulate the dynamics of the process, a recent work adopts cellular automata (CA) - this approach emphasizes the way in which locally-made decisions give rise to global patterns.

3. THE FUZZY INFERENCE MODEL

3.1 Influence factors, development and rules

The module described in this section is a spatial model that uses a set of explanatory geographical variables, designated as spatial influence factors, to predict the potential for development, for a specific consumer class.

Spatial factors to be considered in SLF are local structural factors, relative location factors, and neighborhood factors. The structural factors are variables that affect the site unit they are primarily associated with, and their effect is confined to the geographic unit boundaries (e.g. slope, altitude, land use classification).

The distance to certain geographical features defines the relative location factors; for instance, proximity to roads, proximity to urban centers.

The neighborhood effect represents the influence of entities or features in an adjacent area or in its own geographic unit. Important neighborhood factors are load saturation levels: they models the dynamics of change from stage to stage influencing the results on each following time stage.

Each point in a map is associated with saturation curves S_i for each type i of consumer. A saturation curve describes the number of consumers of a given type, at a certain location, as a function of time. It usually displays a S shape (figure 1).

In our approach, we do not define a priori any saturation curve. Instead, the shape of the saturation curve is built dynamically as a function of all geographical influence factors and the dynamic interaction between different consumer classes. This comes as an output of the inference process, and the rule learning procedure also allows the learning of saturation growth at each location and its weight.

The result of the module is a map for the potential for development (PfD): it is a continuous map of values between zero and a maximum value representing the maximum possible growth in one stage for a geographic unit.

The function that models the growth phenomena for each consumer class i could be represented by

$$\dot{S}_c\Big|_t = f\left(S_1\Big|_{t-1},\ldots,S_m\Big|_{t-1},I_1,\ldots,I_p\right)$$

where $\dot{S}_c$ represents the output of the model (potential-for-development for consumer class c) and is the differential of the saturation curve on stage t. The S_i represent saturation levels for each consumer class at stage t and the I_i represent other geographic influence factors.

In our approach, the functions that relate saturation and its derivative or PfD are established by the rules in the fuzzy inference engine. These functions are represented as a set of thousands of fuzzy rules. Here's one example of a rule:

IF *(distance to road is CLOSE)* ***AND***
(distance to urban center is MODERATE CLOSE) ***AND***
(terrain slope is MODERATE) ***AND***
(domestic saturation is MEDIUM) ***AND***
(industrial saturation is LOW)
THEN
Domestic PfD is 20 consumers per stage per km^2 ***AND***
Industrial PfD is 0.1 consumers per stage per km^2

These rules are automatically generated and used by the spatial model and are easily understood by human specialists. The rules are stored in the GIS database and are used as in a lookup table in the process.

The geographical data are stored as a structure of grids of values (raster structures). In the process the system checks the rules activated in each location. If the rule is activated, an activated value G_j is computed. At the same location several rules will be activated simultaneously and the result will be a weighted sum of their activated values.

The fact that these rules can be generated and understood simultaneously by the system and by human experts is a great advantage because historic and human knowledge can be joined and interpreted.

As we will see in a further section, we employ a cellular automata module to compute the development based on

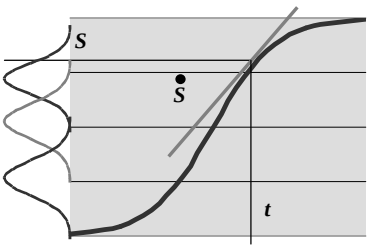

Figure 1 – The number of consumers as a function of time defines the Saturation curve at a location; its derivative is seen as the potential for development PfD. The saturation level at stage t is characterized through fuzzy descriptors.

potential for development maps. Moreover, these development maps are used to compute the new saturation levels for each consumption class.

3.2 Training or tuning the neuro-fuzzy rule model

To capture the regression (function approximation) between geographic influencing factors and preference (development) we have developed a geographical fuzzy inference model. Some conventional SLF methods use simple linear or polynomial regressions, without geographical partitioning [3]. In some cases the user specifies the weights parameters; more advanced methods use geostatistics to compute these relations. Examples of these methods are geographical weight regression and expansion methods [10].

The fuzzy systems deal with qualitative information allowing the implementation of linguistic descriptions for influence factors (close to the road; far from urban center; many houses; few industry). Other advantage of a fuzzy approach is its capacity for generalization, allowing information aggregation and extrapolation to other space-time scenarios with less descriptive information.

Many variants and operations can be used in fuzzy-logic inference [11]. This section will describe briefly the technique we implemented on a GIS; for more detailed information on fuzzy logic see references [12].

A basic component of a fuzzy inference system is a fuzzy rule. The rules are expressed using linguistic labels such as the rule: IF (road is close) AND (urban center is close) THEN (development growth index is 0.8). Fuzzy membership functions (MFs) associate linguistic labels (e.g. close) with a particular area of one of input or output variable. In our case, the THEN-part of each rule does not consist of a membership variable but of a crisp value 0.8. This is called a zero-order Sugeno fuzzy system. In an n-th order Sugeno fuzzy system the THEN-part of each rule consists of a polynomial of degree n in the input variables.

Different shapes of the MFs can be proposed such as triangular, trapezoidal, or Gaussian, for instance:

$$m_{iv}(x_v) = \exp\left(-\frac{(c_{iv} - x_v)^2}{2\sigma_{iv}^{\ 2}}\right) \qquad (1)$$

where i denotes the index of the different MFs defined for variable v and x_v denotes the input for variable v. The parameters c_{iv} and σ_{iv} are the center and the "width". In order to minimize the number of layers used on the GIS implementation, for each input value only two membership functions have values higher than zero and their sum may be one. This can be achieved by adopting an adequate normalizing procedure.

As input variables we can have three classes of influence factors: distance factors (e.g. distance to roads or to urban centers); zone-count factors (number of houses on a 5 km radius); local factors (e.g. terrain slope, urban planning directives). As output we get a map of development potential for each consumer class (number of additional consumers).

After the MFs definition we can formulate the rules j in term of linguistic values. Input variables are combined in

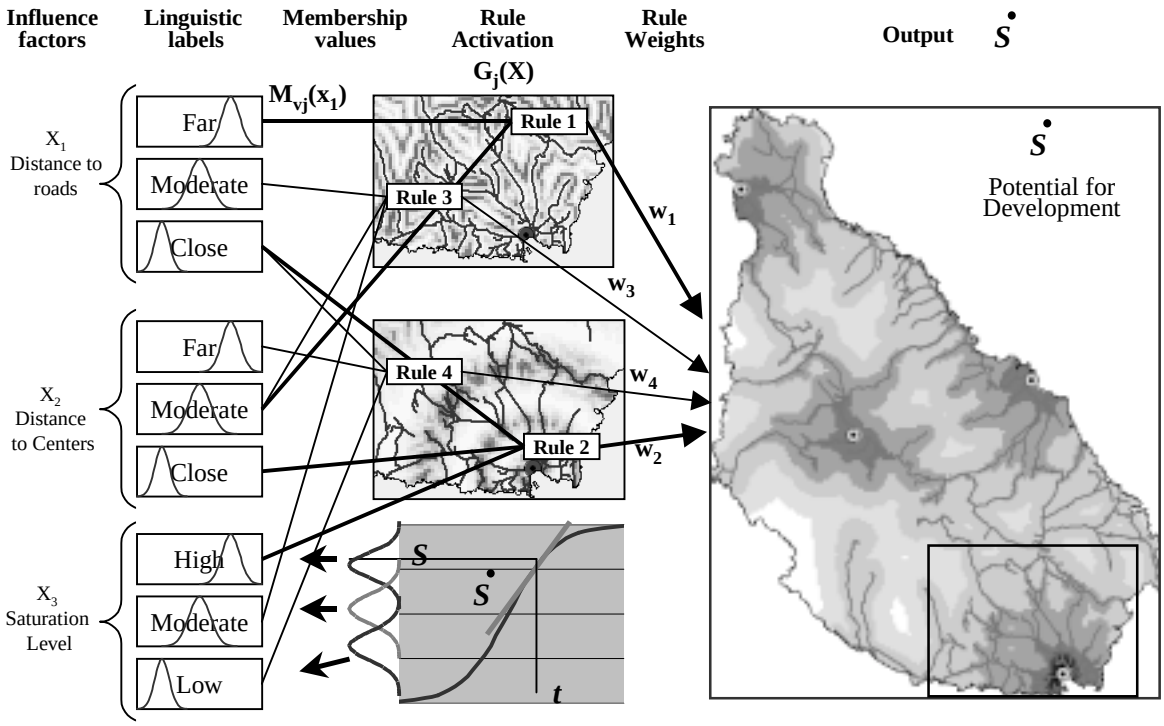

Figure 2- Illustration of the fuzzy inference process - on each map location, membership functions are activated by input values for several influence factors; several layers of rule zones are mapped; the weights of rules are applied to each case and a map of PdF is generated for each consumer class - the darkest zones in the map on the right represent the zones with higher PdF (near centers and near roads).

expressions using fuzzy operators such as fuzzy AND (T-Norm) or fuzzy OR (T-conorm). In the case of Gaussian MF the fuzzy AND can be performed by the arithmetic \product of membership values across the input variables x_v.

$$G_j(x_v) = \prod_v M_{iv}(x_v) \tag{2}$$

For each consumer class *c* the output value is calculated by the OR operation and can be generated by

$$O_c(x_v) = \sum_j w_j G_j(x_v) \tag{3}$$

where w_j is the THEN-part (or output weight) of the fuzzy rule *j*. The output weights w_j can be set manually by domain experts. Alternatively, a given training data set

$$D = \left(\xi^k, \zeta^k\right),\ k = 1,...M,\ \xi^i \in \Re^n,\ \zeta \in \Re^m \tag{4}$$

could be used to perform training, where *M* is the number of training points, *n* is the number of influence factors and *m* is the number of consumer classes. This training would find the output weights that minimize the summed square error.

$$E = \frac{1}{2}\sum_{k=1}^{M}\left(O_k\left(\xi^k\right) - \zeta^k\right)^2 \tag{5}$$

If the IF-part of the fuzzy rules is fixed, the determination of weights w_j can be solved by the method of least squares based on standard matrix techniques

$$\begin{bmatrix} w_1 \\ \vdots \\ w_j \end{bmatrix} = \begin{bmatrix} \sum_k G_1(\xi^k)^2 & \cdots & \sum_k G_1(\xi^k)\cdot G_j(\xi^k) \\ \vdots & \ddots & \vdots \\ \sum_k G_j(\xi^k)\cdot G_1(\xi^k) & \cdots & \sum_k G_j(\xi^k)^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_k G_1(\xi^k)\cdot \zeta^k \\ \vdots \\ \sum_k G_j(\xi^k)\cdot \zeta^k \end{bmatrix} \tag{6}$$

When implemented on GIS the rules are coded as map regions. The number of rules activated in each geographical location is 2^v (the same number of maps is needed to store rule coding), where *v* is the number of input variables. For each rule map we compute a stack of maps with membership values $G_j(x_i)$.

These membership values are functions of the geographical value for the input variables (influence factors). All the calculations associated with each rule *j* are computed based on zonal functions available on GIS, in which the zones are the regions where rules are activated.

The set of rules (variable labels, codification index and weights) is stored on a GIS database to be used on other time-space scenarios.

The obtained maps of development potential are continuous on space and static in time. To solve the SLF problem, which is discontinuous and dynamic, we use a cellular automata model.

4. CELLULAR AUTOMATA

The CA theory was first introduced by Jon Von Neuman [13] and is ideally applied for dynamic discrete modeling [14]. A CA is a discrete dynamical system because space, time and system states are discrete and these states change sequentially over time and space. Each point in a rectangular spatial grid, called a cell, can have any one of a finite number of states. The states of the cells in the lattice are updated according to a local rule, which depends on the cell state and the state of its neighbors at the previous time step. The state of the entire lattice is updated synchronously in discrete time steps.

In our formulation at any specific point of time t, the CA automaton is a collection of binary states e_{ij}^t in cell location (i,j), with value 1 if the new consumer is added to the site and 0 if no consumer is added.

$$CA = \left\{e_{ij}^t\right\}\ \ 0 < i \le r;\ \ 0 < j \le c;\ \ \forall e_{ij}^t \in E \tag{7}$$

where E is the finite set of states, r and c are the number of rows and columns of the map grid.

The CA is an iterative process computing development based on potential for development and computing new potential based on previous iteration development.

The Potential for Development (PfD) is initially set by the fuzzy system. The PfD is represented as a stack of continuous maps, one for each consumer type, representing the potential growth number of consumer per stage and per geographic unit (e.g. 20 domestic consumers per stage and per km^2). The Development, which is the output of the CA, represents the effective number of consumer growth. A global geographical trending controls the global development, the sum of all developments in the region. The CA process finishes when the sum of all cell developments reaches the global trending value (e.g. the growth for year 2001 in the whole region tends to 250 industrial consumers and 5000 domestic consumers).

The iterative process of the CA is based on state transitions $S_i(t)$; in our model, these will be transitions from non developed to developed. The state transition is done according to a set of rules such as

$$\text{if}\ \ P_i(t) > P_b(t)\ \ \text{then}\ \ S_i(t) = 1\ \ \text{else}\ \ S_i(t) = 0$$

In our model a transition exists if the cell has a PfD value $P_i(t)$ higher than a specified boundary value $P_b(t)$. The boundary value is specified by the system by ranking PfD intervals.

The development $D_i(t)$ is recalculated in each iteration incrementing the number of consumers, by steps D_{step} (measured in number of consumers), only on cells marked as developed $S_i(t)=1$.

$$D_i(t) = D_i(t-1) + S_i(t)\cdot D_{step} \tag{8}$$

The new potential $P_i(t+1)$ is recalculated based in three components:

- positive feedback of the cell on the previous iteration, weighted by α;
- neighborhood effect based on the 8 adjacent neighborhoods [15], weighted by β;
- innovation factor modeled as random noise, weighted by λ;

and is given at time t+1 by

$$P_i(t+1) = \alpha \cdot P_i'(t) + \beta \cdot \frac{1}{8} \cdot \sum_{j \subset \Omega_i} P_j'(t) + \lambda \cdot \varepsilon_i(t) \qquad (9)$$

where α, β and λ are the weights for each component, with values [0,1] and $\alpha+\beta+\lambda=1$, and Ω is the set of adjacent neighbors cells. $P'_i(t)$ is the updated potential to development in time stage t on site i, computed based on the output of the fuzzy inference model $P_i(0)$ and on the development computed by the CA on iteration t:

$$P_i'(t) = P_i(0) - D_i(t)$$

At the end of each stage the PfD maps may be recalculated, using the fuzzy inference model and the new geographic data computed with the CA or introduced by the planner.

5. SCENARIO GENERATOR

A scenario is a sequential set of stages (time sequence) including the input data and parameters to be used in each stage. The scenario generation requires the participation of the Planners specifying some of geographical input data (e.g. road coverage, land use impositions) and parameters for each stage (global trending and parameters α, β and λ). Some of the input data are automatically calculated by the system or by other modules. For instance, the saturation level is recalculated used as input on stage t is recalculated with the results of development from stage t-1. Other input data, as road coverage, may come from external planning models or simply by interaction with the planner.

Multiple options on input data for a specific stage give place to a tree of scenarios that can automatically generated by a scenario generator module. For instance, two possible values for global trending at stage t and the event of building or not a bridge at stage t+2 originate 4 different scenarios.

For each scenario a sequence of the fuzzy system module and the CA module will run as many times as the number of stages. The result is a sequence of geographical maps with consumer growth along the several stages. The maps of consumption are computed based on the number of consumers and on End Use models that define the consumption behavior of each consumption class.

6. NETWORK GENERATOR

The SLF model is the data feeder for several models more familiar to network designers. The load scenarios describing maps of forecasted values along time stages are the base data for the following expansion planning modules:

- secondary network routing
- substation siting
- substation service area optimization
- routing for lines interconnecting substation

The main objective of these modules is the pre-processing of geographical data in order to obtain a set of robust feature options to be used in automated planning tools. These options are basically electric facility sizing and location (e.g. sizing and for lines and substations, routing for lines and geographical siting for substations). It is known that facility location may be highly affected by demand uncertainty and its geographical distribution.

The referred modules build a vectorial topology (based on graphs and nodes) and related databases to store attributes associated with each option. In fact, the procedure allows the extraction of information from the geo-environment ant its storage as attributes of the network features.

7. EXAMPLE

In this section we illustrate the application and results of the SLF model implemented in ArcView GIS and programmed in Avenue. This study presents the forecasting value for domestic consumption in the island of Santiago in Cabo Verde (Africa). The result is one forecasted scenario with eleven stages, which were obtained for illustration purposes and cannot be seen as reflecting the actual situation in Santiago.

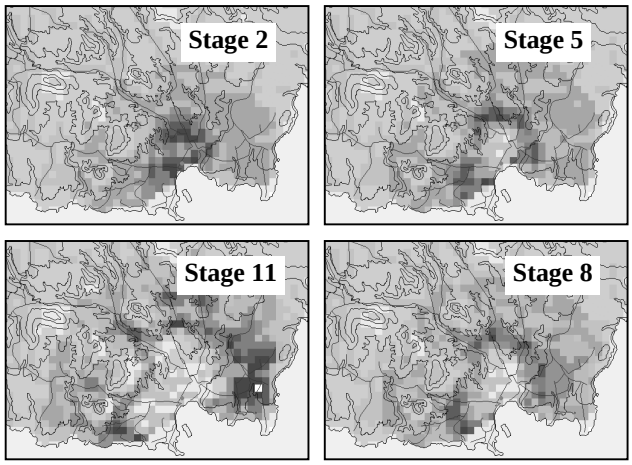

Figure 4 – Maps of PfD, a sample of 4 out of the 11 stages obtained with the fuzzy inference engine. Notice that at the center there is a growing saturation effect and that at later stages the potential for development concentrates mainly in peripheral zones, following roads and avoiding high slopes.

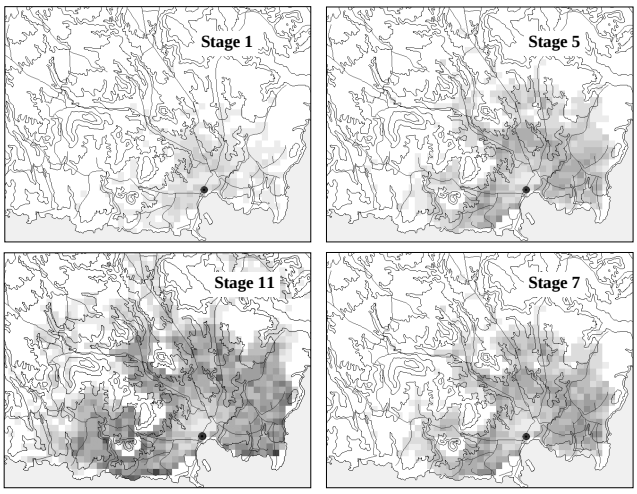

Figure 4– Maps of forecasted number of consumers (of domestic type) after the action of the cellular automata.

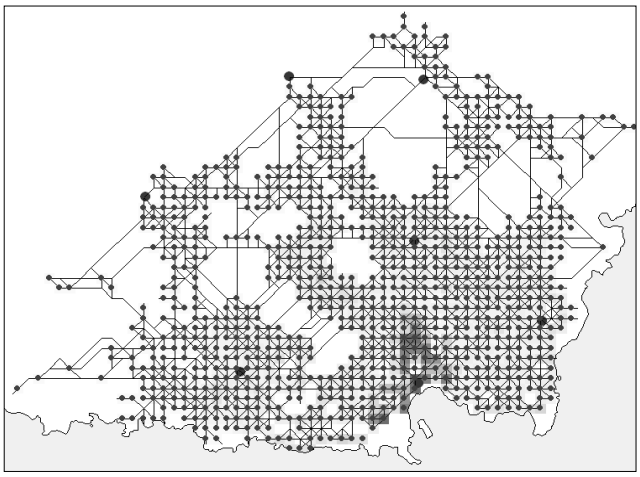

Figure 5– Example of output from the graph generator, for stage 11, taking in account several possible scenarios organized with the scenario generator module. The graph is meshed, and possible sites for substations are identified as large black dots.

The geographical inputs (influence factors), are the following:

- Distance to main urban centers (4 linguistic labels)
- Distance to secondary urban centers (4 linguistic labels)
- Saturation Level (6 linguistic labels)
- Distance to roads (5 linguistic labels)
- Distance to coast line (3 linguistic labels)
- Terrain slope (4 linguistic labels)

Linguistic labels associated with fuzzy membership functions reclassify the influence factor values (e.g. distance to roads between 0 and 2 km: VERY CLOSE; distance to roads between 1 and 3 km: MODERATE CLOSE).

The study region has 2400 km^2 including one main urban center and three secondary centers. The resolution on GIS spatial analysis was 250m which represents cell based maps with 38400 cells. The historical growth is based on the geographical building growth along the last 30 years.

In a 200 MHz Pentium II, the module to generate fuzzy rules (fuzzy inference module) spent approximately 3h30m to generate 2640 fuzzy rules. To compute the eleven stages of the scenario for all study region the fuzzy system and the CA spent approximately 2h40m.

The results are a set of 11 maps representing PfD and 11 maps with the growth in number of consumers. An animation of these results, displaying the time sequence evolution, may be found in http://power.inescn.pt/claudio/.

8. CONCLUSIONS

The research emphasis in automated distribution planning must nowadays be focused on the difficult task of establishing a correspondence between comprehensive geographical representations and the design optimization algorithms, which require data in the form of a graph with nodes and branches.

This paper demonstrates that coupling GIS tools with fuzzy inference engines may allow building the desired integrated environment for engineers, planners and decision makers.

REFERENCES

[1] V. Miranda, L.M. Proença, J.V. Ranito, “Genetic algorithms in optimal multistage distribution network planning", in IEEE Transactions on PWRS, Nov 94

[2] V. Miranda, “Why Risk Analysis outperforms Probabilistic Choice as the effective decision support paradigm for power system planning", IEEE Transact on PWRS, v.13, n.2, May1998

[3] H. L. Willis, Spatial Load Forecasting, Marcel Dekker Inc., New York, 1996. ISBN: 0-8247-9425-7

[4] H. L. Willis, Power Distribution Planning Reference Book, Marcel Dekker Inc., New York, 1997. ISBN: 0-8247-0098-8

[5] M.Y. Chow, H. Tram, "Methodology of Urban Re-Development Considerations in Spatial Load Forecasting", IEEE Transactions on Power Systems, Vol.12, No2, May 1997, pp. 996-1001

[6] M.Y. Chow, H. Tram, "Application of Fuzzy Logic Technology for Spatial Load Forecasting", IEEE Transactions on Power Systems, Vol.12, No3, August 1997, pp. 1360-1366

[7] M.Y. Chow, H. Tram, "Application of Fuzzy Multi-Objective Decision Making in Spatial Load Forecasting", IEEE Transactions on PWRS, Vol.13, n.3, August 1998, pp. 1360-1366

[8] H. Couclelis, "Cellular Worlds: a framework for modelling micro-macro dynamics", Environment and Planning A, No17, 1985, pp. 585-596

[9] M. Batty, “Cellular automata and urban form: a primer”. Journal of Am. Planning Association, No63, 1997, pp. 266-274

[10] A.S. Fotheringham, M.E. Charlton, C. Brunsdon, "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis.", *Environment and Planning A*, No30, 1998, pp. 1905-1927.

[11] D. Driankov et al., *An introduction to fuzzy control*, ed. Springer-Verlag, 2a. edition, 1996

[12] H.J. Zimmermann, Fuzzy Set Theory and Its Applications, Kluwer Academic Publishers, Boston, 1996.

[13] J. Von Neuman, *Theory of Self-Reproducing Automata*, edited by A.W. Burks, University of Illinois Press, 1966

[14] S. Wolfram, *Cellular Automata and Complexity*, Addison-Wesley Publishing Company, 1994, ISBN: 0-201-62664-0

[15] T. Toffoly, N. Margolus, *Cellular Automata Machines*, MIT Press, Cambridge, 1996

Vladimiro Miranda received his Licenciado, Ph.D. and Agregado degrees from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 1977, 1982 and 1991, all in Electrical Engineering. In 1981 he joined FEUP and currently holds the position of Professor Associado Agregado. In 1985 he joined also INESC, a research and development institute. He is currently the President of the Scientific Board of INESC Porto and Executive Board Adviser for Power Systems. He has been a member of several Expert Committees in Power Systems and has had responsibility over several research projects namely within the European Union programmes; he is the chairman of the Luso-Afro-Brazilian Cooperation Network in Power. He has authored or co-authored many papers, namely in his areas of interest, related with Power System planning and the application of fuzzy sets and other soft computing techniques to Power Systems.

Cláudio Monteiro was born in France, on March 14th, 1968. He received his Licenciado and MSc. degrees from FEUP in 1993 and 1996, in Electrical Engineering and Computers. In 1993 he joined INESC as a researcher in the Power System Unit. Presently is working in his Ph.D. related integrating distribution planning models and Geographical Information Systems under uncertain reasoning.