# Short-term electric load forecasting in uncertain domain: A fuzzy decision tree approach

ABÍMBỌ́LÁ R. ÌYÀNDÁ[1]
ỌDẸ́TÚNJÍ A. ỌDẸ́JỌBÍ [1]
A. O. KỌ́MỌLÁFẸ́[2]

Ọbáfẹ́mı Awólọ́wọ̀ University,
Ilé-Ifẹ̀, Ọ́sun State, Nigeria.
Computer Science & Engineering Department[1]
Electrical & Electronic Engineering Department[2]
[1] (abiyanda,oodejobi)@oauife.edu.ng
[2] ookomolafe@oauife.edu.ng

**Abstract.** The objective of the research reported in this paper is the development of a model for short term load forecasting for use in an environment characterized by uncertainty. The fundamental requirement for the proposed model is the production of robust and accurate performance with minimal computational and data resources. Our solution strategy was developed around a computational intelligence method which exploits knowledge using fuzzy logic and decision tree based techniques. The model was developed and evaluated using three years data (i.e. 2004, 2005 and 2006) on electric loads obtained from the National Control Centre (NCC) Òṣogbo, Nigeria and was implemented using the Fuzzy Decision Tree software (FID 4.2). The data was supported by knowledge elicited from experienced power monitoring staff at NCC. The results showed that the average fractional forecast errors for the proposed model on selected data from the three years was 0.17 while that of the conventional multiple regression model was 0.80.

**Keywords:** Short term load forecasting, Fuzzy decision tree, uncertain domain.

(Received October 19th, 2011 / Accepted December 22nd, 2011)

## 1 Introduction

The problem of electric power load forecasting (EPLF) is as old as the electric power technology itself. Unlike the electric power technology, however, the EPLF problem continues to grow in complexity as the challenges of meeting customers' demands continue to increase in dimension and dynamics. The dimensions of physical environment, technology, culture, policy and global trends towards the need for sustainable energy as well as related security issues, combine and culminate to determine the amount of electric power that will be consumed. Therefore, the electric load forecast problem continues to change in form and pattern in an uncertain manner making its description vague and sometimes intractable. An accurate EPLF system can be a powerful tool particularly in environment where the power supply is very limited, such as in many Africa countries.

The result of an accurate load forecast could be used to provide advance information to consumers about possible power outage. Such information could help the consumers to better manage their activities and reorganized their schedules. For example, under extreme emergency conditions, shedding of the non-essential loads (like air conditioning), can be performed in a timely manner, if the predicted load is known well in advance. Electric power consumers in the essential services sector (e.g hospitals) could use the advance knowledge of possible power outage to put in place contingency or alternative plans, such as the re-scheduling of critical operations. In addition, a good short-term load forecasting (STLF) data could assist the electric power provider to more accurately determine the unit price of power supply.

This work makes two important contributions to the present state of knowledge in the area of electric load forecasting: (i) the identification and specification of load forecasting problem in an uncertain domain, and (ii) the use of Fuzzy Decision Tree (FDT) method for modeling the load forecasting problem.

In order to present the proper perspective of our approach, a brief description of the characteristics of the electric load forecasting problem in an uncertain domain is provided in the next subsection. Section 2 contains a review of the literature on the computational approaches to electric load forecasting while Sections 3, 4 and 5 contain the description of the methodology for our fuzzy decision based approach. In Section 6, we discussed the evaluation of our approach and section 7 concludes this paper.

### 1.1 The STLF problem in uncertain domain

Three types of load forecasting have been identified in the literature [4, 31]: *Short-term, Medium-term* and *Long-term*. STLF attempt to predict electric load demand in advance from one hour to one week while medium-term load forecasting attempt to predict for between one month and one year. Electric power load forecast for more than one year are usually termed long-term load forecasting. Our focus here is on the typical 24-Hour STLF problem [8]. STLF is a complex task, because available generation must match customers' demands on an instantaneous basis.

In an uncertain environment, such as Nigeria, the forecast must respond to unforeseen events related to weather, societal and cultural as well as possible power generation instability problems. The current practice of load forecasting in Nigeria, and probably in most technologically developing countries, is to rely on the knowledge of experienced power monitoring operators. In Nigeria, the operators responsible for this task work at the National Control Centre (NCC) and they use mostly manual calculations and intuitive methods with limited data. Based on our interviews with the experienced operators, we found that the electric load forecast is generally based on the ratio of maximum transformer loading of each station to the overall transformer loading of all the stations. This value is scaled up by multiplying it by maximum generated power at a particular time.

Load components come from the electric power demands of houses, shopping centres, markets and administrative centres. For larger cities, demand may include residential districts and industrial zones. The load component associated with these geographical locations are often specified in linguistic terms rather than precisely defined mathematical terms [6]. The fundamental attribute of the STLF problem in this context can be summarized with three terms: uncertain, imprecise and nonlinear. In order to provide a more realistic forecast of the actual load situation we need to exploit all the information provided as much as engineering modeling can permit.

To this end, the STLF system required in this environment should be developed with the following attributes in mind: (i) the historical load data available are not adequate; (ii) load behaviour are highly prone to sudden and uncertain changes which cannot be adequately rendered by numerical data as some salient bahaviour of the STLF problem can only the expressed linguistically based on personal experience; (iii) a model that could be used to generate and document the STLF process as well as assist in obtaining an understanding of the dynamics of load in such domain.

## 2 Related work

Computational approaches to STLF can be grouped into three major classes: data driven, rule driven and hybrid.

### 2.1 Data driven approach

In the data driven approach, historical data of electric load consumption pattern are processed statistically or modeled using an automatic prediction process. This approach usually requires a mathematical model that represents load as a function of different factors such as time of the day, weather, and customer class. Multiple linear regression [36], adaptive and/or general exponential stochastic time series [3, 23] based techniques are usually employed in this approach. More recently, however, data driven techniques from soft computing such as the genetic or evolutionary algorithms (GA) [7] and Artificial Neural Network (ANN) [5, 33] have become popular. The large amount of data require to achieve acceptable performance makes these model unsuitable for our application.

Another very powerful data driven technique which has the desirable attribute of explainability is the decision tree, also called the *Interactive Dichotomizer 3* (ID3) [24]. Its use in various areas of electric power related technologies such as modeling [35, 15], prediction [28], control [18, 30], management [16], amongst others, has been demonstrated. The use of ID3 is motivated by five of its important attributes: (i) only a small set of inputs is required to produce sufficiently accurate prediction; (ii) ID3 handiness for engineering decision making; (iii) ID3 is computationally fast and always provides definite results; (iv) ID3 are simple and inexpensive to develop; and (v) it is ease to extract rule from ID3, making it easy to explain the STLF process. The ID3 technique, however, has some shortcomings which include the binary nature of its knowledge based partitioning and the pruning algorithm complexity. These weaknesses renders it unsuitable for our application. An alternative tree building method is the classification and regression trees (CART) [29] algorithm in which the regression technique is used in tree generation. However, CART has similar shortcomings to those of the ID3 technique.

### 2.2 Rule driven approach

In the rule driven (also called rule or knowledge based) approach, inference are developed around a set of assertions, which collectively form the *working memory*, and a set of rules that specify how to act on the assertion set. A number of STLF system have been developed around the rule driven approach usually in the context of expert system or decision support systems [26, 25]. The following attributes of the rule based approach makes them attractive for our application: (i) ability to manage a few data samples, (ii) availability of experts for consultation, (iii) rule bases are modular and therefore facilitates easy modification.

However, the binary nature of rules makes it unsuitable to model the uncertainty and linguistic concepts that characterizes the STLF domain. To address this problem, the fuzzy logic techniques has been introduced into the basic rule driven construct. Fuzzy logic is a generalization of the basic binary logic [8] but unlike in the binary logic, truth-values are assigned to variables in the range $[0, 1]$. Intuitively, a fuzzy set defines a class which admits the possibility of partial membership in itself. For example, if $X = \{x\}$ denotes a space of objects, the fuzzy set $A \in X$ is a set of ordered pairs $A = \{x, \mu_A(x)\}$, where $\mu_A(x)$ is the degree to which $x$ belongs to $A$. If the function $\mu_A(x)$ returns the value $0.0$ then $x$ does not belong to $A$ at all but if the value returned is $1.0$ then $x$ is totally a member of $A$. Partial memberships to $A$ are model by numbers between $0.0$ and $1.0$, and the closer the number is to $1.0$, the more $x$ belongs to $A$. For example, $\mu_A(x)$ of $0.5$ indicates that $x$ membership in $A$ is 50%.

Economakos [6] is perhaps the first to address the problem of electric power load demand forecasting using the fuzzy logic technique. Since then, a number of works [14, 1, 19] have reported good results in the application of fuzzy logic based approach to different constructs of the STLF problem. Among the advantages of fuzzy logic are the absence of a need for an explicit mathematical model for mapping inputs to outputs. With such generic conditioning rules, properly designed fuzzy logic systems can be very robust when used for forecasting. However, the fuzzy logic based technique, as is the case for all other rule driven approach, lack some important feature such as weak optimization and low learning and explanation capacity.

### 2.3 Hybrid approach

The shortcomings of the data driven and rule driven approaches motivates the development of the hybrid methods. In the hybrid based approach, techniques from the data driven and rule driven approaches are integrated based on their strength or the attributes desired from their application. Researchers have integrated techniques such as the ANN and fuzzy [34], fuzzy and regression [4] fuzzy and genetic algorithm [4, 2], ANN and genetic [17, 11], as well as genetic and decision tree [32], amongst others, in order to exploit and combine the capabilities of the composite techniques in their method.

Two techniques come out clearly, as having the capability to model some aspect of the STLF problem we specified in section 1.1: the ID3 and the fuzzy logic techniques. One of the strengths of the ID3 over the other data driven approaches is the ease with which they can be extended to symbolic or non-numerical domains. Exploiting this attribute, uncertainty can be represented symbolically using the fuzzy logic paradigm and integrate it into the ID3 paradigm. The fusion of the fuzzy logic (FL) with decision trees (ID3) enable us to combine the uncertainty handling and approximate reasoning capabilities of the FL with the comprehensibility and ease of application of the Dtree techniques in the modeling of our STFL problem. This way, we are able to model domain attributes using pre-defined linguistic variables which enhances the representative power of the decision tree.

The techniques of the fuzzy decision tree method has been developed and well reported in the literature [20, 37]. In [27], for example, the fuzzy decision tree method was used in the modeling economic dispatch problem including environmental constraints. To the best of our knowledge, this work is the first to exploit it in the proposed STLF domain. In the next section we discuss the development of the fuzzy decision tree based STLF system.

## 3 Research Methodology

Fuzzy decision trees (FDT) represent classification knowledge more naturally to the way of human thinking and are more robust in tolerating imprecise, conflicting, and missing information [37]. FDT aim at high comprehensibility, normally attributed to ID3, with the gradual and graceful behaviour attributed to fuzzy systems [12]. The development of Fuzzy Decision Trees (FDT) differs from traditional decision trees in two respects: it uses splitting criteria based of fuzzy restrictions and its inference procedures are different [37, 12]. We are adopting the following three steps, adapted from [37], for developing our FDT: (i) Fuzzifying the training data; (ii) Inducing and pruning the generated tree; and (iii) Applying fuzzy rules represented by the tree for forecasting.

## 4 Fuzzifying the training data

The analysis presented here are based on the three years hourly electric load data obtained from NCC for 2004, 2005 and 2006. These are about the only adequate data that can be used in this work. The data covered the two seasons (Rain and Harmattan) experienced in Nigeria and covers from January to December of each year. The data for the hourly weather conditions for the three years were obtained from the Nigerian Meteorological Agency (NIMET).

Based on a careful analysis of the data and information elicited from the domain experts and the literature, we identified two classes of factors that influences short term electric load demand: (i) the environmental or physical factors such as weather, and (ii) the social or human factor. The social factors considered here are: (a) population, (b) user-type, (c) social-event, and (d) day-type. The environmental factors are more amenable to numerical rendering and modeling since they have some definitive numerical measure. However, recent occurrences characterized by global warming is making it difficult to predict them in a definitive manner. The social event are uncertain and very difficult to model by crisp computation framework. The membership functions for the variables in this category are determined based on experts' opinion or common perceptions. In the next subsection, we provide a detailed description of each of the variables used in our STLF model as well as the design of the membership function used to modeled their universe of discourse (UoD).

After exploring the UoD of the variables using the MatLab software, we selected the trapezoidal membership function due to its simplicity and ease of implementation. We, therefore, designed suitable membership functions and the set of linguistic terms for each of the variables in our model. The membership functions used in our model, as implemented in the FID 4.2 software, are shown in Figure 1.

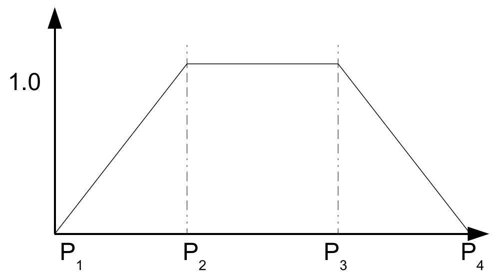



**Figure 1:** Trapezoidal membership function template

Given the points $P_1$, $P_2$, $P_3$ and $P_4$ as shown in Figure 1, the template for the trapezoidal membership function for the linguistic variable $v$ is defined by the equation:

$$\mu_v(x) = \begin{cases} \frac{P_2-x}{P_2-P_1} & P_1 \le x < P_2 \\ 1.0 & P_2 \le x < P_3 \\ \frac{P_4-x}{P_4-P_3} & P_3 \le x \le P_4 \end{cases} . \qquad (1)$$

The membership functions of our model are explained in details in the following subsections.

### 4.1 Weather

The weather is identified by two seasons: Harmattan and Rain which is characterized by three variables: (i) *Temperature*, *Humidity* and *Wind-speed*. The Harmattan season usually occurs between the end of November and lasts until late January or early February. It is usually very dry, cool and windy. The Rainy season, on the other hand, is usually wet, hotter and less winding than the Harmattan season. Between these two extremes, however, there are a range of weather situations which can be described by various degree of strengths of the combinations of the variables identified. For example, the weather in May will have attribute values closer to that of a rainy season while that in October will have attribute values closer to the Harmattan season based on their proximity to the seasons at the two extremes identified. We describe these variables in more details in the following subsections.

#### 4.1.1 Temperature

The temperature variable has direct impact on the daily electric load peak. Three linguistic terms are defined over the UoD for this variable, these are: *cold*, *mild* and *hot*. These instances are defined based on temperature measurements on the degree centigrade scale. Based on the available data, we assigned $20.00^0C$ and $40.00^0C$ as the minimum and maximum temperatures, respectively. Each of the linguistic variables are defined by trapezoidal membership functions specified by four data points $\{P_1, P_2, P_3, P_4\}$ with values: *Cold*= $\{20.00, 20.00, 24.00, 32.00\}$; *Mild*= $\{24.00, 28.00, 32.00, 36.00\}$, and *Hot* =$\{28.00, 36.00, 40.00, 40.00\}$.

#### 4.1.2 Wind-Speed

The wind speed variable identifies the current of air and its UoD is defined in the numerical space assigned the values 0.00 through 10.00 for the minimum and maximum, respectively. Each value for the Wind speed is normalized in the range $[0.00, 1.00]$. Two linguistic terms were defined over the UoD, namely: (i) *Windy* and (ii) not winding (*Nwindy*). The trapezoidal membership functions defined for each of the linguistic terms have four data point values: *Nwindy* = $\{0.00, 0.00, 0.40, 0.60\}$ and *Windy* = $\{0.40, 0.60, 1.00, 1.00\}$.

#### 4.1.3 Humidity

The humidity variables is expected to measure the degree of dampness of the weather and its UoD is defined over the numerical space assigned the values 80.00 through 90.00 for the minimum and maximum, respectively. Two linguistic terms were defined over the UoD, namely: (i) *Humid* and (ii) not humid (*Nhumid*). The trapezoidal membership functions defined for each of the linguistic term having the following four data point values: *Nhumid* = $\{80.00, 80.00, 84.00, 86.00\}$ and *Humid*= $\{84.00, 86.00, 90.00, 90.00\}$.

### 4.2 User-Type

The users are defined by the electric power equipments that they use. For the commercial and domestic users, these equipment will included electric cooker, microwave oven, air conditioner, heater and iron, fridge and freezer as well as fan, radio, television, video and computer. Most user will also have electric bulbs for lighting points. In addition to these, industrial users will have heavy equipments which have very high electric power consumption capacity. To model the user type variables, we define a UoD with numerical values ranging from 0.00 to 120.00. Each group of power equipment are then assigned numerical values in this range depending on their power consumption capacity. We defined three linguistic terms, $\{$*Light*, *Mid*, *Heavy*$\}$ over the UoD and associate each of the terms with numerical values defined by trapezoidal membership functions with data points specified as follows: *light* = $\{0.00, 0.00, 24.00, 72.00\}$; *Mid* = $\{24.00, 48.00, 72.00, 96.00\}$ and *Heavy* = $\{48.00, 96.00, 120.00, 120.00\}$.

### 4.3 Day-Type

From interview with the operators at the NCC and review of the literature, the following day types were identified: (i) Weekdays (Monday through Friday); (ii) Saturday and (iii) Sunday. These day types are assigned values ranging from 0.00 to 10.00 representing the minimum and maximum values, respectively. Three linguistic terms were defined over its UoD which includes: Week-Day (*WD*), Saturday (*Sat*) and Sunday (*Sun*). The trapezoidal membership functions defined for each variable has the data points: *WD*= $\{0.00, 0.00, 2.00, 6.00\}$; *Sat* = $\{2.00, 4.00, 6.00, 8.00\}$, and *Sun* = $\{4.00, 8.00, 10.00, 10.00\}$.

### 4.4 Population

This variable expresses an estimate of the total number of people resident in the particular geographical locations. The population estimate variable becomes important as the censors figures that were provided are not adequate and sometimes contradicts the evidence. In order to keep the model as simple as possible, we partitioned the population UoD into two: *loosely* and *densely* representing loosely populated and densely populated areas, respectively. The trapezoidal membership function defined over the UoD have the following data point values: *loosely*:$\{0.00, 0.00, 0.40, 0.60$ and *densely* = $\{0.40, 0.60, 1.00, 1.00\}$.

### 4.5 Social-Event

Two input variables, namely *Day-type* and *Population* are influenced by the social events taking place at any particular time. There are a number of festivals and events, e.g. sporting and game events, that influences the pattern and amount of power consumption at any point in time. For example, during the Christmas and new year period, people tend to move from densely populated cities to scarcely populated villages to celebrate the occasion. This movement is also a function of the day in which the holiday falls. If the holiday falls into the weekend days, i.e. Saturday and Sunday, the volume of the movements will be more than when it falls on weekdays such as Monday. We however, analyzed the social events separately from day type and population because each social event has different intensity and overall affect on the other two variables. To this end, the UoD defined for the different types of the social events were assigned values ranging from $0.00$ to $25.00$.

Values are assign to social events based on their estimated intensity and effect on electric power consumption. For example, the Christmas and New Year days are assigned values $10.00$; Easter, *Eidel Efik* (*Eid1*) and *Eidel Fiktri* (*Eid2*) are assigned value $8.00$. Days such as the Independence Day, Children's day and Workers' day are assigned numerical values $5.00$. To compute the social event for a particular case we added the numerical values associated to all social events on that day. For example, if the *Eidel Efik* day falls on the Christmas day then the value of the social events would be $18.0 = 8.00 + 10.00$.

Three linguistic terms, *Low*, *Mid* and *High*, were defined over the UoD and the trapezoidal membership functions for the terms have data point values: *Low* = $\{0.00, 0.00, 5.0, 15.00\}$; *Mid*=$\{5.00, 10.00, 15.00, 20.00\}$, and *High* = $\{10.00, 20.00, 25.0, 25.00\}$.

### 4.6 Output variable definition

When these rules are fuzzified, the result will produce the forecasted amount of electric load (*FLoad*) to be consumed at a particular instance. From an analysis of the load distribution data provided we assigned the minimum and maximum loads of 2000 $MW$ and 4000 $MW$, respectively. The forecast load was scaled and categorized using numerical values ranging from $0.00$ (for the minimum) to $15.00$ (for the maximum). Typical forecast load assignments are shown in the sample data in Table 1.

All values were normalized in the interval $[0, 1]$ for our FDT implementation.

## 5 Inducing and Pruning the Fuzzy Decision Tree

The procedure for building the FDT is similar to that for the ID3 in that it comprise three elements: (i) selection of splits at every new node of the tree, (ii) a rule for determining when a node should be considered as the terminal and (iii) a rule for assigning labels to identified terminal nodes. The main difference between FDT and ID3 is that training examples are assigned to nodes in FDT technique based on the degree to which they belong to the classification at a node. Also the fuzzy norms are used to deal with conjunction in the proposition for building FDT. If, however, the node membership is computed incrementally, the computational complexities for the two methods are the same [12].

**Table 1:** Output variable analysis

| Temp | Hum | WindS | Season | User-Type | Social-Event | Day-Type | Population | FLoad |
|---|---|---|---|---|---|---|---|---|
| Hot | Yes | T | H | H | H | H | H | 15 |
| Hot | Yes | F | H | H | H | M | H | 14 |
| Hot | No | T | M | H | M | H | H | 13 |
| Hot | No | F | H | L | H | H | H | 12 |
| Mild | Yes | T | L | H | H | H | H | 11 |
| Mild | Yes | F | M | H | M | M | H | 10 |
| Mild | No | T | L | L | L | L | L | 1 |
| Mild | No | F | L | L | M | L | L | 2 |
| Cold | Yes | T | L | M | L | M | L | 3 |
| Cold | Yes | F | L | L | L | L | H | 4 |
| Cold | No | T | L | M | M | M | L | 5 |
| Cold | No | F | L | M | M | M | H | 6 |
| Hot | Yes | F | H | L | H | L | L | 7 |
| Mild | No | F | L | M | M | M | H | 8 |
| Hot | No | F | H | M | M | H | L | 9 |

A number of heuristics have been proposed for building and pruning fuzzy decision trees [37, 22]. We are adapting the method developed in [12] due to its simplicity and the availability of software for its implementation. The algorithm for building the tree is presented as follows:

Fuzzy decision tree build algorithm

Start With the entire example $E$
E = GetData();
TreeBuilt = *False*
While !TreeBuilt {
$X^{Root} = W$
While !EndofNodes() {
Compute the example counts as:
$P_k^N = \sum_{j=1}^{|E|} f_{k+1}(X_j^N, \mu_{v_k^c}(y_j))$
and $P^N = \sum_{k=1}^{|D_c|} P_k^N$
Compute the standard information content as:
$I^N = -\sum_{k=1}^{|D_c|}(\frac{P_k^N}{P^N} \times \log \frac{P_k^N}{P^N})$.
$\forall$Node $i$ {
search the set of remaining attributes from $V - V^N$ by:
computing $I^{S_{V_i}^N}$
selecting attribute $V_i$ such that the information gain $G_i^N$ is maximal
}
If$((X_j^N > 0.0$ has unique classification) OR $(V^N = V)$){
TreeBuilt = *True*
}
}
Split $N$ into $|D_i|$ sub nodes by making,
child $N|v_p^i$ get samples defined by $X^{N|v_p^i}$ and
computing new memberships using the fuzzy restrictions
leading to $N|v_p^i$ using the equation
$X_j^{N|v_p^i} = f_1(f_0(e_j, v_p^i), X_j^N)$
}

In addition to the algorithm, the parameters, variables, and functions used in our FDT design are listed and described in Table 2.

The training data were generated from the hourly

**Table 2:** FDT variables and parameters

| | Variable/ parameter | Description |
|---|---|---|
| 1 | $V = \{V_1, V_2, V_4, V_5, V_6, V_7\}$ | The set of seven fuzzy input variables defined in section 4. |
| 2 | $D = D_1^i, D_2^i, D_3^i$ | The set of terms defined over variable $V_i$. |
| 3 | $v_p^i$ | The fuzzy term $p$ for variable $V_p$, e.g. $\mu_{Low}^{Temp}$. |
| 4 | $u^i \in U_i$ | The crisp data in the UoD for variable $V_i$. |
| 5 | $D_c$ | The fuzzy terms for the *Fload* decision variable. |
| 6 | $E = \{e_j \mid e_j = u_j^1, u_j^2, \ldots, u_j^n, y_j\}$ | The training example database, each item in the database is viewed as an event. |
| 7 | $W = \{w_j\}$ | The confidence weight. |
| 8 | $w_j$ | The weight of $e_j \in E$. |
| 9 | $N$ | Nodes in the fuzzy decision tree. |
| 10 | $F^N$ | The set of fuzzy restrictions on the path leading to $N$. |
| 11 | $V^N$ | The set of attributes appearing on the path leading to $N$. |
| 12 | $X^N = X_j^N$ | The set of memberships in $N$ for all the training examples. |
| 13 | $N\|v_p^i$ | The particular child of node $N$ created by using $V_i$ to split $N$. |
| 14 | $S_V^N$ | The set of $N$'s children when $V_i \in V - V^N$ is used for the split. |
| 15 | $P^N$ | The total example count for node N. |
| 16 | $I^N$ | Information measure for node N. |
| 17 | $P_k^N$ | The example count for decision $v_k^c \in D_c$ in node $N$. |
| 18 | $G_i^N = I^N - I^{S_V^N I}$ | The information gain when using $V_i$ in $N$. |
| 19 | $\mu(.) : X \to [0.0, 1.0]$ | A mapping from $X$ to $[0.0, 1.0]$. |

electric load for 2004 and 2005 while selected data from 2004 and 2005, as well as the data for 2006 were used as a test data. The entire database $(E)$ of events for the STLF experiment is divided into two disjoint sets: (i) the training and (ii) test sets. The training set is further divided into two other disjoint sets. One of these sets was used for building the tree while the other was used for pruning the resulting tree. The procedure for building the FDT utilises the training data set and the tree nodes are successfully added in a top-down fashion, until the stopping criteria are met. The resulting tree is optimized by a pruning procedure which acts in a bottom-up version, to remove irrelevant parts of the tree.

To deal with missing values, an example is split into all children if the need feature value is not available. The attribute utilization is then reduce based on the percentage of the examples with unknown values. If $P^{u^i_{\text{unknown}}}$ denotes the total count of examples in node $N$ with unknown values for $V_i$, then the split function $f_r()$ at splitting point $r$ is defined as:

$$f_r(e_i,\ [V_i\ is\ v_p^i]) = \begin{cases} \frac{1.0}{|D_i|} & \text{if } u_j^i \text{ unknown,} \\ \\ \mu_{v_p^i}(u_j^i) & \text{otherwise} \end{cases} .$$

Note that $|D_i|$ is the cardinality of set $D_i$, which is the number of elements in the set. In our model, the cardinality for the linguistic terms *Temp, User-type, Social-Event* and *Day type* is 3 while those of *Population*, *WindS* and *Hum* is 2. The FDT pruning algorithm as implemented in the FID 4.2 software [12, 22] was used.

### 5.1 Model Implementation

The Fuzzy Decision Tree model was implemented using the FDT 4.2 software developed by [13]. Three files, namely: the event file, the attribute file and the parameter template file were generated.

The event or data file contains the training example data (TE). The values of each variable ( whether numeric, linguistic, or missing), for each case of event are generated and the decision value (numeric or linguistic) as well as the corresponding weight (W) value (weight of the event) are assigned. A sample set of data extracted from our data file is shown in Table 3. The attribute file contains data relating to the attributes of our STLF model. This includes definitions of partitioning sets and the corresponding names as well as attributes of predefined partitions, with some restrictions on the partitioning preprocessing. The file also contains the definition of the decision class. The parameter file contains default values which we modified to customize the options for STLF problem.

A sample of rules obtained using the above procedure are shown below:

**IF** Temperature is mild $(30^0C)$ and User-Type is mid (60) and Social-Event is mid (12.5) and Wind-Speed is notwindy (0.5) and Humidity is nothumid (85) and Day-Type is Saturday (5) and Population is loosely (0.5) **THEN** Fload is 3067MW (8)

**IF** Temperature is cold $(22^0C)$ and User-Type is heavy (86) and Social-Event is low (8.5) and Wind-Speed is windy (0.6) and Humidity is humid (87) and Day-Type is weekday (2) and Population is Loosely (0.2) **THEN** Fload is 2399MW (2.99)

## 6 Model results, evaluation and brief discussion

In order to evaluate our model, we compared its prediction results ("Fload") with the actual load ("A") demand as well as the results obtained from conventional multiple regression ("F") [9, 10] method. It is important to note that the actual load is not the full load that consumers demanded but the load consumed from the amount of energy provided. The January $1^{st}$ 2004 and

**Table 3:** Sample data

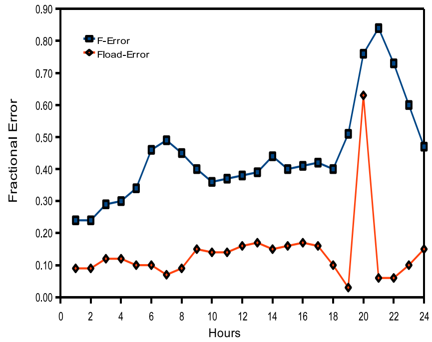
| TE | Popn | Temp | Hum | WS | UT | SE | DT-WD | Classi-fiation | W |
|---|---|---|---|---|---|---|---|---|---|
| 1 | A | 0.10 | 1.00 | 0.10 | 0.40 | 0.60 | 0.70 | 0 | 1 |
| 2 | B | 0.20 | 0.10 | 0.20 | 0.60 | 0.40 | 0.70 | 0 | 1 |
| 3 | A | 0.30 | 0.10 | 0.30 | 0.40 | 0.50 | 0.80 | 0 | 1 |
| 4 | B | 0.40 | 0.20 | 0.40 | 0.50 | 0.20 | 0.90 | 0 | 1 |
| 5 | A | 0.50 | 0.50 | 0.50 | 0.20 | 0.10 | 1.00 | 1 | 1 |
| 6 | B | 0.60 | 1.00 | 0.60 | 0.10 | 0.10 | 0.10 | 1 | 1 |
| 7 | B | 0.70 | 0.10 | 0.70 | 0.20 | 0.20 | 0.10 | 1 | 1 |
| 8 | B | 0.80 | 0.50 | 0.80 | 0.30 | 0.30 | 0.20 | 0 | 1 |
| 9 | A | 0.90 | 0.10 | 1.00 | 0.40 | 0.40 | 0.40 | Yes | 1 |
| 10 | B | 1.00 | 0.20 | 0.10 | 0.50 | 0.50 | 0.60 | 0 | 1 |
| 11 | A | 0.10 | 0.30 | 0.50 | 0.60 | 0.60 | 0.40 | 0 | 1 |
| 12 | A | 0.10 | 0.40 | 0.10 | 0.70 | 0.70 | 0.50 | 0 | 1 |
| 13 | B | 0.20 | 0.50 | 0.50 | 0.80 | 0.80 | 0.20 | 1 | 1 |
| 14 | B | 0.50 | 0.60 | 0.30 | 1.00 | 1.00 | Sat | No | 1 |
| 15 | A | 1.00 | 0.70 | 0.60 | 0.10 | 0.10 | 1.00 | 1 | 1 |
| 16 | A | 0.10 | 0.80 | 0.70 | 0.50 | 0.50 | 0.10 | 1 | 1 |
| 17 | B | 0.50 | 0.90 | 0.80 | 0.10 | 0.10 | 0.50 | 1 | 1 |
| 18 | B | 0.10 | Humid | 0.90 | 0.50 | 0.50 | 0.10 | 0 | 1 |
| 19 | A | 0.50 | NHumid | Windy | 0.30 | 0.30 | 0.50 | 0 | 1 |
| 20 | B | 0.30 | 0.10 | 0.10 | 0.60 | 0.60 | 0.30 | 0 | 1 |
| 21 | A | 0.60 | 0.60 | 0.90 | 0.70 | 0.70 | 0.60 | 1 | 1 |
| 22 | B | 0.70 | 0.40 | 0.50 | 0.80 | Heavy | 0.50 | No | 1 |
| 23 | A | 0.80 | 0.50 | 0.70 | 0.90 | 0.10 | 0.70 | 1 | 1 |
| 24 | B | 0.90 | 0.70 | 0.30 | Heavy | 0.20 | 0.30 | 1 | 1 |
| 25 | A | Cold | 0.90 | 0.80 | 0.80 | 0.30 | 0.80 | 1 | 1 |
| 26 | B | Mild | 0.80 | 0.20 | 0.20 | 0.40 | 0.20 | 1 | 1 |
| 27 | A | 0.60 | 0.50 | 0.40 | 0.40 | 0.50 | 0.40 | 0 | 1 |
| 28 | A | 0.40 | 0.10 | 0.60 | 0.60 | 0.60 | 0.60 | 1 | 1 |
| 29 | A | 0.50 | 0.30 | 0.40 | 0.40 | 0.70 | 0.40 | 0 | 1 |
| 30 | A | 0.70 | 0.20 | 0.50 | 0.50 | 0.80 | 0.50 | 1 | 1 |
| 31 | B | 0.90 | Nhumid | 0.20 | 0.20 | 1.00 | 0.20 | 0 | 1 |
| 32 | A | 0.80 | Humid | 0.40 | Low | 0.10 | 0.40 | 1 | 1 |
| 33 | B | 0.50 | 0.10 | 0.60 | 0.50 | 0.50 | 0.60 | 0 | 1 |
| 34 | B | 0.10 | 0.90 | 0.40 | 0.60 | 0.10 | 0.40 | 0 | 1 |
| 35 | B | 0.30 | 0.50 | 0.50 | 0.70 | 0.50 | 0.50 | 0 | 1 |
| 36 | A | 0.20 | 0.70 | 0.20 | 0.80 | 0.30 | 0.20 | 0 | 1 |
| 37 | B | Cold | 0.30 | 0.50 | 0.90 | 0.60 | 0.50 | 1 | 1 |
| 38 | B | Mild | 0.80 | 0.60 | 1.00 | 0.70 | 0.60 | 1 | 1 |
| 39 | A | 0.10 | 0.20 | 0.70 | 0.10 | 0.70 | 0.70 | 1 | 1 |
| 40 | B | 0.90 | 0.40 | 0.80 | 0.10 | 0.80 | 0.80 | 0 | 1 |
| 41 | A | 0.50 | 0.60 | 0.90 | 0.20 | 0.90 | 0.90 | 1 | 1 |
| 42 | B | 0.70 | 0.40 | 1.00 | 0.40 | 1.00 | 1.00 | 1 | 1 |
| 43 | B | 0.30 | 0.50 | 0.10 | 0.60 | 0.10 | 0.10 | 0 | 1 |
| 44 | A | 0.80 | 0.20 | 0.10 | 0.40 | 0.10 | 0.10 | 1 | 1 |
| 45 | A | 0.20 | 0.10 | 0.20 | 0.50 | 0.20 | 0.20 | 1 | 1 |
| 46 | B | 0.40 | 0.20 | 0.40 | 0.20 | 0.40 | 0.40 | 0 | 1 |
| 47 | B | 0.60 | 0.30 | 0.60 | 0.60 | 0.60 | 0.60 | 0 | 1 |
| 48 | B | 0.40 | 0.40 | 0.40 | 0.40 | 0.40 | 0.40 | 0 | 1 |
| 49 | B | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 1 | 1 |
| 50 | A | 0.20 | 0.60 | 0.20 | 0.20 | 0.20 | 0.20 | 0 | 1 |

March $2^{nd}$ 2005 data are from our training sets while those for January $1^{st}$ 2006 and September $4^{th}$ 2004 are from the test set.

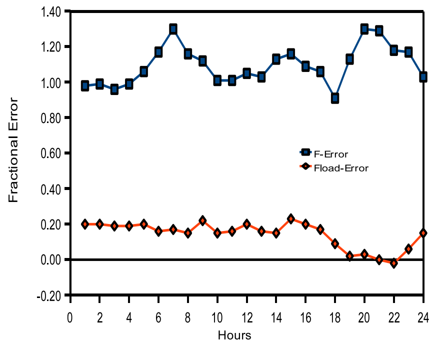
We compared the forecasts produced by our model with those produced by the experts who use manual techniques in their prediction. The fractional errors were computed by the formula:

$$\text{Fractional Error}(n) = \frac{\text{Forecasted Load}(n) \; - \; \text{Actual Load}(n)}{\text{Actual Load}(n)} \quad (2)$$

For January $1^{st}$ 2004, the average fractional error obtained using the proposed model was $0.11$, while that of the conventional multiple regression model (MRM) was $0.45$. For March $2^{nd}$ 2005, the average fractional errors obtained using the proposed model was $0.14$, while that of the conventional model was $1.10$. For January $1^{st}$ 2006, the average fractional error obtained using the proposed model was $0.21$, while that of the conventional (MRM) was $0.78$. For September $4^{th}$ 2005, the average fractional error obtained using the proposed model was $0.22$, while that of the conventional (MRM) was $0.88$. The variables *F-Error* represents the forecast load error for the conventional (MRM) while *FLoad-Error* represents the forecast load error for our FDT based model. The results are presented graphically in Figures 2 and 3. Our results agreed with what obtains in the literature [21].




(a) January $1^{st}$ 2004

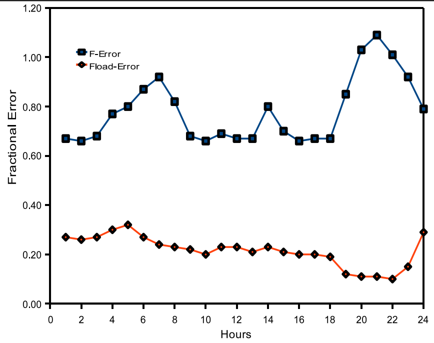



(b) March $2^{nd}$ 2004

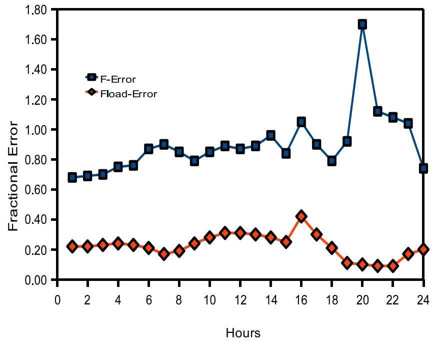
**Figure 2:** Fractional errors for January and March 2004 evaluation experiments

## 7 Summary and Conclusion

The aim of this communication is to present the application of FDT technique in the computational modeling and simulation of the STLF problem in an uncertain domain. We stated that the load components, as exhibited by the Nigerian electric power environment, are imprecisely defined and their variations are not known exactly. Thus the knowledge about this STLF problem is rather vague as it does not depend on any formal crisp system but on the experiences of the expert electric power load operators. We also showed that modeling such problem using the Fuzzy Decision Tree (FDT) method seems to produce relatively accurate prediction when compared with the conventional (MRM). Our application of the FDT method in the context of STLF is




(a) January $1^{st}$ 2005




(b) September $4^{th}$ 2005

**Figure 3:** Fractional errors for January and September 2005 evaluation experiments

novel.

The proposed model can be the basis for implementing a commercial software for electric load forecasting. We, however, think that, for a more thorough testing, electric load of a longer period than the 3 year used in this work should be carried out. Another area of further research would be the development of an online software that embodies the FDT model for short-term electric load forecasting. Such system will be expected to have a real-time data logging module that could help in reducing the non-availability of adequate data for use in load forecasting.

## References

[1] Al-Kandari, A. M., Soliman, S. A., and El-Hawary, M. E. Fuzzy short-term electric load forecasting. *International Journal of Electrical Power & Energy Systems*, 26:111–122, 2004.

[2] Basu, M. An interactive fuzzy satisfying method based on evolutionary programming technique for multiobjective short-term hydrothermal scheduling. *Electric Power Systems Research*, 69:277–285, 2004.

[3] Basu, T. K., Bhattacharya, T. K., and Purkayastha, P. Medium range forecasting of hourly power system load by time series analysis using the Walsh transform. *International Journal of Electrical Power & Energy Systems*, 13(4):193–200, 1991.

[4] Bunnoon, P., Chalermyanont, K., and Limsakul, C. A computing model of artificial intelligent approaches to mid-term load forecasting : a state-of-the-art- survey for the researcher. *International Journal of Engineering and Technology*, 2010.

[5] Catalão, J. P. S., Mariano, S. J. P. S., Mendes, V. M. F., and Ferreira, L. A. F. M. Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research*, 77:1297–1304, 2007.

[6] Economakos, E. Application of fuzzy concepts to power demand forecast. *IEEE Transactions On Power Systems*, SMC-9(10):651–660, 1979.

[7] EL-Naggar, K. M. and AL-Rumaih, K. A. Electric load forecasting using genetic based algorithm, optimal filter estimator and least error squares technique: Comparative study. *Transactions On Engineering, Computing And Technology*, 6:138–142, 2005.

[8] Feinberg, E. A. and Genethliou, D. *Load Forecasting: Chapter 12*. Available from http://www.ams.sunysb.edu, 2006.

[9] Ho, K.-L., Hsu, Y.-Y., Chen, C.-F., Lee, T.-E., Liang, C.-C., Lai, T.-S., and Chen, K.-K. Short term load forecasting of Taiwan power system using a knowledge-based expert system. *IEEE Transactions On Power Systems*, 5(4):1214–1221, 1990.

[10] Ho, K.-L., Hsu, Y.-Y., and Yang, C.-C. Short term load forecasting using a multilayer neural network with an adaptive learning algorithm. *IEEE Transactions On Power Systems*, 7(1):141–149, 1992.

[11] Huang, S.-J. and Huang, C.-L. Genetic-based multilayered perceptron for Taiwan power system short-term load forecasting. *Electric Power Systems Research*, 38:69–74, 1996.

[12] Janikow, C. Z. Exemplar learning in fuzzy decision trees. *FUZZIEEE*, 2(8-11):1500–1505, 1996.

[13] Janikow, C. Z. User's manual for fuzzy decision trees FID3.4. Department of Mathematics and Computer Science University of Missouri St. Louis, 2007.

[14] Juan, M. and Pedro, E. A new approach for power system online DSA using distributed processing and fuzzy logic. *Electrical System Research*, 77:106–118, 2006.

[15] Karapidakis, E. S. and Hatziargyriou, N. D. Online preventive dynamic security of isolated power systems using decision trees. *IEEE Transactions On Power Systems*, 17(2):297–304, 2002.

[16] Leonidaki, E. A., Georgiadis, D. P., and Hatziargyriou, N. D. Decision trees for determination of optimal location and rate of series compensation to increase power system loading margin. *IEEE Transactions On Power Systems*, 21(3):1303–1310, 2006.

[17] Maifeld, T. and Sheblé, G. Short-term load forecasting by a neural network and a refined genetic algorithm. *Electric Power Systems Research*, 31(3):147–152, 1994.

[18] Mei, K. and Rovnyak, S. M. Response-based decision trees to trigger one-shot stabilizing control. *IEEE Transactions On Power Systems*, 19(1):531–537, 2004.

[19] Miranda, V. and Monteiro, C. Fuzzy inference and cellular automata in spatial load forecasting. *in Proc. IEEE PES Winter Meeting*, 2, 2000.

[20] Mitra, S., Konwar, K. M., and Pal, S. K. Fuzzy decision tree, linguistic rules and fuzzy knowledge-based netwrok: Generation and evaluation. *IEEE Transactions On SYSTEMS, Man, and Cybernetics- Part C: Applications and Reviews*, 32(4):328–339, 2002.

[21] Mohamad, E. A., Mansour, M. M., EI-Debeiky, S., and Mohamed, K. G. Egyptian Unified Grid hourly load forecasting using artificial neural network. *International Journal of Electrical Power & Energy Systems*, 20(7):495–500, 1998.

[22] Olaru, C. and Wehenkel, L. A complete fuzzy decision tree technique. *IEEE Transactions On Power Systems*, 138:221–254, 2003.

[23] Paarmann, L. D. and Najar, M. D. Adaptive online load forecasting via time series modeling. *Electric Power Systems Research*, 32:219–225, 1995.

[24] Quinlan, J. R. Decision trees and decisionmaking. *IEEE Transactions On SYSTEMS, Man, and Cybernetics*, 20(2):339–346, 1990.

[25] Rahman, S. and Hazim, O. Load forecasting for multiple sites: development of an expert system-based technique. *Electric Power Systems Research*, 39:161–169, 1996.

[26] Ramachandra, T., George, V., Vamsee, K., and Purnima, G. Decision support system for regional electricity planning. *Energy Education Science and Technology*, 17(1):7–25, 2006.

[27] Roa-Sepulveda, C. A., Herrera, M., Pavez-Lazo, B., Knight, U. G., and Coonick, A. H. Economic dispatch using fuzzy decision trees. *Electric Power and Systems Research*, 66:115–122, 2003.

[28] Rovnyak, S., Kretsinger, S., Thorp, J., and Brown, D. Decision trees for real-time transient stability prediction. *IEEE Transactions On Power Systems*, 9(3):1417–1426, 1994.

[29] Safavian, S. R. and Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Transactions On Power Systems*, 21(3):660–674, 1991.

[30] Senroy, N., Heydt, G. T., and Vittal, V. Decision tree assisted controlled islanding. *IEEE Transactions On Power Systems*, 21(4):1790–1797, 2006.

[31] Tamimi, M. and Egbert, R. Short term electric load forecasting via fuzzy neural collaboration. *Electrical Power Systems Research*, 2000.

[32] Teeuwsen, S. P., Erlich, I., El-Sharkawi, M. A., and Bachmann, U. Genetic algorithm and decision tree-based oscillatory stability assessment. *IEEE Transactions On Power Systems*, 21(2):746–753, 2006.

[33] Vahidinasab, V., Jadid, S., and Kazemi, A. Day-ahead price forecasting in restructured power systems using artificial neural networks. *Electric Power Systems Research*, 78:1332–1342, 2008.

[34] Vellasco, M. M. B. R., Pacheco, M. A. C., Neto, L. S. R., and de Souza, F. J. Electric load forecasting: evaluating the novel hierarchical neuro-fuzzy BSP model. *International Journal of Electrical Power & Energy Systems*, 26:131–142, 2004.

[35] Voumvoulakis, E. M. and Hatziargyriou, N. D. Decision trees-aided self-organized maps for corrective dynamic security. *IEEE Transactions On Power Systems*, 23(2):622–630, 2008.

[36] Yang, J. and Stenzel, J. Short-term load forecasting with increment regression tree. *Electric Power Systems Research*, 76:880–888, 2006.

[37] Yuan, Y. and Shaw, M. J. Induction of fuzzy decision treees. *Fuzzy Sets and Systems*, 69:125–139, 2005.