HYBRID EVOLUTIONARY DATA ANALYSIS TECHNIQUE FOR ENVIRONMENTAL MODELING

Jesús Acosta\textsuperscript{1*}, Àngela Nebot\textsuperscript{2} and Josep M. Fuertes\textsuperscript{3}

1: Departamento de Instrumentación Industrial
Instituto Universitario Tecnología “Alonso Gamero” (IUTAG)
4101 - Coro, Venezuela.
e-mail: jacostav@cantv.net, web: http://www.lsi.upc.edu/dept/investigacion/seccion/soco/

2: Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
08034 - Barcelona, Spain
e-mail: angela@lsi.upc.edu, web: http://www.lsi.upc.edu/dept/investigacion/seccion/soco/

3: Departament d’Enginyeria de Sistemes, Automàtica i Informàtica Industrial
Universitat Politècnica de Catalunya
08034 - Barcelona, Spain
e-mail: josep.m.fuertes@upc.edu

Key words: Ozone forecast, environmental modeling, evolutionary algorithms, genetic algorithms, fuzzy inductive reasoning, genetic fuzzy systems

Abstract. In this work an evolutionary fuzzy system (EFS) is presented and applied to an environmental problem, i.e. modeling ozone concentrations. The hybrid system is composed by a FIR methodology and a genetic algorithm (GA) that take charge of determining, in an automatic way, the fuzzification parameters in that fuzzy system. The obtained results are compared with some of the most popular and classical modeling methods, neural networks and other FIR models.

1. INTRODUCTION

Recently a lot of research efforts have been directed towards the combination of different methods for Data Analysis (DA). In this way, evolutionary algorithms (EAs) have been combined with different techniques either to optimize their parameters acting as evolutionary tuning processes or to obtain hybrid DA methods, i.e. evolutionary-neural processes \cite{[1]}, evolutionary regression models \cite{[2]} and evolutionary fuzzy systems (EFSs) \cite{[3]}. In this paper, we propose a EFS that improves a modeling and simulation technique, the Fuzzy Inductive Reasoning (FIR). The main goal of the EFS is to take advantage of the potentialities of GAs to learn at the same time the fuzzification parameters of FIR, i.e. the
number of fuzzy sets (classes) per variable and the membership functions that define its semantics. Due to the fact that it is a methodology based on fuzzy logic, FIR modeling and prediction performance is influenced by these discretization parameters.

FIR has been applied to different kinds of applications (e.g. control, biomedicine, ecology), usually obtaining good results [4,5,6]. In these studies, default values have been used to determine the number of classes and the associated membership functions. The default value for the number of classes per variable is set to three and the equal frequency partition (EFP) is used as the default method to obtain the membership functions of the classes. However, experience has shown that in some applications, i.e. biomedical and ecological, the determination of the parameters needed in the discretization step becomes significant for the identification of a good model that captures systems behavior in an accurate way. Therefore, the automatic determination of good fuzzification parameters in the FIR methodology is an interesting and useful alternative to the use of heuristics and/or default values.

In this research the EFS developed has been used to model the behavior of the ozone contaminant in an Austrian region. The problem of estimating the ozone levels is an important task because this toxic gas can produce harmful effects on the population’s health such as eyes irritation, respiratory problems and makes worse cardiovascular diseases. In order to provide adequate early warnings, it is valuable to have accurate and reliable forecasts of future high ozone levels. Therefore, the construction of ozone models that capture as precisely as possible the behavior of this gas in the atmosphere is of great interest not only for environmental scientists but also for government agencies. Previous works [7, 8] have deal with it using several classical methods, neural networks and fuzzy systems.

2. FIR METHODOLOGY

The conceptualization of the FIR methodology arises of the General System Problem Solving (GSPS) approach proposed by Klir [9]. This methodology of modeling and qualitative simulation is based on systems behavior rather that on structural knowledge. It is able to obtain good qualitative relations between the variables that compose the system and to infer future behavior of that system. It has the ability to describe systems that cannot easily be

![Figure 1. Fuzzy Inductive Reasoning (FIR) scheme](image)
described by classical mathematics, i.e. systems for which the underlying physical laws are not well understood. FIR is composed of four main processes, namely: fuzzification, qualitative model identification, fuzzy forecast and defuzzification. Figure 1 describes the processes of FIR methodology.

The fuzzification process converts quantitative data stemming from the system into fuzzy data, i.e. qualitative triples. The first element of the triple is the class value, the second element is the fuzzy membership value, and the third element is the side value. The side value indicates whether the qualitative value is to the left or to the right of the peak value of the associated membership function (see Figure 2).

![FIR fuzzification process of ambient temperature variable](image1)

Figura 2. FIR fuzzification process of ambient temperature variable

The qualitative model identification process is the responsible of finding causal and temporal relations between variables and therefore of obtaining the best model that represents the system. A FIR model is composed of a mask (model structure) and a pattern rule base (behavior matrix). A mask denotes a dynamic relationship among qualitative variables. An example of a mask is presented in equation 1.

![Behavior Matrix (Pattern Rule Base)](image2)

Figura 3. FIR Pattern rule base obtaining

<table>
<thead>
<tr>
<th>x</th>
<th>u1</th>
<th>u2</th>
<th>u3</th>
<th>u4</th>
<th>y1</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>t - 2δt</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>t - δt</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>t</td>
<td>0</td>
<td>-4</td>
<td>0</td>
<td>0</td>
<td>+1</td>
</tr>
</tbody>
</table>

(1)

Each negative element in the mask is called a m-input (mask input). It denotes a causal and temporal relation with the output, i.e. it influences the output up to a certain degree. The enumeration of the m-inputs is immaterial and has no relevance. The single positive value denotes the output. In position notation, the mask of equation 1 contains four m-inputs, it can be written as (1,4,10,12,15), enumerating the mask cells from top to bottom and from left to right. The qualitative model identification process evaluates all the possible masks and
concludes which one has the highest prediction power by means of an entropy reduction measure, called the quality of the mask, $Q$, that takes a value between 0 and 1, where 1 indicates the highest quality. The mask with the maximum $Q$ value is the optimal mask. Once the best mask has been identified, it can be applied to the qualitative data obtained from the system resulting in a particular pattern rule base, see Figure 3.

Once available the FIR model, the prediction system can take place using FIR inference engine, called fuzzy forecast process. FIR inference engine is a specialization of the k-nearest neighbor rule, commonly used in the pattern recognition field. Defuzzification is the inverse process of fuzzification. It allows converting the qualitative predicted output into quantitative values that can then be used as inputs to an external quantitative model. For a deeper and more detailed insight into the FIR methodology, the reader is referred to [10].

3. THE EFS APPROACH

The main components of the genetic learning process developed in the context of the FIR methodology to learn the number of classes of each system variable and the membership function for each class are described next.

3.1. Genetic representation

The goal is to find the number of classes (granularity level) per variable and the membership functions that define its semantics (identified by the landmarks). Therefore, each chromosome, $C$, is composed by: $C = C_1C_2$.

- Number of classes ($C_1$): The number of linguistic terms (classes) for $N$ variables is codified using a vector of $N$ integers in the range [2..9]. We consider a maximum class value of 9 in order to get relatively simple models. Therefore, if we denote by $E_i$ the number of classes for the variable $i$, the number of classes representation for a system of $N$ variables, $C_1$, is defined by: $C_1 = (E_1,E_2,...,E_N)$.

- Membership functions ($C_2$): A specific variable is represented by the proportion of data samples that contains each class, codified in the range [0.1]. An example of chromosome representation for a unique variable that has 4 classes could be (0.3,0.4,0.1,0.2), meaning that the membership function of the first class contains the 30% of the data samples available for this variable, and the second, third and forth membership functions contain 40%, 10% and 20% of the data records, respectively. Therefore, if we denote by $D_{ij}$ the data proportion of the variable $i$ and class $j$, and $C_{2i}$ the information of the data proportion for all the classes of the variable $i$, the membership representation, $C_2$, for a system of $N$ variables (including inputs and outputs), is defined by: $C_2 = (C_{21},C_{22},...,C_{2N})$, where $C_{2i} = (D_{i1},...,D_{iN})$. Of course, the sum of the proportions for each variable must be 1.

3.2. Initial gene pool

Conformed by four unique blocks with the same number of individuals each, except the first one. The considered groups are: 1) First group: each chromosome has the same number of classes in all its variables and the membership functions are uniformly distributed across the
variable working range (EFP method), 2) Second group: each chromosome has different granularity per variable and the membership functions are uniformly distributed as in group one, 3) Third group: each chromosome has the same number of classes in all its variables and the membership functions are no-uniformly distributed across the variable working range (the data proportion is generated randomly), and 4) Fourth group: each chromosome has different number of classes per variable, as in group two, and the membership functions are established in the same way as in the third group.

3.3. Fitness or objective function

The evaluation of the chromosomes is done following the next steps:
1) Decode the information of the chromosome, building the associated fuzzy partition in the FIR structures.
2) Execute the qualitative model identification process of the FIR methodology with the training data set, using the partition built in the previous step. Therefore, the mask associated to that partition with the highest quality measure is obtained.
3) Compute the objective function. In this research two objective functions are proposed: a) using the quality of the optimal mask, i.e. \( l(Q) \), or b) computing the predicted error of a portion of the training data set, not used in the training process, i.e. \( MSE_{train} \). The size of the portion of the training data set used for objective function evaluation purposes is defined with respect to the size of the whole training data set. The normalized mean square error in percentage (MSE), given in Equation (2), is used for this purpose,

\[
MSE = \frac{E[(y(t) - \hat{y}(t))^2]}{VAR[y(t)]} \cdot 100\%
\]  

(2)

where \( \hat{y}(t) \) is the predicted output, \( y(t) \) the system output and \( VAR \) denotes variance.

3.4. Genetic operators

In this work the Stochastic Universal Sampling proposed by [11] is used, including an elitist selection. The considered operators are:
1) Crossover operators: Two different crossover operators are used:

- Crossover when both parents have the same granularity level per variable: In this case, the genetic search has located a promising space zone that has to be adequately exploited. This task is developed by applying the non-uniform arithmetic crossover operator [12] in \( C_2 \) and maintaining the parent \( C_1 \) values in the offspring.
- Crossover when the parents encode different granularity levels: This second case highly recommends the use of the information encoded by the parents for exploring the search space in order to discover new promising zones. Hence, when \( C_1 \) is crossed at a certain point, the values in \( C_2 \) corresponding to the crossed variables are also crossed in the two parents. In this way, a standard crossover operator is applied over the two parts of the chromosomes. This operator performs as follows: a crossover point \( p \) is randomly generated in \( C_1 \) and the two parents are crossed at the \( p-th \) variable in \( C_1 \) and \( C_2 \), producing two meaningful descendents.
2) **Mutation operators:** The selected mutation operators are similar to the ones proposed by [13]. A brief description of them is given below.

- **Mutation on C1:** In this case, a local modification is done by changing the number of classes of the variable to the immediately upper or lower value (the decision is made randomly). When the value to be changed is the minimum, 2, or the maximum, 9, the only possible change is done. Once a new value is selected, a uniform fuzzy partition for this variable is stored in its corresponding zone of C2.

- **Mutation on C2:** Here, the data proportion associated to the gene of the selected chromosome is increased or decreased (the decision is made randomly) by a factor in-between the range \([V_{min}...MAX]\) set, also, randomly. Where \(MAX = 0.5 - V_{min}(N_{label} - 1)\). The other proportions of the same variable are adjusted in order to maintain the addition to 1. When the value to be changed plus the factor get out of the limits of the range \([V_{min}...V_{max}]\), the only possible change is done, i.e. increase or decrease by the proportion factor, respectively.

### 3.5. Genetic parameters

The values of the probabilities have been established according to [14]. The values of the parameters for the GA are: a) Population size: 40 individuals; b) Crossover Probability: 0.6; c) Mutation Probability (per individual): 0.1; d) Chromosomes evaluations number (stop criteria): \(\{500, 1000, 2000, 4000, 8000, 16000\}\).

### 4. OZONO MODELS

To deal with the problem under study we were provided with the data used by [8], stemming from the Viennese basin in the East Austrian region. It was registered, mostly, in the summer season, since ozone, \(O_3\), occurs in highest concentrations. The ozone values (output variable), measured in ppb (parts per billion) are the average of five measurement points providing three hours fixed average values. The input variables correspond to weather data originated from the weather prediction model of the European Center for Medium Range Weather Forecasts, i.e. the temperature (T) in °C, the cloud cover (CC) that take values ranging from 0 (no clouds) to 1 (completely cloudy) and the wind speed (WS) in m/s. Ozone and weather data were available for the periods: 07-07-1995 to 25-09-1995 (81 values), and 01-05-1996 to 30-09-1996 (149 values).

In the previous works, the available data was divided into the training set (all data from 1996) and the test set (all data from 1995). This distribution of data constitutes the denominated Initial Partition [7,8] and it is used also in the present study to comparison purposes. Also, the root mean square error (RMS) described in Equation 3 is used for the computation of each model prediction error.

\[
RMS = \sqrt{\frac{\sum_{t=1}^{N}(y(t) - \hat{y}(t))^2}{N}}
\]  

where, \(\hat{y}(t)\) is the predicted output, \(y(t)\) the system output and \(N\) the number of samples.
4.1. Previous works

Table 1 contains the best RMS prediction errors achieved by [7,8] when different types of neural networks (NNs) are used for the same problem. In [8] the NNs used were a Multilayer Perceptron (MLP), an Elman Networks (EN) and a Modified Elman Networks (MEN). The Wieland and Wotawa's NN models [8] were compared with statistical models, IMPO models and Santiago models, showing that the prediction performance of these NNs was better. The IMPO models use a chemical/physical approach developed by the Institute for Meteorology and Physics of the Universität für Bodenkultur Wien [15]. On the other hand, in [7] the authors used a MEN network with hyperbolic tangent activation function and a powerful type of recurrent neural network called Long Short Term Memory (LSTM).

<table>
<thead>
<tr>
<th>NN model</th>
<th>#HU</th>
<th>LR</th>
<th>Bias</th>
<th>Steps</th>
<th>RMS&lt;sub&gt;test&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>2</td>
<td>1</td>
<td>N</td>
<td>3000</td>
<td>15.0537</td>
</tr>
<tr>
<td>MLP</td>
<td>5+1</td>
<td>Y</td>
<td>1000</td>
<td>11.2004</td>
<td></td>
</tr>
<tr>
<td>MLP</td>
<td>5</td>
<td>0.4</td>
<td>Y</td>
<td>100</td>
<td>11.1768</td>
</tr>
<tr>
<td>MLP</td>
<td>2+1</td>
<td>0.2</td>
<td>Y</td>
<td>1000</td>
<td>10.8125</td>
</tr>
<tr>
<td>MEN (α=0.2)</td>
<td>5+1</td>
<td>0.2</td>
<td>Y</td>
<td>3000</td>
<td>10.5150</td>
</tr>
<tr>
<td>EN</td>
<td>8+1</td>
<td>0.2</td>
<td>Y</td>
<td>5000</td>
<td>10.2186</td>
</tr>
<tr>
<td>MEN-HYP (α=0.2)</td>
<td>5+1</td>
<td>0.2</td>
<td>Y</td>
<td>3000</td>
<td>9.9979</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.00125</td>
<td>---</td>
<td>3500</td>
<td>9.7961</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. NNs models for the prediction of maximum ozone concentration (Initial Partition) taken from [7,8].

The first column of Table 1 described the method evaluated. The second stands for the number of hidden units. The third column shows the learning rate. The fourth column indicates whether bias neurons had been used or not. The fifth column gives the necessary number of learning steps needed by the NN in order to obtain the best results. The last column indicates the RMS error of the test set.

It is also presented in [7] the results of the FIR methodology when is used to obtain the bests models for the problem at hand. In this study default and heuristic fuzzification parameters values were used. The best result were obtained with the EFP method and discretizing the temperature into 3 classes whereas the rest of the variables were discretized into 2 classes. Table 2 shows the optimal and the best suboptimal masks for the problem under study using the fuzzification parameters described before. The first column of this table indicates whether it is the optimal mask or a suboptimal one. The second column describes the causal relations of each mask in a position notation. The third column prints the quality measure of each mask. The last column shows the RMS error obtained when the test set is predicted. As can be seen from Tables 1 and 2 the MEN network with hyperbolic tangent activation function has the lower prediction error. A more detailed discussion of these results can be found in [7].

<table>
<thead>
<tr>
<th>Mask type</th>
<th>Mask relations</th>
<th>Q</th>
<th>RMS&lt;sub&gt;test&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>(7,12,14,15,16)</td>
<td>0.4366</td>
<td>12.6768</td>
</tr>
<tr>
<td>Subopt</td>
<td>(6,12,13,16)</td>
<td>0.5680</td>
<td>10.5908</td>
</tr>
<tr>
<td>Subopt</td>
<td>(1,4,13,16)</td>
<td>0.3158</td>
<td>10.5383</td>
</tr>
<tr>
<td>Subopt</td>
<td>(4,4,13,16)</td>
<td>0.4033</td>
<td>10.1837</td>
</tr>
<tr>
<td>Subopt</td>
<td>(7,12,13,16)</td>
<td>0.5456</td>
<td>10.1110</td>
</tr>
<tr>
<td>Subopt</td>
<td>(1,1,13,16)</td>
<td>0.3566</td>
<td>10.0220</td>
</tr>
<tr>
<td>Subopt</td>
<td>(9,12,13,16)</td>
<td>0.2330</td>
<td>9.5770</td>
</tr>
<tr>
<td>Subopt</td>
<td>(8,8,13,16)</td>
<td>0.2466</td>
<td>9.7577</td>
</tr>
</tbody>
</table>

Table 2. FIR mask for the prediction of maximum ozone concentration (Initial Partition) taken from [7]. EFP Method. Granularity (3,2,2,2)
4.2. Learning the granularity level and the membership function

In this section the utility of the EFS developed is evaluated for the problem at hand. Thirty executions were performed for each objective function and stop criteria. Tables 3 and 4 show a summary of the best results obtained for each objective function, i.e. 1-Q and MSE_{train}. In this application the last 8% of the training signal is used for MSE_{train} objective function evaluation and the first 92% of the signal is used to obtain the FIR models (mask and pattern rule bases). The reduced amount of data available does not allow increasing the percentage of the signal used to compute the MSE_{train} objective function.

<table>
<thead>
<tr>
<th>Gran.</th>
<th>Data Proportion</th>
<th>Opt. Mask</th>
<th>Q</th>
<th>1-Q</th>
<th>RMS_{test}</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (3.3.6)</td>
<td>(T:0.28,0.42,0.20); (CC:0.26,0.27,0.22,0.25)</td>
<td>(3.16)</td>
<td>0.9823</td>
<td>0.0177</td>
<td>20.1400</td>
</tr>
<tr>
<td>B (3.3.2)</td>
<td>(T:0.15,0.15,0.15,0.16,0.08,0.16,0.16,0.18); (O3:0.05,0.05)</td>
<td>(3.16)</td>
<td>0.9823</td>
<td>0.0177</td>
<td>20.1400</td>
</tr>
<tr>
<td>C (5.4.2)</td>
<td>(T:0.33,0.33,0.34); (CC:0.15,0.15,0.15,0.15,0.15,0.15,0.15,0.15)</td>
<td>(3.16)</td>
<td>0.9823</td>
<td>0.0177</td>
<td>20.1400</td>
</tr>
<tr>
<td>D (5.4.2)</td>
<td>(T:0.31,0.22,0.16,0.16,0.27); (CC:0.45,0.57)</td>
<td>(3.16)</td>
<td>0.9823</td>
<td>0.0177</td>
<td>20.1400</td>
</tr>
<tr>
<td>E (3.2.2)</td>
<td>(T:0.23,0.46,0.31); (CC:0.72,0.28)</td>
<td>(3.16)</td>
<td>0.9823</td>
<td>0.0177</td>
<td>20.1400</td>
</tr>
</tbody>
</table>

Table 3. Results when the FIR methodology enhanced with AG is used to learn the granularity and the membership functions. 1-Q cost function is used.

Both tables are organized as follows. The first column is divided in 5 sections. Sections A and B correspond to the results obtained using the EFS suggested. In row A the best solutions obtained with both objective functions are shown, i.e. the solutions with lowest 1-Q and MSE_{train} values. Row B presents the solutions obtained that have the lowest RMS_{test} values. Sections C, D and E present the solutions obtained when different genetic algorithms (GAs) are used to set the fuzzification parameters separately. For instance, section C shows the best obtained prediction results when a GA is used to obtain the number of classes of each variable and the EFP method (default value) is used to determine the membership function of each class. Section D presents the best obtained prediction results when two GAs are used to set the number of classes and the membership function parameters separately. Finally, section E shows the best obtained prediction results when a GA is used to obtain the membership functions given the granularity (3.2.2.2), set in the previous FIR work described in [7]. Sections C, D and E were designed in order to take their results as a reference. The second column presents the granularity level. The third column shows the data proportion for the input variables \(T, CC, WS\) and the output variable \(O3\). The granularity level and data proportion per variable are the output of the EFS and the input parameters of the FIR fuzzification process. The fourth column presents the optimal mask, in position notation, encountered by FIR when the parameters obtained are used. The fifth column corresponds to the quality associated to the optimal mask. The sixth column is the value of the objective function, i.e. 1-Q or MSE_{train}. The seventh column is the equivalent error in RMS of the MSE_{train} objective function. The last column shows the RMS prediction error obtained for test data set.
Although the EFS does not assure the optimal solution, all solutions have a high mask quality with respect to the FIR models of Table 2. The prediction errors of the test sets (last column) are also presented to show the accuracy of each model obtained. Notice that although the $\text{RMS}_{\text{test}}$ is usually smaller when the EFS is used, this is not always true. This is due to the fact that the test data sets had not been used in the FIR model identification process. Only the suboptimal solutions (row $B$ for each objective function) present lower prediction errors than those shown in Tables 1 and 2. The best result of 9.1930 RMS obtained by the MEN network with hyperbolic tangent activation function is higher than the 8.8196 and 8.8591 RMS errors obtained by FIR methodology enhanced with EFS.

<table>
<thead>
<tr>
<th>Gran.</th>
<th>Data Proportion</th>
<th>Opt. Mask</th>
<th>$Q$</th>
<th>$\text{MSE}_{\text{train}}$</th>
<th>$\text{RMS}_{\text{train}}$</th>
<th>$\text{RMS}_{\text{test}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(2,3,2)</td>
<td>T=0.36,0.54; CC=0.34,0.36,0.36</td>
<td>(11,12,14,16)</td>
<td>0.4170</td>
<td>37.96</td>
<td>6.4434</td>
</tr>
<tr>
<td></td>
<td>(6,4,6,6)</td>
<td>T=0.10,0.15,0.19,0.22,0.27,0.08; CC=0.22,0.26,0.27,0.25</td>
<td>(13,14,16)</td>
<td>0.4298</td>
<td>41.14</td>
<td>6.7327</td>
</tr>
<tr>
<td></td>
<td>(3.3,4,3)</td>
<td>T=0.09,0.11,0.14,0.16,0.18,0.24,0.25</td>
<td>(32,45,50,0.33)</td>
<td>0.5947</td>
<td>46.68</td>
<td>7.1457</td>
</tr>
<tr>
<td></td>
<td>(3,2,6,2)</td>
<td>T=0.13,0.23,0.24,0.25</td>
<td>(32,0.35,0.89)</td>
<td>0.4429</td>
<td>57.13</td>
<td>7.9055</td>
</tr>
<tr>
<td></td>
<td>(3,2,2,2)</td>
<td>T=0.16,0.16,0.17,0.17,0.17</td>
<td>(32,0.36,0.47,0.47)</td>
<td>0.4118</td>
<td>64.84</td>
<td>8.4229</td>
</tr>
<tr>
<td>C</td>
<td>(2,4,2,2)</td>
<td>Equal Frequency Partition Method</td>
<td>(12,13,14,16)</td>
<td>0.4206</td>
<td>57.61</td>
<td>7.9383</td>
</tr>
<tr>
<td>D</td>
<td>(2,4,2,2)</td>
<td>T=0.31,0.59; CC=0.34,0.23,0.18,0.25</td>
<td>(12,13,14,16)</td>
<td>0.4851</td>
<td>59.51</td>
<td>8.0651</td>
</tr>
<tr>
<td>E</td>
<td>(3.2,2,2)</td>
<td>T=0.55,0.45,0.19; CC=0.35,0.45</td>
<td>(13,14,16)</td>
<td>0.4851</td>
<td>59.51</td>
<td>8.0651</td>
</tr>
</tbody>
</table>

Table 4. Results when the FIR methodology enhanced with AG is used to learn the granularity and the membership functions. Prediction error of the last 8% of the training data set ($\text{MSE}_{\text{train}}$) cost function is used.

If we look closer to Tables 3 and 4, it can be seen that the best $\text{RMS}_{\text{test}}$ obtained by the EFS for the 1-Q objective function is lower than the errors of reference C, D and E. This is also true for the $\text{MSE}_{\text{train}}$ objective function except for the result obtained in the column of reference E, although both are very similar. The EFS proposed adjust the fuzzy partitions globally, allowing the generation of better fuzzy systems. However, it presents the inconvenience that works with a very big search space making difficult and slow the learning process. Also, as expected, the $\text{MSE}_{\text{train}}$ objective function is more expensive from the CPU time point of view. For example, the computational time needed to perform 30 executions for 16000 evaluations when the 1-Q and $\text{MSE}_{\text{train}}$ objective functions are studied is 41:19 and 58:05 hours, respectively, in a Pentium IV computer (2.66 Hz). It is important to remark, here, that the determination of the discretization parameters is an offline process that does not interfere to time prediction performance when the FIR model is already available.

The user should decide which objective function to use taking into account the characteristics of the problem to be solved, the size of the optimization problem and his/her own needs.

5. CONCLUSIONS

We have presented an EFS approach for forecasting the maximum ozone concentration in the East Austrian region. The performance of the best EFS inferred models is satisfactory since the errors are lower than those obtained by other previously tested techniques. However, the errors obtained are still high from the application perspective. We think that the available data
is very scarce in order that a data analysis methodology can obtain better results. In order to improve the results it is necessary to have a larger set of data.

REFERENCES


