

Indian Weather Forecasting using ANFIS and ARIMA based Interval Type-2 Fuzzy Logic Model

Ayush Agrawal*, Qureshi M F.**

*Department of Information Technology, Dr C V Raman University,
Kota, Bilaspur, Chhattisgarh 495001-India

**Department of Electrical Engineering, Government BRP Polytechnic,
Rudri, Dhamtari, Chhattisgarh, 493776- India
(ayushagrwal369@gmail.com; mfq_pro@rediffmail.com)

Abstract:

This paper presents a comprehensive study of ANFIS+ARIMA+IT2FLS models for forecasting the weather of Raipur, Chhattisgarh, India. For developing the models, ten year data (2000-2009) comprising daily average temperature (dry-wet), air pressure, and wind-speed etc. have been used. Adaptive Network Based Fuzzy Inference System (ANFIS) and Auto Regressive Moving Average (ARIMA) models based on Interval Type2 Fuzzy logic System (IT2FLS) have been applied. To ensure the effectiveness of ARIMA+IT2FLS and ANFIS techniques, different models employing a different training and test data set have been tested. The criteria of performance evaluation are calculated for estimating and comparing the performances of ARIMA+IT2FLS and ANFIS models. The performance comparisons of ANFIS and ARIMA+IT2FLS models due to MAE (Moving Average Error), RMSE- R^2 (Root-Mean-Square error) criteria, indicate that ANFIS yields better results. Interval Type-2 Fuzzy time series models have been proposed for forecasting temperature, pressure, wind speed and other weather parameters. In this paper a hybrid fuzzy time series model is proposed that will develop Interval type 2 fuzzy models based on ARIMA. The proposed model will use ARIMA to select appropriate coefficients from the observed dataset. IT2-FLS is utilized here for handling the uncertainty in the time series data so that it may yield a more accurate forecasting result.

Keywords

Autoregressive Integrated Moving Average (ARIMA) models, Time Series Forecasting, Adaptive Network Based Fuzzy Inference System (ANFIS), Interval Type-2 Fuzzy Logic Systems (IT2-FLS).

1. Introduction

Weather forecast systems are among the most complex equation systems that computer has to solve. A great quantity of data, coming from satellites, ground stations and sensors located around our planet send daily information that must be used to foresee the weather situation in next hours and days all around the world. Weather reports give forecast for next 24, 48 and 72 hours for wide areas (Pasero 2004). Weather forecasts provide critical information about future weather. There are various techniques involved in weather forecasting, from relatively simple observation of the sky to highly complex computerized mathematical models. Data recorded by a weather station at the prominent meteorological center of Raipur, India have been used for the analysis and forecast applying ANFIS and ARIMA+IT2FS, they are evaluated and compared. Being nonlinear, the fuzzy logic systems (FLS) have been combined with time series models to get rid of the problems occur with linear data and can also handle linguistic chaos and ambiguity. Autoregressive (AR), Multiple Regressions, Exponential Smoothing and Autoregressive Integrated Moving Average (ARIMA) are a few examples of time series models that have been combined with FLS. These Fuzzy time series models are used by many researchers to solve linear and nonlinear problems. Researchers are finding their way to solve the chaotic and uncertain problems using the extensions of classical fuzzy model. At present Interval Type-2 Fuzzy logic Systems (IT2-FLS) are extensively used after the thriving exploitation of Type-2 FLS. The non-interval secondary membership function of a general Type 2 Fuzzy sets (T2-FS) made it computationally more complex. The reason that IT2 fuzzy sets are enjoying the center stage these days; is the non-fuzziness of its secondary function. It has been deduced from their result that the hybrid models outperform the individual models. Keeping this concept, in this paper, we have proposed an efficient fuzzy time series model based on ARIMA and IT2-FLS to improve forecasting accuracy and handle forecasting problems. The result is compared with ANFIS Model. Among scientific research works related to the weather forecasting some should be mentioned. Hu (1964) initiated the implementation of ANN, an important Soft Computing methodology in weather forecasting. Cook and Wolfe (1991) developed a neural network to predict the average air temperatures. Fuzzy logic can also be of great use in the atmospheric data analysis and prediction. Being capable of dealing with linguistic variables, this methodology can be utilized in analyzing atmospheric variables. The local regression models concentrated on fitting using a small portion of the data, whilst the weighted fuzzy time series models established fuzzy relationships using the promising data from the entire database. ARIMA is integrated with fuzzy regression model for

forecasting the weather condition. The forecasted results were made good by narrowing the fuzzy interval (Upper and lower bound). When fuzzy intervals were wide deleting its upper or lower bound the interval were made narrow. This gave better performance. However deleting the bounds may cause incomplete or missing interval/data. Results with better accuracy were obtained by proposing a heuristic approach of fuzzy metric. T2-FLS is applied to forecast Mackay-Glass time series. Using T1-FLS, a T2-FLS is formed by incorporating noise information. The FLSs were designed on a single realization for practice. However, for different realizations of same data set, different FLSs parameters must be choose to obtain improved forecasting.

The rest of the paper is organized as follow. In section 2, a data analysis is presented. In section 3 and 4, ANFIS model and methodology are presented respectively. A hybrid model simulation (Proposed Model based on IT2-FLS and ARIMA Models) is proposed in section 5 and discussion is described in section 6. The work is concluded in section 7.

2. Data Analysis

Data recorded by a weather station at a prominent meteorological center of Raipur, India are used for the analysis in this study. Their initial analysis has shown that the most important weather parameters are the average temperature (dry and wet), wind-speed and pressure variables. These variables also represent a strong correlation with the other weather parameters.

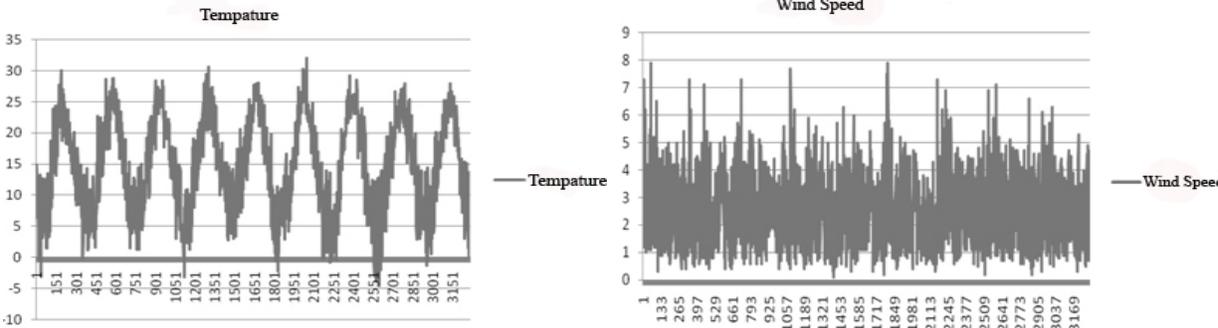


Fig.1. Average temperatures in period 2000-2009 Fig.2. Wind-speed in period 2000-2009

Temperature

Temperature, in general, can be measured to a higher degree of accuracy relative to any of the other weather variables. Historical temperature data recorded in Raipur meteorological center are used. The average temperatures in the period of ten years (2000-2009) are plotted in Figure 1.

Wind-Speed

Wind is often considered as one of the most difficult meteorological parameters to forecast. It is significant in winter whether the temperature is low, which is roughly below 4.4°C. When the temperature is below a freezing point, wind is a major factor determining the cooling rate. Figure 2 presents the wind-speed recorded in Raipur, India between years 2000 and 2009.

Pressure

Barometric pressure changes with local weather conditions, making the barometric pressure an important and useful weather forecasting tool. High pressure zones are generally associated with fair weather, while low pressure zones are generally associated with poor weather. For forecasting purposes, the absolute barometric pressure value is generally less important than the change in the barometric pressure. In general, the rising pressure indicates improving weather conditions, while the falling pressure indicates deteriorating weather conditions. The pressures in the period of ten years (2000-2009) are plotted in Figure 3.

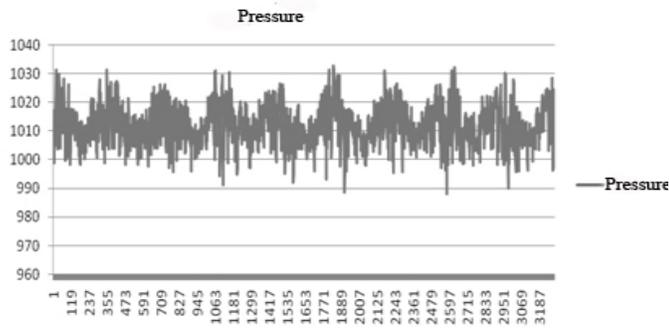


Fig. 3. Pressure in period 2000-2009

3. Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

ANFIS is a hybrid of two intelligent systems: Artificial Neural Networks (ANNs) and Fuzzy

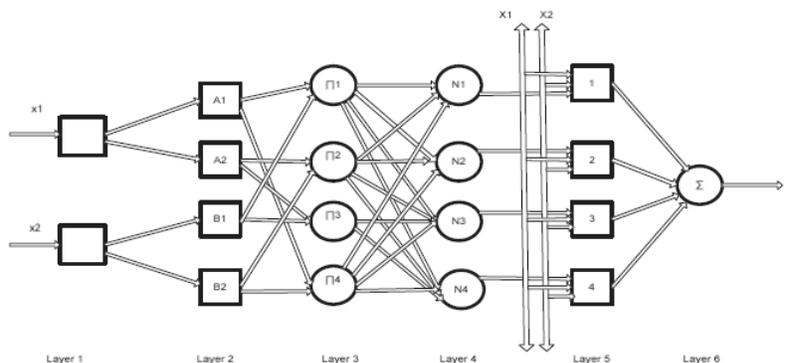


Figure 4 ANFIS Structure

Inference Systems (FISs). The ANFIS of interest is of a 6 layer feed forward neural network and of the Sugeno (Type-2) FIS type. To understand the structure and operation of ANFIS in forecasting, a 2 input- 1 output ANFIS model is presented and its structure and operation is related to a generalized model. Figure 4 shows the ANFIS structure and Equations 2 to 5 are the rules for this model where the IF part of the rule is referred to as the antecedent and the THEN part is the consequent.

$$\text{Rule 1: If } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } f_1 = p_1x + q_1y + r_1 \quad (1)$$

$$\text{Rule 2: If } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{ then } f_2 = p_2x + q_2y + r_2 \quad (2)$$

$$\text{Rule 3: If } x \text{ is } A_3 \text{ and } y \text{ is } B_3, \text{ then } f_3 = p_3x + q_3y + r_3 \quad (3)$$

$$\text{Rule 4: If } x \text{ is } A_4 \text{ and } y \text{ is } B_4, \text{ then } f_4 = p_4x + q_4y + r_4 \quad (4)$$

In general, an n-input, 1-output ANFIS model is an n + 1 dimensional input-output space. Therefore, a 2 inputs-1 output ANFIS model is a 3-dimensional input-output space. In order for ANFIS to be used to model a system, data that is representative of the target system must be presented to ANFIS. The entry of raw data or **crisp inputs** from the target system into ANFIS corresponds to layer 1 – the **input layer** in Figure 4. Since the Neural Network classifies data and looks for patterns within it, then when the input data is in the 3-dimensional space, it is classified into groups called fuzzy spaces. To do this, the crisp inputs are compared with membership functions in the antecedent of the rules of ANFIS, to determine the degree to which the inputs, in this case, X1 and X2 belong to fuzzy sets Ai and Bi respectively. The degree to which the inputs lie within the fuzzy space is given a value between 0 and 1. This process is known as **fuzzification** and takes place in layer 2, the **fuzzification layer**. Each node in this layer is adaptive and is given by:

$$O_{2,i} = \mu_{A_i}(x_1) \text{ for } i = 1, 2 \quad (5)$$

$$O_{2,i} = \mu_{B_{i-2}}(x_2) \text{ for } i = 3, 4 \quad (6)$$

where i represents the node of ANFIS and, μ_{A_i} and $\mu_{B_{i-2}}$ are the antecedent membership functions which can be any parameterized function such as the bell-shaped function given in Equation 7:

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\frac{(x - c_i)^2}{a_i} \right]^{b_i}} \quad (7)$$

a_i , b_i and c_i are the **non-linear or premise** parameters of the antecedent of the rules.

Once the locations of the inputs in the fuzzy spaces are identified, then the product of the degrees to which the inputs satisfy the membership functions is found and can be mathematically expressed by

Equation 8. This product is called the **firing strength** of a rule and is represented by layer 3, the **rule layer** where each node in this layer is fixed and represented by a rule.

$$O_{3,i} = w_i = \mu_{A_i}(x_1) \mu_{B_{i-2}}(x_2) \text{ for } i = 1, 2, 3, 4 \quad (8)$$

Each fuzzy space is governed by an ANFIS rule where the antecedent of the rule defines a fuzzy space in the input space. For ANFIS, there are Mn fuzzy rules where M is the number of membership functions per input and n is the number of inputs. Hence, for the ANFIS model of study in this paper where there are 2 membership functions per input, there are 4 fuzzy rules. The implication of the role of this layer is that the firing strengths give an indication of which rule may be most applicable to the inputs.

In layer 4, the **normalization** layer, the ratio of each rule's firing strength is calculated with respect to the sum of the firing strengths of all the rules. Each node in this layer is fixed and its output is given by Equation 9:

$$O_{4,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2 + w_3 + w_4}, i = 1, 2, 3, 4 \quad (9)$$

In layer 5, the **defuzzification** layer, the output of each node is the **weighted consequent value** given by Equation 10:

$$O_{5,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i), i = 1, 2, 3, 4 \quad (10)$$

where p_i , q_i and r_i are called **linear or consequent parameters** of the fuzzy rules. The consequent of the ANFIS rule defines the output in the fuzzy space. Each neuron in this layer is connected to the respective normalization neuron and inputs X_1 and X_2 . Layer 6 is the **summation layer** and its output which is the sum of all the outputs of the layer 5 is represented by:

$$O_{6,i} = \sum \bar{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i}, i = 1, 2, 3, 4 \quad (11)$$

This gives the overall output for the respective inputs within the fuzzy space.

The most fundamental component of ANFIS is its IF-THEN rules which are defined by its premise and consequent parameters. The IF-THEN rules tell where in the fuzzy space the output lies for the given inputs.

Developing the ANFIS Weather Forecasting Model

In order to determine the characteristics of ANFIS that can ultimately provide the wind speed forecasting model with the best prediction accuracy the System Identification method proposed by J.S.R. Jang was used. System Identification defines a mathematical model for a target system for which there isn't any *a priori* knowledge by observing its input-output data pairs. System Identification is broken down into two main steps: **Structure and Parameter Identification**. Structure Identification involves determining the type and number of inputs and membership functions that can best represent the target system to be identified while Parameter Identification involves finding the parameters of the rules that make up the ANFIS system. The application of System Identification to finding the best model involves repeatedly performing Structure and Parameter Identification. To develop the wind speed forecasting models using ANFIS, the structure and parameter identification phases were broken down into a series of steps. The choice of these steps was influenced by the software (ANFIS Editor in MATLAB 7.0.1) that was available for the development process of the ANFIS wind speed forecasting models.

A. Structure Identification

This is the preparation phase for parameter identification and it primarily involves establishing the inputs for the ANFIS models and generating the initial FIS or the initial parameters that would be the starting point to finding the optimum model that would best fit the target system. Therefore, structure identification for this application was separated into (1) selecting the inputs, training and checking data for ANFIS and (2) generating its initial FIS.

Selection of inputs, training and checking data

Generally, there are two types of time series models that can be developed: **Univariate** and **Multivariate**. Therefore, ANFIS was designed to model univariate and multivariate time series where the **Univariate time series models** used current and past data on wind speed data alone while the **Multivariate time series models** used wind speed data along with data for variables that correlate with wind speed. For the ARIMA time series approach, data was collected from the target wind speed system and was divided into subsets where the number of components in each subset was dependent on the combined number of inputs and output.

For ANFIS to properly model the wind system, it must be able to capture as many characteristics of the wind speed over time. To do this, a large training data set must be used. To evaluate the sensitivity of the ANFIS with respect to different training sizes, various amounts of data were used,

starting with 6 months up to all the data available. Careful thought was given to the number of inputs used for the development of the ANFIS models as using a large number of inputs can considerably increase the length of time taken to train while at the same time allow a greater amount of information into the model with each epoch. The latter would result in more selective training and hence provide better prediction results. Therefore, different ANFIS configurations with 7, 6, 5, 4 and 3 inputs were chosen

B. Parameter Identification

Parameter Identification is the training of the ANFIS models through the application of an optimization scheme on the generated initial FIS. Training involves tuning the parameters of the ANFIS models, which determine the shape of the membership functions and ultimately dictates how the rules for the trained prediction models would behave. To train the ANFIS models, the optimization technique, error tolerance and the number of epochs are chosen. The fuzzy logic toolbox provided two optimization methods: hybrid and back propagation. To develop the ANFIS wind speed forecasting models, the hybrid technique was used since it is more popularly used with ANFIS than the back propagation. In addition, it is regarded as the faster of the two techniques.

4. Methodology

Data Preprocessing

The case data for the study was obtained from the metrological station in Raipur, India and it covers a period of 125 months from January 2000 to May 2010. The preprocessing of the data was first carried. Missing values were replaced with zeros. The meteorological dataset had seven (7) attributes; their type and description are presented in Table 1.

Table 1: Attributes of Meteorological Dataset

Attribute	Type	Description
Year	Numerical	Year considered
Month	Numerical	Month considered
Min temperature	Numerical	Monthly Maximum temperature
Max temperature	Numerical	Monthly minimum temperature
Rainfall	Numerical	Total monthly rainfall (mm)
Wind run	Numerical	Wind Run in km
Relative humidity	Numerical	Monthly Relative Humidity in %

A graph of all the attributes was plotted to see the trend of the data using Matlab software before it was normalized to an appropriate scale by using the ARIMA Model. The scaled data was used to train the networks.

Software (ANFIS Editor in MATLAB 7.0.1)

In the ANFIS model, crisp input series are converted to fuzzy inputs by developing membership functions for each input series. The Neuro-fuzzy model builds intelligence and reasoning into the system by performing Subtractive Clustering on the fuzzy sets to determine the number and type of fuzzy membership function. The membership function pattern used for the input series is of the Bell shape. The goal of ANFIS is to find a model, which will simulate correctly the inputs with the outputs data. The ANFIS Matlab Fuzzy Logic Tool Box was used for the study. The ANFIS network was trained using a hybrid learning algorithm that uses least squares method in the forward pass to identify the consequent parameters of the layer 4, while in the backward pass the errors are propagated backwards and the premise parameters are updated by gradient descent method. Figure 5 shows the screen shot of the ANFIS GUI editor.

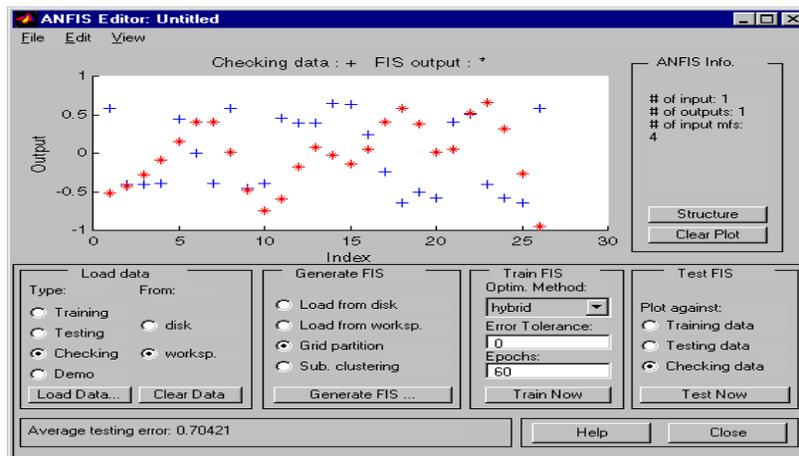


Figure 5: ANFIS GUI Editor showing test data with the FIS Output

The ANN and ANFIS models used for the analysis of the weather attributes were trained for $t = 100$ epoch each.

Auto Regressive Integrated Moving Average (ARIMA)

Time series is a unique sequence of data taken at a specific interval of time. This data may be stationary or non-stationary. The data must first be made stationary by numerous methods, as forecasting cannot be done with data comprising non-stationary behavior. ARIMA is a non-stationary model that can be reduced to a stationary time series by differencing. The model has three components i.e. Autoregressive, Integrated and Moving Average(see Fig.6.)

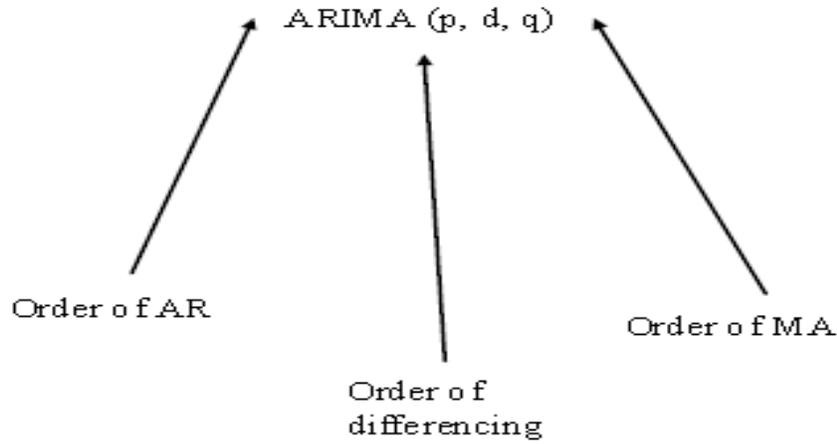


Fig 6: A General View of ARIMA models.

The ARIMA Model

In real world applications, many processes can be represented using the time series as follows:

$$X(t-p), \dots, x(t-2), x(t-1), x(t) \quad (12)$$

For making a prediction using time series, a great variety of approaches are available. Prediction of scalar time series $\{\hat{x}(n)\}$ refers to the task of finding estimate $\hat{x}(n+1)$ of next future sample $x(n+1)$ based on the knowledge of the history of time series, i.e. samples $x(n)$, $x(n-1)$ (Rank 2003). Linear prediction, where the estimate is based on a linear combination of N past samples, can be represented as below:

$$\hat{x}(n+1) = \sum_{i=0}^{N-1} a_i \cdot x(n-i) \quad (13)$$

with prediction coefficients, $a_i = 0, 1, \dots, N-1$.

Introducing general nonlinear function $f(\cdot), f(\cdot): \mathbb{R}^n \rightarrow \mathbb{R}$ applied to vector $x(n) = [x(n), x(n-M), \dots, x(n-(N-1)M)]^T$ of the past samples, nonlinear prediction approach $\hat{x}(n+1) = f(x(n))$ (Rank 2003) is reached. Traditionally, a time series forecasting problem is tackled using linear techniques such as Auto Regressive Moving Average (ARMA) and Auto Regressive Integrated Moving Average (ARIMA) models popularized by Box and Jenkins (1976). General form of ARMA(p,q) model can be written as below:

$$X_t - \sum_{r=1}^p \phi_r X_{t-r} = \sum_{s=0}^q \theta_s \cdot \varepsilon_{t-s} \quad (14)$$

Where $\{e_t\}$ is white noise . This process is stationary for appropriate ϕ, θ (Box and Jenkins 1976). General form of the ARIMA model is given by:

$$Y_t = a_0 + \sum a_i \cdot y_{t-i} + \sum b_j \cdot e_{t-j} \quad (15)$$

$i=1,2, \dots, p$ and $j=0,1, \dots, q$

Where Y_t is a stationary stochastic process with non-zero average, a_0 - constant coefficient, e_t - white noise disturbance term, a_i represents autoregressive coefficients and b_j denotes moving average coefficients. Performance analysis of the ARIMA model is done using the SPSS's package program. The Model obtained is ARIMA (2,1,1).

Interval Type-2 Fuzzy Logic Systems (IT2-FLS)

Type-1 fuzzy sets cannot handle the uncertainty exist in data since its membership grades are crisp value. T2-FS sets are introduced as an extension of T1-FS with a third dimension. The additional dimension in T2-FS helps in handling more uncertainties than T1-FS. Instead of using T2-FS with complex computation, IT2-FS can be preferred with much simplified computation. This paper is not supposed to give a background of T2 FS, that can be found. An IT2 FS can be defined as

Definition: A T2-FS denoted by \tilde{A} will be an IT2-FS, when all membership functions $\mu_{\tilde{A}}(x, u) = 1$, where $x \in X$ and $u \in J_x \subseteq [0, 1]$, i.e.

$$\tilde{A} = \{((x, u), 1) \mid x \in X, u \in J_x, J_x \subseteq [0, 1]\}, \quad (16)$$

\tilde{A} can also be expressed as

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} 1/(x, u), J_x \subseteq [0, 1]. \quad (17)$$

The rule base and inference engine are activated by these sets to produce IT2 FS. A type-reducer in the IT2-FLS reduces these fuzzy sets to IT1 FS, which are further defuzzified to get a crisp value. The defuzzified value can be obtained by taking the average of the two end-points of the interval.

5. Simulation of Proposed Model based on IT2-FLS and ARIMA Models

To tackle various forecasting problems and improve forecasting accuracy an efficient fuzzy time series model based on IT2-FLS and ARIMA is proposed in this section. The novel IT2- FARIMA model will estimate the parameters using ARIMA (see Fig 7). These estimated parameters will be fuzzified by IT2-FLS and will be then used for forecasting.

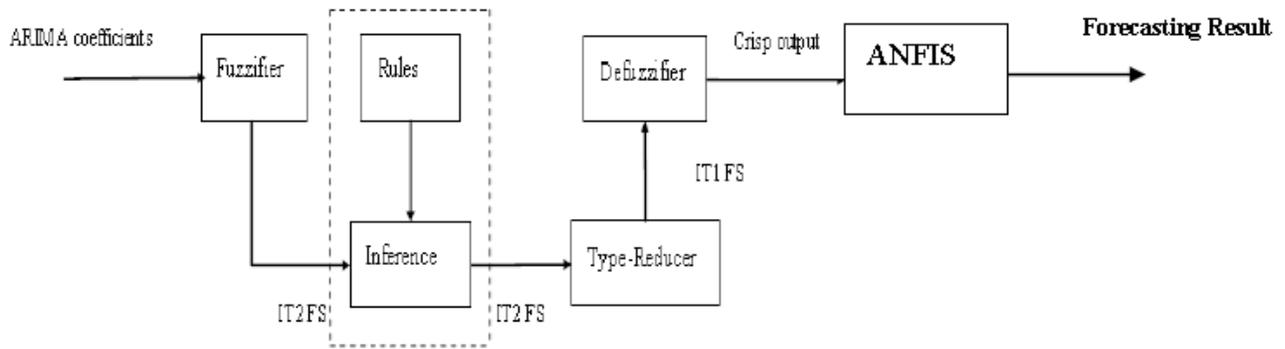


Fig 7: Proposed Interval Type 2 Fuzzy Model based on ARIMA and ANFIS (ARIMA+IT2FLS+ ANFIS)

The proposed model consists of following steps

A. Fit variable/coefficient through ARIMA model

The observed time series data must first be transformed into a reduced data set since dealing with large data requires high processing power and large amount of memory. To select an optimum set of parameters a predictive model is used. ARIMA in the proposed model will ensure the appropriate parameters for the available dataset. With the help of various statistical measurements an appropriate model and its coefficient will be estimated.

B. Fuzzify the fitted coefficients

The estimated parameters will be fuzzified in the fuzzifier block of IT2 FLSs.

C. Establish the Fuzzy logic relationship

The model will be evaluated using Takagi-Sugeno fuzzy rule based model, which is better than Mamdani’s model [24] in the sense that it reduces the number of required rules.

D. Reduce the fuzzified set

Type-reduction technique extracts an interval for the output value from the uncertain region. Among various type-reduction methods the output will be reduced by a centroid type-reduction technique. The centroid of IT2 FS was developed so as to measure and compute the uncertainty of that FS.

Simulation Process: The ANFIS+ARIMA+IT2FLS Model

Using the Matlab GUI Text viewer the following parameters were observed for the ANFIS network:

1. Number of nodes: 204,
2. Number of linear parameters: 100,
3. Number of nonlinear parameters: 150

4. Total number of parameters: 250, 5. Number of training data pairs: 100, 6. Number of checking data pairs: 25, 7. Number of fuzzy rules: 50

Figure 8 shows the ARIMA+IT2FLS+ ANFIS training result for the minimum temperature data, figure 9 shows the ARIMA+IT2FLS+ ANFIS validation result for minimum temperature data, figure 10 shows the ARIMA+IT2FLS+ ANFIS training result for maximum temperature data, figure 11 shows the ARIMA+IT2FLS+ ANFIS validation result for maximum temperature data, figure 12 shows the ARIMA+IT2FLS+ ANFIS training result for rainfall data, figure 13 shows the ARIMA+IT2FLS+ ANFIS training result for rainfall data, figure 14 shows the ARIMA+IT2FLS+ ANFIS training result for wind data, figure 15 shows the ANFIS validation result for wind data, figure 16 shows ARIMA+IT2FLS+ ANFIS training result for relative humidity data and figure 17 shows the ARIMA+IT2FLS+ ANFIS validation result for relative humidity data.

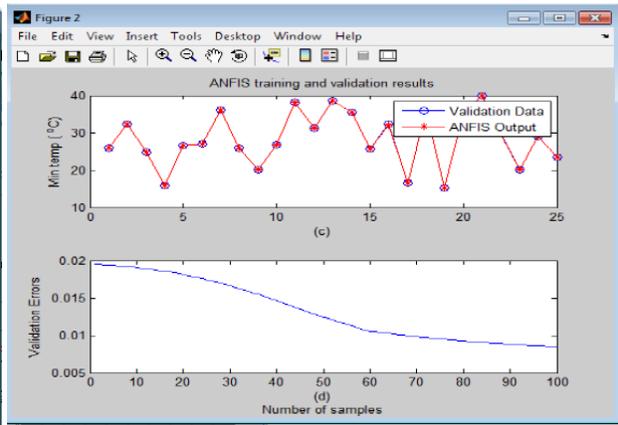
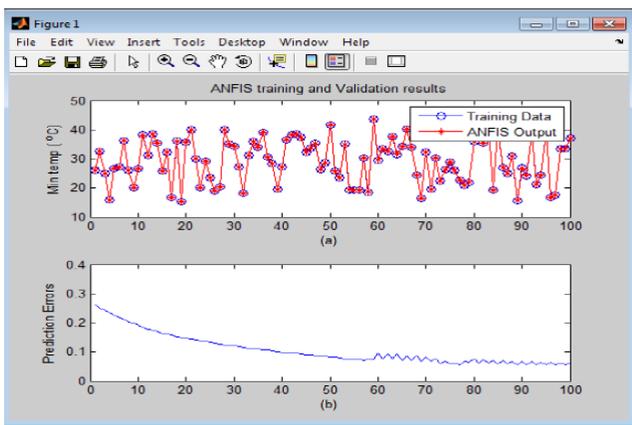


Figure 8: ARIMA+IT2FLS+ ANFIS Minimum Temperature Training Result

Figure 9: ARIMA+IT2FLS+ ANFIS Minimum Temperature Validation Result

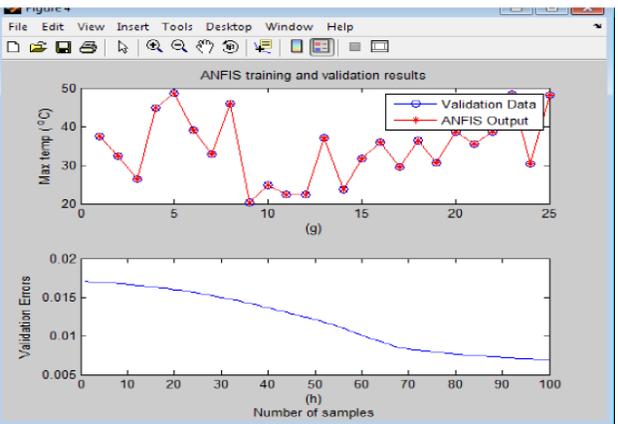
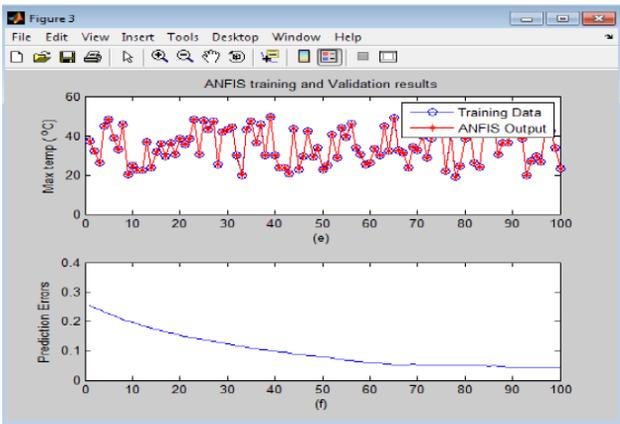


Figure 10: ARIMA+IT2FLS+ ANFIS Maximum Temperature Training Result

Figure 11: ARIMA+IT2FLS+ ANFIS Maximum Temperature Validation result

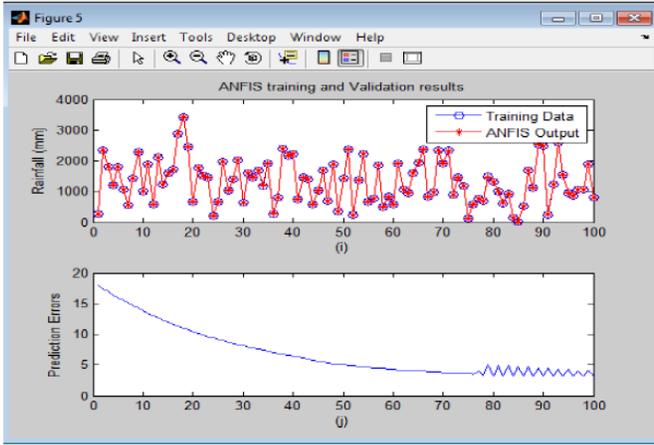


Figure 12: ARIMA+IT2FLS+ ANFIS Rainfall Training result data

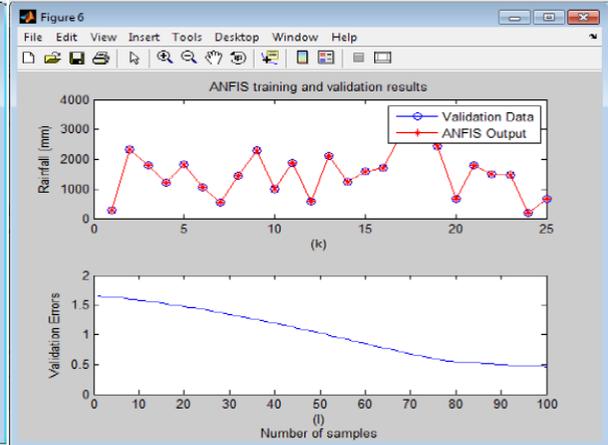


Figure 13: ARIMA+IT2FLS+ ANFIS Rainfall Validation result data.

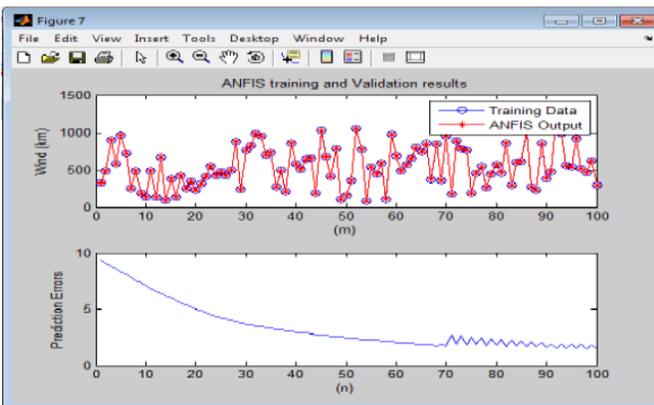


Figure 14: ARIMA+IT2FLS+ ANFIS Training result for Wind Data

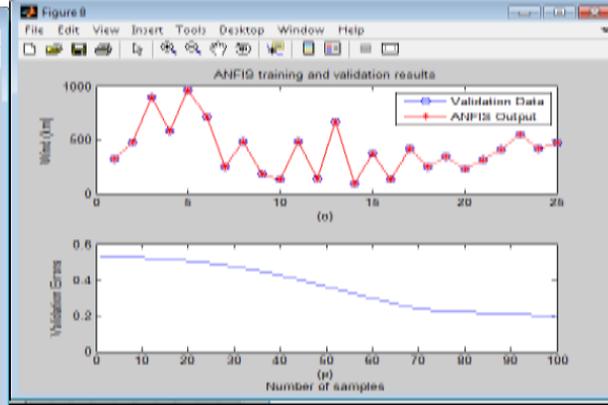


Figure 15: ARIMA+IT2FLS+ ANFIS Validation result for the Wind data

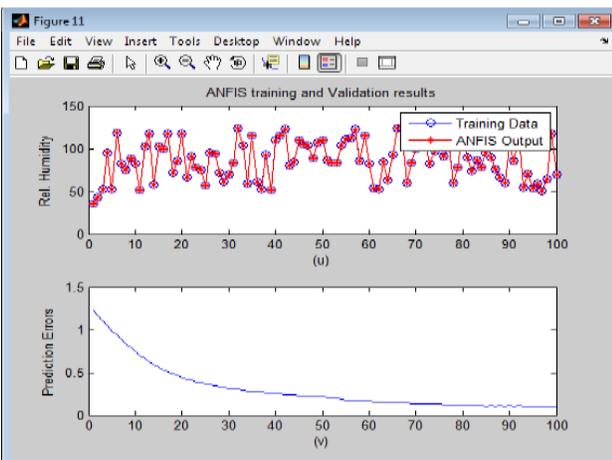


Figure 16: ARIMA+IT2FLS+ ANFIS Training Result for the Relative Humidity data

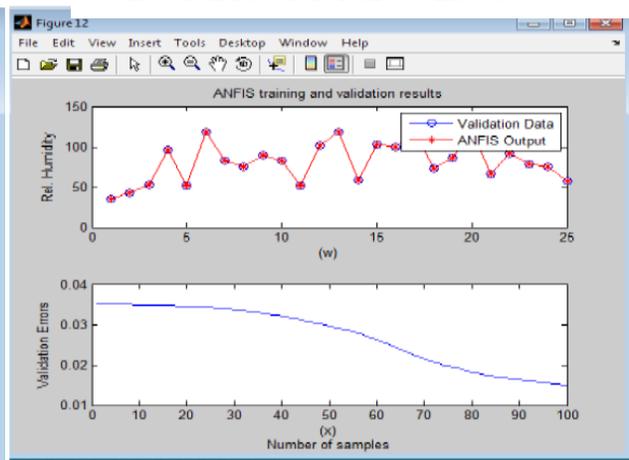


Figure 17: ARIMA+IT2FLS+ ANFIS Validation Result for the Relative Humidity

It can be seen that the network prediction of the training data of the ANFIS+ARIMA+IT2FLS and validation data also closely match the original training data. This also shows that the trained model and the validation model generally captures and approximate the system accurately. The Mean square error can be seen in table 2. Table 3 presents a comparison of the best ANFIS and ANFIS+ARIMA+IT2FLS models.

Table 2: Result of the ANFIS Model

Weather attribute	Minimum performance index	MSE	RMSE	Kstep Ahead Prediction Error
Min Temp	0.0075	0.0123	0.1093	0.0959
Max Temp	0.00065	0.0115	0.1078	0.0934
Rainfall	0.4433	1.0164	1.0089	6.2545
Wind	0.0013	0.0019	0.0377	3.1879
Relative Humidity	0.0142	0.0258	0.1575	0.2964

Table 3: Comparison of the ANFIS+ARIMA+IT2FLS and ANFIS models performance

Weather Attributes	ANFIS		ANFIS+ARIMA+IT2FLS	
	MSE	RMSE	MSE	RMSE
Minimum Temperature	-0.000	0.000	0.0133	0.1153
Maximum Temperature	0.000	0.000	0.0118	0.1086
Rainfall	3.976	1.993	1.0185	1.0092
Wind	0.3611	0.6009	0.002	0.0447
Relative Humidity	0.663	0.8142	0.0271	0.1645

A Research Approach for Handling These Problems

The above all three issues can be handled by the proposed hybrid model of ANFIS and Interval Type 2 Fuzzy model based on ARIMA (IT2FARIMA). Issue 1 and 2 are the limitations of ARIMA which would be effectively handled by IT2-FS. The last issue of improving the accuracy would be solved by combining the two dissimilar models. The propose model will combine the ARIMA model with IT2-FLS, where the former is a linear model and later is a nonlinear one. The sequence for hybridizing the model will follow the conventional hybrid model, which starts with a linear model, followed by a nonlinear model to model the residual. The model will be evaluated by Mean Squared Error (MSE), RMSE and MAE. Measurements formulas discuss in this paper are,

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}$$

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} = 1 - \frac{E}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}$$

where y_i , \hat{y}_i , \bar{y}_i are actual, forecasted and mean values respectively, and N is the forecast number.

6. Discussion

From the comparison of the two models used in this study both the ANFIS and ANFIS+ARIMA+IT2FLS model were able to capture the dynamic behavior of the weather data, resulting in a more compact and natural internal representation of the Temperature, Rainfall, Wind and Relative humidity information contained in the weather profile. Both the ANFIS and ANFIS+ARIMA+IT2FLS models performed better than the conventional models. Of the two models the ANFIS+ARIMA+IT2FLS model gave a better performance.

Model Integration

Models are integrated for the purpose of obtaining improved results than the individual. Since the real time series data contain linear and non linear patterns, the best approach would be the combination of linear and nonlinear model. The resulted single model would be able to deal with both types of relationships. Both the models are non-linear. The same data were used by ARIMA (2,1,1)+IT2FLS model to evaluate performance criteria. It is evidence from the result that ANFIS+ARIMA+IT2FLS performed well at testing data where the performance of the ANFIS model was good in training data. At feature extraction linear models perform well, where as in ANFIS model it is done by nonlinear model. That is the reason that testing data done by ANFIS+ARIMA+IT2FLS gave good results as compared to ANFIS. Another weakness of ANFIS is its complex computations. The performance

comparisons of ANFIS and ANFIS+ARIMA+IT2FLS models is done on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2), are shown below: In this study, four inputs and one output were used to forecast the weather in the ANFIS model. 5% of the data were used in testing and 95% of data were used in a training process. The same data were used by ANFIS=ARIMA(2,1,1)=IT2FLS method to evaluate performance criteria. The performance comparisons of the ANFI and ARIMA models due to MAE, RMSE, R^2 CRITERIA are shown in Table 4 and Table 5.

Table.4. Performance comparison of ANFIS and ANFIS+ARIMA+IT2FLS models

	MSE		RMSE	
	ANFIS	ANFIS+ARIMA+IT2FLS	ANFIS	ANFIS+ARIMA+IT2FLS
Training Data	1.2552	1.123	1.3312	1.575
Testing Data	1.3212	1.178	1.4176	1.309

Table.5. Performance comparison of ANFIS and ANFIS+ARIMA+IT2FLS models

	R^2
ANFIS	0.912
ANFIS+ARIMA+IT2FLS	0.988

As can be understood from Tables 4 and Table 5, compared to the ANFIS, the ANFIS+ARIMA+IT2FLS due to MAE, RMSE, R^2 criteria gives better results.

7. Conclusions

ANFIS has gained a great popularity in time-series prediction because of its simplicity and reliability. The performance of ANFIS and ARIMA+IT2FLS is compared. Compared to the ARIMA+IT2FLS, the ANFIS can more efficiently capture dynamic behavior of the weather temperature, resulting in a more compact and natural internal representation of the temporal information contained in the weather profile. It is important to have reliable and accurate techniques to forecast the wind speed in the very short term. One such technique is the use of ANFIS to develop wind speed forecasting models. The approach that was taken in this development was one proposed by J.S.R. Jang called

System Identification. This is closer to a trial and error process that involved training different models and comparing them until a model that produced satisfactory results was obtained. Univariate and multivariate models were developed. 4 models that gave predictions with errors in the range of 25.5% a 32.5% were obtained. When the output of these models was compared to the actual data, the results proved to be very good.

References

1. R. John, S. Coup land.(2007) "Type-2 fuzzy logic : A historical view," *IEEE Computational Intelligence Magazine*, vol. 2, pp. 57-62.
2. F.M. Tseng, G.-H. Tzeng, H.-C. Yu, B.J.C. Yuan,(2001)" Fuzzy ARIMA model for forecasting the foreign exchange market," *Fuzzy Sets and Systems*, vol. 118, pp 9-19.
3. Q. Song, B.S. Chissom,(1993) "Forecasting enrollments with fuzzy time series-part 1," *Fuzzy Sets and Systems*, vol. 54, pp. 1-9.
4. Q. Song, B.S. Chissom, (1994)"Forecasting enrollments with fuzzy time series-part 2," *Fuzzy Sets and Systems*, vol. 62, pp. 1-8.
5. K. Yu, (2005)"Weighted fuzzy time series models for TAIEX forecasting, " *Physical A*, vol. 349, pp. 609-624.
6. K. Huang, Hui-Kuang Yu,(2005) "A Type 2 fuzzy time series model for stock index forecasting," *Physical A*, vol. 353, pp. 445-462.
7. S.-M. Chen, J.R. Hwang,(2000) "Temperature prediction using fuzzy time series," *IEEE Trans. Syst. Man, Cybern, Part B* vol. 30 (2), pp. 263-275.
8. M. Tektaş, (2010)"Weather forecasting using ANFIS and ARIMA models. A case study for Istanbul,"*Environmental Research, Engineering and Management*, vol. 51, pp. 5-10.
9. N. Merh, V.P. Saxena, K. R. Pardasani,(2010) "A comparison between Hybrid Approaches of ANN and ARIMA for Indian stock trend forecasting, " *Business Intelligence Journal*, vol. 3, No 2, pp. 23-43.
10. E. Cadenas, W. Rivera,(2010) "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA_ANN model," *Renewable Energy*, vol. 35, pp 2732-2738.
11. J.M. Mendal, R.I. John and F. Liu,(2006) "Interval type 2 fuzzy logic systems made simple," *IEEE Trans. On Fuzzy Systems*, vol. 14, pp. 808-821.

12. F.M. Tseng, G.H. Tzeng, H.C. Yu, B.J.C. Yuan,(2000) “Fuzzy ARIMA model for forecasting the foreign exchange market,” *Fuzzy Sets and Systems*, vol. 118, pp. 9-19.
13. N.N. Karnik and J.M Mendel,(1999) “Applications of type-2 fuzzy logic systems to forecasting of time-series,” *Information Sciences*, vol. 120, pp. 89-111.
14. T.A. Jilani and S. Burney(2008) “Multivariate stochastic fuzzy forecasting models,” *Expert Systems with Applications*, vol.35, pp. 691–700.
15. T.A. Jilani and S. Burney (2008) “A refined fuzzy time series model for stock market forecasting”, *Physica-AStatistical mechanics and its applications*, 387, pp. 2857-2862.
16. H. Wu, J.M. Mendel,(2002) “Uncertainty bounds and their use in the design of interval type-2 fuzzy logic systems,” *IEEE Transactions on Fuzzy Systems*, pp. 622- 639.
17. W. Yuan-kang and J.-S. Hong. A literature review of wind forecasting technology in the world.[Online].Available:
<http://www.labplan.ufsc.br/congressos/PowerTech07/papers/246.pdf>.
18. J.-S. R. JANG, C.T. Sun and E. Mizutani,(1997)”Neuro-Fuzzy and Soft Computing, A Computational Approach to Learning and Machine Intelligence, New Jersey: Prentice Hall, pp. 73,74, 86, 95-97, 86-87, 26-28, 74-85.
19. MATHWORKS, Fuzzy Logic Toolbox – anfis and the ANFIS Editor GUI, MATLAB 7.0.1.
20. Qureshi Mohd. Farukh, Jha Manoj, Sao Gopi,(2009)“Fuzzy interval theory based governing control and excitation control for stability of power system.” *AMSE Journals, Advances C, Automatic Control (theory and applications)*, Vol. 64, Issue 1,pp1-19.
21. Qureshi M. F.,Bharti I.C. (2006) “Fuzzy based study and simulation of local heat transfer coefficient at Circumference of horizontal tube in free board region of fluidized bed” *AMSE Journals, Modelling-B*, Vol.75,Issue 2, pp1-20.
22. Qureshi M.F.,Devangan P. and Devangan N.P. (2010) “Mamdani-ANFIS and its application of student’s value” *AMSE Journals,Modelling-D*, Vol 31,issue 1,pp 65-74.
23. Qureshi et.al. (2006) “Design of fuzzy rule-based classifier for data mining from library data warehouse using EMO algorithms”.*AMSE Journals, Advances-D*, Vol.11, Issue 4,pp 31-42.