Chapter – 4

Adaptive neuro-fuzzy inference system to forecast the peak gust speed during thunderstorms

4.1 Summary

The aim of the present study is to develop an adaptive neuro-fuzzy inference system (ANFIS) to forecast the peak gust speed associated with thunderstorms during the pre-monsoon season (April – May) over Kolkata (22° 32' N, 88° 20' E), India. The pre-monsoon thunderstorms occurred during 1997 to 2008 are considered in this study to train the model. The input parameters are selected from various stability indices using statistical skill score analysis. The most useful and relevant stability indices are taken to form the input matrix of the model. The forecast through the hybrid ANFIS model is compared with non-hybrid Radial Basis Function Network (RBFN), Multi Layer Perceptron (MLP) and Multiple Linear Regression (MLR) models. The forecast error analyses of the models in the test cases reveal that ANFIS provides the best forecast of the peak gust speed with 3.52% error whereas the error with RBFN, MLP and MLR models are 10.48%, 11.57% and 12.51% respectively. During the validation with the 2009 observation of India Meteorological Department (IMD), the ANFIS model confirms its superiority over other comparative models. The forecast error during the validation of ANFIS model is observed to be 3.69%, with lead time of less than 12 hours, whereas the errors with RBFN, MLP and MLR are 12.25%, 13.19% and 14.86% respectively. ANFIS model may, thus, be used as an operational model for forecasting the peak gust speed associated with thunderstorms over Kolkata during the pre-monsoon season.

Chaudhuri, S. and Middey, A., 2011, Adaptive neuro-fuzzy inference system to forecast peak gust speed during thunderstorms, Meteorology and Atmospheric Physics, (Accepted)
4.2 Materials and Methods

4.2.1 Meteorological Data

The Radiosonde (RS) / Rawinsonde (RW) data used in the present study are collected from the Department of Atmospheric Sciences, University of Wyoming (http://www.weather.uwyo.edu) during the pre-monsoon season (April - May) for the years 1997 to 2009. The location of the study is Kolkata (22° 32'N, 88° 20'E), India. The raw data are the Radiosonde (RS) / Rawinsonde (RW) sounding observations of both 00 UTC and 12 UTC. If thunderstorm occurred between 00 UTC and 12 UTC then the RS/RW data of 00 UTC are taken for the analysis, whereas 12 UTC data are taken for the thunderstorms that occurred after 12 UTC.

The raw data are processed for the computation of the stability indices; Lifted Index (LI), Showalter Index (Sw), Total Total Index (TT), K- Index (KI), Convective Available Potential Energy (CAPE), Convective Inhibition Energy (CINE), Bulk Richardson Number (BRN), Boyden Index (BI), Severe Weather Threat (SWEAT), Storm Relative Helicity (Srh), Vorticity Generation Parameter (VGP) and Surface Dew Point Temperature (Td). During the period from 1997 to 2008, 140 thunderstorm records over Kolkata are collected from Regional Meteorological Office (RMO) of India Meteorological Department (IMD) at Kolkata. There were more thunder cloud activity events documented with just light to moderate rain but these 140 thunderstorm days were recorded with significant peak gust speeds (ranging from 40 to 120 kmh⁻¹). So, these 140
thunderstorm events with peak gust speed records along with corresponding RS/RW derived products are taken up for the present study.

4.2.2 Methodology

The methodology comprised in the present study contains the application of ANFIS and other neural network models. Earlier in the 90’s, J-S Roger Jang (1993) pioneered the method of developing ANFIS models and suggested those models as promising tools for automatic control and signal processing. One such ANFIS model is constructed in this paper to forecast the surface peak gust speed associated with pre-monsoon thunderstorms and the model forecast is compared with other models RBNN, MLPN and MLR (Maqsood et al. 2005).

Neuro-fuzzy hybrid system combines the fuzzy systems, which deal with explicit knowledge that can be expounded and realized, and neural networks deal with implicit knowledge that can be acquired by learning. Neural network learning provides an excellent mode to adjust the skillful knowledge and automatically generate additional fuzzy rules and membership functions to meet certain specifications and reduce the errors. The Trapezoidal-shaped membership function is used in this paper for the input parameters. The trapezoidal curve is a function of a vector, x and depends on four scalar parameters a, b, c and d;
This particular type of function is chosen for the study because of the nature of input vector parameters, which avail a large range in maxima (i.e. at membership function 1) and minima (i.e. at membership function 0). To map such input vectors, four scalar parameters (a, b, c and d) are required. The crisp input parameters are mapped into fuzzy membership function (Figures 4.2 – 4.5) using MATLAB software.

**Architecture of ANFIS**

ANFIS structure with n inputs and one output is shown in figure 6. Model initialization is done for common rule set with n inputs and m IF-THEN rules as follows (El-Shafie et al., 2007):

- **Rule 1:** If $X_1$ is $A_{11}$ and $X_2$ is $A_{12}$ and ... and $X_n$ is $A_{1n}$
  Then $u_1 = p_{11}X_1 + p_{12}X_2 + \cdots + p_{1n}X_n + q_1$

- **Rule 2:** If $X_1$ is $A_{21}$ and $X_2$ is $A_{22}$ and ... and $X_n$ is $A_{2n}$
  Then $u_2 = p_{21}X_1 + p_{22}X_2 + \cdots + p_{2n}X_n + q_2$

- **Rule m:** If $X_1$ is $A_{m1}$ and $X_2$ is $A_{m2}$ and ... and $X_n$ is $A_{mn}$

\[
f(x,a,b,c,d) = \begin{cases} 
    0, & x \leq a \\
    \frac{x-a}{b-a}, & a < x \leq b \\
    \frac{b-a}{b-a}, & b < x < c \\
    \frac{d-x}{d-c}, & c < x \leq d \\
    0, & x > d 
\end{cases}
\] (4.1)

```
0, x ≤ a
\]
\[
\frac{x-a}{b-a}, a ≤ x ≤ b
\]
\[
\frac{b-a}{b-a}, b < x < c
\]
\[
\frac{d-x}{d-c}, c < x ≤ d
\]
\[
0, x ≥ d
```

---

87
Then \[ u_n = p_mX_1 + p_mX_2 + \cdots + p_mX_n + q_m. \]

The corresponding ANFIS structure is illustrated in Fig. 6. The output of a fuzzy axon is computed using the formula;

\[ f_j (X,w) = \min \forall_i (MF(X_i,w_{ij})) \]

(4.2)

Where

\[ i = \text{input index}, \]
\[ j = \text{output index}, \]
\[ X_i = \text{input } i, \]
\[ w_{ij} = \text{weights (MF parameters) corresponding to the } j \text{th MF of input } i \]
\[ MF = \text{membership function of the particular subclass of the fuzzy axon}. \]

All layers in ANFIS structure are either adaptive or fixed.

The first layer is the input layer with four input nodes where the raw values of the four input parameters are stored. The next layer is the fuzzy membership layer of the input parameters. Three nodes (high, medium and low) define the membership grade of a fuzzy set \((A_{ij})\) and specify the degree to which the given input belongs to one of the fuzzy sets. The third layer is the logical operation layer with AND (multiplication) rule. The system acquires its adaptability by utilizing a hybrid learning method that combines back propagation and error optimization algorithms. There are basic learning methods like pattern learning (Jang, 1993) and hybrid learning rule. In the present study a hybrid learning rule is used which combines the gradient descent and the least squares estimate (LSE) to identify the
parameters. This layer receives the input in the form of product of all the output pairs from the previous layer:

$$w_j = \mu_{A_1}(X_1)\mu_{A_2}(X_2)\ldots\mu_{A_n}(X_n) \text{ for } (1 \leq i \leq m)$$  \hspace{1cm} (4.3)

In the same layer the weights ($w_j$) are normalized as follows;

$$\sum_{j=1}^{m} w_j f_j = \frac{\sum_{j=1}^{m} w_j f_j}{\sum_{j=1}^{m} w_j} \text{ for } (1 \leq j \leq m) \hspace{1cm} (4.4)$$

The multiplication of normalized weight and the network output gives;

$$w_j u_j = \bar{w}_j(p_{j1}X_1 + p_{j2}X_2 + \cdots p_{jn}X_n + q_j) \text{ for } (1 \leq j \leq m) \hspace{1cm} (4.5)$$

The overall output of ANFIS network is;

$$O = \sum_{j} \bar{w}_j u_j \hspace{1cm} (4.6)$$

There are forward and backward passes. In the forward pass of the hybrid learning algorithm, the functions operate forward and the ensuing parameters are identified by LSE. In the backward pass, the error rates propagate backward and the premise parameters are updated by the gradient descent method.

**Architecture of Artificial Neural Network (ANN)**

Artificial Neural Network (ANN) model (Bishop, 1995; Lawrence and Jeanette 1994) was designed to mimic the characteristics of the biological neurons in the human brain and nervous system.
In back propagation algorithm, an initial weight vector $w_0$ of a feed forward neural network (Figure 4.7) is iteratively adapted to find an optimal weight vector according to the recursion;

$$w_{k+1} = w_k + \eta d_k$$  \hspace{1cm} (4.7)

The positive constant $\eta$ represents the learning rate. The direction vector $d_k$ is the negative gradient of the output error function and is defined as;

$$d_k = -\nabla E(w_k)$$  \hspace{1cm} (4.8)

3.2.1 Multi Layer Perceptron (MLP)

A multilayer feed forward network (Wasserman and Schwartz, 1988) with 1 input neurons $m^1$ neurons in the first hidden layer, $m^2$ neurons in the second hidden layer and $n$ output neurons in the output layer is written as $1 - m^1 - m^2 - n$.

The hidden layer is computed as;

$$B_j = \sum_{i=1}^{m^1} W_{yi} A_j + W_{0j} A_0$$ \hspace{1cm} (4.9)

$B_j (j=1, 2... n) =$ hidden nodes

$A_i (i=1, 2... m) =$ input nodes

$W_{yi} =$ weights between nodes $A^i$ and $B^j$

The bias node ($A^0$) typically has a constant input of 1 with matching weight $W^{0j}$.
The model developed in the present study is one hidden layer neural net with back propagation algorithm to forecast the peak gust speed associated with severe thunderstorms which are described as;

\[ y_t = f(x_1, x_2, \ldots, x_n) \]

where \( (x_1, x_2, \ldots, x_n) \) represent the data on a given time series and \( y_t \) is the predicted value. In general, for each discrete \( t \), a different function \( f_t \) is used. The form of this function is obtained after adjusting a set of parameters that defines it. Usually, the available data are separated into a training set and a test set. The network learns by adjusting the inter connection between the input and hidden layers. When the learning or training procedure is completed, a suitable output is generated at the output layer. The number of neurons in the hidden layer can vary depending on the complexity of the problem and the size of input set (Gardner and Dorling, 1998).

Each neuron \( j \) receives incoming signals from every neuron \( i \) in the previous layer. Each incoming signal \( (y_i) \) is associated with a weight \( (w_i) \). The net input \( x^i_j \) to neuron \( j \) is a sum of the incoming signal times the weight;

\[ x_j = \sum_i y_i w_{ij} \]  

(4.10)

The most commonly used transfer or activation function is a sigmoid function. The transfer function \( f' \) is defined as;
\[ f(X) = \frac{1}{1 + \exp(-X)} \] (4.11)

3.2.2 Radial Basis Function Network (RBFN)

The Radial Basis Function Network (RBFN) is embedded in a two layer neural network where each hidden unit implements a radial activation function. In the present study, Gaussian activation function is used for RBFN;

\[ \Phi_j(X) = \| (X - \mu_j)^T \sum_j^{-1} (X - \mu_j) \| \] (4.12)

For \( j = 1, \ldots, L \); where \( L \) is the number of hidden unit and \( X \) in the input matrix.

\( \mu_j \) and \( \sum_j \) are the mean and covariance matrix of the \( j \)th Gaussian function.

The output layer implements a weighted sum of the hidden units output as;

\[ \Psi_k(X) = \sum_{j=1}^{L} \lambda_{jk} \Phi_j(X) \] (4.13)

For \( k = 1, \ldots, M \); where \( M \) is the number of output unit and \( \lambda_{jk} \) are the output weights.

The output of the RBFN is limited to the interval (0, 1) by the sigmoid function given by

\[ Y_k(X) = \frac{1}{1 + \exp[-\Psi_k(X) \}} \] (4.14)

The quality of prediction is assessed from the performance of the test set of data.

Percent errors of prediction (PE) are calculated as (Perez and Reyes 2001);
Where, \( \{ \} \) denotes the average of test cases. The predicted and actual values of the parameters are denoted by \( Y^{dp} \) and \( Y^{da} \) respectively. The root mean square error (RMSE) and mean absolute error (MAE) are computed for testing (validation) the output data;

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_{da} - Y_{dp})^2}
\]  

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_{dp} - Y_{da}|
\]

The predicted and actual values of the parameters are denoted by \( Y^{dp} \) and \( Y^{da} \) respectively.

### 4.2.3 Implementation Procedure

Skill score are more important than the accuracy-based prediction of detection (POD) and false alarm ratio (FAR) in evaluating the ability of the various stability indices and other parameters to forecast peak gust speed associated with thunderstorms. The Heidke Skill Score (HSS) uses random forecasts as the control set. However, Jolliffe and Stephenson (2003) demonstrated that Odds Ratio Skill Score (Yule's Q) is more preferable than HSS in evaluating the performance of binary forecasts. Nevertheless, either HSS or Yule's Q should provide a better assessment of performance than POD and FAR for binary events.
Critical Success Index (CSI) is the number of correct forecasts divided by the total number of occasions on which the events have been predicted. The Hit Rate (HR) or Percentage of Forecast Correct (PFC) is simply the fraction of the number of forecast occasions when the categorical forecast correctly anticipated the subsequent event or non event. The peak gust speeds over the threshold value of 60 kmh$^{-1}$ during the thunderstorm days are attempted to forecast in this study. Here the peak gust speed is the record of the highest instantaneous wind speed at the meteorological station (here Kolkata) during the period of thunderstorm activity (genesis to dissipation) for a given day. As per the records of India Meteorological Department the thunderstorm gust speed varies between 40 and 120 kmh$^{-1}$ over the region of study. The damage potential increases as the gust speed exceeds 60 km h$^{-1}$. The threshold value is thus chosen to be 60 kmh$^{-1}$. The input parameters of the regression model (or other models like ANFIS and neural network) are selected on the basis of skill score analysis which demands discrete data. This dichotomous (Yes/No) skill score analysis is done over a threshold value of 60 kmh$^{-1}$. The POD, CSI, HR, FAR and Yule’s Q of the selected stability indices are computed. It is observed that among all the stability indices, LI, CIN, CAPE and BRN are more consistent for evaluating the peak gust speed during thunderstorms over Kolkata (Figure 4.1). There are some other methods like principle component analysis and principled input selection methodology using ANFIS.
Anirban Middey

(http://www.mathworks.com/products/fuzzylogic/demos.html?file=/products/demos/shipping/fuzzy/gasdemo.html) for selecting the regressors. The major threat due to thunderstorm over the study region is strong gust speed. CAPE is directly related to the maximum potential vertical speed within an updraft. Thunderstorms are fueled by strong rising air, thus LI is a good indicator to produce severe thunderstorms. CIN is the amount of energy that acts as a barrier and must be supplied to a parcel to raise the level of free convection. BRN evaluates the balance between the instability (CAPE) and wind shear in a thunderstorm environment. The stability indices selected for the study might not be able to predict the intensity of thunderstorms individually but are useful while perform in a combination. During the period from 1997 to 2008 in the pre-monsoon season (April-May), 140 thunderstorm days are selected for the present study. Among these 140 thunderstorm days, 70 days are selected to train the models and the remaining 70 cases are kept for testing or validation of the models. The test set is totally independent of the training set. Four most consistent stability indices (LI, CIN, CAPE and BRN) computed from the sounding data before the occurrence of thunderstorm are taken as the input parameters.

4.3 Results and discussion

The ANFIS model output is found to have good resemblance with the observed peak gust speed during the pre-monsoon thunderstorms (Figure 4.8). The scatter plot (Figure 4.9) of observed and predicted peak gust speed shows high value of
R² (0.98) with ANFIS model. The prediction efficiency of ANFIS model is compared with other two non-hybrid AI models, RBFN and MLPN and a conventional statistical model, Multiple Linear Regression (MLR) analysis. The RBF network is prepared with four input single hidden layer and single output system. The hidden layer contains four hidden nodes. The actual peak gust speed and RBFN predicted peak gust speed also have good similarity (Figure 4.10). In this case the R² value of the scatter plot between observed and predicted peak gust speed is found to be 0.92 (Figure 4.11). MLP network is framed as 4:1(4):1 system, that is, four inputs with single hidden layer having four hidden nodes and single output system. The output of MLPN deviates slightly from the actual observation (Figure 4.12). The R² value of the scatter plot between observed and predicted peak wind speed is observed to be 0.86 (Figure 4.13). A conventional statistical model, Multiple Linear Regression (MLR) analysis is formulated to compare the prediction quality with the AI model outputs. Multiple regression provides a powerful method to analyze multivariate data. There are various methods and criteria for selecting variables for regression (Linhart, 1960; Wilks, 2006). The regression weights and are computed in a way that minimizes the sum of squared deviations (SSD) which is formulated as;

$$SSD = \sum_{i=1}^{N} (Y_i - Y'_i)^2$$  

(4.18)

The MLR equation is expressed as;
\[ Y_p = 42 - 2 \times LI + 0.13123 \times CIN + 0.0005488 \times CAPE + 0.001343 \times BRN \] (4.19)

Where \( Y_p \) is the predicted peak wind speed with the MLR analysis.

The prediction using MLR deviates more from actual observation than the AI models. The \( R^2 \) value of the scatter plot between observed and predicted peak wind speed using MLR analysis is found to be 0.845 (Figure 4.15). Error bars are estimated using the method of standard error computation from the selected data set. Standard error is computed as;

\[
\text{Standard Error (SE)} = \frac{s}{\sqrt{n}} \quad (4.20)
\]

where \( s \) is the standard deviation of the population and \( n \) is the number of observations.

Error bars, showing the standard errors associated with predicted data, are estimated (Figures 4.9, 4.10, 4.12 and 4.14). The error bars all having same size i.e. all are in overlapping region and hence statistically not significant. Error matrices with PE, RMSE and MAE are computed for different methods using equations 4.15, 4.16 and 4.17 (Figure 4.16). The prediction errors PE, RMSE and MAE during training the ANFIS models are found to be 3.5%, 3.05 and 2.18 respectively. The prediction error PE, RMSE and MAE with RBFN model are 10.48%, 9.42 and 7.92 respectively. The prediction errors PE, RMSE and MAE with MLPN are coming out to be 11.57%, 10.09 and 8.62 respectively. The overall prediction error PE, RMSE and MAE with MLR analysis are observed to
be 12.51%, 9.86 and 8.97 respectively. The hybrid ANFIS model shows its superiority over the other non-hybrid and conventional statistical method in the test cases (validation) also (Figure 4.17).

4.4 Validation

To check the potential and efficiency of the ANFIS predictive model, 50 cases are randomly picked up from the total data set of 140 cases. The overall prediction error with the different predictive models in the test cases are evaluated (Figure 4.16). During the validation, the hybrid ANFIS model shows to be superior to other comparative models in forecasting the peak wind speed associated with pre monsoon thunderstorms. The forecast errors are found to be slightly higher during the validation for all the models than during training the models. The overall prediction errors PE, RMSE and MAE with ANFIS model are observed to be 3.69%, 2.46 and 2.37 respectively, which is much better than the RBFN (12.25%, 11.56 and 8.86), MLPN (13.19%, 12.32 and 11.24) and MLR (14.86%, 13.57 and 13.45) model forecast. Error bar analysis is also provided in the graph of error matrices using SE computation from equation 4.20. The error bars show lowest variations in forecasting peak wind speed with ANFIS model (Figures 4.16 and 4.17). The skill scores of the wind event prediction for the final ANFIS model along with other models are shown in figure 18. As mentioned earlier in section 2, that the damage potential of thunderstorm gusts increases as the peak gust speed value exceeds 60 kmh⁻¹. So the analysis of wind event prediction skill scores is
done over the threshold value of 60 kmh\(^1\). The result depicts clear preeminence of ANFIS model over other models. It can thus, be stated that the ANFIS model is capable of forecasting the surface peak gust speed during thunderstorms over Kolkata with considerable accuracy and performs better than other models like RBFN, MLPN and MLR.

4.5 Applicability and limitations

The robustness of ANFIS model for accurate forecast depends on the training and testing of the model. ANFIS model can be applied to forecast any weather. In the present study, only four most useful stability indices (LI, CIN, CAPE and BRN) are considered to forecast peak wind speed through skill score analyses which are used to form input matrix of the model. The performance of ANFIS model has to be ascertained using more input parameters. There may be possibility that the RBF, MLP or MLR techniques might have performed better with different parameter choices which are yet to ensure.

5. Conclusion

In the present study, the attempt has been made to develop a hybrid AI model to predict the surface peak gust speed during pre-monsoon thunderstorms over Kolkata. Four most useful and relevant stability indices are selected as predictors from a broad range of stability indices by forecast skill score analysis. As a hybrid soft computing model, ANFIS has higher flexible adaptive network than the other non-hybrid ANN models. The hybrid learning procedure in ANFIS can adopt
complex behavior of input-output system. Other neural network methods like RBF and MLP are lacking of such adaptive hybrid learning procedures. Performance of ANFIS over a nonlinear, complex function simulating and chaotic statistics prediction will be much better than RBF, MLP and MLR methods. The ANFIS model is recommended to be a dependable predictive model for forecasting surface peak gust speed during thunderstorms over Kolkata. A comparative study with other models reveals the robustness and reliable performance of the ANFIS model.

ANFIS techniques are very familiar method in signal processing, data classification, time series forecast etc. But in atmospheric science field, the functioning of ANFIS methods are yet to explore in full fledges. The ANFIS model can also perform well in simulating cloud growth, constructing rainfall-runoff models. Wherever complex nonlinear functioning or chaotic time series predictions are there, ANFIS technique can be applied of course the success largely depends on the proper training and testing of the model.
Figure 4.1 The diagram showing the variations of statistical skill score parameters of the stability indices during pre monsoon season over Kolkata from 1997 to 2008.
Figure 4.2 The diagram showing the membership function for different ranges of Lifted Index (LI) in degree Celsius during pre-monsoon thunderstorms over Kolkata.
Figure 4.3 Diagram showing the membership function for different ranges of Convective Inhibition Energy (CIN)(J/kg) during pre-monsoon thunderstorms over Kolkata.

- Low
- Medium
- High
Figure 4.4 Diagram showing the membership function for different ranges of Convective Available Potential Energy (CAPE) (J/kg) during pre-monsoon thunderstorms over Kolkata.
Figure 4.5 Diagram showing the membership function for different ranges of Bulk Richardson number (BRN) during pre-monsoon thunderstorms over Kolkata.
Figure 4.6  The architecture of the Adaptive Neuro Fuzzy Inference System (ANFIS) with four input parameters.
Figure 4.7 Artificial Neural Network with four input-single hidden layers with four hidden nodes and one output system.
Figure 4.8 Diagram shows the variation in the observed and the forecast of wind speed with ANFIS during validation.
Figure 4.9 Diagram shows the scatter plot of observed and ANFIS model forecast of wind speed associated with pre-monsoon thunderstorms over Kolkata.

\[ R^2 = 0.9894 \]
Figure 4.10 Diagram shows the variation in the observed and forecast of wind speed with RBFN model during validation.
Figure 4.11 Diagram shows the scatter plot of the observed and RBFN model forecast of wind speed associated with pre-monsoon thunderstorms over Kolkata.
Test cases

- Observed
- Forecast with MLP

Figure 4.12 Diagram shows the variation in the observed and forecast of wind speed with MLP model over Kolkata during pre monsoon thunderstorms.
Figure 4.13 Diagram shows the scatter plot of observed and MLP model forecast of wind speed associated with pre-monsoon thunderstorms over Kolkata.

$R^2 = 0.8627$
Test cases

---

Figure 4.14 Diagram shows the variation in the observed and forecast of wind speed with MLR model over Kolkata during the pre monsoon thunderstorms.
Figure 4.15 Diagram shows the scatter plot of observed and MLR model forecast of wind speed associated with pre-monsoon thunderstorms over Kolkata.
Figure 4.16 Diagram shows different error matrices in forecasting peak gust for pre-monsoon thunderstorms using different models during training of the models.
Figure 4.17 Diagram shows different error matrices in forecasting peak gust for pre-monsoon thunderstorms using different models during validation (testing) of the models.
Figure 4.18 Diagram showing the skill scores of the wind event prediction for the final ANFIS model along with other models.