International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-3, Issue-4, July-2015

Forecasting the Water Vapour Distribution Over India Using Artificial Neural Networks

Venkata Sheshanna Kongara, Dr. D. Punyasesudu

Abstract—**Atmospheric water vapour plays an important role in radio communications for both terrestrial and earth space communication systems as well as for global climate change studies. Hence the water vapour data and its information plays vital role for scientists and researchers to evaluate the hidden patterns, trends for analysis and forecasting. In general radiosonde observations are the primary source of upper air water vapour data and are being used for estimation of different atmospheric parameters; on the other hand, different statistical and scientific methods are using to process the meteorological datasets in water vapour studies. However data assessment and usage is critical and challenging with many data level limitations to mitigate the accurate results. Currently there are extraordinary prospects in the Information Systems to process these data and explore. The data mining applications are the most promising features for radiosonde water vapour data forecasting easily and efficiently. As a part of this study, radiosonde observations over 34 Indian stations pertaining to a period of 17 years from 1997 to 2014 obtained from the British Atmospheric Data Centre(BADC) are pre-processed from the different met parameters of air temperature, dew point temperature and pressure levels for water vapour concentrations estimation at different height (pressure levels). This paper converse various analytics of surface level water vapour distribution patterns and trends with validation of the forecasting model based on the current trends by using improved version of artificial neural network (ANN) method. The proposed ANN model has been trained with training data and tested with test data for accuracy and better performance. The forecasted results are presented comparatively with the actual trends in the form of Physiographic Divisions, annual and seasonal charts water vapour distribution over India for better decision supporting systems**

Index Terms—**Atmospheric water vapour, climate change, Hidden Patterns, Data warehousing, Data mining, Decision supporting, artificial neural network, Forecasting model, Physiographic Divisions and radio communication**

I. INTRODUCTION

A. *Importance of the Water Vapour and Objectives of the present work*

The Water vapor (WV) is the most significant greenhouse gas and plays an important role in radio communications for both terrestrial and earth space communication systems as well as for global climate change studies and weather forecasting. The Water Vapour and Oxygen needs to be considered in the effective estimation performance of earth space communication connections. In the clear air the Water Vapour and oxygen absorption causes frequency dependant signal attenuation, propagation delay, ray bending and medium noise. Therefore it is an essential area of practice and research about Atmospheric water vapour [1] [2]; consequently the water vapour datasets usage and the method of data computation are the key factors in evaluation of atmospheric water vapour trends and usage. There are different statistical and scientific methods are using to process the water vapour datasets and measure the correlated innovations [3][4] [5][6][7].However due to the data limitations researchers are facing challenges in storing, retrieving, managing and exploration of these structured and un-structured data which is very large in size. Hence the approaches of data maintenance and conversion involves in time consuming, expensive, and complex to mitigate the accurate results; consequently most of the researchers choosing these datasets specific to the locations or regional levels with the limited time periods and limiting the scope of the study. In recent years there are extraordinary prospects in the Information Systems to process these data and explore. The data mining applications are the most promising features for radiosonde water vapour data forecasting easily and efficiently for effective decision supporting.

Data mining is defined as a technique to get useful, previously unknown information from databases and data warehouses. Data mining techniques are include association Rules, clustering, classification, prediction, sequence mining, web mining, text mining and spatial data mining [8] [9]. As a part of this study, radiosonde observations over 34 Indian stations pertaining to a period of 17 years from 1997 to 2014 obtained from the British Atmospheric Data Centre(BADC) are pre-processed from the different met parameters of air temperature, dew point temperature and pressure levels for water vapour concentrations estimation at different height (pressure levels). This paper converse various analytics of surface level water vapour distribution patterns and trends with validation of the forecasting model based on the current trends by using improved version of artificial neural network (ANN) method. The proposed ANN model has been trained with training data and tested with test data for accuracy and better performance. The forecasted results are presented comparatively with the actual trends in the form

Manuscript received 3rd July,2015

Venkata Sheshanna Kongara, Research Scholar, Department of Computer Science & Technology, Sri Krishnadevaraya University,Anantapur-515 003,Andhra Pradesh, India. +91 9573544822.

Dr. D.Punyasesudu, Professor, Department of Physics, Rayalaseema University, Kurnool, 518 003, Andhra Pradesh, India,

of Physiographic Divisions, annual and seasonal charts water vapour distribution over India for better decision supporting systems. In fact it is an improved version of the detailed studies related to water vapour distribution over India.

B. Physiographic features of India

India is a country of great geographical extent with a diversity and rich cultural heritage. The physical features of India are not same across the country. They are varying from place to place with significant features which are highest mountains in the north, wide plains in the centre, desert region in the northwest, narrowing peninsula to the south, mountain ranges to the northeast and a lower range running along the west coast of the peninsula. The vast expanse of the Indian Ocean lies to the south of the land mass with the Bay of Bengal and Arabian Sea pushing north engulfing the peninsula [10], Hence Physiographic features plays an important role on the climatic variations of the country. And also the magnitude of these climatic changes varies from different regions and point in time which are observed that different climatic seasons in a year commonly known as winter, summer, monsoon and post monsoon. The seasons are categorized from various months incorporated as winter [December to February], summer [March to May], Monsoon [June to September] and the post monsoon [October to November]. The physiographic divisions as shown in the map [11] figure 1 and the water vapour seasonal trends with physiographic divisions mentioned in table A: are analyses in this study

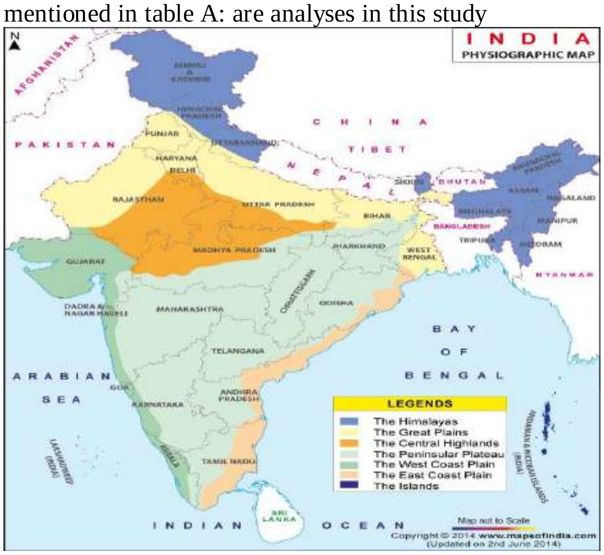


Figure 1: The physiographic divisions by Maps of India

Following are the different categories of the Physiographic divisions over India are considered in the study.

i. The Kashmir Region: These are highest mountain ranges are located in North Western Region of India. This region is uninhabited precipitation due to the Himalayan Mountains

ii. The Assam Region: Covers different states in North Eastern Region of India, the climate is cold and rainfall most of the months

iii. The Northern Plains: The Indo-Gangetic plains located in Northern and Eastern Region of India. The climate is very hot in summers and very cold in winters

iv. The Central Highlands: These are composed of different Plateaus which are lies in North Central Region of India. The climate is with hot summers and mild winters

v. The Desert Region: These are lies in the North Western region of India, The climate is very hot, and rainfall is inadequate and unpredictable

vi. The Peninsular Plateau: The Peninsular plateau lies to the south of the northern plains of India with different weather

vii The Eastern Coastal Plains: These are lies along the east coast of the Southern Region of India and washed by Bay of Bengal.

viii. The Western Coastal Plains: These are lies along the west coast of Southern Region of India and washed by Arabian Sea. The rainfall is heavy in this region

ix. The Island: The one in Bay of Bengal called as Andaman and Nicobar islands which is in South Eastern Region of India and Arabian Sea called as Lakshadweep islands which is in South Western Region of India

The rest of the paper is organized as follows. The literature review and related work for water vapour studies are discussed in section II. The details of the water vapour data processing and Methodology for training a neural network are described in Section III. The data analysis and results are presented in section IV and concluded the paper with accomplishment in the study and future work in Section V.

II. RELATED WORK

Smita.Nirkhi et al [12] reviewed on how to apply Artificial Neural Network in Data mining techniques. The Neural network features are suitable for solving the problems of data mining with its characteristics of good robustness, self organizing adaptive, parallel processing, distributed storage and high degree of fault tolerance. Hence they have been successfully applied in a wide range of supervised and unsupervised learning applications in various studies.

WANG Yong and XU Hong et al [13] discussed on Water vapor and its changes directly affected the weather. It is one of key factors about severe weather formation and evolution. Accurate, and timely rainfall forecast is also important factors which increased forecast accuracy of storms, floods and other disastrous weather. Hence they build models of the data for training and simulation based on neural network technology, and analyzed the results of rainfall forecast by using GPS precipitable water vapor and other meteorological parameters. Through data preprocessing, BP neural network modeling and analysis it has been completed the design of rainfall forecast.

Stefania Bonafoni et al [14] proposed a method based on neural networks to retrieve integrated precipitable water vapor (IPWV) over land from brightness temperatures measured by the Advanced Microwave Scanning Radiometer Earth Observing System (AMSR-E). Water vapor values provided by European Centre for Medium-Range Weather Forecasts (ECMWF) were used to train the network. The performance of the network was demonstrated by using a separate data set of AMSR-E observations and the corresponding IPWV values from ECMWF. Their study was optimized over two areas in Northern and Central Italy. In addition, results were compared with the IPWV values obtained from ground-based radiometer, and a global positioning system (GPS) receiver respectively.

Bimal Dutta et al [15] artificial neural network model discussed and has been developed to run humidity forecast for a day based on the previous day's data. The humidity data from weather stations and the meteorological data from Kolkata Meteorological department were collected during 1989-1995 with several meteorological parameters like temperature, humidity, air pressure and vapour pressure as an input for training the model. The novel architecture of the proposed model contains several multilayer perceptron network (MLP) to realize better performance. The model is enriched by analysis of several alternative models like online feature selection MLP (FSMLP) and self organizing feature map MLP (SOFMMLP). The improvement of the performance in the prediction accuracy has been demonstrated by the selection of the appropriate features and two alternative ANN models were tested with continuous humidity data and compared. So feature selection technique (FSMLP) can be used to increase the predicted result of neural network based prediction system.

Sridevi Jade et al [16] conducted the study on Indian GPS network data along with the interpolated National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis meteorological data to estimate the precipitable water vapor (PWV) over the Indian atmosphere for a 4 year period (2001–2004) at 21 Indian GPS and 7 International GNSS Service stations.

Wayan Suparta et al [17] conducted the study on the estimation of precipitable water vapor (PWV) values over Malaysia especially Bangi and Sabah region using adaptive neuro fuzzy inference system (ANFIS) and multi layer perceptron (MLP) models. Parameters used to develop these models were the surface meteorological data and PWV data from GPS. Data observations were used for training and testing model starting from 1 – 30 November 2012 of University Kebangsaan Malaysia (UKMB) station and 1-30 September of University Malaysia Sabah (UMSK) station. Two kinds of training data set model were provided for ANFIS and MLP models. Comparison between the result shows that ANFIS techniques have a better performance in estimating and prediction the PWV value compared to MLP model.

[18] A. J. Litta et al studied and evaluated the utility of ANN for estimating hourly surface temperature and relative humidity, different experiments are conducted with artificial neural network model to predict severe thunderstorms that occurred over Kolkata during May 3, 11, and 15, 2009, using thunderstorm affected meteorological parameters. The capabilities of six learning algorithms, namely, Step, Momentum, Conjugate Gradient, Quick Propagation, Levenberg-Marquardt, and Delta-Bar-Delta, in predicting thunderstorms and the usefulness for the advanced prediction were studied and their performances were evaluated by a number of statistical measures. The results indicate that Levenberg-Marquardt algorithm well predicted thunderstorm affected surface parameters and 1, 3, and 24 h advanced prediction models are able to predict hourly temperature and relative humidity adequately with sudden fall and rise during thunderstorm hour. The developed ANN model with LM algorithm was used to predict surface temperature and relative humidity at hourly intervals with 1, 3, 6, 12, and 24 h ahead during same severe thunderstorm cases.

Dr. S. Santhosh Baboo et al [19] discussed about an efficient weather forecasting system using Artificial Neural Network. In their study the back propagation neural network is used for predicting the temperature based on the training set provided to the neural network. Through the implementation of the system, it is illustrated, how an intelligent system can be efficiently integrated with a neural network prediction model to predict the temperature. This algorithm improves convergence and damps the oscillations. This method proves to be a simplified conjugate gradient method. When incorporated into the software tool the performance of the back propagation neural network was satisfactory as there were not substantial number of errors in categorizing. Back propagation neural network approach for temperature forecasting is capable of yielding good results and can be considered as an alternative to traditional meteorological approaches.

III. DATA PROCESSING & METHODOLOGY

A. Source of the data

The global atmospheric radiosonde data has millions of soundings from worldwide upper air stations over the period 1996 to present. Each sounding is composed of records of pressure, geo-potential height, temperature, dew-point temperature, wind-speed and wind-direction at standard and significant pressure levels proposed by World meteorological Organization (WMO) from the surface to approximately 20-30 km. This data was extracted from the daily occurrences of atmospheric soundings compiled by British Atmospheric Data Centre (BDAC) and is made available for individuals and organizations for atmospheric studies[20].The radiosonde takes measurements at intervals of approximately 2 seconds. The high resolution data files contain all such data. The standard resolution data files contain measurements taken at particular levels of the atmosphere. We obtained different stations details, air pressure, air temperature, dew point temperature, wind speed, wind direction and various other parameters for the time period 1997-2014 over India. Further based on the data availability we have selected the necessary stations for various geographical locations specified in Table A and data cleansing activity is conducted to focus on the target data. As per the accessibility all the datasets are collected in GMT and converted the dates from GMT time to India IST for 00:00 GMT as 5:30 AM IST and 12:00 GMT as 17:30 PM IST for all the datasets. The Pressure levels are considered at the surface levels while validating the altitude in Meters to identify Surface Pressure for various stations based on the station height. Extracted the rules and map them into different dimensions and measures to produce understandable and useful knowledge for analysis

B. Compute the Integrated Water Vapour Concentration

H. Sarkar et al [21] conducted Comparative Study of Integrated Water Vapor (IWV) and of Attenuation of 94

GHz Signal from Radiometer and Radiosonde Observations during Monsoon Period over Kolkata, India. The Radiosonde data they have calculated the integrated water vapor content at different heights for the specific dates using regular empirical formula. However Khamphoui et al [22] discussed on the impact of water vapor in atmosphere and clear sky attenuation is mainly due to the absorption caused by water vapour and oxygen molecules. It increases with the relative humidity as well as the temperature and alleged that the water vapour concentration is strong function temperature and humidity. In their studies they have used most revised version of the equation to calculate the water vapour concentration (M) by using following equation. We are also using the identical equation for our studies for evaluation of the integrated water vapour as below.

The VAPOUR_CONCENTRATION (M) are calculated based on the Vapour Pressure (i.e.) → 1

Vapour Concentration (M) in grams/meter cube (g/m^3)

$$= 216.7 * \left[\frac{Vapour\ Pressure\ (E)}{Air\ Temperature\ (K)}\right]$$

The VAPOUR_PRESSURE (E) are calculated based on the Empirical formula of vapour pressure (E) (i.e.) → 2

$$Vapour\ Presssure\ (E) = 6.11 * Exp\left[\frac{19.7 * Dew\ Point\ Temperature\ oC}{Dew\ Point\ Temperature\ oC + 273.15}\right]$$

The DEW_POINT_TEMPERATURE are converted from Kelvin to Degree Centigrade for analysis (i.e.) → 3

$$Dew\ point\ Temperature\ oC = Dew\ point\ Temperature\ (K) - 273.15$$

The calculated water vapour concentration datasets are arranged in the dimensional tables of the data warehouse for our analyses, which may be reusable for the further studies.

C. Water Vapour Data Pre-Processing and Classification

Data preprocessing is a fundamental stage of data analysis for better performance and good quality of results. In general missing data are unintended and uncontrolled by the researchers, but the overall result is that the observed data cannot be analyzed because of the incompleteness of the data sets. Missing Values and its problems are very common in the data cleaning process. Several methods have been proposed [23] [24] so as to process missing data in datasets and avoid problems caused by it. In the radiosonde source data also there are some missing values for various parameters for different stations and different dates of the years. The BDAC filled those missing data sets with 999999999 in their data, Hence we have handled those missing values filled with précised one by using cluster based missing values algorithm is consider to process the water vapour data for better results.

Building effective classification systems is one of the central tasks of data mining. Many different types of classification techniques are available that includes Decision Trees, Naive-Bayesian methods, Neural Networks, Logistic Regression. We conducted a comparative review for various methods of classification and measured the accuracy of the results. It show that using Neural Networks obtains the best result among the other methods therefore in the water vapour studies we have used improved version of neural networks algorithm to classify the water vapour data for better classification and prediction.

D. Artificial Neural Network Models:

As acknowledged in the Introduction of this paper data mining Artificial Neural Network applications are the best suitable for atmospheric water vapour studies. An artificial neural network (ANN) is a mathematical model based on biological neural networks, in other words, is an emulation of biological neural system. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. In the Figure 2 shown the Biological and structural aspects of the ANN model for better understanding

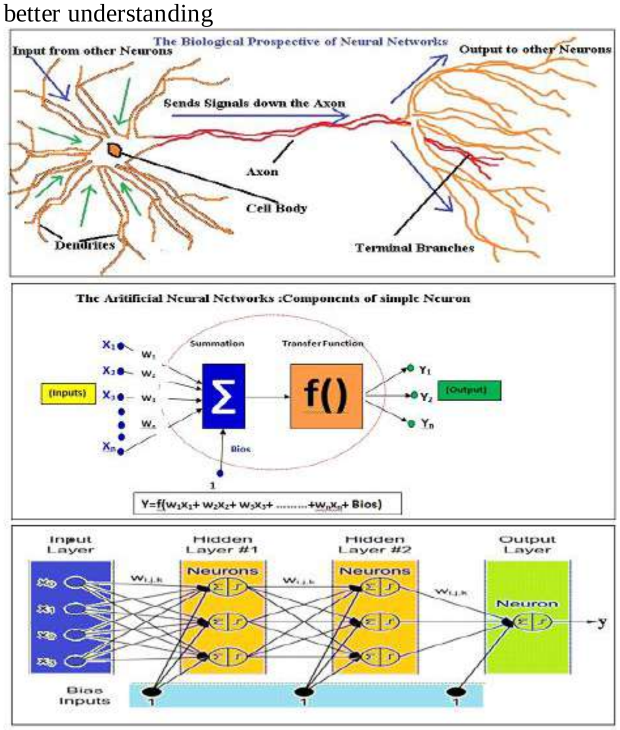


Figure 2: The Biological and structural aspects of the ANN model

i. Back Propagation Neural Networks

Back propagation, or propagation of error, is a common method of teaching artificial neural networks how to perform a given task. The back propagation algorithm is used in layered feed forward ANNs. This means that the artificial neurons are organized in layers, and send their signals “forward”, and then the errors are propagated backwards. The back propagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then the error (difference between actual and expected results) is calculated. The idea of the back propagation algorithm is to reduce this error, until the ANN learns the training data.

The Back propagation Algorithm was proposed by Rumelhart, McClelland in 1985.It has been one of the most studied and used algorithms for neural networks learning ever since.

Below pseudocode illustrates The Back Propagation Algorithm

1. Feed the Input data values X0,X1,X2,X3……Xi of into the Network (l) (Feed Forward)
2. Initialize the weights as random numbers of [-1,+1] intervals for each Input value (Wij)l
3. Apply the summation of all such inputs and weights as a forward Computation of all Xi(l)
4. Propagate the input values of all the forward Computations using Activation Function for normalization (as compute the net input and the output of each unit in the hidden and output layers of Yi for Y0,Y1,Y2…Yn
5. Back propagate the error from output values Yi to input Xi like Y1-X1=?(as calculate error gradients)
6. Update the network input value's weights to reflect the propagated error
7. Iterate the next step until it is time to stop to fit the best feasible error minimization
8. Return the final weights and terminate

ii. Building a Water Vapour Distribution Forecasting Neural Network Model (WVDFNN) using Gradient Descent Training

Valmik B Nikam et al [25] conducted study after extracting knowledge from weather historical data collected from Indian Meteorological Department (IMD) Pune, for Modeling Rainfall Prediction Using Data Mining Method-A Bayesian Approach. The collected weather data comprising of 36 attributes, only 7 attributes (like Surface Level Pressure (SLP), Mean-Sea Level Pressure (MSLP), Dew Point temperature (DPT), Relative Humidity (RH), Vapor Pressure (VP), Wind Speed (FFF) etc.) are most relevant to rainfall prediction. They made data preprocessing and data transformation on raw weather data set, so that it shall be possible to work on Bayesian, the data mining, prediction model used for rainfall prediction. The model is trained using the training data set and has been tested for accuracy on available test data. And they concluded that the model has simplicity, good prediction performance, and can be used for both binary and multiclass prediction problems. The Bayesian prediction model can easily learn new classes. The accuracy will grow with the increase of learning data.

Deepak Ranjan Nayak et al [26] conducted a survey on different prediction methods of MLP, BPN, RBFN, SOM and SVM and said that these are methods are suitable to predict rainfall than other forecasting techniques such as statistical and numerical methods. They concluded that some limitation of those methods has been found in their studies. Their survey also reports that rainfall prediction using ANN technique is more suitable than traditional statistical and numerical methods

However we are also conducted a comparative study for various methods of forecasting and measured the accuracy of the results. It shows that using Neural Networks gradient descent based method of training obtains the best result among the other methods for the water vapour distribution forecasting studies over India.

Smita Kulkarni et al [27] also conducted similar studies to identify a non-linear methodology to forecast the time series of average summer monsoon rainfall over India. Three advanced back propagation neural network learning rules namely, momentum learning, conjugate gradient descent (CGD) learning, and Levenberg–Marquardt (LM) learning, and a statistical methodology in the form of asymptotic regression are implemented for this purpose. Monsoon rainfall data pertaining to the years from 1871 to 1999 are explored. After a thorough skill comparison using statistical procedures the study reports the potential of CGD as a learning algorithm for the back propagation neural network to predict the said time series.

In our studies we establish that improved version of neural networks Back Propagation algorithm of gradient descent with momentum and adaptive learning rate back propagation method. The neural network training was done using MATLAB R2015a applications with effective forecasting of the water vapour distribution for different physiographic divisions and seasonal variations in prospect trends, which are presented in section IV Results and Analysis.

Following illustration and algorithm will explain the better understanding

The gradient descent technique is come within reach of to find a local minimum of a function in the training. The approach it works is where we start with an initial guess of the solution and we take the gradient of the function at that point, then step the solution in the negative direction of the gradient and we repeat the process.

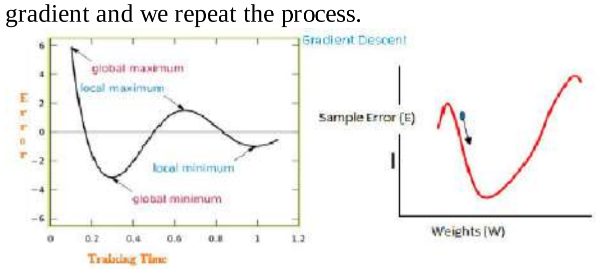


Figure 3: Gradient descent method for error evaluation with training time and weights to evaluate the local/Global minimum

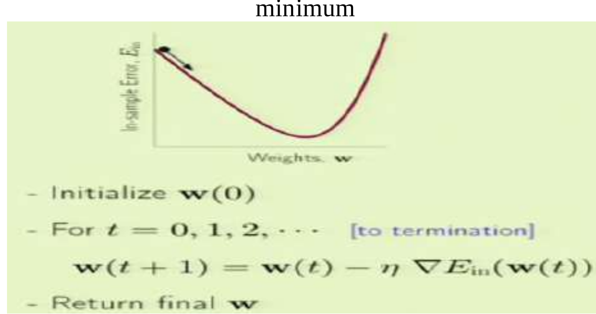


Figure 4: Gradient descent algorithm

Gradient Descent (GD) method:

We have Gradient Descent (GD) minimizes which is minimizes Error Function

GD minimizes E (W) that is function of W minimizes W that happen in sample of error in mind

In order to compute the gradient of the error, we need to evaluate the hypothesis at every point in the data sample of Xn input, Yn vectors

$$Error\,(Weight\ W) = \frac{1}{N}\sum_{n=1}^{N} e(h(Xn), Yn)$$

By iterative steps along $-\nabla E$

$\Delta(W) = -\eta \nabla E(W)$

The W space $\Delta Error\,(Weight)$ is based on all examples of (Xn, Yn), here the η is the learning rate

Step1: Pick one (Xn, Yn) example at a time randomly.

Step2: Apply GD to **e** (h (Xn), Yn) not to the in sample error all the points, but the sample error at that point

Step3: Think of the "Average" Direction that we are going to descent along (It means if we take the gradient of the error that measure that we are going to minimize in this case for just 1 example $-\nabla e(h(Xn), Yn)$ and we take the expected value on the experiment that we picked up from the examples of the entire training set at random which is

$$En[-\nabla e(h(Xn), Yn)] = \frac{1}{N}\sum_{n=1}^{N} -\nabla e(h(Xn), Yn)$$

(for every example probability of 1 or N and the expected value would 1 or N and summation of those values). It should be the average direction. So we are going every step along with direction plus noise

Step4: look at the quantity at the right hand side of (step 3) happens to be identically $-\nabla E$ total in sample error. So at if an expected value going along in the direction we want except that we know the one example involved in the computation is big advantage and we have stochastic aspect to the gain.

Step5: keep repeating until to the expected value in that direction, by that time we did it lot of time and the noise will be average out and actually we are going in the ideal direction

iii. The Evaluation criteria for WVDNN Forecasting

Following are the different evaluation methods to validate the network in the training from input values to the output computed and the obtained error after training

Mean Square Error (MSE) =

$$Error(Weight\ W) = \frac{1}{N}\sum_{n=1}^{N} (\mathrm{Yn} - \mathrm{Xn})^2$$

Where *Yn* is observed values and *Xn* is trained values in the sample

Root Mean Square Error (RMSE):

The RMSE is a frequently used measure of the difference between values predicted by a model and the values actually input from the sample that is being trained. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power.

The RMSE of a model prediction with respect to the estimated variable X_{input} is defined as the square root of the mean squared error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{target} - X_{input,i})^2}{n}}$$

Where X_{target} is observed values and X_{input} is trained values in the sample

The correlation coefficient (R-value):

The correlation coefficient (R-value) performing a linear regression between the ANN-predicted values and the targets and is computed by Regression R

$$R = \frac{\sum_{i=1}^{N} t_i p_i}{\sqrt{\sum_{i=1}^{n} t_i^2}\sqrt{\sum_{i=1}^{n} p_i^2}}$$

where R is correlation coefficient; N is the number of samples; ti =Ti −T; Pi = Pi − P and Ti and Pi are the target and predicted values for i=1,….,N and T and P are the mean values of the target and predicted data set, respectively.

i).While validation case with R is equal to 1 refers to a perfect correlation and the predicted values are either equal or very close to the target values

ii).The R and RMSE are used in the Network validation phase

iii). MSE is used for network training

iv. Training the WVDF Neural Network algorithm of gradient descent with momentum and adaptive learning rate back propagation method

As explained in the introduction the ANN network consisting set of Input layers, Hidden Layers and the output layers, while training the network following steps are using in MATLAB nntool

Step 1: Arrange the Inputs and the Output values to prepare a Neural Network for winter, summer, monsoon and post monsoon seasons of the Actual Year

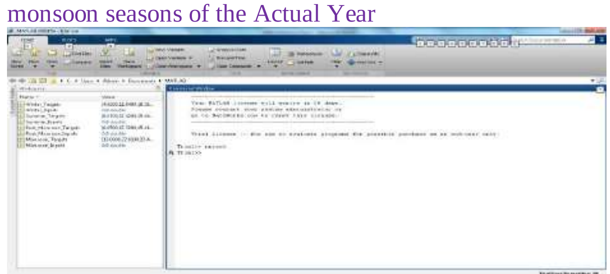


Figure 6: MATLAB R2015a application interface and command window to set the input and the out layers data for training

Step2: Choose the Neural Network tool (nntool) to design an ANN model with required Inputs and Outputs to fit into the Network for training

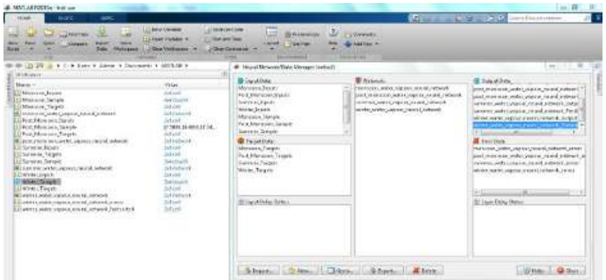


Figure 5: Neural Network/Data Manager (nntool) tool to arrange the input/output data and the build the networks

Step 3: Build the Network with different parameters and training

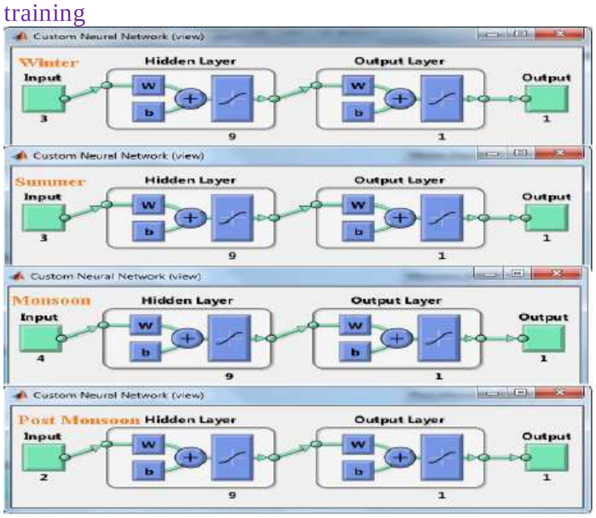


Figure 7: different seasonal neural networks build for training.

While training a network, followed below properties and options

1. Selected the Network Type as Feed-forward backprop
2. Given set of Input vales as required for Input Data
3. Given Set of expected output vales as required for Target Data
4. Selected the training function as TRAINGDX which is Gradient descent with momentum and adaptive learning rate back propagation
5. Selected the Adaption learning function as LEARNGDM which is Gradient descent with momentum weight and bias learning function
6. Selected the performance function as MSE which is Mean squared normalized error performance function
7. Select the number of layers 1 for output layer
8. Selected the 9 hidden layer neurons for 3*9 sized data as a input
9. Selected the transfer function as LOGSIG which is Log-sigmoid transfer function

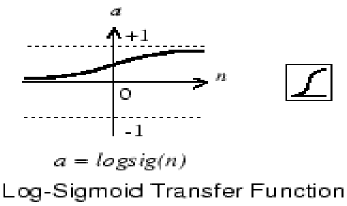


Figure 8: Sigmoid activation for logsig (n) = 1 / (1 + exp (-n))

10. Created the network for training

Step4: Train the Network using with different parameters of Training Error Performance, Training state and the Network Training Regression until for the better results

1. Set the input data and target data values for the network
2. set the training results outputs and error parameters
3. set the training parameters like no of epochs as 1000
4. set the max fail epochs to 1000 to complete and exit
5. applied Train Network to retrieve the results

Training Error Performance:

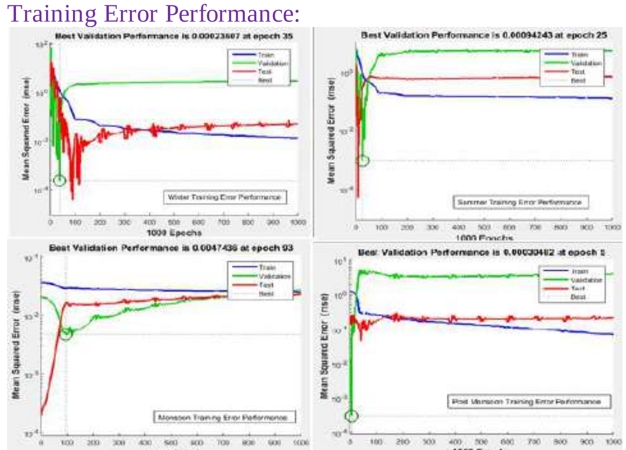


Figure 9: shows the Training performance as different hidden neurons chose for training network for the best validation out of 1000 epochs

Training state

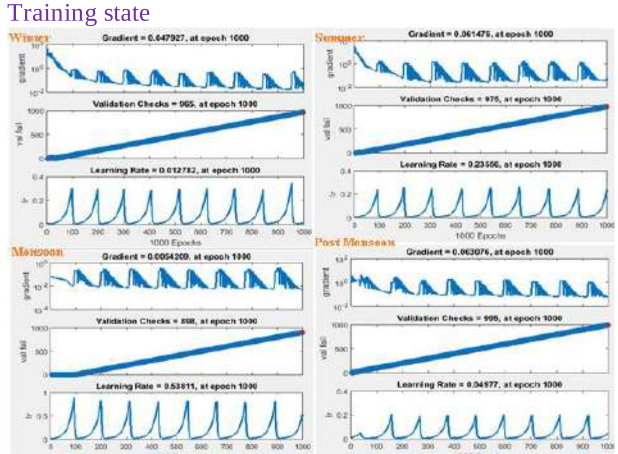


Figure 10: shows the changing curve of error in different fitting options of gradients, validation checks and the learning rates at epoch 1000

Network Training Regression

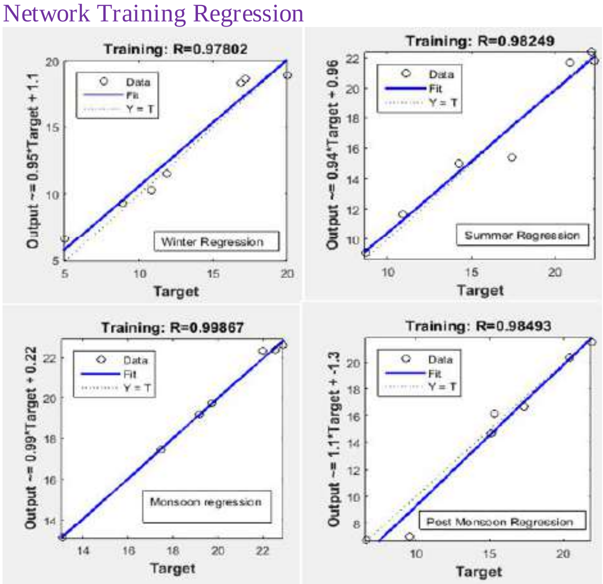


Figure 11: Shows that network training regression R Values measure the correlation between outputs and targets.

All the plots have a high value better than 0.98 to 0.99. The outputs of the training network are quite close to the targets. Hence the network model has a good training performance.

Step5: Evaluate the Input data and the output /forecasted data return from the network along with the simulation results for better analysis

Network training is carried out while evaluating the following

1. Selected the Data Divisions at Random
2. Training method used as Gradient Descent with Momentum & Adaptive LR (traingdx)
3. The Performance is validated with Mean Square Error (MSE)
4. The computation are used as MEX compiler used for validation
5. While training set the epochs progress with in 1000 iterations and validated the best fit for the stopping criteria
6. Fit the minimum accurate training time for the network for validation
7. Fit the performance of network for better training and validation
8. Validated the best fit of the gradient to mitigate the error of the network
9. Trained the network until to fit the feasible parameter for the better results
10. Build the plots for the performance, Training state and regression as shown in the Figure 11

IV. ANALYSIS AND RESULTS

We have conducted a comparative study for various methods of ANN forecasting and measured the accuracy of the results. It shows that using Neural Networks obtains the best result among the other methods for the water vapour distribution forecasting studies over India.

The work is anticipated for validation of the forecasting model based on the current trends with water vapour distribution by using neural networks for different physiographic divisions and seasonal variations in prospect trends

The graphs Illustrates the various physiographic divisions monthly mean water vapour density trends in grams per cubic meter (gm/m3) at Surface Level during 2014 winter, summer, monsoon and post monsoon seasons of the forecasted year that is for the year 2015 and the correlation coefficient given the best fit of the forecasted values for analysis and decision supporting.

The following table describes the actual water vapour input values Vs. Forecasted water vapour distribution over India.

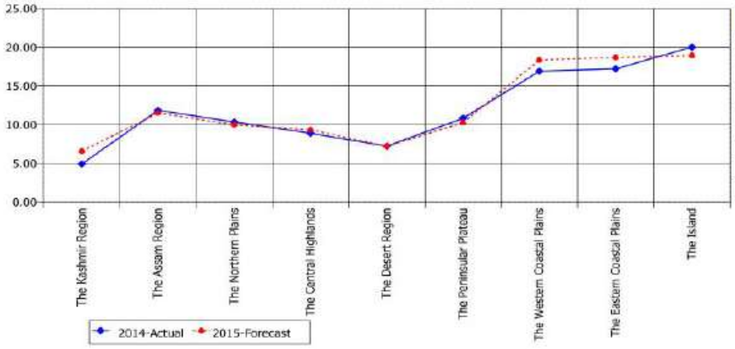
Season	Physiographic Division	2014-Actual	2015-Forecast
Winter	The Kashmir Region	4.92	6.57
Winter	The Assam Region	11.84	11.50
Winter	The Northern Plains	10.35	9.98
Winter	The Central Highlands	8.89	9.33
Winter	The Desert Region	7.22	7.20
Winter	The Peninsular Plateau	10.83	10.26
Winter	The Eastern Coastal Plains	17.23	18.69
Winter	The Western Coastal Plains	16.92	18.35
Winter	The Island	20.02	18.93
Summer	The Kashmir Region	8.67	9.08
Summer	The Assam Region	17.42	15.39
Summer	The Northern Plains	15.44	15.47
Summer	The Central Highlands	10.91	11.68
Summer	The Desert Region	9.76	9.23
Summer	The Peninsular Plateau	14.27	15.01
Summer	The Eastern Coastal Plains	22.38	21.77
Summer	The Western Coastal Plains	20.86	21.68
Summer	The Island	22.2	22.36
Monsoon	The Kashmir Region	13.08	13.12
Monsoon	The Assam Region	22.91	22.61
Monsoon	The Northern Plains	22.4	22.47
Monsoon	The Central Highlands	19.73	19.71
Monsoon	The Desert Region	17.49	17.45
Monsoon	The Peninsular Plateau	19.21	19.18
Monsoon	The Eastern Coastal Plains	22.62	22.49
Monsoon	The Western Coastal Plains	22.02	22.31
Monsoon	The Island	22.55	22.38
Post Monsoon	The Kashmir Region	6.65	6.72
Post Monsoon	The Assam Region	17.32	16.67
Post Monsoon	The Northern Plains	15.11	14.73
Post Monsoon	The Central Highlands	13.04	13.06
Post Monsoon	The Desert Region	9.62	7.00
Post Monsoon	The Peninsular Plateau	15.32	16.13
Post Monsoon	The Eastern Coastal Plains	20.4	20.38
Post Monsoon	The Western Coastal Plains	19.9	20.30
Post Monsoon	The Island	21.86	21.53

Table A: The actual water vapour distribution Vs. Forecasted water vapour distribution over India for different seasons.

Following are the analytical reports from the neural network provided target data of the forecasted water vapour distribution

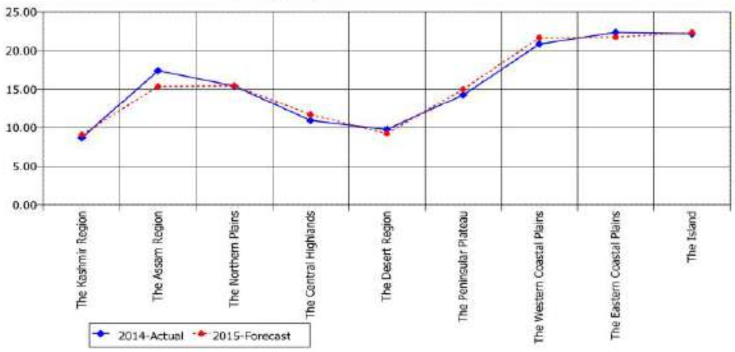
A) Seasonal Water Vapour (grams/cubic meter g/m^3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting for different Physiographic Divisions Over India
B) Physiographic Divisions and 4 Seasonal Water Vapour (g/m^3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India

Winter Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting for different Physiographic Divisions Over India



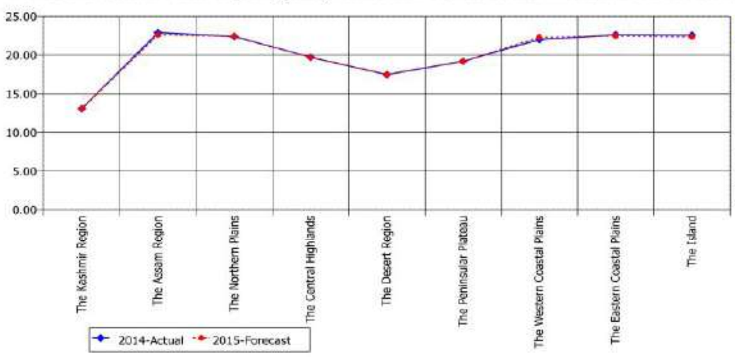
Physiographic Division	2014-Actual	2015-Forecast	Accurate %
The Kashmir Region	4.92	6.57	133.52
The Assam Region	11.84	11.50	97.17
The Northern Plains	10.35	9.98	96.44
The Central Highlands	8.89	9.33	104.93
The Desert Region	7.22	7.20	99.79
The Peninsular Plateau	10.83	10.26	94.77
The Western Coastal Plains	16.92	18.35	108.48
The Eastern Coastal Plains	17.23	18.69	108.45
The Island	20.02	18.93	94.54

Summer Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting for different Physiographic Divisions Over India



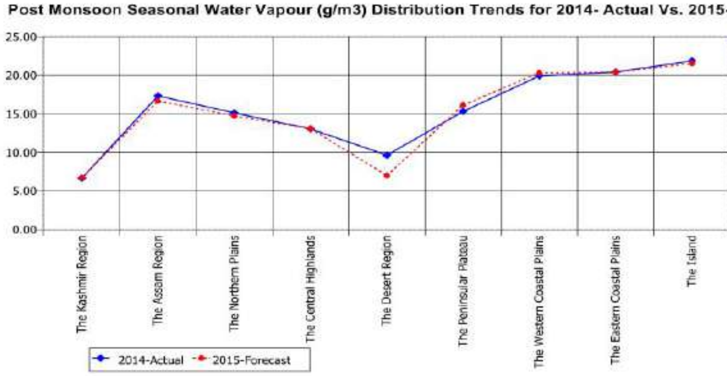
Physiographic Division	2014-Actual	2015-Forecast	Accurate %
The Kashmir Region	8.67	9.08	104.72
The Assam Region	17.42	15.39	88.37
The Northern Plains	15.44	15.47	100.20
The Central Highlands	10.91	11.68	107.08
The Desert Region	9.76	9.23	94.52
The Peninsular Plateau	14.27	15.01	105.22
The Western Coastal Plains	20.86	21.68	103.93
The Eastern Coastal Plains	22.38	21.77	97.29
The Island	22.20	22.36	100.71

Monsoon Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting for different Physiographic Divisions Over India



Physiographic Division	2014-Actual	2015-Forecast	Accurate %
The Kashmir Region	13.08	13.12	100.30
The Assam Region	22.91	22.61	98.71
The Northern Plains	22.40	22.47	100.31
The Central Highlands	19.73	19.71	99.91
The Desert Region	17.49	17.45	99.80
The Peninsular Plateau	19.21	19.18	99.82
The Western Coastal Plains	22.02	22.31	101.34
The Eastern Coastal Plains	22.62	22.49	99.43
The Island	22.55	22.38	99.24

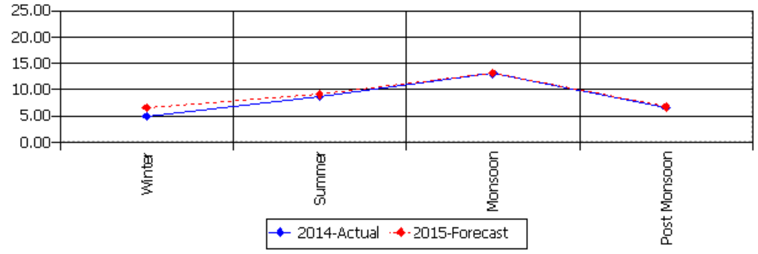
Post Monsoon Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting for different Physiographic Divisions Over India



Physiographic Division	2014-Actual	2015-Forecast	Accurate %
The Kashmir Region	6.65	6.72	101.02
The Assam Region	17.32	16.67	96.23
The Northern Plains	15.11	14.73	97.47
The Central Highlands	13.04	13.06	100.13
The Desert Region	9.62	7.00	72.72
The Peninsular Plateau	15.32	16.13	105.29
The Western Coastal Plains	19.90	20.30	102.03
The Eastern Coastal Plains	20.40	20.38	99.92
The Island	21.86	21.53	98.50

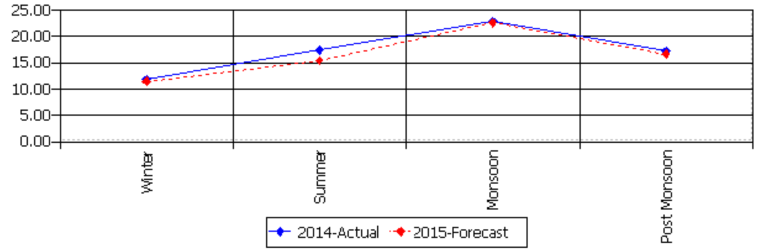
Figure 12: The Seasonal Water Vapour (g/m^3) distribution forecasting trend for different Physiographic Divisions over India

The Kashmir Region Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



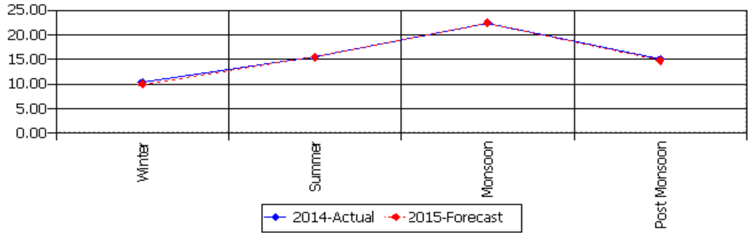
Season	2014-Actual	2015-Forecast	Accurate %
Winter	4.92	6.57	133.52
Summer	8.67	9.08	104.72
Monsoon	13.08	13.12	100.30
Post Monsoon	6.65	6.72	101.02

The Assam Region Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



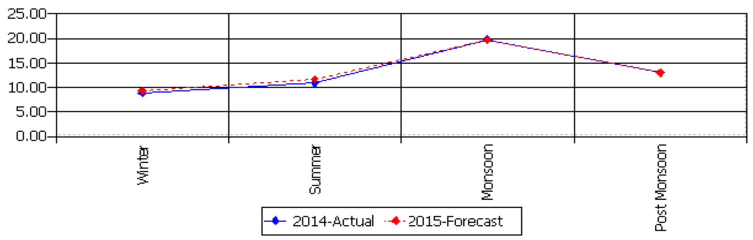
Season	2014-Actual	2015-Forecast	Accurate %
Winter	11.84	11.50	97.17
Summer	17.42	15.39	88.37
Monsoon	22.91	22.61	98.71
Post Monsoon	17.32	16.67	96.23

The Northern Plains Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



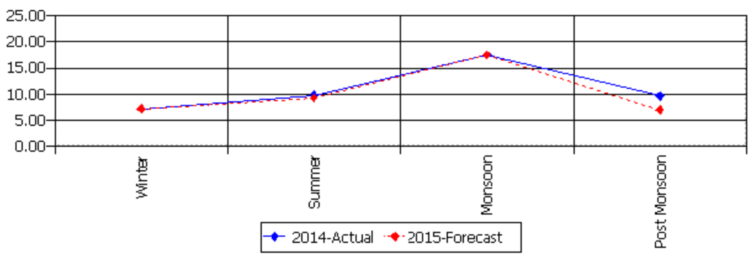
Season	2014-Actual	2015-Forecast	Accurate %
Winter	10.35	9.98	96.44
Summer	15.44	15.47	100.20
Monsoon	22.40	22.47	100.31
Post Monsoon	15.11	14.73	97.47

The Central Highlands Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



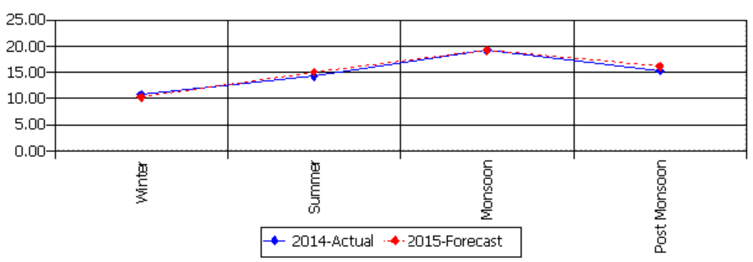
Season	2014-Actual	2015-Forecast	Accurate %
Winter	8.89	9.33	104.93
Summer	10.91	11.68	107.08
Monsoon	19.73	19.71	99.91
Post Monsoon	13.04	13.06	100.13

The Desert Region Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



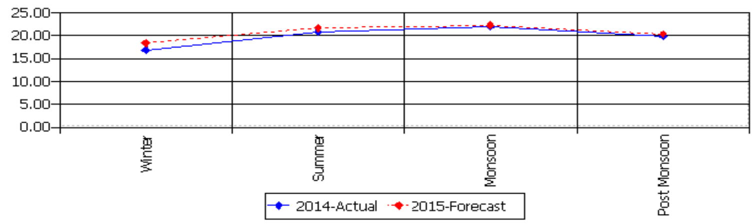
Season	2014-Actual	2015-Forecast	Accurate %
Winter	7.22	7.20	99.79
Summer	9.76	9.23	94.52
Monsoon	17.49	17.45	99.80
Post Monsoon	9.62	7.00	72.72

The Peninsular Plateau Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



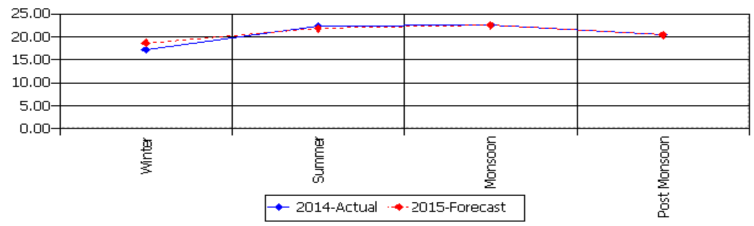
Season	2014-Actual	2015-Forecast	Accurate %
Winter	10.83	10.26	94.77
Summer	14.27	15.01	105.22
Monsoon	19.21	19.18	99.82
Post Monsoon	15.32	16.13	105.29

The Western Coastal Plains Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



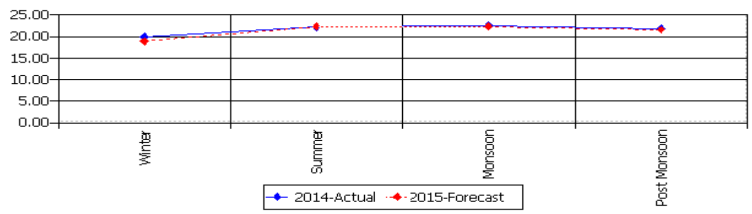
Season	2014-Actual	2015-Forecast	Accurate %
Winter	16.92	18.35	108.48
Summer	20.86	21.68	103.93
Monsoon	22.02	22.31	101.34
Post Monsoon	19.90	20.30	102.03

The Eastern Coastal Plains Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



Season	2014-Actual	2015-Forecast	Accurate %
Winter	17.23	18.69	108.45
Summer	22.38	21.77	97.29
Monsoon	22.62	22.49	99.43
Post Monsoon	20.40	20.38	99.92

The Island Seasonal Water Vapour (g/m3) Distribution Trends for 2014- Actual Vs. 2015-Forecasting Over India



Season	2014-Actual	2015-Forecast	Accurate %
Winter	20.02	18.93	94.54
Summer	22.20	22.36	100.71
Monsoon	22.55	22.38	99.24
Post Monsoon	21.86	21.53	98.50

Figure 13: The Physiographic Divisions Water Vapour (g/m^3) distribution forecasting trend for different Seasons over India

V. Conclusion and Future work

The water vapor is the most significant greenhouse gas and plays an important role in global climate change studies and weather forecasting. Hence in this paper we discussed the effectiveness of using neural networks to forecast Atmospheric water vapour distribution parameters from existing research and with an experimental dataset. The learning of our neural network research discussed that four different seasonal neural network architectures are applied to experimental datasets in order to forecast the water vapour distribution, These experiments are conducted by gradient descent with momentum and adaptive learning rate back propagation algorithm, the trained neural networks are predicted water vapour distribution values very effectively for winter, summer, monsoon and post monsoon seasons of the various physiographic divisions over India

The simulation results are also showing the training algorithm performing the best accurate % of the training error Performance and the Network Training Regression, along with the learning rate and the momentum with a satisfactory approximation while forecasting.

However on the other hand there are still many hard hitting problems are facing in ANN studies like

1. Back Propagation algorithm working in accordance with the direction of the mean square error gradient descent, hence there are many local and global minimum mean square error gradients, so which makes neural network is easy to fall into local minimum as shown in the figure 3

2. Back Propagation algorithm network training rate is very slow and the iterative, hence it could waste a lot of time for fitting the best network

3. While training ANN the input data and the target data considered easily, however the selection of the number of hidden layer neurons of the network is still not able to manage, and also the generalization ability of the learning network is poor.

Hence as part of our continues learning and further development in future work we would like to explore various ANN method for different inter disciplinary research areas for better decision supporting systems

Acknowledgment

We wish to thank the British Atmospheric Data Centre (BADC), which is part of the NERC National Centre for Atmospheric Science (NCAS), for providing the radiosonde data and the supporting documentation to understand the data and usage for our water vapour distribution studies over India.

References

[1] IM Held and BJ Soden, "Water vapor feedback and global warming 1", Annual Review of Energy and the Environment 25 (1), 441-475

[2] S.K. Sarkar, H.N. Dutta, M.V.S.N. Prasad and B.M. Reddy," TROPOSPHERIC WATER VAPOUR OVER INDIA", Antennas and Propagation, 1989. ICAP 89.,

Sixth International Conference on (Conf. Publ. No.301) 302 - 306 vol.2

[3] R.M. Gairola, and V.K. Agarwal, "Measurements of Water Vapour and Rainfall Over Indian Ocean and Sub-Continent with a TERRA-MODIS and TRMM"www.ursi.org/Proceedings/ProcGA05/pdf/F10P.6(01549).pdf

[4] P. Kishore, M.VenkatRatnam , S.P.Namboothiri , IsabellaVelicogna , GhouseBasha , J.H.Jiang , K. Igarashi, S.V.B.Rao and V.Sivakumar, "Global (501S–501N) distribution of water vapor observed by COSMIC GPS RO Comparison with GPS radiosonde, NCEP, ERA-Interim, and JRA-25 reanalysis datasets", Journal of Atmospheric and Solar-Terrestrial Physics (2011) 1849–1860

[5] Sanjay Kumar, A.K. Singh, Anup K. Prasad and R.P. Singh, "Variability of GPS derived water vapor and comparison with MODIS data over the Indo-Gangetic plains", Physics & Chemistry of the Earth, doi:10.1016/j.pce.2010.03.040, 2010

[6] Swastika Chakraborty and Animesh Maitra, "A Comparative Study of Cloud Liquid Water Content from Radiosonde Data at a Tropical Location", International Journal of Geosciences, 2012, 3, 44-49

[7] Swastika Chakraborty and Animesh Maitra, "A Comparative Study of Cloud Liquid Water Content from Radiosonde Data at a Tropical Location", International Journal of Geosciences, 2012, 3, 44-49

[8] Meghali A. Kalyankar,Prof. S. J. Alaspurkar "Data Mining Technique to Analyse the Metrological Data", (IJARCSSE), 2013, Volume :3, Page(s) :114-118

[9] Gaurav J. Sawale, Dr. Sunil R. Gupta, "Use of Artificial Neural Network in Data Mining For Weather Forecasting", (IJCSA), 2013, Volume:6, Page(s) 383-387

[10] http://www.yourarticlelibrary.com/geography/the-main-physiographic-divisions-of-india-geography/5461/

[11] http://www.mapsofindia.com/maps/india/physiographic.htm

[12] [Ms.Smita.Nirkhi, "Potential use of Artificial Neural Network in Data Mining",2010, IEEE, Volumn 2, PP 339-343

[13] WANG Yong and XU Hong, "The Study of Rainfall Forecast Based on Neural Network and GPS Precipitable Water Vapor",2010, IEEE, ESIAT, pp 17-20

[14] Stefania Bonafoni,Vinia Mattioli, Patrizia Basili,Piero Ciotti and Nazzareno Pierdicca,"Satellite-Based Retrieval of Precipitable Water Vapor Over Land by Using a Neural Network Approach",2011,IEEE,TGARS, Vol 49, No 9, pp 3236-3248

[15] Bimal Dutta and Susanta Mitra, "Better Prediction of Humidity using Artificial Neural Network", 2011, IEEE, pp 59-64

[16] Sridevi Jade and M. S. M. Vijayan, "GPS-based atmospheric precipitable water vapor estimation using meteorological parameters interpolated from NCEP global reanalysis data",JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 113, D03106, doi:10.1029/2007JD008758, 2008, pp 1-12

[17] Wayan Suparta and Kemal Maulana Alhasa, "A Comparison of ANFIS and MLP Models for the Prediction of Precipitable Water Vapor",Proceeding of the 2013 IEEE International Conference on Space Science and Communication (IconSpace), 1-3 July 2013, Melaka, Malaysia, pp 243-248

[18] A. J. Litta,Sumam Mary Idicula,and U. C.Mohanty, "Artificial Neural Network Model in Prediction of Meteorological Parameters during Premonsoon Thunderstorms", International Journal of Atmospheric Sciences,Hindawi,http://dx.doi.org/10.1155/2013/525383 pp 1-14

[19] Dr. S. Santhosh Baboo and I.Kadar Shereef, "An Efficient Weather Forecasting System using Artificial Neural Network",International Journal of Environmental Science and Development, Vol. 1, No. 4, October 2010, pp 322-326

[20] Met Office (2006): Met Office Global Radiosonde Data. NCAS British Atmospheric Data Centre (BDAC), date of citation.http://catalogue.ceda.ac.uk/uuid/f2afaf808b61394b78bd342ff068c8cd

[21] H. Sarkar, S.K. Midya and S. Goswami," A Comparative Study of Integrated Water Vapor (IWV) and of Attenuation of 94 GHz Signal from Radiometer and Radiosonde Observations during Monsoon Period over Kolkata, India."The Pacific Journal of Science and Technology, Volume 13. Number 1. May 2012 (Spring)

[22] Khamphoui Southisombath, Toshio Wakabayashi and Yoshiaki Moriya, "The impact of water vapor in atmosphere on Ku band satellite DTV reception in a sub-tropical region", Proceedings of ISAP'04, Sendai, JAPAN, IEICE , 1381-1384

[23] Ding Youqing, Yumei Fu, Zhu Fang, and Zan Xinwu," Comparison of Missing Data Filling Methods in Bridge Health Monitoring System", Proceedings. 12th IEEE International Conference on Cognitive Informatics & Cognitive Computing 978-1-4799-0783-0/13/2013 IEEE

[24] B. Mehala, P. Ranjit Jeba Thangaiah, and K. Vivekanandan, "Selecting Scalable Algorithms to Deal With Missing Values" International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009, 80-83

[25] Valmik B Nikam and B.B.Meshram, "Modeling Rainfall Prediction Using Data Mining Method-A Bayesian Approach", 2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation,DOI 10.1109/CIMSim.2013.29, pp 132-136

[26] Deepak Ranjan Nayak,Amitav Mahapatra and Pranati Mishra, "A Survey on Rainfall Prediction using Artificial Neural Network",International Journal of Computer Applications (0975 – 8887),Volume 72– No.16, June 2013,pp 32-40

[27] Smita Kulkarni and Dr.Milind Mushrif, "Comparative study among different neural net learning algorithms applied to rainfall predication",2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies,2014,IEEE, DOI 10.1109/ICESC.2014.104, pp 209-216