MODELING OF AIR TEMPERATURE USING HYBRID WAVELET TRANSFORM - ANFIS - SUPPORT VECTOR MACHINE COMPUTING TECHNIQUES

Thesis

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

KARTHIKA B. S.

DEPARTMENT OF APPLIED MECHANICS AND HYDRAULICS

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA,

SURATHKAL, MANGALORE – 575 025

JUNE 2016

MODELING OF AIR TEMPERATURE USING HYBRID WAVELET TRANSFORM - ANFIS - SUPPORT VECTOR MACHINE COMPUTING TECHNIQUES

Thesis

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

KARTHIKA B. S.

(Register Number: AM12P03)

DEPARTMENT OF APPLIED MECHANICS AND HYDRAULICS

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA,

SURATHKAL, MANGALORE – 575 025

JUNE 2016

D E C L A R A T I O N

By the Ph.D. Research Scholar

I hereby *declare* that the Research Thesis entitled **Modeling Of Air Temperature Using Hybrid Wavelet Transform - ANFIS - Support Vector Machine Computing Techniques,** Which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfilment of the requirements for the award of the Degree of **Doctor of Philosophy** in **Applied Mechanics and Hydraulics Department,** is a *bonafide report of the research work* carried out by me. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

AM12P03, KARTHIKA B. S.

(Register Number, Name & Signature of the Research Scholar)

Department of Applied Mechanics and Hydraulics

Place: NITK-Surathkal

Date:

C E R T I F I C A T E

This is to *certify* that the Research Thesis entitled **Modeling Of Air Temperature Using Hybrid Wavelet Transform - ANFIS - Support Vector Machine Computing Techniques,** submitted by KARTHIKA B. S. (Register Number: AM12P03) as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfilment of the requirements for the award of degree of **Doctor of Philosophy**.

Dr. Paresh Chandra Deka
Associate Professor
Research Guide
(Name and Signature with Date and Seal)

Chairman - DRPC
(Signature with Date and Seal)

ACKNOWLEDGEMENTS

A successful project is a dream comes true for any student and to fulfill this dream, efficient teamwork is required. I am thankful to a lot of people who are directly and indirectly involved in this project.

The satisfaction that accompanies the successful completion of the project work would be incomplete without the mention of those people who made it possible through their guidance, support and encouragement.

It gives me a great pleasure to acknowledge and sincerely express my deep sense of gratitude to my research supervisor **Dr. Paresh Chandra Deka**, Associate Professor Department Applied Mechanics and Hydraulics, NITK for his encouraging, logical and constructive suggestions in the execution of this project work. The interaction and the time spent in discussions are imprinted in my memory permanently. Only with his moral support and stimulating guidance, this research work could be completed and I could publish my work in many international journals and conferences.

I express my deep sense of gratitude to the Committee members of N.E.S.Shimoga, Principal of J.N.C.C.E and Dr. A. V. Pradeep Kumar, H.O.D, Dept. of Civil Engg. J.N.N.C.E. Shimoga for permitting me to carry out my research work.

I am greatly indebted to Research Progress Appraisal Committee members, Prof. A Mahesha and Prof. G. Ram Mohana Reddy, for their useful suggestions during the progress of the work.

I sincerely acknowledge the help and support of Prof. Dr. Lakshman Nandagiri, Prof. M. K. Nagaraj and Prof. Subba Rao, former Head of the Department and Prof. Dwarakish G.S, present Head of the Department of Applied Mechanics and Hydraulics, NITK, Surathkal, for permitting me to use the departmental computing and laboratory facilities and his continuous support in completing the work.

I express my deep sense of gratitude to the friends Dr. Prasanth J. Dr. Kumar raju. B C, Dr. Harish, Vikas A.C, Akshy B.J, Sangeev Gadad, Sujay and Amith Patil who helped me in every mode of action to complete my research work successfully. Supriya B.J, Sandesh B.J, Dr. Kiran M, Amar raju and Vijay Bhaskar Koudaki also inspired me to carry out my research work.

I sincerely acknowledge the help and support of all the Professors, Associate Professors, Assistant Professors and Teaching and non teaching staff of Department of Applied Mechanics and Hydraulics, NITK, Surathkal, in completing the work.

I sincerely acknowledge the invaluable help rendered by college colleague A.N Srinivas Pandith, S.M.Hiremat, Rabinandan .J, Akshatha.M, students Akruth Gowda, Rahul Biligi, Shankar ,Manu and Nagaraj, Manager, MPM, Bhadravathi .

I sincerely acknowledge Mr. B. Jagadish, Foreman, Department of Applied Mechanics and Hydraulics, and Mr. Balakrishna, Literary Assistant at NITK, Surathkal, for their support and help during the research work.

I express my heartfelt gratitude to the authors of all those research publications which had been referred to in preparing this thesis.

Finally, I wish to express love and affection to my beloved family members. I would like to put my regards and thanks to my loving parents, brother, wife ADVI.B.V and daughter KASHVI, who were a constant source of moral support, love and encouragement throughout my studies.

Karthika.B.S.

ABSTRACT

The accurate modeling of average air temperature is a significant and much essential parameter in frame of reference for decision-making. Therefore, the characterization of such parameter is an important task. The information about the air temperature also helps in planning and management of water resources, irrigation, drought detection, tourism, health and issues of day to day life.

In this study, a hybrid model consists of Wavelet - ANFIS has been developed for air temperature modeling. The results are compared with Wavelet - SVM, single ANFIS, and single SVM to confirm the superiority of the proposed model.

To model average air temperature, ANFIS models were developed with different membership, namely generalized bell-shaped built-in membership function (GBELLMF), and Gaussian curve built-in membership function (GAUSSMF). Additionally, to check the result of modeling of average air temperature, SVM model was developed. To enhance the accuracy of modeling performance, single ANFIS and single SVM is integrated along with wavelet transformations were tested. Here wavelet transformation was used as pre-processing the data by capturing valuable information on various resolution levels.

This study extends for seven stations in Karnataka state of India (Shimoga station, Raypura station, Linganmakki station, Honnali station, Hiriyur station, Bhadra station (B. R. Project) and Davanagere station) observed data of meteorological data like rainfall, wind speed, humidity and sunshine hour as input and as target average air temperatures are used for all the models. In the next phase, the influence of air pollutants along with the meteorological parameters has been investigated for average air temperature modeling for a specific Bhadra station in Karnataka state, India, which is near to industrial city. The obtained results were evaluated using Correlation Coefficient, Root Mean Square Error and Scatter Index.

The performance of ANFIS, SVM, hybrid Wavelet - ANFIS and hybrid Wavelet - SVM is analyzed for modeling of average air temperature. Out of seven stations, station Linganamakki showed better performance with CC of 0.954, RMSE is 0.71and SI is 0.027 with hybrid Wavelet- ANFIS model (Gbell membership). Also for single Bhadra station, Hybrid Wavelet - ANFIS model with the parameter combination (rainfall, wind speed, humidity, sunshine hour) for Db5 with level4 (2MF) and Gauss membership function is having the results of CC is 0.98, which is best in case of accuracy. The study reveals the higher accuracy of hybrid Wavelet - ANFIS in modeling air temperature for various meteorological and air pollutants input scenarios.

Keywords: Average Air Temperature, Air pollutant, Modeling, ANFIS, SVM, Wavelet - ANFIS and Wavelet - SVM.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF NOTATIONS

Symbol	Description
nsv	Number of Support Vector
Db	Daubechies Mother Wavelet
Gbell	GBELL membership function
Gauss	GAUSS membership function
Greek	
ε	Error in error tube
σ	Standard deviation
Abbreviations	
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
CC	Correlation Coefficient
erbf	Exponential radial basis function
FIS	Fuzzy Inference System
GP	Genetic programming
GA	Genetic algorithm
GA-SVMR	Genetic algorithm based support vector machine regression
NITK	National Institute of Technology Karnataka
RMSE	Root mean square error
rbf	Radial basis function
SI	Scatter index
SVM	Support Vector Machines
SA	Sensitivity analysis

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The Sun is the primary source of heat energy supplied to the earth's atmospheric surface. Even, small change in the solar radiations creates major changes in the weather on the earth, which may lead to variation in the air temperature. These fluctuations in the air temperature lead to variations in water evaporation, air saturation and so on.

Air temperature being one of the most important meteorological parameters, it is recorded worldwide on a regular basis at the meteorological stations. The quality of life in the semi urban areas is profusely instigated by the impact of average air temperature. Valuable information about air temperature is used by many applications like weather forecasting, climate changes and other environmental issues. Therefore, the characterization of such parameter needs more in depth study specially in urban way of life.

Since, average air temperature is one of the importantparameters which directly influence climatic variation; an intense study is earmarked in the field of average air temperature on spatial and temporal scale. Subsequent information about the air temperature also helps in planning and management of water resources, irrigation, tourism and issues of day to day life.

Due to the dynamic nature of the atmosphere, it is difficult to model ambient air temperature data accurately (Pal et al., 2003). Air temperature mainly depends on many meteorological and air pollution parameters. Air pollutants can be either natural or may be the result of various anthropogenic activities. The industrial contaminates are normally products of external combustion like smoke, dust, SO_2, NO_2 and others. Also increasing number of automobiles on the road, aircraft on atmosphere generates huge air pollution. Various techniques like linear regression, auto regression, Multi Layer Perceptron, Radial Basis Function networks are applied to predict atmospheric parameters like temperature, wind speed, rainfall, air pollution etc. (Mohandes et al., 2004; Min and Lee, 2005; Osowski and Garanty. 2007; Radhika and Shashi, 2009; and Pires et al., 2011).

1.2 AIR POLLUTANTS

Air pollutants can be either natural or may be the result of various activities of man like industrial operations. The industrial contaminates can be either by products of external combustion like smoke, dust, and sulphur oxides or by products of internal combustions like the reactions in petrol and diesel engines.Further, the emission can be either primary pollutants or secondary pollutants.

1.2.1 Suspended Particulate Matter

The aerodynamic diameter <10 μm Suspended particles in the outdoor air are considered as PM_{10}.They are in rigid and liquid molecular forms (Brunelli et al., 2007), concentration of particles may be dust of deserts, burning of fossil fuels and chemical pollutants reactions. The solid particles are irregular in shape and liquid particles are in spherical in shape. Particles size larger than 100 μm tends to settle out of the air by gravity.

1.2.2 Sulphur dioxide in ambient air

The keyconstituent of air pollutant is Sulphur dioxide. The sources of sulphur dioxide are combustion of fuel and coal. The main reason for concentration of sulphur dioxide in the earth atmosphere is due to fuel used for heating and power generation. Depending upon the fuel contents of sulphur varies from 1% for good quality anthracite to over 4% for bituminous coal. The crude petroleum products contain less than 1% sulphur. Refining process tends to concentrate sulphur compounds in the heavier fractions. Fuel gases also contain sulphur in small quantity.Generally the concentration of sulphur dioxide in the flue gases ranges between from 0.05-0.4.Another source of sulphur dioxide is metallurgical operations,when the process of smelting is done toores; sulphur dioxide is evolved in stack concentrations of 5-10 %(SO_2).

1.2.3 Nitrogen dioxide in ambient air

Next to sulphur dioxide, the second most abundant atmospheric pollutant is oxides of nitrogen. Chemical reactionsfor the production of nitric acid are the prime donors of the nitrogen oxides to environment. After that,emissions from the automobiles are the next producers of nitrogen oxides. Out of seven oxides of nitrogen (N_2O, NO, NO_2, NO_3, N_2O_3, N_2O_4, N_2O_5), only nitric oxide and nitrogen dioxide arise from many human activities and are classified as pollutants. They are usually reported as “total oxides and nitrogen” or NOx in atmospheric analyses. It is a standard practice in the chemical industry to absorb and recover significant quantities of oxides of nitrogen.

1.2.4 Sources of outdoor Air pollution

Sources of ambient air pollution are both natural and man-made. Oxides of sulphur and nitrogen are from volcanoes, oceans, biological decay, lightning strikes and forest fires.All the time natural pollutions are normal in condition. However, due to dynamic nature, because of volcanic eruption or forest fire, concentration can increase dramatically in nature.

The burning of fossil fuels, such as oil, and coal, in industries and vehiclesare the mostcommon source of man-made air pollution for nature. Depending on the different nature of the fuel and the different type of combustion process, pollutants like nitrogen oxides, sulphur dioxide, carbon monoxide, particulate matter, lead and volatile organic compounds are released into the atmosphere. Forest burning, chemical, fertilizer and paper manufacture are the other source of pollutants. These pollutants are known as primary pollutants because they have direct sources. Amount of pollution from different sources are represented in the Fig 1.1.

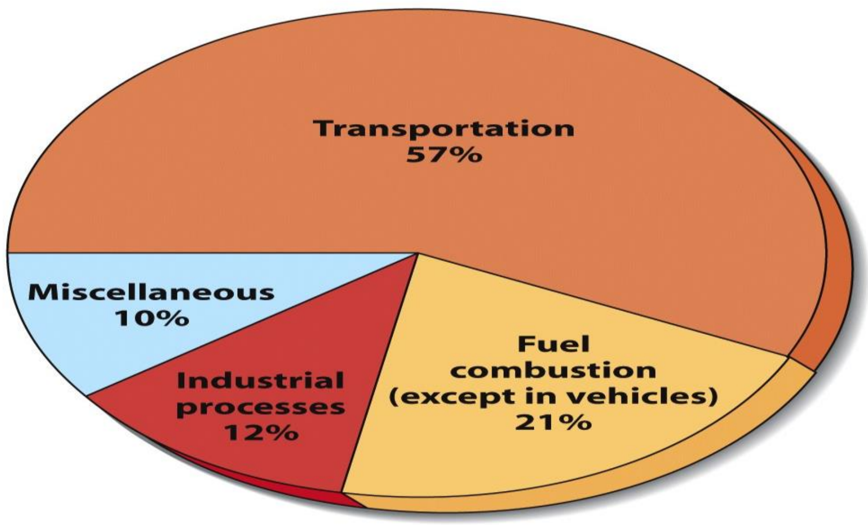


Fig.1.1Contribution of pollution from different source to atmosphere

1.3 SCOPE OF THE PRESENT INVESTIGATIONS

Air temperature strongly influences agriculture process like many metabolic processes in plants (respiration, photosynthesis, development, etc.) and it is one of the principal environmental factors affecting their growth and yield. In the field of tourism, planning is done on the basis of climate especially temperature. For the clothing industries temperature play an important role in designing material and pattern. In agriculture equipment related industries demand of materials (pipe, water pumping motors) are majorly depend on temperature and also epidemic diseases also depending fluctuation of temperature. So modeling of air temperature helps in decision making in advance. So modeling of air temperature helps in decision making in advance. The output of developed model can be used for future planning of cropping activities, guidelines for prevention of temperature related diseases in human as well as in animal, management of tourism activities. Based on relationships between temperature and meteorological parameters observed at the evening time or other hours of the day, some of the models work on statistical analysis. These models work for particular site specific condition. Based on past works, researchers have developed semi empirical and theoretical models. Basically theoretical models require parameters that depend on both soil conditions (measure is a complex work), and site conditions (which are estimated by statistical analysis of experimental data). Therefore, although models are based on an analytical approach, but practically these models are semi- empirical. Some analytical models, for the prediction of minimum temperature on an hourly basis are difficult to use because of lack of data concerning the parameters employed.

In recent years, many methods have been applied to time series analysis. One among them is soft computing, a branch of computer science, which tries to build an intelligent and sophisticated machine. Intelligence means offer power to derive the answer and not simply arrive to the answer. Purity of thinking, machine intelligence, freedom to work, dimensions, complexity and fuzziness handling increase, as we go higher and higher in the hierarchy. The main aim is to develop a computer or a machine which will work in a similar way as human beings can do (Chaturvedi, 2008).

Soft computing is the fusion of methodologies designed to model and enables the solutions to real world problems, which are not modeled or too difficult to model mathematically. The aim of soft computing is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve close resemblance with human like decision making and overall low cost. Soft computing techniques have been recognized as attractive alternatives to the standard, well-established hard computing paradigms. Traditional hard computing methods are often too cumbersome for today's problems. They always require a precisely stated analytical model and often a lot of computational time.One of the important features of soft computing is the acquisition of knowledge/information from inaccurate and uncertain data. Soft computing is often robust under noisy input environments and has high tolerance for imprecision in the data on which it operates.

Soft computing methods like Adaptive Neuro-Fuzzy Inference System (ANFIS) can serve a specific approach of modeling nonlinear functions. In ANFIS, the membership function parameters are extracted from a dataset that describes the system behaviour (Guler et al. 2005). The ANFIS learns the features by example dataset and adjusts the system parameters according to a given error criterion (Jang, 1992; Jang, 1993). Another method is Support Vector Machine (SVM) which works on Structural Risk Minimization

(SRM) principle (Gunn, 1998). On the expected risk, SRM minimizes an upper bound, as contrasting to ERM that minimizes the error on the training data. This nature equips SVM with an improved capability to simplify, which is the objective in statistical learning.

Wavelet transform, better version of Fourier transform, gives knowledge about the physical form of the data. It supplies a time frequency representation of a signal at many different periods in the time domain (Daubechies, 1990). Wavelet transformed data of the original time series improve the ability of a predicting model by capturing useful information on various resolution levels (Kim, et al. 2003; Rajaee, et al. 2010). Wavelet decomposition is the one which decomposes time series data into a different time and scale of wavelet transformation, and thus one can get the property of time series in different frequency bands as time goes by (Strang et al. 1996). This method has been widely used in multi-scale analysis of time-series (Shao et al. 2006; Friedman et al. 2001). Regularities of short-term (high frequency) and long-term (low frequency) data reflect in different frequency bands after wavelet decomposition of time-series which includes many processes under various time scales (Liu et al. 2012). If wavelet coefficients in different frequency bands are applied to ANFIS, the result of the hybridization has a possibility of improved time-series of different time scales, which can reflect its natural information in a better way.

Even though many numerical models are available for modeling of average air temperature, the limited input for the models and the complexity of the equations available have made it difficult to use for the modeling.

1.4 PROBLEM FORMULATION

The air temperature process is highly nonlinear and exhibits seasonal variation. Even though researchers have carried out number of studies for predicting air temperature by considering meteorological data and air pollution data using some popular soft computing techniques like ANN, ANFIS and SVM, it is noticed that these models showed very poor agreement with in-situ data. The wavelet transformation is getting attention of researchers due to the analysing capability of non-linearity and non-stationary time series. Till today, as per the author's knowledge, the potential of hybridizing the soft computing techniques like Wavelet with ANFIS or Wavelet with SVM are not explored fully and also the performance evaluation among various developed models are scanty. Further, for modeling of air temperature, using meteorological along with air pollution parameters SPM, NO_2 and SO_2 are not incorporated for input-output mapping to assess the influence of these parameters.

In view of the above aspects, it is decided to take up a study on utilizing various artificial intelligence techniques along with wavelet transform in modeling of air temperature leading to new approach to the enhancement of model accuracy with various input scenarios. Here, it is proposed to develop highly efficient hybrid models such as Wavelet with ANFIS and Wavelet with Support Vector Machine (SVM) to model the air temperature. Also, finally uncertainty analysis has been carried out using Bootstap method for best model.

1.5 ORGANIZATION OF THE THESIS

The thesis is presented in five chapters:

Chapter 1 – Introduction: Introduction to Air temperature, as well as air pollution data with different types of air pollution data and scope of the present investigations has been discussed.

Chapter 2 – Literature review: The literature review applications of soft computing techniques in modelling of averageair temperature with meteorological and Air pollution data, problem formulation and objectives of the present work have been discussed.

Chapter 3 – Study area and Methodology: Briefly explained study area and data used for developing soft computing models. Also, theoretical background of research methods used to developed soft computing models to Model average air temperature, such as, ANFIS, SVM, hybrid Wavelet-ANFIS and hybrid Wavelet- SVM has been discussed.

Chapter 4 – Results and Discussion: The results obtained from the soft computing models, such as, ANFIS, SVM, hybrid Wavelet-ANFIS and hybrid Wavelet- SVM in modeling of average air temperature are analysed, interpreted and discussed. Also, the performance of these models is compared with each other.

Chapter 5 – Summary and Conclusions: Brief summary of the research work and the conclusions drawn based on the results of soft computing models and suggestions for future work have been presented.

The Appendix I is followed by references, list of publications based on the present work, and a brief resume of the researcher.

CHAPTER 2

LITERATURE REVIEW

2.1GENERAL

Since two decades, many works were carried out for air temperature modeling using various techniques in different climatic regions with various meteorological and other parameters. In the following section, a detailed review of relevant literature has been included.

2.2. REVIEW OF STUDIES ON AIR TEMPERATURE

Ninyerola et al.(2000) used empirical methodology for modelling and mapping the air temperature (mean maximum, mean and mean minimum) and total precipitation and used multiple regression analysis using the backward stepwise method for choosing the independent variables(geographical variables (altitude (ALT), latitude (LAT), continentality (CON), solar radiation (RAD) and a cloudiness factor (CLO)) included in the model .Independent variables were elaborated from a 180 m resolution digital elevation model (DEM). The results concluded that, this model is acceptable for modelling.

Jarvis and Stuart (2001) try to explore the derivation and selection of a comprehensive set of continuous topographic and land cover–related variables to guide the interpolation of daily maximum and minimum temperatures over England and Wales, for an entire annual cycle to a resolution of 1 km. They used of digital elevation data and land cover data, using the modeling capabilities of geographical information systems. They concluded the work by saying incorporation of coastal shape and situation, land cover, and soils data might further improve the modeling of local-scale influences on maximum and minimum temperature.

Bilbao et al.(2002) tried a comparative assessment of air temperature models, using hourly and daily air temperature measurements in 34 different stations in the north Mediterranean belt. For their work Double cosine model, Knight's model, Erbs's model and the ''standard'' models were compared. They concluded that Erbs's model works better for estimating hourly air temperature from monthly mean values and the ''standard'' model gave the best performance for estimating daily mean air temperature from daily minimum and maximum air temperature values.

Jusuf and Hien (2009) developed an empirical model for air temperature prediction to evaluate the impact of estate development by means of Geographical Information System (GIS).They used daily minimum (Ref Tmin), average (Ref Tavg) and maximum (Ref Tmax) temperature at reference point, average of daily solar radiation (SOLAR), percentage of pavement area over R 50m surface area (PAVE), average height to building area ratio (HBDG), total wall surface area (WALL), Green Plot Ratio (GnPR), sky view factor (SVF) and average surface albedo (ALB) as a input for the web based Screening Tool for Estate Environment Evaluation (STEVE) model.They carried out Sensitivity analyses to observe the dependence of the air temperature due to the variations of each variable. They concluded that calculated air temperature have shown a good fit and their differences are with the acceptable range.

Alvares et al.(2012) attempted to model air temperature which is one of the main weather variables influencing agriculture around the world. For their research they tried with multiple regression and geographic information system techniques. They used the independent variables latitude, longitude, altitude, and their combinations for modelling. They are recommended to above mention models for predict air temperature in all Brazilian territories.

Rashid (2014) has tried address the common problem in construction industry production and timing in Baghdad region. He developed an empirical formula by using average daily total solar radiation, building area percentage over radius 50m, 100m and 150m surface from the building center, average building height to area ratio, total wall area to green area ratio, sky view factor, and albedo. The results show that model working better for prediction.

2.3 ARTIFICIAL NEURAL NETWORKS MODELS

Several researchers have adopted soft computing techniques to solve complex associated with estimation/prediction of meterlogical parameters like air temperature are discussed below:

Tasadduq et al. (2002) used artificial neural networks for the prediction of ambient temperature on hourly basis in advance of 24 h. To train the model full year hourly values of ambient temperature was used. Back propagation and a batch learning scheme are used to train the network. For hourly temperature prediction in particular, neural network was testified as valuable tool.

Smith et al. (2009) for accurate air temperature prediction developed and implemented respective tools based on Artificial Neural Network (ANN) model using the data collected through 2005. Model evaluation over various instants managed to perform better in parameter selection by presenting more accurate comparisons of distinct models than those afforded by single-network evaluation.

De andDebnath (2009) tried to investigate the strength of ANN to forecast the Maximum and Minimum Temperature for Monsoon month. Using temperature data of January to May maximum and minimum temperature for the months of June, July and August was predicted. The developed model performed better for August month and worst for month June.

Baboo and Shereef (2010) proposed a neural network model to forecast temperature with quantitative data about the current state of the atmosphere. The main parameters used are Temperature (°C), Dew Point (°C), Humidity (%), Sea Level Pressure (hPa), Visibility (km), Wind (km/h), Gust Speed (km/h) and Precipitation (cm). The comparison of obtained values and measured values confirm that model have the potential for temperature forecasting.

Bilgili and Sahin (2009) predicted long term monthly air temperature and rainfall based on ANN using Geographical variables (latitude, longitude, and altitude) and time, which were used as an input data for this approach. By such analysis it proves that ANN based model works better for prediction.

Singh et al. (2011) used genetic algorithm and back propagation technique to predict time series based temperature. Back propagation integrated with genetic algorithm is the most important algorithm used to train neural networks. In the proposed technique, the effect of under training and over training the system is also shown. The proposed technique can learn efficiently by combining the strengths of GA with BP. It is good at time series data, global search and it works with a population of points instead of a single point.

Kadu et al. (2012) attempted to forecast temperature using ANN. For hourly temperature forecasting back propagation neural network were developed. The experimental results show that the model works effectively without excessively compromising the performance.

2.4 HYBRID MODELS

Application of hybrid model got research attention in recent years in different fields like estimation of potential evaporation, prediction of rainfall, hydrological modeling, mean sunshine hour prediction and modeling of SO_2 (Kumar et al. (2012); Deka et al. (2005); Mellit et al. (2007); Aldrian et al. (2008);Savic et al. (2013)).

Daneshmand et al. (2015) carried on a research to predict monthly minimum temperature data using ANFIS model with 42 years of data. They implemented the same approach by formulating the same model along with spectral analysis, correlation coefficient, and the knowledge of experts were used to select needed input parameters. The results concluded that, this model is acceptable for prediction.

Cobaner et al. (2014) tried to estimate maximum, minimum and average temperature using artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), and multiple linear regression (MLR) models with latitude, longitude, and altitude of the location, and the month number as the input variables for 20 years data set of 275 stations. They compared different model performances and concluded that ANFIS performed better than other models.

Kumar (2012) formulated weekly temperature forecasting using ANFIS (Adaptive Neuro Fuzzy Inference system). This model was used for forecasting the results at least a minimum of one week ahead of weekly temperature with the help of current week's maximum mean weekly temperature as an input for prediction. This formulation also included gbell type of membership function considering ten years data. By such analysis, it is proved that this model could be effectively used to obtain one-week ahead weekly temperature prediction.

2.5 SUPPORT VECTOR MACHINE (SVM) MODELS

Recently use of SVM got wide attention in modeling of air temperature due to Structural risk minimization principal (Gunn, 1998) which works better compared to traditional Empirical Risk Minimization (ERM) principle as by conventional neural networks. Applications of SVM are in fields like prediction of maximum temperature and forecasting of weather etc.

Moser and Serpico (2009) tried to estimate land and sea surface temperatures. Using SVM, they developed a model to estimate the surface temperature from satellite radiometry. To incorporate temperature estimates into meteorological data or hydrological -assimilation schemes, additional pixel wise error statistics were added as input. They used maximum-likelihood or confidence-interval supervised estimators. The error contribution modeling due to intrinsic random variability in the data was achieved in the both cases.

Radhikaand Shashi(2009) developedSVM model to predict maximum temperature of the next day. For the study, they used location based daily maximum temperatures for a span of previous 'n' days as input. The SVM results were compared with Multi LayerPerceptron (MLP) trained with back-propagation algorithm and the results of SVM are found to be better compared to Multi Layer Perceptron.

Prediction of peak energy consumption gives an idea of usage of heating or air conditioning system. To have this record, daily maximum temperature is required. Tineo et al.(2011) took the help of meteorological data of different station of Europe, which was applied on support vector machine. With this, they included two more variables like synoptic situation of the day and monthly cycle. They compared the outcome with the Multi Layer perceptron and Extreme learning machines. A comparison with alternative neural methods based on statistical test, have shown that the SVM performed better than a multi-layer perceptron and an extreme learning machine in this prediction problem. The SVMr approach is able to obtain accurate prediction for the one-day ahead maximum temperature.

Rao et al. (2012) attempted to forecast weather using SVM. They compared SVM results with MLP. They used data of 5 years maximum temperature of previous ‘n’ days for forecasting maximum temperature in a day. From results, they concluded that proper selection of the different parameters; Support Vector Machines replace neural network based models for forecasting weather application.

Bertiniet et al. (2010) applied hybrid approach to model monthly and daily ambient temperature. In order to train the Artificial Neural Networks (ANN) model effectively, they tried the combination ofBack Propagation (BP) algorithm and the simple Genetic Algorithm (GA).The model output showed remarkable improvement as compared with traditional methods.

2.6 USING WAVELET TRANSFORM WITH OTHER MODELS

However, to deal with nonlinearity, non-stationarity due to the climate change effect and anthropogenic influences and seasonality behaviour of data like air temperature in terms of accurate estimation single model approach is found weak. To enhance the accuracy for modeling of air temperature, hybridization is very much needed. This could be achieved by making necessary changes by means of data pre-processing such as stabilizing mean and stationary time series before applying to any model. One of the recently popular methods is Wavelet transform using this pre-processing of data by stabilizing mean and creating stationary time series before putting them to any predictive model. Recently, Wavelet transform is widely used in data pre-processing techniques in non-stationary and noisy time series.

Mellit et al. (2007) concentrated on Artificial Neural Networks (ANN) with Discrete Wavelet Transform (DWT) to determine and modeling of total solar radiation data from sunshine duration and mean temperature. They used 5 years total solar radiation, mean temperature and sunshine duration for model. These data are transformed by using the DWT and used for training and testing the ANN. They used LM (Levenberg- Marquardt) algorithm for training ANN model. Test results show that the correlation coefficient is 96.2%, which is satisfactory. In addition, estimated data by this model ANN-DWT were compared with different ANN architecture in order to present the performance of this method. The advantage of this model is to estimate the total solar radiation data from only mean temperature and sunshine duration.

Deka and Prahlada (2012) developed hybrid Wavelet-ANN model for forecasting of ocean wave height. They used Wavelet decomposed data as in put for the ANN model. Results revealed that WLNN works accurate than single ANN model.

Dadu and Deka (2013) used hybrid Wavelet-ANN model to forecast riverflow. With Daubechies wavelets of order 4 (Db4) and 5 (Db5)up to seven level, data is decomposed. Output of wavelet was used as input for ANN. It was found that WANN was working better with Db4 and Db5.

Ding-cheng et al.(2010) experimented with a new hybrid model to predict Air temperature. For their work they approached with a new model based on EMD (Empirical Mode Decomposition) and LS-SVM (Least Squares Support Vector Machine). Using EMD, it decomposed time series into a series of different scales. Then, decomposed data used as input for the LS-SVM model to predict the temperature. They compared hybrid output with single LS-SVM and neural network prediction method. Results confirmed that models having higher accuracy compare to single models.

Using Wavelet and ANFIS, Ashish and Rashmi (2011) tried to forecast the daily air pollution. For the study, they used daubechies 8 Wavelet at level 3 for decomposition daily averaged value of air pollution parameters like Carbon Monoxide (CO). Here ANFIS acts as a basis for constructing a set of fuzzy rules to generate the stipulated input-output pairs. On the basis of these predicted values, the final forecasting was prepared. From the model performance, they concluded that Wavelet decomposition (approximation and details) plays a vital role in the prediction.

Kisi et al. (2011) employed a hybrid Wavelet- Genetic programming to model for predicting the short-term and long-term air temperature. They tried to compare hybrid Wavelet- Genetic programming with the single Genetic programming model. The results showed that hybrid Wavelet- Genetic programming worked better in term of predicting air temperature.

Many works are carried out using Wavelet- Neuro in different application like forecasting of precipitation, prediction of suspended sediment load, short term load casting and forecasting of steam flow.

Partal and Kisi (2007) used Wavelet and Neuro-Fuzzy for precipitation forecasting with the conventional single model. The Wavelet-Neuro-Fuzzy model provided better performance especially for time series which had zero precipitation in the summer months and for the peaks in the testing period. In terms of forecasting performance, hybrid model evidently outperformed other models.

Rajaee et al. (2010) worked out a model Neuro-Fuzzy, conjunction of Wavelet -Neuro fuzzy and conventional sediment rating curve models for prediction of suspended sediment load in a gauging station in USA. They decomposed river discharge and suspended sediment load at different scales and used as input for the model. They concluded that wavelet analysis and Neuro-Fuzzy model performance was better for prediction of suspended sediment load.

Chaturvediet al. (2013) carried out work for short term load forecasting using wavelet in combinations with neuro fuzzy modules. In that approach, data was decomposed into Daubechies wavelets Db8 and outcome of the decomposition was used as an input for ANFIS. The results showed that wavelet decomposition plays a vital role in the analysis of load forecasting.

Yarar (2014) developed a model for water usage policy by forecasting the stream flow. Wavelet-Neuro Fuzzy (WNF) was developed to forecast the stream flow data. The study was conducted for 5 Flow Observation Stations (FOS) which belonged to Sakarya Basin in Turkey. Obtained results showed that hybrid WNF model was more accurate than Auto Regressive Integrated Moving Average (ARIMA) model for forecasting the stream flow.

Mohammadi et al. (2015) attempted to estimate the total monthly mean daily solar radiation using SVM-WT. Performance of the model was compared with the other models like ANN, GP and ARMA. They concluded that a new hybrid model was more reliable in estimating solar radiation compared to other models.

Salcedo-Sanz et al. (2015) tried to predict monthly air temperature using two Machine Learning algorithms (Support Vector Regression (SVR) and Multi-layer Perceptron (MLP)).These two models were compared and found SVR was working better.

In recent time many researchers used meteorological and pollution parameters to forecast other meteorological and pollution data using hybrid model.

Osowski and Garanty (2007) used meteorological, NO_2, CO, SO_2 and dust to predict daily air pollution forecasting by Wavelet decomposition and support vector machine. Using Wavelet decomposition daily, data was decomposed and then SVM was used for forecasting. Application of SVM instead of classical MLP has enabled to obtain much better accuracy in terms of forecasting.

Shaharuddin et al. (2008) using Wavelet decomposition tried to investigate the relationships between $PM_{10,}$ rainfall, temperature and wind. For the study they decided to use Non–decimated Wavelet transform, because it has better characteristics in the statistical point of view. Meteorological parameters have great influences to suspended particulate variation. They observed positive correlation between PM_{10} and temperature, at the same time negative correlation between PM_{10} and rainfall and both positive negative correlation for PM_{10} and wind.

Keeping in mind, the health impact caused by air pollution, a forecasting model was developed (Vong et al., 2012). For this study, they used meteorological and pollution data like suspended particulate matters (SPM), nitrogen dioxide (NO_2), sulphur dioxide (SO_2) and ozone (O_3) collected daily at the monitoring stations. The predictionresults of Linear model and RBF model showed a relative good fit to observed test set of over one year of data, particularly for SO_2 and NO_2.

Comparison of hybrid models and comparison of single model with hybrid models has nowadays attracted the attention of researchers in the other areas.

Moosavi et al. (2013) carried out work using various soft computing techniques to forecast the groundwater level for different prediction periods. They compared several data-driven models like ANN, ANFIS, Wavelet - ANN and Wavelet - ANFIS models. It was demonstrated that wavelet transform in both the hybrid models can improve accuracy of groundwater level forecasting. They concluded that, the forecasts made by Wavelet-ANFIS models are more accurate than those by ANN, ANFIS and Wavelet-ANN models.

Patil and Deka (2015) tried to estimate evapotranspiration in arid regions of India. They used Wavelet transform with ANFIS and Wavelet transform with ANN and these models are compared with Single ANFIS and ANN model. Study concluded that hybrid Wavelet-ANFIS and Wavelet-ANN working better than Single ANFIS and ANN model.

Raghavendra and Deka (2015) developed a hybrid Wavelet packet - Support vector regression model to forecast monthly groundwater level fluctuations and performance of the hybrid model is compared with single Support vector regression. Results reveal that WP - SVR model outperform classic Support vector regression model.

2.7 SUMMARY OF LITERATURE

From the literature review, it was observed that, few researchers have used hybrid models like Wavelet transform with ANFIS and Wavelet transform with SVM which can handle the above mentioned limitations such as nonlinearity, noisy, uncertainty or even non-stationary in other areas and very few to certain extent for the modeling of air temperature. Also, in urban and industrial areas, very few publications available till date regarding use of pollutants along with meteorological parameters in modeling air temperature. Hence, a novel approach is proposed for accurate modeling of air temperature using hybridized Wavelet-ANFIS and Wavelet-SVM model for modeling the air temperature for different input scenarios. These hybrid methods are expected to do better in prediction models for the scenarios of uncertainty, incompleteness and noisy data as mentioned in the past study.

2.8. RESEARCH OBJECTIVE

Based on the literature review, following objectives were finalized. The main objectives of the proposed research are:

1. To investigate the potential and applicability of soft computing techniques like Support Vector Machines (SVM), Adaptive Neuro-Fuzzy Inference Systems (ANFIS) to effectively address various tasks and issues associated with the modeling of air temperature using meteorological variables.
2. Various hybrid models to be developed by integrating Wavelet transform with Support Vector Machines (SVM) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) techniques.
3. To evaluate the performance of various hybrid model results for selection of best model.

Secondary objective of the proposed research is:

To investigate the influence of air pollution parameters along with meteorological parameters in air temperature modeling.

CHAPTER 3

STUDY AREA AND METHODOLOGY

3.1 INPUT DATA

In the present work, in the first stage, monthly recorded meteorological data from seven stations located in and around the Shimoga district (Fig. 3.1) (Shimoga station, Raipura station, Linganmakki station, Honnali station, Hiriyur station, Bhadra station (B. R. Project) and Davanagere) which covers a time period of 11 years (from January 2001 to December 2011) was used. Shimoga district covers an area of 8477.84 sq. km and lies in the western part of the Karnataka state between 13° 27' to 14° 39" North latitude and 74° 38´to 75° 45′ East longitudes. The area belongs to the tropical climate region. Generally, the weather is hot and humid in the eastern part and very pleased with the remaining parts of the area. The evapotranspiration is normally higher in the ghat section as compared to plane areas in the east. Summer prevails between March to early June, the wet months start from early June to September and during October and November months scanty rain is experienced by N-E monsoon. The winter commences in mid-November and ends in the middle of February. The maximum and minimum temperature of this area is 31^{0}C and 18 ^{0}C. The relative humidity ranges from 27 to 88% and receives an average annual rainfall of 535 - 2828 mm.Rapid growth of urbanization and vehicles are contributing in increase of temperature in recent time. The various inputs for the model are expressed in time series as shown in Figs 3.2 to 3.6.

In the next stage, to check influence of air pollution parameters (SPM, SO_2, NO_2) along with the meteorological data in the modeling of air temperature, models are applied to Bhadra station (B. R. Project) (13° 42′ N and 75° 38′ 24″ E) for a period of January 2009 to July 2012. The Mysore Paper Mills and Vishweshvarayya Iron and Steel Industries are the major industrial activity noted in the command area of the Bhadra station. Every year, growth of vehicles are exponential, this may lead to higher emission rate of pollution. The air pollutant records are shown in Fig3.7.From the figure, it was observed that all the three pollutants are having similar trend of variation and also within permissible limit.

Fig. 3.1 Study area

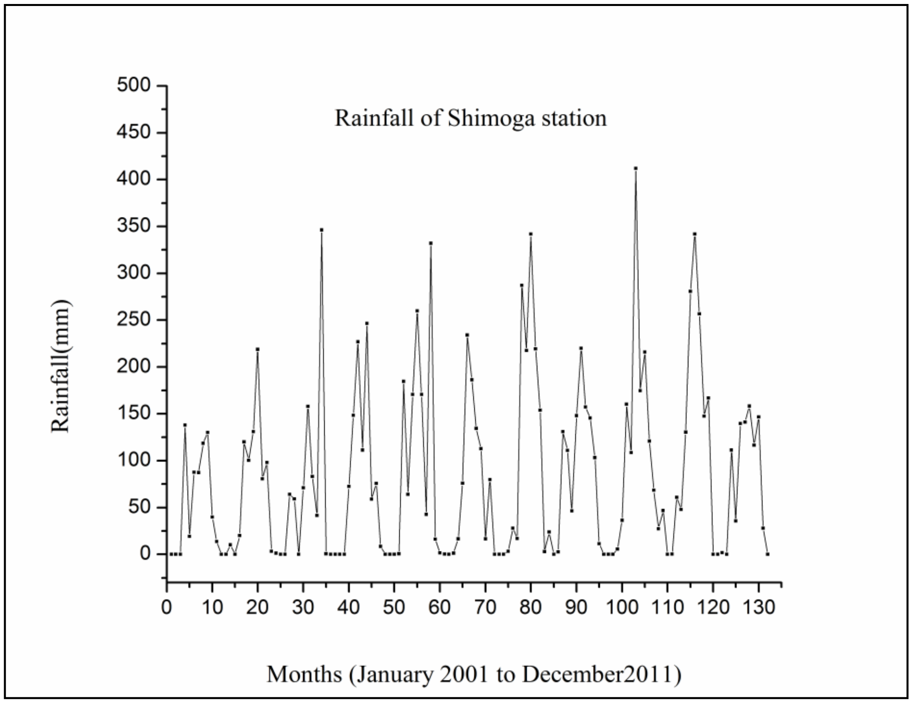


Fig.3.2 Rainfall of Shimoga station

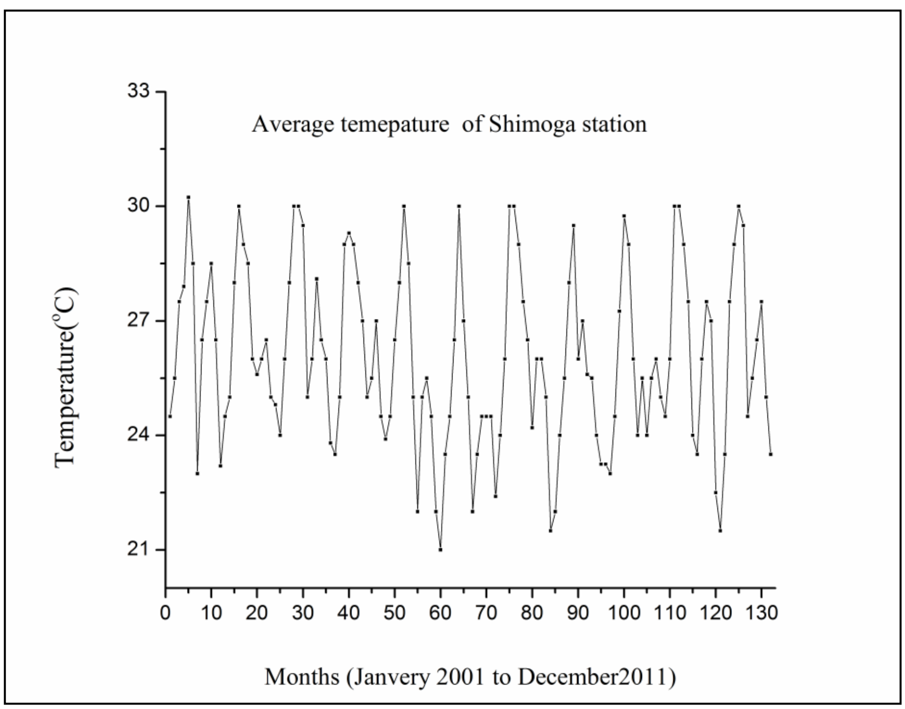


Fig. 3.3 Average air temperature of Shimoga station

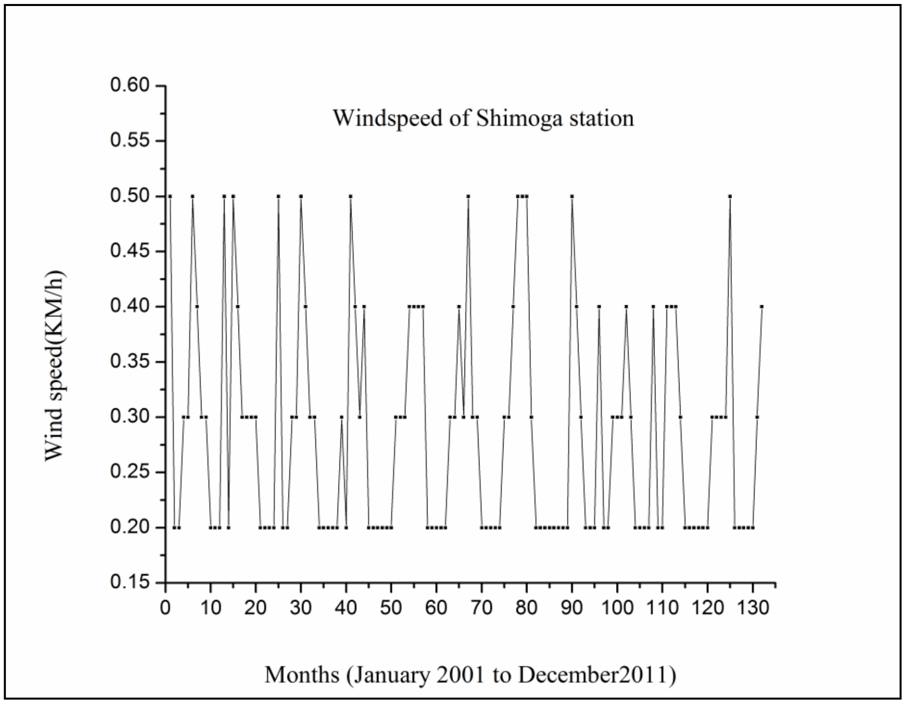


Fig. 3.4 Wind speed of Shimoga station

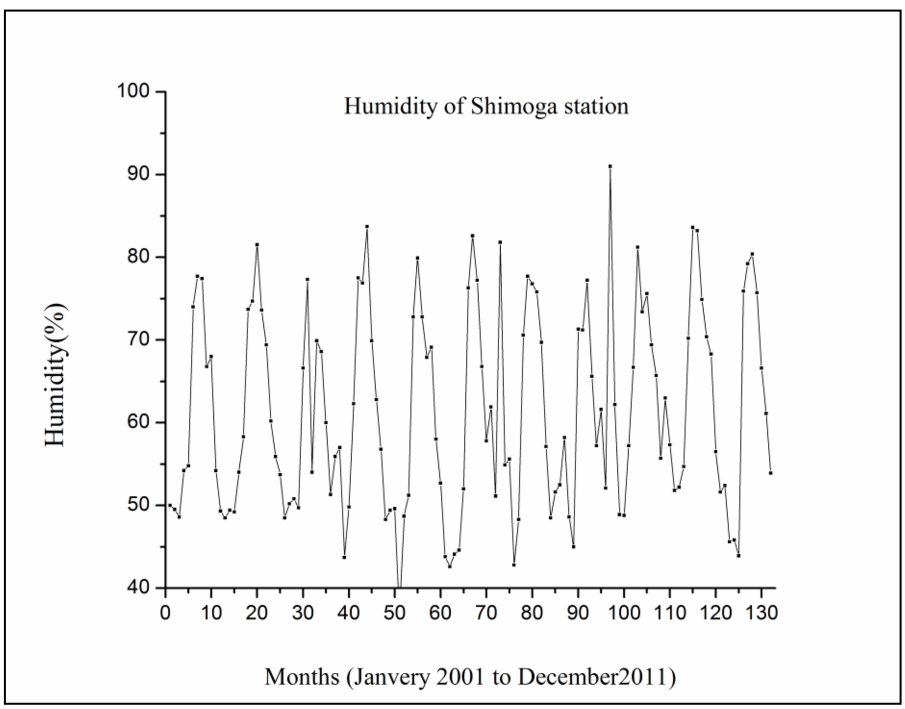


Fig. 3.5 Humidity of Shimoga station

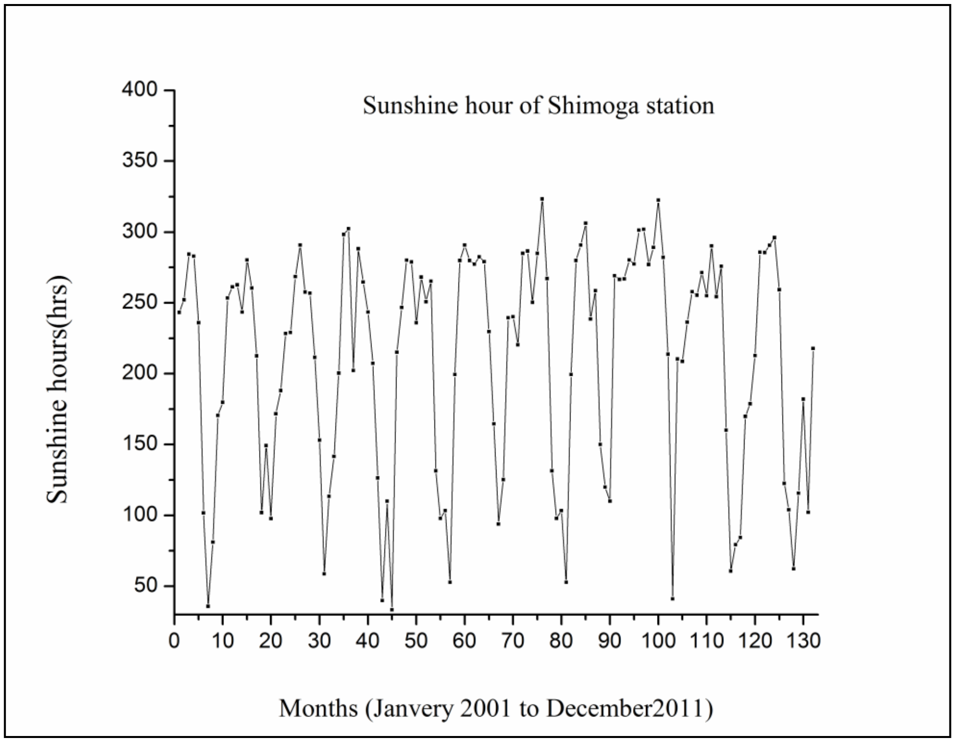


Fig. 3.6 Sunshine hour of Shimoga station

Table 3.1 Statistical analysis of data of seven stations.

STATIONS		SHIMOGA	HONNALI	B.R.PROJECT	DAVANGERE	LINGANAMAKKI	HIRIYUR	RAIPURA
RAINFALL(mm)	MIN	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	MAX	412.00	253.80	380.00	256.00	1433.40	267.30	250.00
	MEAN	86.02	61.31	93.14	58.80	275.27	52.80	49.30
	STD	94.46	65.86	101.30	69.66	396.86	63.50	61.64
AIR TEMPERATURE (^oC)	MIN	21.00	21.00	13.80	20.50	21.50	21.50	14.50
	MAX	30.23	32.60	32.35	31.60	29.70	32.00	31.00
	MEAN	26.05	25.63	24.28	26.58	25.87	26.51	24.25
	STD	2.30	2.32	3.35	2.45	1.51	2.65	3.21
WIND SPEED(m/s)	MIN	0.06	0.04	0.01	0.01	0.02	0.10	0.01
	MAX	0.50	0.80	0.10	0.42	4.00	0.40	1.10
	MEAN	0.28	0.23	0.05	0.23	1.33	0.26	0.38
	STD	0.11	0.21	0.04	0.09	1.33	0.07	0.25
HUMIDITY(%)	MIN	36.50	67.00	63.50	47.70	66.50	43.60	67.10
	MAX	91.00	87.50	92.00	86.30	99.30	81.80	91.90
	MEAN	61.46	78.41	81.23	68.87	83.85	64.65	79.00
	STD	12.16	4.87	6.93	9.52	8.22	8.84	5.73
SUNSHINE HOUR(hrs)	MIN	33.40	24.90	30.70	23.70	11.70	51.00	96.70
	MAX	323.30	304.90	303.70	336.40	302.00	408.90	310.80
	MEAN	213.78	191.93	180.72	192.72	188.87	218.87	230.83
	STD	78.74	82.76	72.36	78.19	82.40	67.93	50.77

Statistical analyses of seven stations are given in Table 3.1. By comparing analysis of all seven staions, it was observed that similar trend prevailing in all sations except for linaganamakki station . In lianganamakki station, standerdivation are 396.86, 1.51, 1.33, 8.2 and 82.40 for rainfall, temperature, wind speed, humidity and shine shine hour respectively which shows higher variability compared to other stations.

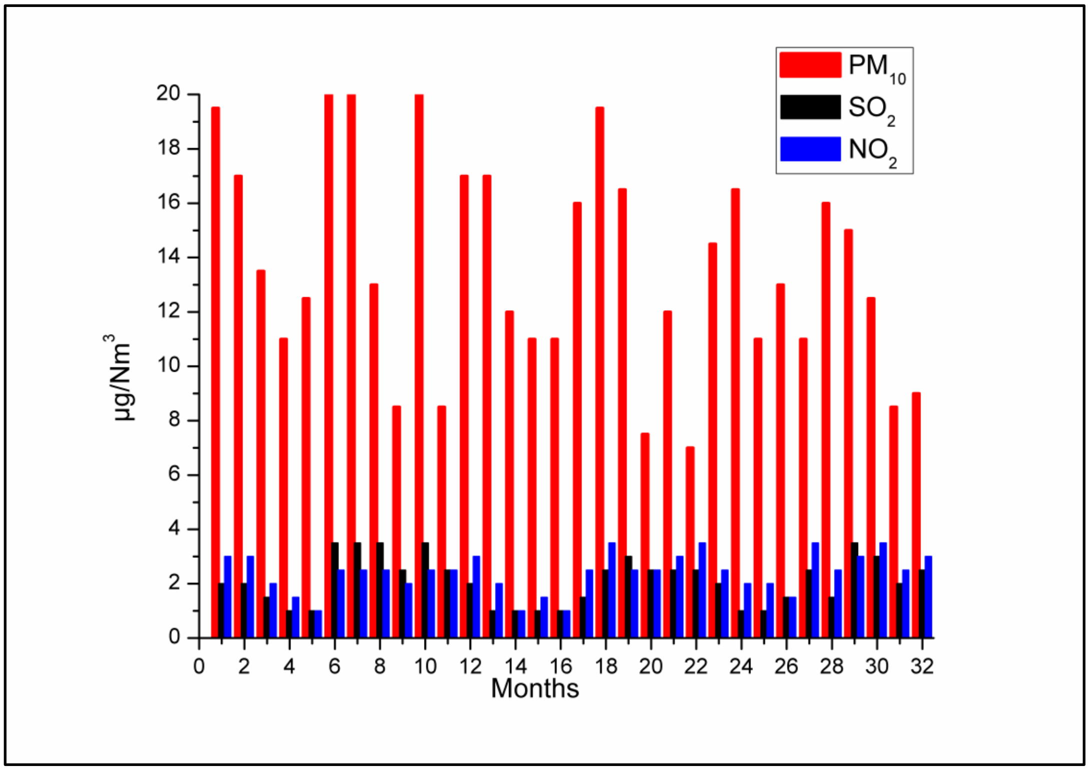


Fig. 3.7 Air pollutant concentration of Bhadra station.

3.2 RESPIRABLE DUST SAMPLER:

Respirable Dust Sampler is the instruments, works on the simple design standardized by USEPA for monitoring pollutant gases like SO_2, NO_2 and the Suspended Particles (TSP) in ambient air (Fig 3.8). To know the concentration of specific pollutants, these gases are analysed by simple chemistry method.

3.2.1 Respirable Suspended Particulate Matter

Ambient air laden with suspended particulates enters the system through the inlet pipe in the sampler. As the air passes through the system, non-respirable dust is separated from the air stream by centrifugal forces acting on the solid particles. These separated particulates fall through the conical hopper and get collected in the sampling bottle placed at its bottom. The fine dust forming the respirable fraction of the Total Suspended Particulate (TSP) passes through the cyclone and is carried by the air stream to the filter paper clamped between the top cover and filter adaptor assembly. The respirable dust (RSP) is retained by the filter and the carrier air exhausted from the system through the blower.

Fig. 3.8 Respirable Dust Sampler

3.2.2 Sulphur dioxide in ambient air

Sulphur dioxide from air is absorbed in a solution of potassium tetrachloro-mercurate (TCM).As such, dichlorosulphitomercurate complex is formed. This complex is made to react with pararosaniline and formaldehyde to form the intensely colouredpararosanilinemethylsulphonic acid. The absorbance of the solution is measured by means of a suitable spectrophotometer at 560nm.

3.2.3 Nitrogen dioxide in ambient air

Ambient NO_2 was collected by bubbling air through a solution of sodium hydroxide and sodium arsenite. The concentration of nitrite ion produced during sampling was determined calorimetrically by reacting the nitrite ion with phosphoric acid, and sulphanilamide , and N-(1-napthyl)-ethylenediamine di-hydrochloride (NEDA)

3.3 METHODOLOGY

3.3.1 General

The present research work processed in two stages. In the first stage(Fig.3.9) measured data like meteorological parameters (rainfall, humidity, wind speed, sunshine hour) were used as a input for the both ANFIS and SVM model for all the seven station (Shimoga station, Raypura station, Linganmakki station, Honnali station, Hiriyur station, B. R. Project and Davanagere station) and observed average air temperature as output for the model. In the next step, all the observed data like meteorological parameters (rainfall, humidity, wind speed, sunshine hour) of all seven station were pre-processed through Daubechies mother Wavelet up to 3rd level of decomposition and then this pre-processed data was used as input for both ANFIS model and SVM model with observed average air temperature as output for the model. Further, the performance of the models in terms of accuracy were compared between single ANFIS, single SVM, hybridized Wavelet-ANFIS and hybridized Wavelet-SVM models, and the same are presented in in next chapter.

In the second stage(Fig 3.10), observed meteorological parameters (rainfall, humidity, wind speed, sunshine hour) and air pollution parameters (SPM, Nitrogen dioxide (NO_2) and Sulphur dioxide (SO_2)) of Bhadra station, Karnataka, India were together used for input for models like ANFIS and SVM and observed temperature was the output for the model. In extension of the work, observed meteorological parameters (rainfall, humidity, wind speed, sunshine hour) and air pollution parameters (SPM, Nitrogen dioxide (NO_2) and Sulphur dioxide (SO_2)) were applied to Daubechies mother Wavelet (Db1, Db2, Db3, Db4, Db5) up to fifth level. These decomposed data was used as input for ANFIS and SVM model and original measured air temperature is used as output for the model. In the final phase the accuracy assessment of single ANFIS, single SVM, hybridized Wavelet-ANFIS and hybridized Wavelet-SVM models are done and the same are presented in next chapter**.**

3.3.2 Sensitivity analysis;

The parameter values and assumptions of any model are subject to change and error. Sensitivity analysis (SA), broadly defined, is the investigation of these potential changes and errors and their impacts on conclusions to be drawn from the model.

SA can be easy to do, easy to understand, and easy to communicate. It is possibly the most useful and most widely used technique available to modelers who wish to support decision makers. In case of testing the robustness of an optimal solution, identifying critical values, thresholds or break-even values where the optimal strategy changes, investigating sub-optimal solutions and assessing the "riskiness" of a strategy or scenario sensitivity analysis is useful.

3.3.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. It was adopted to find out the input variables with percentage of influences on modeling of air temperature.

The number of components extracted (created) in a principal component analysis is equal to the number of observed variables being analysed. However, in most analyses only the first few components account for meaningful amounts of variance so only these first few components are interpreted and used in a subsequent analyses such as a multiple regression. The first principal component accounts for the most possible variance in the data. The second component accounts for the most variance not accounted for by the first component, and so on until all variables are accounted for. The first few components account for most of the total variation in the data, and can be used for subsequent analysis.

The first principal component extracted in a principal component analysis accounts for a maximal amount of total variance in the observed variables. Under typical conditions, this means that the first component is correlated with at least some of the observed variables. In fact, it is often correlated with many of the variables.

The second principal component extracted has two important characteristics.

- The second component accounts for a maximal amount of variance in the data not accounted for by the first component. Under typical conditions, this means that the second component is correlated with some of the observed variables that did not display strong correlations with the first component.

- The second characteristic of the second component is that it is uncorrelated with the first component. If you compute the correlation between component 1 and component 2, that correlation is zero.

The remaining components extracted in the analysis display these same two characteristics each component accounts for a maximal amount of variance in the observed variables that was not accounted for by the preceding components and is uncorrelated with all of the preceding components. A principal component analysis proceeds in this manner with each new component accounting for progressively smaller amounts of variance. This is why only the first few components are retained and interpreted. When the analysis is complete, the resulting components display varying degrees of correlation with the observed variables, but are completely uncorrelated with one another.

3.3.3. Wavelet Analysis

The Wavelet Series has been just like Continuous Wavelet Transform (CWT) and it requires a significant amount of time and resources, depending on the results required. Ψ (t) is the mother wavelet or the basis function (Eq. 3.1). The Continuous Wavelet Transform (CWT) is provided by Eq. (3.2), where f (t) is the signal to be analyzed. The transformation used in the wavelet functions are derived from the mother Wavelet through translation (shifting) and scaling (dilation or compression) (Yarar 2014)

$$\psi_{a,b}(t)=\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \tag{3.1}$$

$$X_{WT}(a,b)=\frac{1}{\sqrt{|a|}}\int f(t).\psi^*\left(\frac{t-b}{a}\right)dt \tag{3.2}$$

Where $\psi_{a,b}(t)$ is the successive wavelet, a is the frequency factor, b is the time factor and ψ^* is the complex conjugate functions of ψ (t).

The Discrete Wavelet Transform (DWT), which is based on sub-band coding, is found to be best for computation of Wavelet Transform (Wei et al.2012). Implementation of this method is easy and works better in terms of computation time and resources required. DWT of f(t) can be written as (Eq.3.3);

$$X_{WT}(j,k)=\frac{1}{\sqrt[-j]{1a1}}\int f(t).\psi^*\left(\frac{t}{a_0^{-j}}-kb_0\right)dt \tag{3.3}$$

The most frequent choice of the parameters a_0 and b_0 is 2 and 1 time steps, respectively (Wei et al.2012). This power of two logarithmic scaling of the time and scale is known as a dyadic grid arrangement and is the simplest and the most efficient case for practical purposes (Mallat 1989)

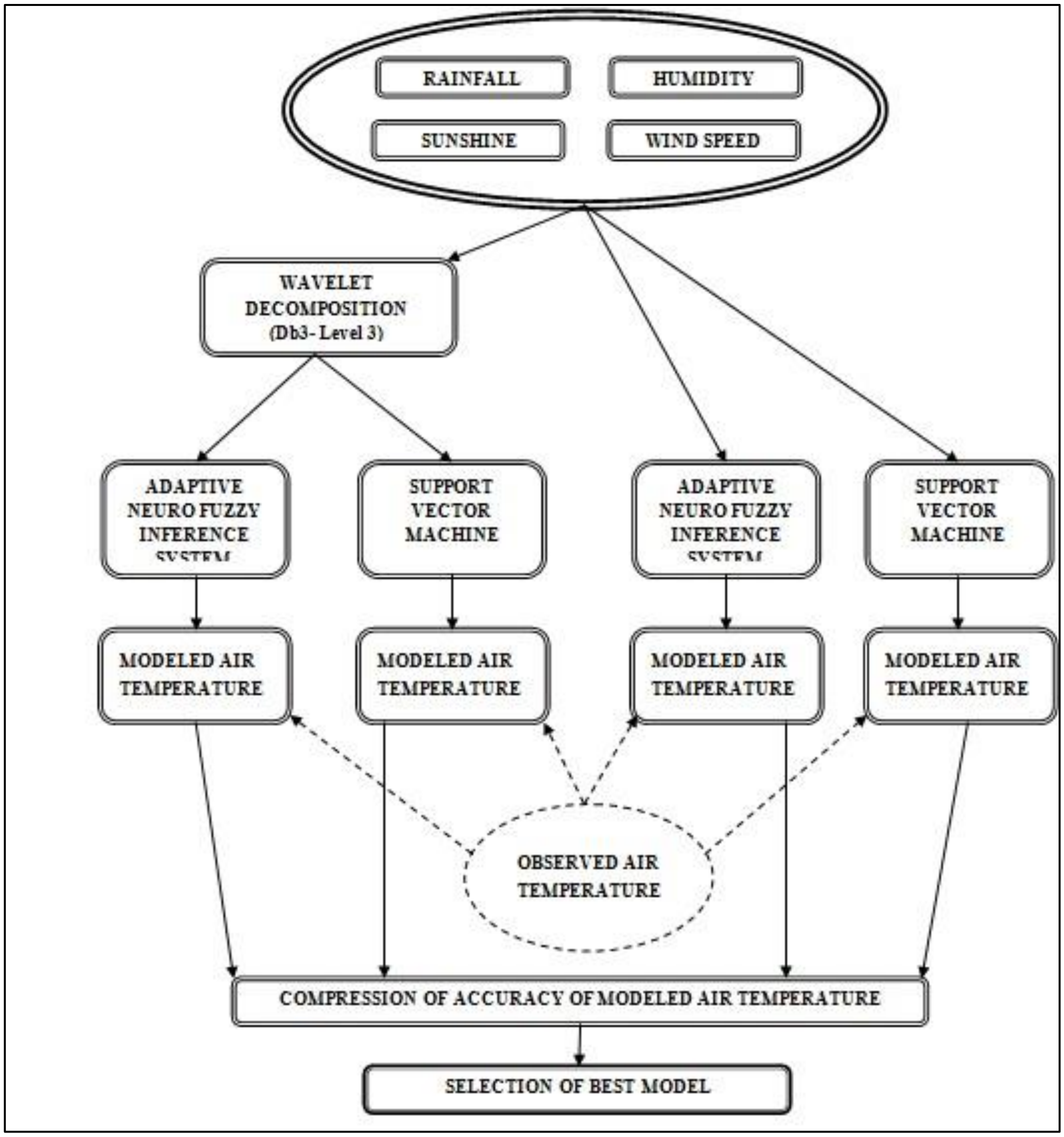


Fig.3.9 Flowchart of the work (First phase with meteorological data)

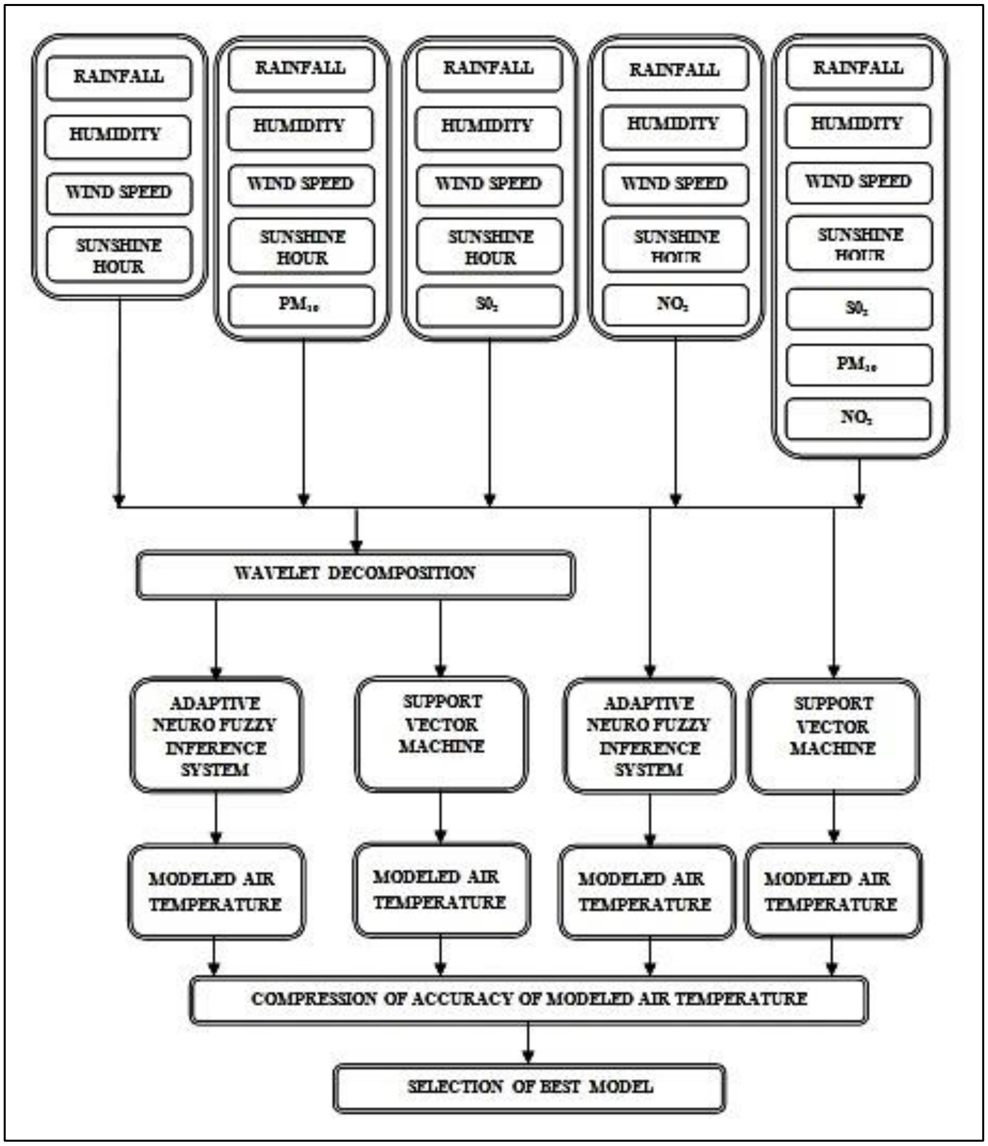
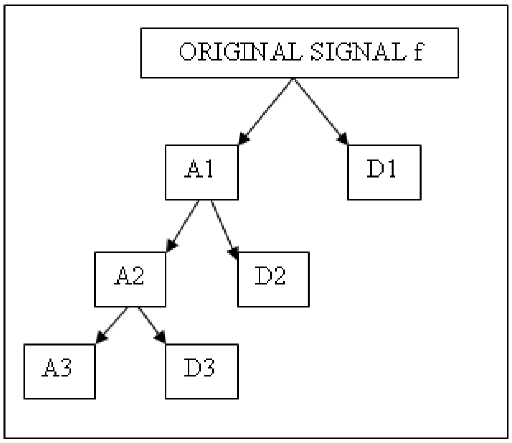


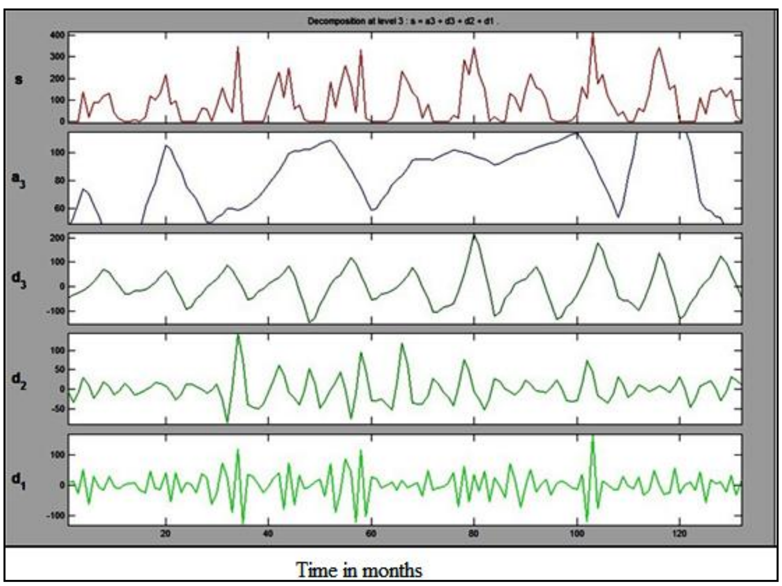
Fig.3.10 Flowchart of the work (Second phase with meteorological and air pollution data)

DWT operates on two sets of function like high-pass and low-pass filters. The original time series is passed through high-pass and low-pass filters and separated at different scales. The time series is decomposed into one comprising its trend (the approximation) and other comprising the high frequencies and the fast events (the detail) (Fig 3.11 and Fig 3.12)



A1,A2,A3… are approximation, D1, D2,D3… are detailing

Fig. 3.11 Architecture of Wavelet Decomposition model



S: Original raw signal input(e.g. rainfall)

a3: approximation and d1, d2, d3: detailing at level1,2,3

Fig. 3.12 Architecture of Wavelet Decomposition model

3.3.4. The Adaptive Neuro-Fuzzy Inference System (ANFIS) model

ANFIS is a combination ANN and Fuzzy Inference System. Bring in the advantages of both self-learning neural network and fuzzy reasoning similar to human reasoning to arrive at reasonable decision. (Yun et al.(2008)) It is a class of adaptive multilayer feedforward networks, applied to forecasting of non-linear, non-stationary data with the help of historical data.

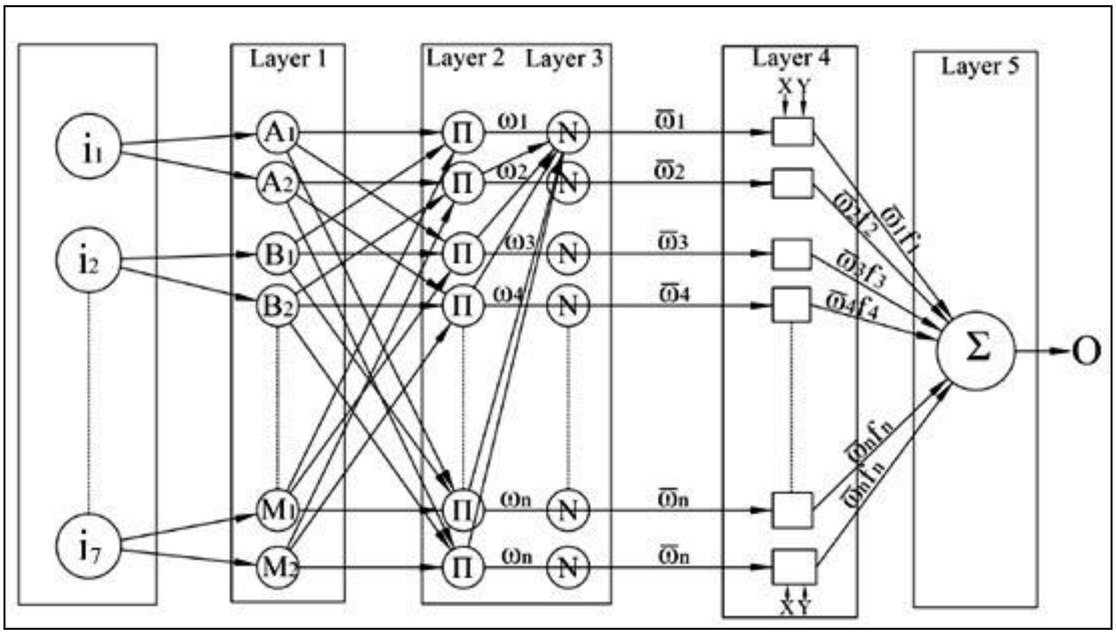


Fig. 3.13 Architecture of ANFIS model

The ANFIS architecture is shown in Fig. 3.13. The ANFIS network contains five layers. Each layer contains several nodes described by the node function. Let O_i^j denote the output of the i^{th} node in layer j.

Here, in 1st layer, $i_1, i_2, \ldots\ i_7$ are input variables like rainfall, wind speed, humidity, sunshine hour, PM_{10}, NO_2 , SO_2 .In the 2nd layer, A_1, A_2,.. B_1, B_2,.. and M_1,M_2 may represent membership function type and finally 5th layer is the output such as air temperature.

Layer 1 Each node in this layer is an adaptive node with node output defined as

$$O_i^1 = \mu_{Ai}(x), \text{for i=1,2...} \tag{3.4}$$

$$O_i^1 = \mu_{Bi-2}(y), \text{ for i=3,4...} \tag{3.5}$$

Where x (or y) is the input to the node; and A_i (or B_i_2) is a linguistic label associated with this node. The membership functions for A and B are usually described by generalized bell functions.

In layer 2, each node π multiplies incoming signals and the output is the product of all the incoming signals.

$$O_i^2 = \omega_i = \mu A_i(x)\mu B_i(y), \qquad \text{for i=1,2} \tag{3.6}$$

Each node output represents the firing strength of a rule.

In layer 3, each node N calculates the ratio of the i^{th} rules firing strength to the sum of all rules's firing strengths.

$$O_i^3 = \overline{\omega_i} = \frac{\omega_i}{\omega_1 + \omega_2}, \qquad \text{for i=1,2} \tag{3.7}$$

The normalized firing strengths are the output from this layer.

In layer 4, each node calculates the contribution of the i^{th} rule to the overall output

$$O_i^4 = \overline{\omega_i} f_i = \overline{\omega_i}(a_i x + b_i y + c_i) \qquad \text{for i=1,2} \tag{3.8}$$

Where $\overline{\omega_i}$ is the output of layer 3 and $\{a_i,b_i,c_i\}$is the parameter set. The parameters of this layer are known as consequent parameters.

In layer 5, the single node Σ calculates the final output as the summation of all input signals

$$O_i^5 = overalloutput = \sum_i \overline{\omega_i} f_i = \frac{\sum_i \omega_i f_i}{\sum_i \omega_i} \tag{3.9}$$

Thus, an adaptive network is functionally equivalent to a Sugeno-type fuzzy inference system(Catalao et al. 2011; Jang and Sun. 1995).

3.3.5. Support Vector Machines

Vapnik (1995) laid foundation for Support Vector Machines (SVM). It is becoming a popular choice because of many attractive features, and its potential in empirical performance. When compared to traditional Empirical Risk Minimization (ERM) principle, employed by conventional neural networks, Structural Risk Minimization (SRM) principle(Gunn, 1998), proves to be finer. On the expected risk SRM minimizes an upper bound, as contrasting to ERM that minimizes the error on the training data. This different nature equips SVM with a superior capability to simplify, which is the objective in statistical learning. To resolve the classification problems SVMs were developed, but in recent time they have been extended to the field of regression problems (Vapnik et al., 1995).

The salient features of SVM are:

(i) SVM is a fully data based nonlinear modeling paradigm.

(ii) SVM approach is based on the principle of structural risk minimization, which helps in larger amount to generalize.

(iii) The parameters of SVM model can be derived by solving a quadratic optimization problem.

(iv) Quadratic form possesses a single minimum which is an objective function of SVM, thus avoiding the heuristic procedure involved in locating the global or the deepest local minimum on the error surface.

(v) In the beginning inputs are nonlinearly mapped into a high dimensional feature space which is then interrelated linearly with the output.

3.3.6. Mathematics behind SVM algorithm for regression

A training data set $g = \{(x_1, y_1), (x_2, y_2),......,(x_p, y_p)\}$ is considered such that $x_i \in \upsilon^{\boldsymbol{N}}$is a vector of input variables and $x_i \in \upsilon$ is the corresponding scalar output (target) value. Here, the objective of modeling is to find a regression function, $y = f\ (x)$, such that prediction of outputs $\{y\}$ accurately, which is corresponding to a new set of input-output examples, *{(x, y)}*, which are drawn from the same underlying joint probability distribution as the training set. To achieve the desired aim, the following linear estimation function is considered by support vector regression (SVR).

$$f(x) = (w.x) + b \tag{3.10}$$

Where, weight vector is represented by w; b refers to a constant known as "bias"; *f(x)*repersents a function termed feature, and $(w.x)$ refers the dot product in the feature space, l , such that $\emptyset$: $x \rightarrow l$, w$\in l$. The basic concept of support vector regression is to map nonlinearly the original data x into a higher dimensional feature space and solve a linear regression problem in this feature space.

The regression problem is equivalent to minimize the following regularized risk function:

$$R(f) = \frac{1}{n}\sum_i^n L(f(x_i) - y_i) + \frac{1}{2}\|w\|^2 \tag{3.11}$$

where,

$$L(f(x) - y) = \begin{Bmatrix} \|f(x) - y\| - \varepsilon, \text{for } \mid f(x) - y \mid x \geq \varepsilon \\ 0,\ otherwise \end{Bmatrix} \tag{3.12}$$

Eq. 3.12 is also called ε-insensitive loss function. This function defines a ε-tube. The loss is zero or else the loss is equal to the magnitude of the difference between the radius ε of the tube and the predicted value, if the predicted value is within the ε-tube. The radius of the tube located around the regression function (Fig 3.14) is represented by a precision parameter ε and the "ε-intensive zone" is the region enclosed by the tube.

The SVM algorithm tries to keep the position the tube around the data as shown in Fig 3.14. The e-insensitive loss function is substituted into Eq. (3.11), the optimization object becomes:

$$\text{Minimize } \frac{1}{2}\|w\|^2 C \ + \ \sum_i^n (\xi_i + \xi_i^*) \tag{3.13}$$

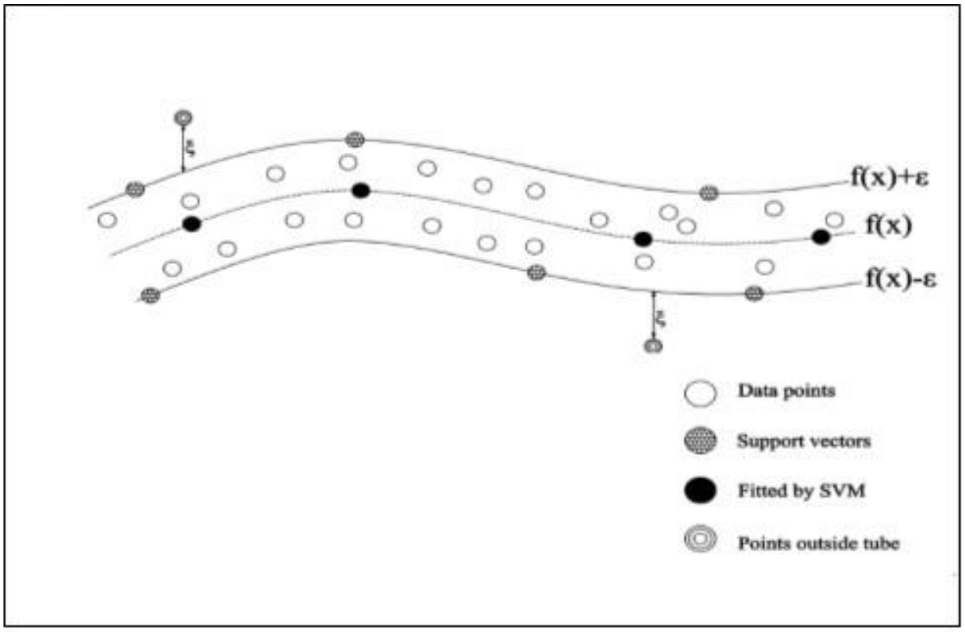


Fig. 3.14 A schematic diagram of support vector regression using e-insensitive loss function

With the constraints,

$$Subjected \begin{Bmatrix} y_i - (w.x) - b \leq \varepsilon + \xi_i \\ (w.x) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i , \xi_i^* \geq 0 \end{Bmatrix} \quad (3.14)$$

Where, the penalty degree of the sample with error exceeding epsilon if the constant C>0.Two slack variables represents the distance from actual values to the corresponding boundary values of ε-tube.

The SVM fits *f(x)* to the data in a manner such that:

(i) Minimizing the slack variables i.e,ξ_i, ξ_i^* the training error is minimized and,

(ii) To increase the flatness of function *f(x)* or to penalize over complexity of the fitting function $||w||^2$ is minimized.

A dual problem can then be derived by using the optimization method to maximize the function,

Maximize

$$-\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_j) - \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \tag{3.15}$$

$$\text{Subject to } \sum_{i=1}^{n}(\alpha_i + \alpha_i^*) = 0 \text{ and } 0 \le \alpha_i, \alpha_i^* \le C \tag{3.16}$$

Where, α_i, α_i^* are lag range multipliers. Owing to the specific character of the above-described quadratic programming problem, support vectors (SVs) are the non-zero coefficients, $(\alpha_i - \alpha_i^*)$ corresponding to input vectorsx_i. The SVs can be thought of as the most informative data points that compress the information content of the training set. The coefficients α and α* have an intuitive interpretation as forces pushing and pulling the regression estimate *f*(x_i) towards the measurements, y_i.

The SVM for function fitting obtained by using the above-mentioned maximization function is then given by,

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\,(x_i.x) + b \tag{3.17}$$

As for the nonlinear cases, the solution can be found by mapping the linear problems with the original ones in a characteristic space of high dimension, in which dot product manipulation can be substituted by a kernel function, i.e $K(x_i, y_i) = \emptyset(x_i)\emptyset(y_i)$. Substituting $K(x_i, y_i) = \emptyset(x_i)\emptyset(y_i)$ in Eq 3.15 allows us to reformulate the SVM algorithm in a nonlinear paradigm. Finally, we have,

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\,K(x_i.x) + b \tag{3.18}$$

3.3.7. Tuning parameters of SVM

The performance of SVM is assessed with leave-one-out error and validation commonly. For the research work 75% for training and 25% for testing of available data which was chosen randomly to assess the performance of SVM.

SVM is considered to be successful only if the system can perform well on test data on which the system is not trained. The generalization performance of SVM is mainly dependent upon the good setting of meta-parameters C, ε , and kernel parameters such as kernel type and loss function type. The complexity of SVM model mainly depends on parameters such as C, ε, γ, and d. The selection such parameter to the optimal is a complicated.

The selection of kernel type and kernel function parameters and distribution of input (x) values of the training data are based on the application-domain knowledge. The trade-off between the model complexity (flatness) and the degree to which deviations larger than ε are tolerated in optimization formulation is controlled by the parameter C. For example, the errors penalized are higher when the C value is too large (infinity) and in that situation the SVM is trained to minimize error with lower generalization ability and the errors penalized are less when the C value is too small which allows the minimization of margin, thus higher generalization ability. The SVM model complexity (and hence its generalization performance) depends on all five parameters due to the selection of optimal parameter is further a complicated problem. Parameter ε controls the width of the ε -insensitive zone which is used to fit the training data [Vapnik, 1995; Vapnik, 1998].

The number of support vectors used to construct the regression function is mainly affected by the value ε. Fewer support vectors and the bigger ε are selected. On the other hand, bigger ε -values result in more 'flat' estimates. Hence, both C and ε -values affect model complexity.

The parameters should be properly optimized to minimize the generalization error. The choice of C and ε can be summarized as follows by Existing practical approaches (Cherkassy and Ma, 2004):

- Based on a prior knowledge and /or of user parameter C and ε are finalised. (Vapnik, 1998; Scholkopf and Burges (1999)).For a non-expert user this approach will not be an appropriate one. Based on observation that support vectors lie outside the ε -tube and the number of support vector strongly controls the complexity of the SVM .Scholkopf and Burges (1999) instead of controlling ε suggest to control another parameter ν (i.e., the fraction of points outside the ε -tube). Parameter ν has to be user-defined by this approach. Mattera and Haykin, [1999] propose to choose ε - value so that the percentage of support vectors in the SVM regression model is around 50% of the number of samples. However, one can easily show examples when optimal generalization performance is achieved with the number of support vectors larger or smaller than 50%.

- In agreement with general sources on SVM [Vapnik,1998], Smola et al., (1998) proposed asymptotically optimal ε -values proportional to noise variance. These approaches do not reflect sample size; this will be the main practical drawback of this approach. Instinctively for a large sample size the value of ε should be smaller than for a small sample size (with the same level of noise).
- Selection of parameter ε. The value of ε should be proportional to the input noise level, that is $\varepsilon \propto \sigma$ [Vapnik,1998; Smola et al.,1998].

Cherkassky and Ma [2002] propose the following (empirical) dependency:

$$\varepsilon = \tau\sigma\sqrt{\frac{\ln n}{n}} \tag{3.19}$$

- For diverse data set sizes, noise levels and target functions for SVMregression they recommend constant value τ = 3 which works better. Here n is refers to the number of training data sample. They presume that the standard deviation of noise σ is known or can be estimated from the data which is again a difficult task for non expert user.
- Selecting parameter C equal to the range of output values [Mattera and Haykin, 1999]. This is a reasonable proposal, but it does not take into account possible effect of outliers in the training data. Cherkassky and Ma (2002) propose following prescription for regularization parameter:

$$C = \max(|\, \bar{y} + 3\sigma_y \,|,|\, \bar{y} - 3\sigma_y \,|) \tag{3.20}$$

Where, $\bar{y}$ is the mean of the training responses (outputs), and σ_y is the standard deviation of the training response values. They claim that this prescription can effectively handle outliers in the training data. However the proposed value of C-parameter is derived and applicable for RBF kernels only.

- Using cross-validation for parameter choice as suggested by, Scholkopf et al., [1999] which is very computational and data-intensive.
- Several recent references present statistical account of SVM regression [Smola and Scholkopf, 1998] where the C parameter is interpreted as a traditional regularization parameter in formulation that can be estimated for example by cross validation whereas the ε - parameter is associated with the choice of the loss function (and hence could be optimally tuned to particular noise density).
- On the basis of prior application-domain knowledge a particular kernel type and kernel function are selected and also should reflect distribution of input(X) values of the training data. Very little literature is available to throw light on this.

To obtain optimal parameters trial and error is done. The trial and error method can be adopted in such situation which will take more time and may not really obtain best possible results.

3.3.8. Training, testing and Generalizability

An iterative process in which the SVM is mapped with inputs-outputs pairs to train the support vector machine.Here, in the process altering of margin (w) and bias (b) are done to produce the correct output (within a reasonable error margin).By the above process, model produces acceptable results,then it is trained and ready to act upon previously unseen data or else it re-reads the input and again model tries to produce the acceptable output. The margins and bias are considerably adjusted through the training set for each iteration (known as a training cycle). A lot of training cycles may be needed to identify the training set correctly for SVM depending upon the complexity of the task to be learned. Once the performance is optimal for training data, then the same model structure is used to examine the performance of unseen data. If the system works better for the test data on which the system has not been trained, then only SVM learning is considered successful. This capability of a SVM is called generalizability.

3.3.9. Uncertainty analysis:-

Lack of understating of phenomenon leads to uncertainty in selection of model input and consequently associated parameters (Srivastav, 2007). Many researchers have been inclined to simpler way of modeling assuming the model input and processes to be deterministic. As a consequence most of these models have been applied in deterministic way (Christiaens and Feyen, 2002) assuming that input variables and the parameters (After calibration) represents the reality in the accurate way.

Generally, in almost all applications, an ANFIS model is tested for its generalization properties by means of statistical evaluation measures, and no quantification of its predictive uncertainty is reported. The quantification of the uncertainty associated to the results provided by ANFIS models is essential for their confident and reliable use in practice.

The primary sources of uncertainty are input data, the model parameters, and the structure, in addition to the measured data used during calibration. The uncertainty evaluation provides the degree of behaviour of each set of input parameters, which in turn are translated into confidence interval estimates on the output of the model (Wagener, 2003). There are various methods available for quantifying the uncertainty in physical models. Also, in various AI models, three main approaches exist for the estimation of accuracy of models the delta method, the bootstrap method, and the Bayesian approach

The bootstrap method is the simplest approach since it does not require the complex computations of derivatives and Hessian-matrix involved in the delta method or Monte Carlo solutions involved in the Bayesian approach.

Abrahart (2003) employed bootstrap technique to continuously sample the input space in the context of rainfall-runoff modeling and reported that it offered marginal improvement in terms of greater accuracies and better global generalizations. He suggested further research involving bootstrap technique for estimating confidence interval of the outputs.

The presented method can be employed to quantify the uncertainties in parameters and predictions arising from the choice of data used for the model calibration, while other sources of uncertainty are assumed to be minimized (through trial and error procedure) during the calibration of the model.

The Bootstrap method that works under joint stochastic-deterministic modeling framework, and deals with the uncertainty associated with the model input as well as the parameters that results in an uncertainty band around the deterministic simulations, is to be considered in evaluating the uncertainty associated with various hybrid models such as Wavelet-ANFIS.

When sufficiently large sets of examples (training patterns) are available, the sampling variability in weights can be approximated by bootstraps (Stone, 1974). The bootstrap is a computational procedure that uses intensive resampling with replacement, in order to reduce uncertainty (Efron and Tibshirani, 1993).

3.3.10. Bootstrap approach for uncertainity analysis:-

Bootstrap assumes that the training dataset is a representation of the population, and multiple realizations of the population can be simulated from a single dataset. This is done by repeated 'sampling with replacement' of the original dataset of size of N, to obtain B bootstraps datasets, each with size of N. Each bootstrap dataset contains a different data, resulting in B neural networks. A model $\hat{f}(x)$ is fitted to each of the generated bootstrap datasets and bootstrapping estimate $\hat{f}_{bootstrap}(x)$ is calculated as the mean of each model:

equation(3.21)

$$\hat{f}_{bootstrap}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}(x) \tag{3.21}$$

A 100 × (1 − α)% confidence interval (CI) covering the range of the model predictions can be estimated by the following equation (Efron and Tibshirani, 1993):

$$\mathrm{CI} = \hat{f}_{bootstrap}(x) \pm t_{n-p}^{\alpha/2}\sigma(x) \quad (3.22)$$

where , σ(x) is the standard deviation of S bootstrapped estimates , $t_{n-p}^{\alpha/2}$ is the α/2 percentile for the Student t distribution with n − p degrees of freedom, n is the total number of observation and p is the total of parameters in the model. Bootstrapping approach of generating different models and aggregating them to produce an estimate has been found to increase the accuracy of model (Breiman, 1996).

3.3.11. Multiple linear regression (MLR)

Multiple linear regression attempts to model the relationship among two or more independent variables and a dependent variable by fitting a liniear equation to the following form:

$$Y= a+ b_1x_1 + b_2x_2 + \ldots +b_nx_n \quad (3.23)$$

Where Y is the dependent variable, *a* is a constant and b_1 to b_n are multipliers for x_1 to x_n independent variable. Constant and multipliers are estimated through minimizing the sums of square of deviation between each data point and the regression line.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 GENERAL

To model air temperature, monthly averaged observed meteorological parameters were collected from the seven different meteorological stations in and around Shimoga region of same climatic zone from January 2001 to December 2011. In the next stage to examine the influence of air pollution on air temperature estimation, combination of meteorological and air pollution parameters are used as input for a selective station Bhadra (B.R.Project). In the present work meteorological parameters and air pollution parameters are used to analyse the performance of computing techniques like ANFIS, SVM, hybrid Wavelet-ANFIS and hybrid Wavelet-SVM to estimate air temperature. Methodologies of these techniques were briefly explained in the Chapter 3. Further, the collected data was randomly divided into two set, with 75% for training and remaining for testing. In a data set every 4^{th}, 8^{th}, 12^{th} and so on is selected for testing.

To study the potential and applicability of the proposed approach, statistical comparison of measured and estimated values of training and testing data were done. The Root Mean square error (RMSE) between desired output and network estimated outputs were calculated using Eq. 4.1. The Correlation Coefficients (CC) and Scatter Index (SI) between target output and network estimated output is calculated by using Eq. 4.2 and Eq. 4.3. The CC and SI are dimensionless parameter whereas RMSE is in $^{\circ}$C. If the RMSE is lowest, CC is near or equal to one and SI is close to zero is considered as the best model for work.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(o_i - p_i\right)^2} \tag{4.1}$$

$$CC = \frac{\sum_{i=1}^{n} p_i o_i}{\sqrt{\sum_{i=1}^{n} p_i^2 \sum_{i=1}^{n} o_i^2}} \tag{4.2}$$

$$SI = \frac{RMSE}{\overline{o_i}} \tag{4.3}$$

Where, O_i and P_i are observed and modeled air temperature respectively, n is the number of data set used and $\bar{P}_i$ & $\bar{O}_i$ are average modeled air temperature and observed air temperature respectively.

4.2. PERFORMANCE ANALYSIS OF MODELS FOR SEVEN STATIONS (METEOROLOGICAL PARAMETERS ONLY)

4.2.1 PRINCIPAL COMPONENT ANALYSIS

PC analysis was carried out using statistiXL software (www.statistixl.com/downloads/ files). Analysis part consisted of calculation of principal components for the given set of data. In the present case, four parameters were considered for analysis for all the seven stations namely rainfall, wind speed, relative humidity and sunshine hour. The principal components were calculated for all these parameters. The results of the PC analysis are tabulated in Tables 4.1 and 4.2.

It can be observed in Table 4.1 that the Eigen values and the percentage of variance are greater than one and 10% respectively for PC1, PC2, PC3 and PC4. Hence, further PCA loading analysis was carried out for these four principal components. The components with loading greater than or equal to 70% were considered to be influencing the output i.e., modeling of air temperature. The results of the loadings calculated for the four principal components are shown in Table 4.2 which clearly shows that all the parameter has a loading more than 70% and hence the influence of that parameter on estimation of air temperature was more. Henceforth, four input parameters were considered for the further analysis using various soft computing techniques.

Table 4.1 Eigen values for all PCA and percentage of variance

Sl. No	Station	Value	PC1	PC2	PC3	PC4
1	**SHIMOGA**	Eigen value	2.415	0.969	0.389	0.226
		% of Var.	**60.386**	**24.234**	9.721	5.658
		Cum. %	60.386	84.620	94.342	100.000
2	**HONNALI**	Eigen value	1.589	1.280	0.632	0.499
		% of Var.	**39.717**	**32.010**	**15.804**	**12.469**
		Cum. %	39.717	71.728	87.531	100.000
3	**B.R.PROJECT**	Eigen value	2.038	0.982	0.761	0.219
		% of Var.	**50.947**	**24.553**	**19.023**	**5.478**
		Cum. %	50.947	75.499	94.522	100.000
4	**DAVANGERE**	Eigen value	2.002	0.835	0.613	0.551
		% of Var.	**50.038**	**20.865**	**15.320**	**13.777**
		Cum. %	50.038	70.903	86.223	100.000
5	**LINGANAMAKKI**	Eigen value	2.278	0.986	0.465	0.271
		% of Var.	**56.962**	**24.641**	**11.614**	**6.784**
		Cum. %	56.962	81.602	93.216	100.000
6	**HIRIYUR**	Eigen value	1.866	0.996	0.711	0.427
		% of Var.	**46.660**	**24.892**	**17.782**	**10.666**
		Cum. %	46.660	71.552	89.334	100.000
7	**RAIPURA**	Eigen value	1.751	0.936	0.841	0.472
		% of Var.	**43.787**	**23.400**	**21.025**	**11.788**
		Cum. %	43.787	67.187	88.212	100.000

Table 4.2 Principal Component loadings

Sl.No	Station	Value	PC1	PC2	PC3	PC4
1	**SHIMOGA**	Rain fall	**0.860**	-0.055	0.492	-0.124
		Wind speed	0.269	**0.961**	-0.006	0.062
		Relative humidity	**0.902**	-0.204	-0.104	0.366
		Sunshine hour	**-0.889**	0.031	0.368	0.270
2	**HONNALI**	Rain fall	**0.830**	0.207	0.102	0.507
		Wind speed	-0.097	**0.846**	-0.525	-0.001
		Relative humidity	-0.498	0.659	0.561	0.062
		Sunshine hour	**-0.801**	-0.297	-0.179	0.487
3	**B.R.PROJECT**	Rain fall	**0.905**	0.068	-0.256	0.334
		Wind speed	-0.289	**0.913**	-0.286	-0.020
		Relative humidity	0.577	0.378	**0.724**	-0.026
		Sunshine hour	**-0.896**	0.018	0.300	0.327
4	**DAVANGERE**	Rain fall	**-0.779**	-0.160	0.046	-0.604
		Wind speed	-0.535	**0.843**	0.030	0.045
		Relative humidity	**-0.744**	-0.249	0.517	0.343
		Sunshine hour	**0.745**	0.189	0.586	-0.257
5	**LINGANAMAKKI**	Rain fall	**0.901**	0.124	0.077	-0.409
		Wind speed	-0.177	**0.984**	0.002	0.036
		Relative humidity	**0.860**	0.029	0.428	0.275
		Sunshine hour	**-0.834**	-0.045	0.524	-0.165
6	**HIRIYUR**	Rain fall	**-0.742**	-0.007	0.609	0.280
		Wind speed	0.122	**0.990**	0.076	-0.014
		Relative humidity	**-0.748**	0.122	-0.578	0.303
		Sunshine hour	**0.861**	-0.041	0.012	0.506

7	**RAIPURA**	Rain fall	-0.617	-0.356	**0.652**	0.258
		Wind speed	-0.520	**0.817**	-0.012	0.250
		Relative humidity	-0.615	-0.376	**-0.644**	0.257
		Sunshine hour	**0.850**	-0.031	0.001	0.526

4.2.2. ANFIS model

In this study, monthly average observed meteorological data of 7 stations (132 months) were used. Among those 132 months of data of each parameter, 99 months data are used for training and 33 months of data are used for testing. Here, the original raw data of parameters like rainfall, wind speed, humidity and sunshine hour are used for the input of ANFIS model (Sugeno first order with 16 fuzzy rules and Gbell membership function) for both training and testing. Average air temperature data was used as output for both training and testing. Results are shown in Table 4.3. In case of ANFIS model testing; CC is less than 0.5 and also RMSE values are more than 6.6 which reveals its poor performance. Also in SI for testing, values are more than 0.2 which is beyond acceptable limit in terms of accuracy. When ANFIS model results are compared among other station Hiriyur station having better results in terms of CC, RMSE and SI as shown in Fig 4.1a and Fig 4.1b. For ANFIS, station Shimoga and Honnali having less CC and high RMSE value, this may be due to the higher degree of nonlinearity and presence of noisy data.

In Fig 4.1a,the scatter points are sparsely located from 45^{o} line. It reveals the poor agreement between observed and estimated air temperature. Also it was observed that for lower air temperature, model values are relatively closer to 45^{o} line. However, widely deviated model data were found during higher air temperature.

Atime series plot is shown in Fig 4.1b. Here also poor agreement between estimated and measured air temperature are observed throughout the testing period. Also during summer months (April to May), there is a huge error between observed and estimated is identified. This can be termed as highly over estimated. For a lower, temperature region estimated temperature was found to be somewhat nearer to measured air temperature.

Table 4.3 ANFIS model performance

SL.NO	STATIONS	ANFIS model		
		CC	**RMSE(°C)**	**SI**
1	**SHIMOGA**	0.097	20.980	0.810
2	**HONNALI**	0.438	85.200	3.310
3	**B.R.PROJECT**	0.280	6.640	0.260
4	**DAVANGERE**	0.520	10.160	0.380
5	**LINGANAMAKKI**	0.160	31.890	1.200
6	**HIRIYUR**	**0.409**	**6.700**	**0.270**
7	**RAIPURA**	0.409	6.705	0.270

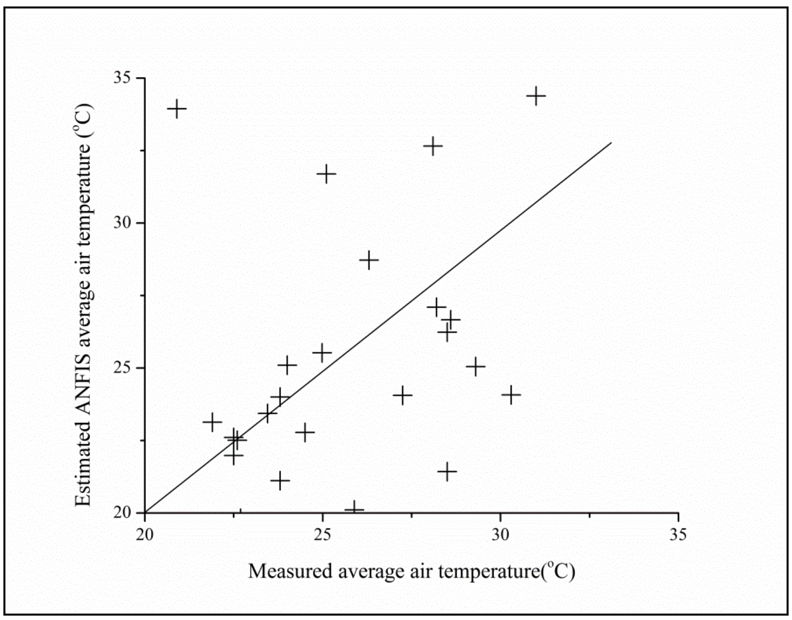


Fig. 4.1a Scatter plot of ANFIS model performance for Hiriyur station

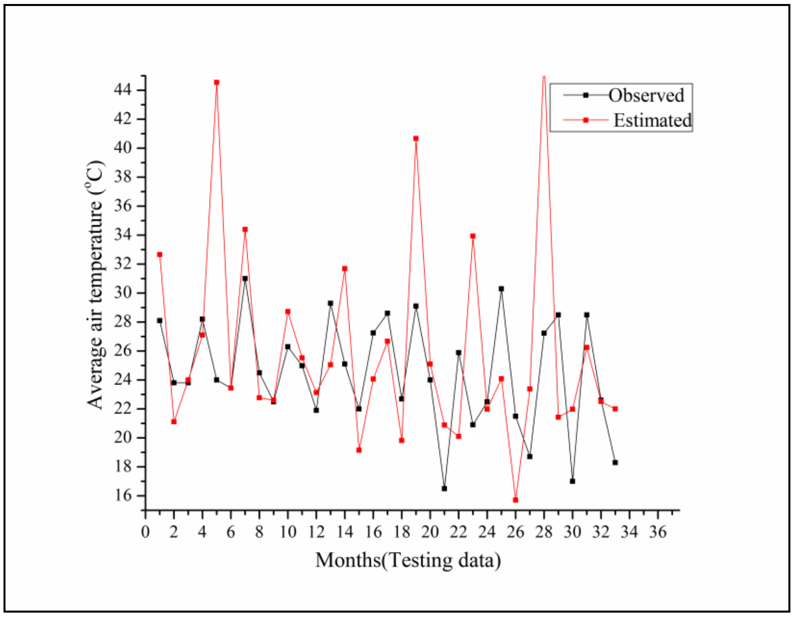


Fig. 4.1b ANFIS model performance for modeled air temperature of test data for Hiriyur station

4.2.3. SVM model

Statistical performance indices computed using the modeled and observed values of testing data for the SVM models are presented in Table.4.4. The performance of SVM depends on the good setting of SVM and kernel parameters. In developing SVM models, initially parameters are randomly selected by coarse grain search (i.e. for C=100,200, 300….2000; ε = 0.5, 1…2; and d = 1,2,...6) to identify the near optimal values, and then a fine grain search (i.e. for C=50,100,500, ….5000; ε = 0.000001,…2; and d =1,2....6) is done to identify the final optimal values. The final optimum values (i.e. for nsv=99; C=50; ε = 0.1; and d = 0.5) of SVM and Polynomial as kernel function. In case of SVM model for seven meteorological stations; CC value was found less than 0.5, with a RMSE more than 2.310 for testing and SI values are less than 0.1, which shows inferior performance. In Fig.4.2a, the scatter points are sparsely located from 45^{o} line. It reveals the high disagreement between observed and estimated air temperature. Also, it was observed that for lower air temperature, model values are relatively closer to 45^{o} line. However, far and wide deviated model data were found during higher air temperature.

Again time series plots are shown in Fig 4.2b. Here also, poor agreement between model and observed air temperature are observed throughout the testing period. Also during summer months (April to May), a huge deviation between observed and estimated is identified which can be termed as highly over estimated. For a lower temperature region estimated temperature is found to be somewhat nearer to measured air temperature.

Table 4.4 SVM model performance

SL.NO	STATIONS	SVM model		
		CC	RMSE(°C)	SI
1	**SHIMOGA**	0.180	2.720	0.100
2	**HONNALI**	**0.530**	**2.310**	**0.080**
3	**B.R.PROJECT**	0.010	4.470	0.180
4	**DAVANGERE**	-0.200	3.290	0.100
5	**LINGANAMAKKI**	-0.030	3.010	0.110
6	**HIRIYUR**	-0.070	3.840	0.140
7	**RAIPURA**	0.310	3.630	0.140

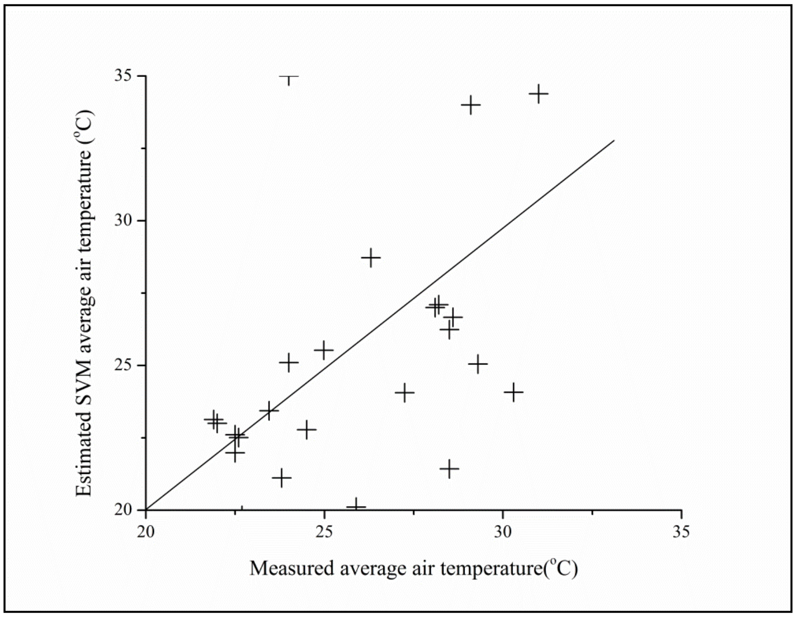


Fig. 4.2a Scatter plot of SVM model performance for Honnali station

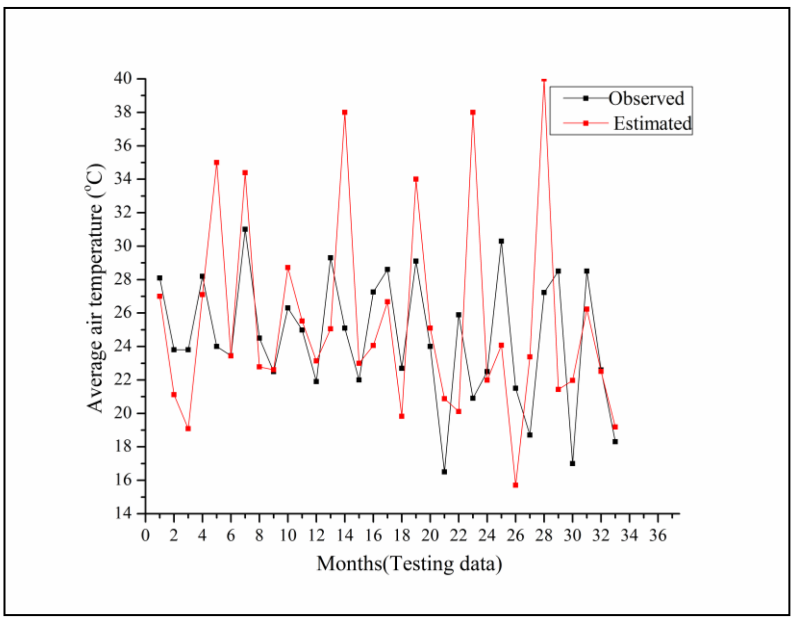


Fig. 4.2b SVM model performance for modeled air temperature of test data for Honnali station

4.2.4 Architecture of hybrid Wavelet-ANFIS model and performance of model

Monthly average temperature data were used for modeling. Input data (Rainfall, Humidity, Wind Speed and Sunshine hour) sets were obtained using DWT for ANFIS modeling. For the input, original time series data was decomposed three level sub-time series by DWT and Approximation of the time series was also obtained. Average temperature data was used for the output of the model. Hybrid modeling process consists of two parts. One of the parts is training and the other one is testing. The model was implemented by using MATLAB R2012a computer programming. 132 monthly data were used for modeling. 99 monthly data were selected for training process which consists of approximation series A as input layer and monthly average temperature data as output layer. Training process was performed with different epoch number, various membership function, number of fuzzy rules based on minimizing RMSE for optimized structure.

In the testing procedure, the 33 monthly data which were having same character with the training process were utilized in the ANFIS models obtained from the training procedure. The best model depends on the type and number of input variable, minimization of optimizing internal parameters such as number of epoch, number of fuzzy rules, type and number of membership function and minimization of Root Mean Squared Errors (RMSE) of the models. The agreements between the observed average temperature values and the estimated values using the hybrid model WNF. Further evaluation on the performance of the model has been done by comparing with a different model's performance like single ANFIS, Single SVM and Hybrid Wavelet-SVM.

Here the performances of hybrid Wavelet - ANFIS model are tabulated in the Table.4.5, which shows hybrid Wavelet - ANFIS works better. All the station CC values are more than 0.9, RMSE was less than 1.3 and SI is less than 0.03 which are within acceptable limit. Out of seven stations, Linganamakki station is having best performance of CC is 0.954, RMSE and SI is 0.710 and 0.27 respectively. The better performance of the model may be due to the combination of Fuzzy Logic, Neural Network and Wavelet Transformation.

Table 4.5 Results of hybrid model (Wavelet - ANFIS) of seven stations for air temperature estimation

SL.NO	STATIONS	Hybrid (Wavelet - ANFIS) model		
		CC	RMSE ($^{\circ}$C)	SI
1	**SHIMOGA**	0.939	0.960	0.037
2	**HONNALI**	0.958	0.852	0.033
3	**B.R.PROJECT**	0.952	1.362	0.055
4	**DAVANGERE**	0.950	0.942	0.035
5	**LINGANAMAKKI**	0.954	0.710	0.027
6	**HIRIYUR**	0.942	1.012	0.038
7	**RAIPURA**	0.950	1.339	0.054

For all the seven stations shown in Figs. 4.3a, 4.4a, 4.5a, 4.6a, 4.7a, 4.8a, and 4.9a, it is observed that the scatter points are uniformly distributed along 45° line. This exhibits the model's best performance over the wide range of observed air temperature measurements and the capability of the model to statistically estimate real time air temperature.

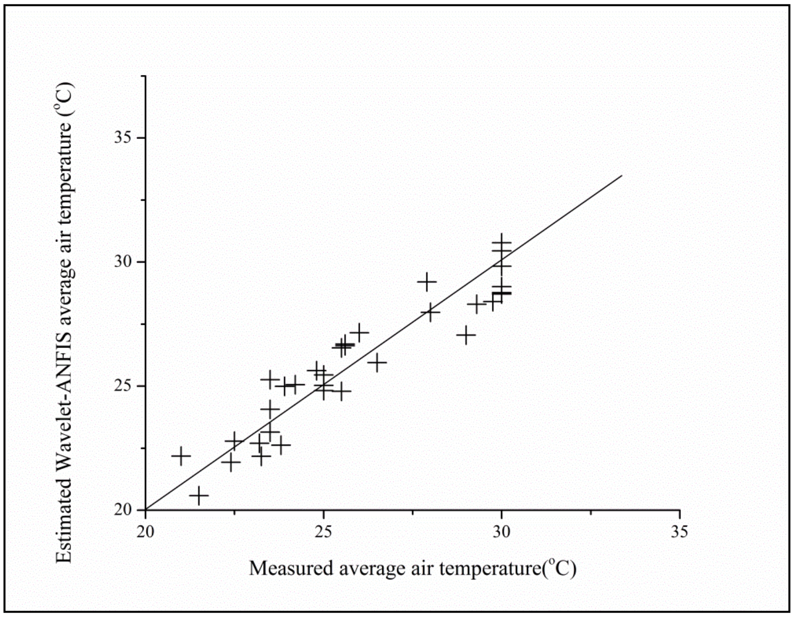


Fig. 4.3a Scatter plot of hybridized Wavelet-ANFIS model performance for Shimoga station

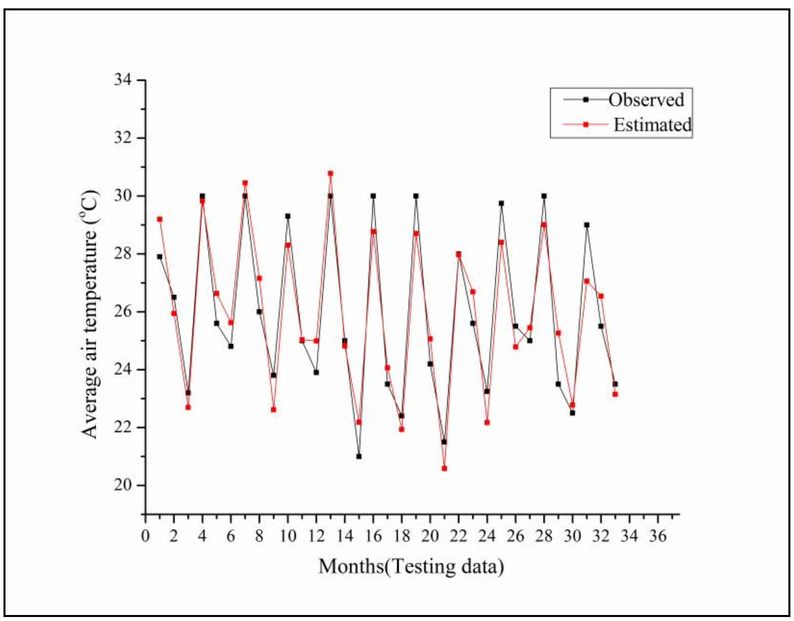


Fig. 4.3b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Shimoga station

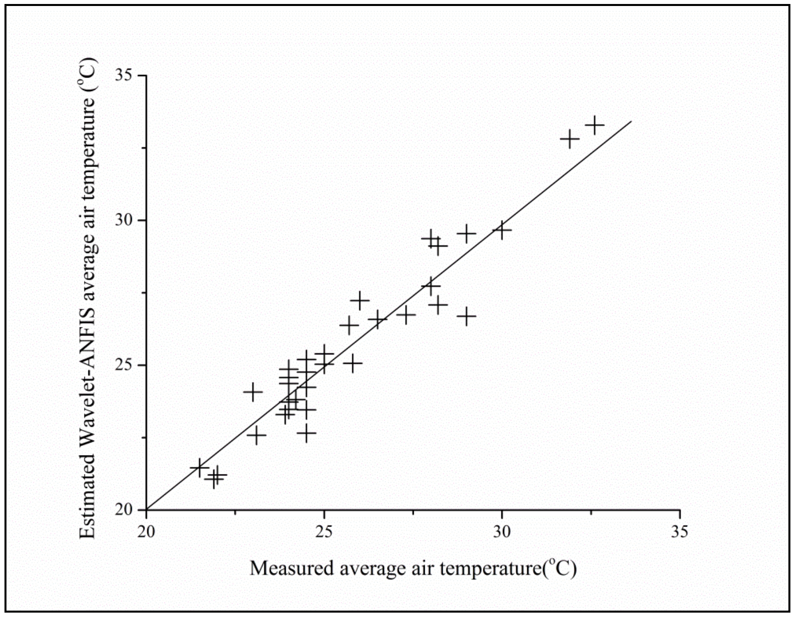


Fig. 4.4a Scatter plot of hybridized Wavelet-ANFIS model performance for Honnali station

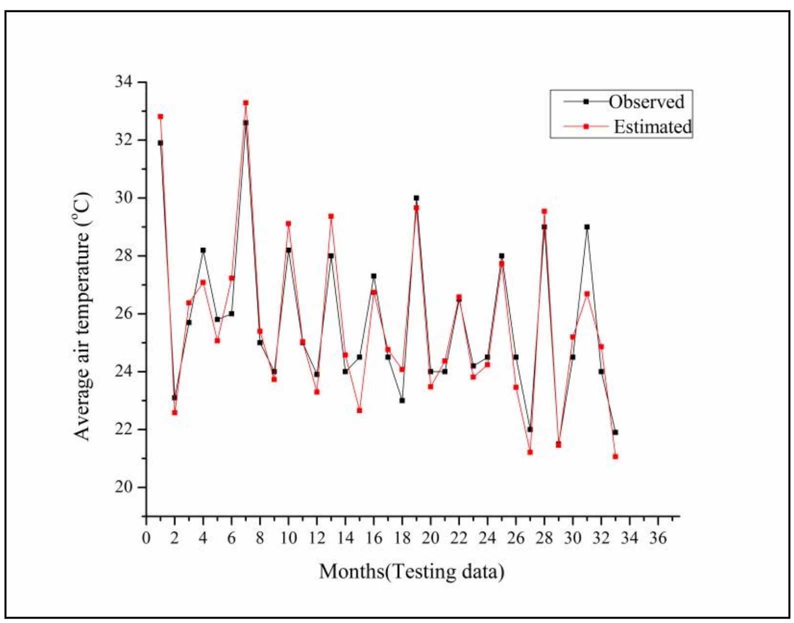


Fig. 4.4b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Honnali station

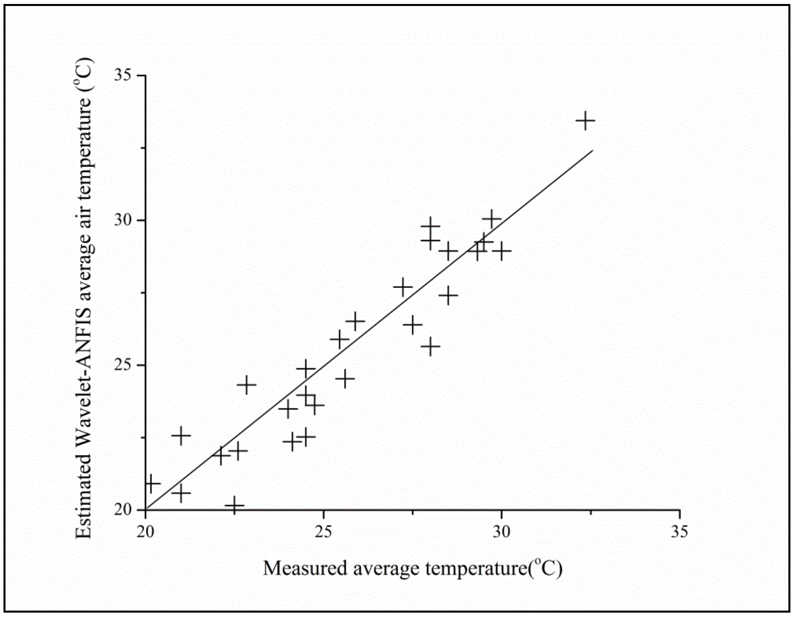


Fig. 4.5a Scatter plot of hybridized Wavelet-ANFIS model performance for B.R.Project station

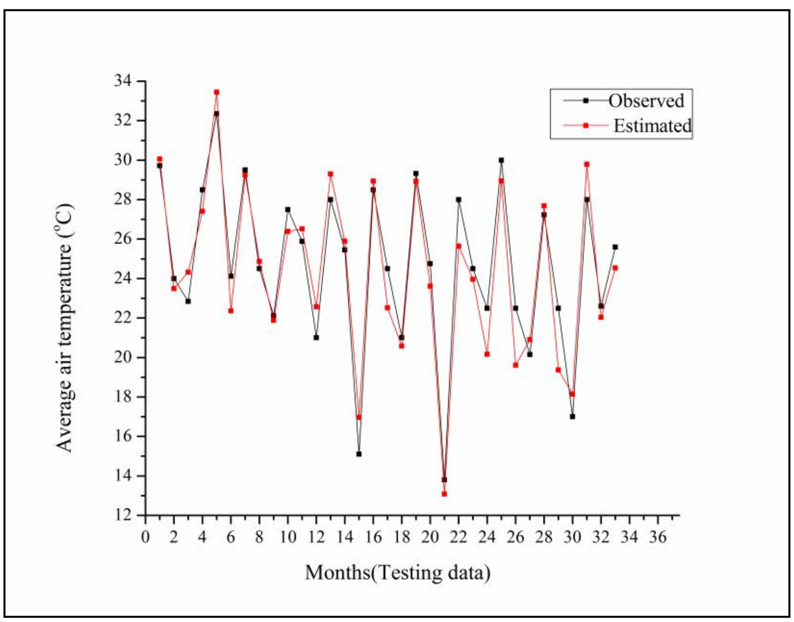


Fig. 4.5b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for B.R.Project station

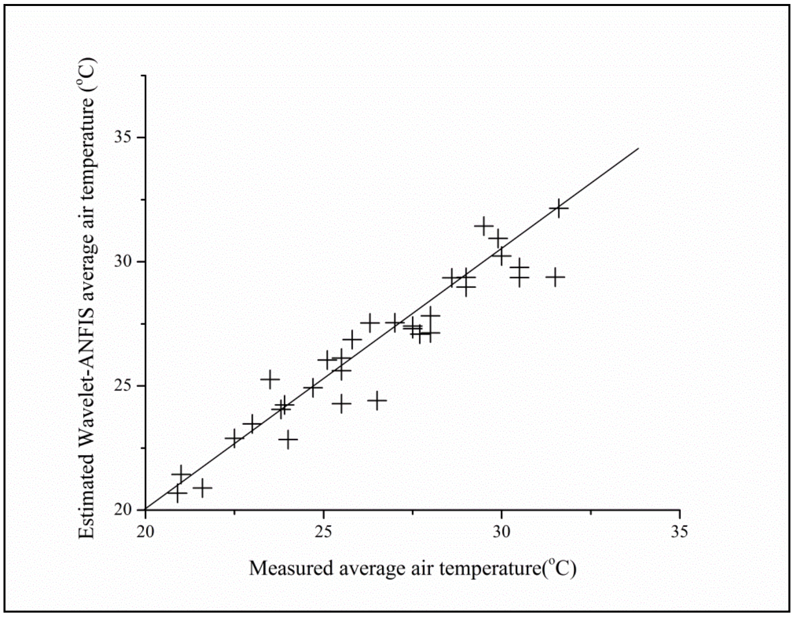


Fig. 4.6a Scatter plot of hybridized Wavelet-ANFIS model performance for Davanagere station

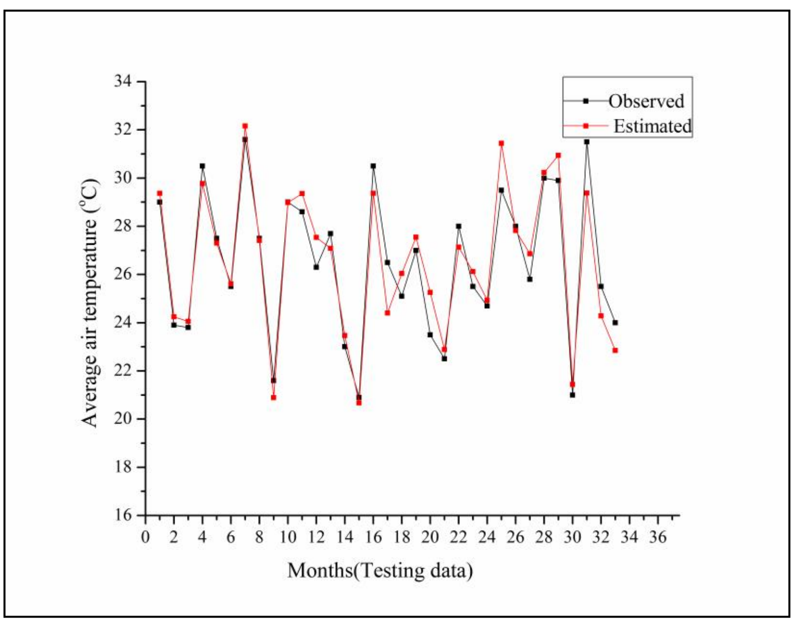


Fig. 4.6b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Davanagere station

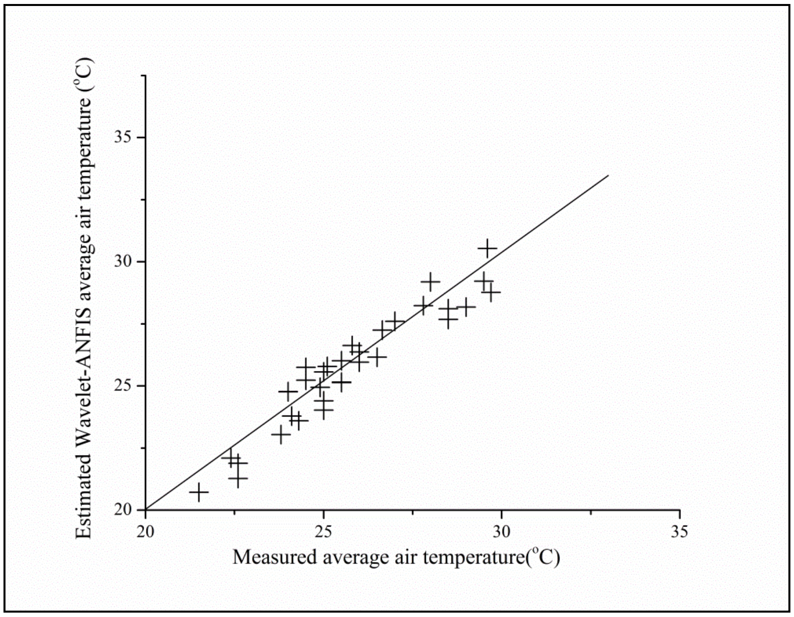


Fig. 4.7a Scatter plot of hybridized Wavelet-ANFIS model performance for Linganamakki station

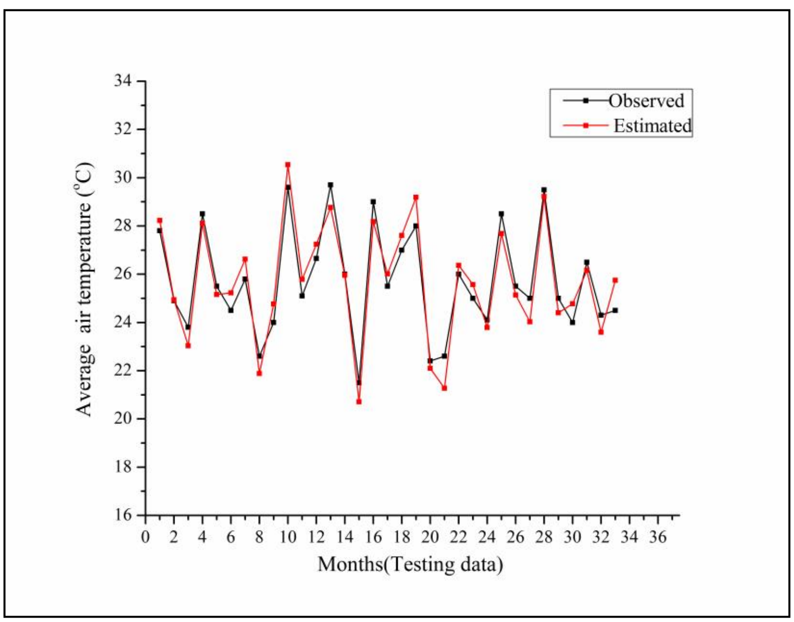


Fig. 4.7b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Linganamakki station

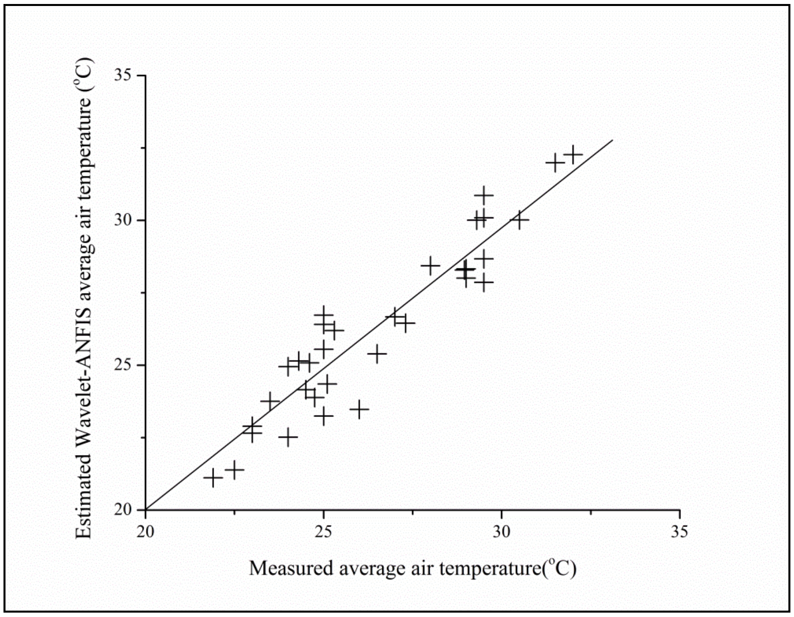


Fig. 4.8a Scatter plot of hybridized Wavelet-ANFIS model performance for Hiriyur station

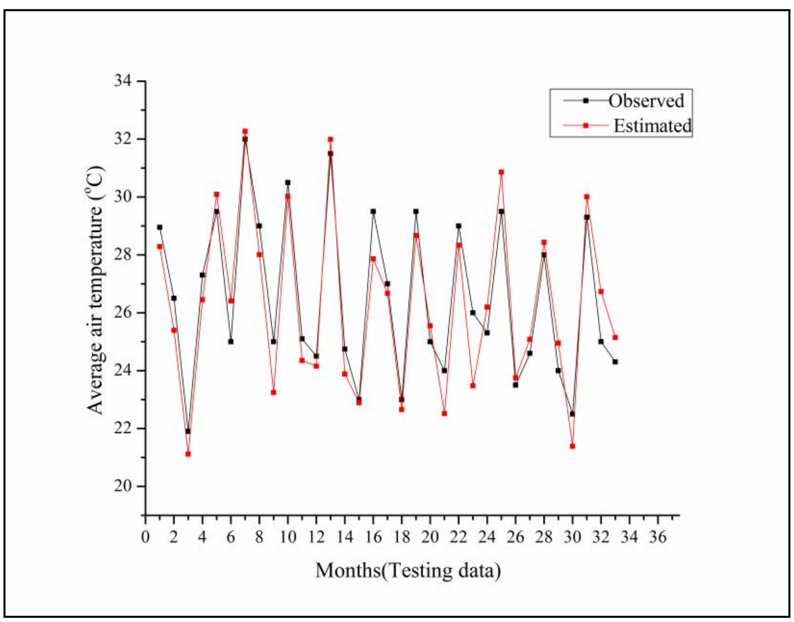


Fig. 4.8b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Hiriyur station

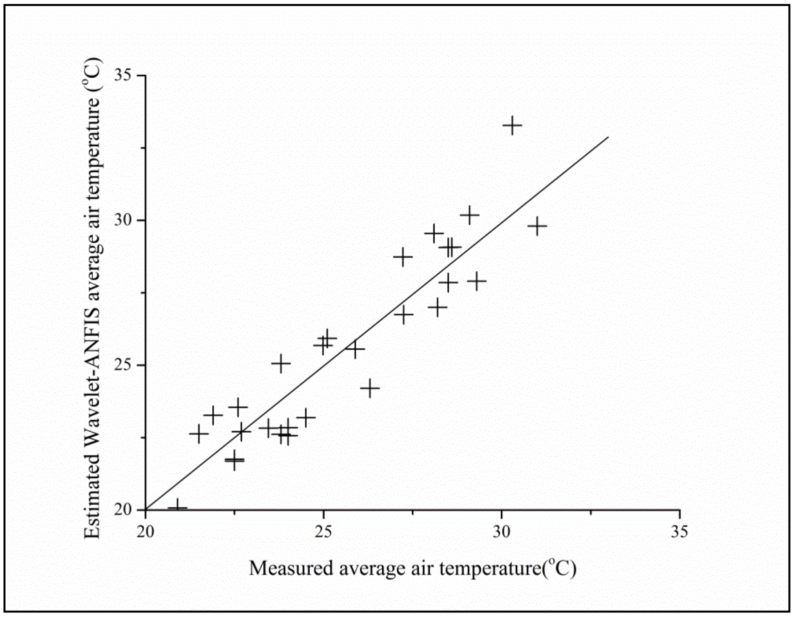


Fig. 4.9a Scatter plot of hybridized Wavelet-ANFIS model performance for Raipura station

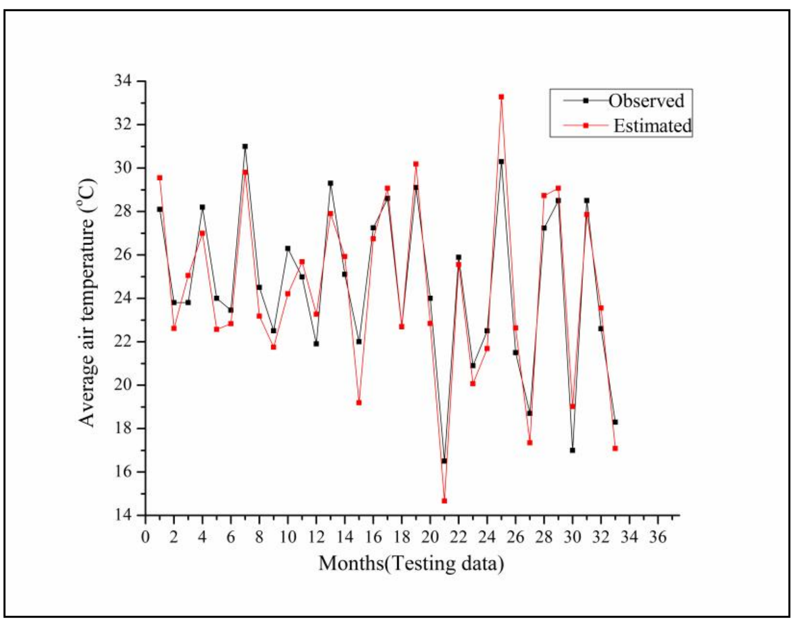


Fig.4.9b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Raipura station

The time series plots are shown in Figs. 4.3b, 4.4b, 4.5b, 4.6b, 4.7b, 4.8b, and 4.9b. For all the testing data set (33 observations per station), the error between measured and estimated air temperature was found less (<1°C), representing the potential of the proposed hybrid technique to efficiently model real time air temperature.

4.2.5. Hybrid Wavelet- SVM model

After analyzing single SVM model, for above mentioned seven stations hybrid Wavelet-SVM model were also tested. Original 132 months data are decomposed by Db3 and level-3 function. The final optimum values (i.e. for nsv=99; C=50; ε = 0.1; and d = 0.5) of SVM and Polynomial as kernel function are used. The wavelet-SVM model with polynomial kernel function shows generalization performance with low CC values, high RMSE and SI values. When compared to other station results, Davanagere station was showing better results among other station with CC 0.30, RMSE 2.68, and SI 0.1 (Table 4.6).

Table 4.6 Results of Hybrid model (Wavelet-SVM) of seven stations for air temperature estimation

SL.NO	STATIONS	Hybrid Wavelet - SVM model		
		CC	RMSE	SI
1	**SHIMOGA**	0.120	2.900	0.110
2	**HONNALI**	0.180	2.720	0.100
3	**B.R.PROJECT**	0.110	4.320	0.170
4	**DAVANGERE**	0.300	2.680	0.100
5	**LINGANAMAKKI**	0.090	2.270	0.080
6	**HIRIYUR**	-0.200	3.280	0.120
7	**RAIPURA**	0.160	3.720	0.150

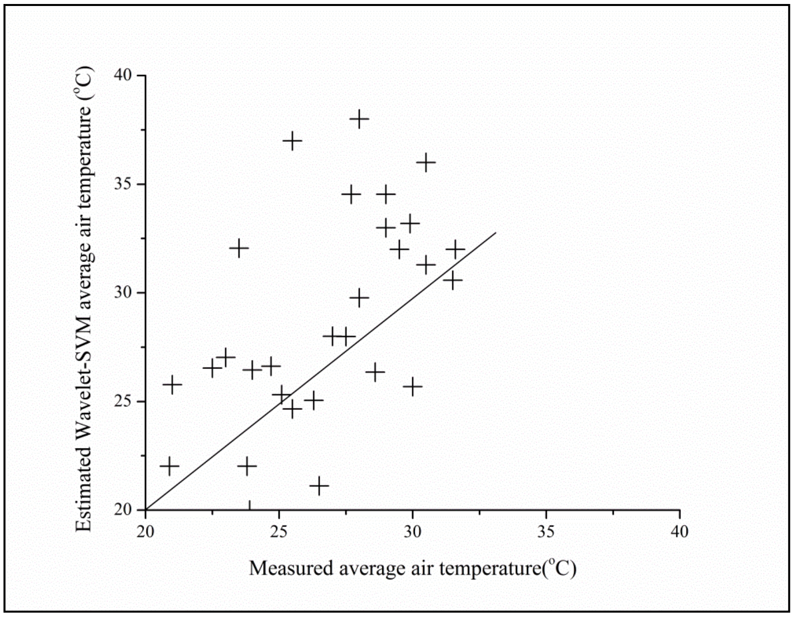


Fig. 4.10a Scatter plot of hybridized Wavelet-SVM model performance for Davanagere station

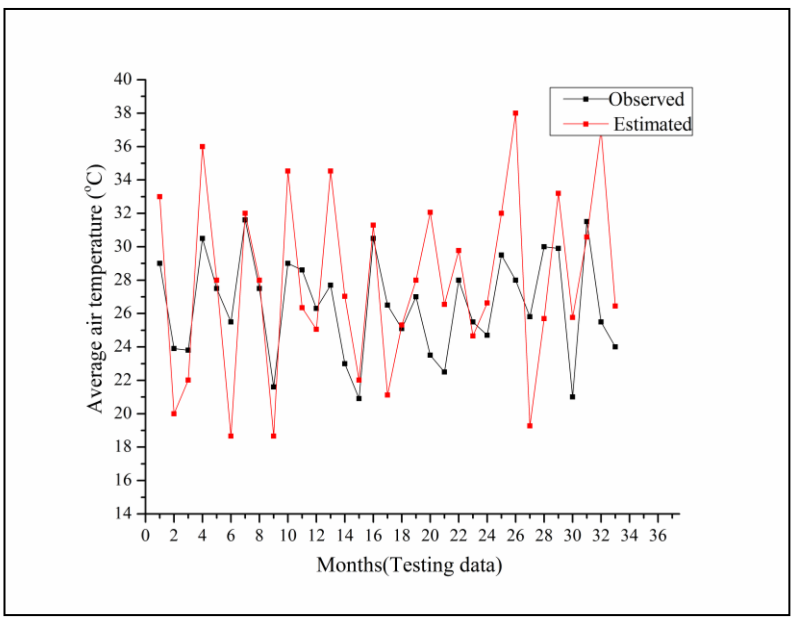


Fig. 4.10b Hybridized Wavelet-SVM model performance for modeled air temperature of test data for Davangere station

In Fig.4.10a, it is observed that along the 45° line, estimated data points are sparsely located. It reveals the poor performance of the model to estimate air temperature. However, far and wide deviated model data were found during lower and higher air temperature. Also it was observed that for middle range of air temperature, model values are relatively closer to 45^{o} line.

Time series plot is shown in Fig 4.10b.In this case also a huge disagreement between estimated and measured air temperature was observed throughout the testing period. Also during all period, there was a more error gap ($>2^{\circ}C$) between measured and estimated data. This reveals that models over estimated air temperature.

4.2.6. Comparison of ANFIS and SVM model

In SVM model, the estimation of air temperature was comparatively better with Honnali station data, showing higher CC of 0.53 and lesser RMSE of 2.31 and SI of 0.08. In ANFIS model, Hiriyur showed higher CC of 0.409 with RMSE and SI of 6.7 and 0.27 respectively when compared to other seven stations. Further, as per Table 4.3 and Table 4.4 the comparison of SVM and ANFIS models showed that SVM model performed better. This may be due to SVM model works on a principle of structural risk minimization. For ANFIS, station Shimoga and Honnali having a lower CC and high RMSE value, this may be due to higher degree of nonlinearity and presence of noisy data.

4.2.7. Comparison of ANFIS model and Hybrid Wavelet-ANFIS model

Among hybridized Wavelet-ANFIS and ANFIS model, hybridized Wavelet-ANFIS showed good results in terms of statistical performance indices like a CC, RMSE and SI for observed and estimated data. By comparing Table 4.3 and Table 4.5, shows the comparative study of two proposed models and the table clearly depicts that hybridized Wavelet-ANFIS model outperforms ANFIS model with higher CC of 0.9 when compared to ANFIS model which is 0.5. Hybridized wavelet-ANFIS model has RMSE of 1.3 which is low; whereas ANFIS model has RMSE of 6.6 which is more. Also, the SI is less than 0.03 for the testing data of hybridized Wavelet-ANFIS model, but SI is more than 0.03 for ANFIS model. Wavelet transforms reduces the noise in the non-stationary data series and makes the periodic information more readily understandable by the model. The wavelet transform has given the strength of generalization to neural network and specialization to Sugeno inference fuzzy logic for training the non-stationary data and predicting the output.

4.2.8. Comparison of SVM model and Hybrid Wavelet-SVM model:

Within the SVM model and Hybrid Wavelet-SVM model, hybridized Wavelet-SVM shows comparatively good results in terms of statistical measures like CC, RMSE and SI for observed and estimated data of seven stations. By comparing Table 4.4 and Table 4.6, shows the comparative study of proposed SVM model and Hybrid Wavelet-SVM model and the table clearly shows that hybridized Wavelet-SVM model performs comparatively better than SVM model. Out of seven stations, Honnali station showed better performance with SVM model with a CC of 0.53, RMSE and SI of 2.31 and 0.080 respectively.

4.2.9. Comparison of Hybrid model (Wavelet-ANFIS) and Hybrid model (Wavelet-SVM) for seven stations

Comparison of performance analysis of two hybrid models like hybrid Wavelet-ANFIS model and hybrid Wavelet-SVM model are shown in Table 4.5 and Table 4.6 . From the table it shows that the hybrid Wavelet-ANFIS model works better compared to hybrid Wavelet-SVM model. Performances of all the seven station are best in terms hybrid Wavelet-ANFIS model with CC, which is more than 0.9 but in hybrid Wavelet-SVM model which is less than 0.3. In terms of RMSE and SI, hybrid Wavelet-ANFIS model is having less than 1.4 and 0.03 respectively and hybrid Wavelet- SVM model is showing more than 2.2 and less than 0.1 respectively.

The better performance of the Wavelet-ANFIS model may be due to the combination of Fuzzy Logic, Neural Network and Wavelet Transformation, Which are complimentary to each other. In ANFIS fuzzy logic handles uncertainty of vagueness and NN is accommodating Non-Linearity. Again non stationary series decomposed to stationary by wavelet helps in reducing variance. Lots of coefficients are generated by wavelet which is better handled by NN component compare to SVM as it regressed only. In performance of SVM, mainly depended on the selection of kernel function and parameter σ, ε, and C.

4.2.10 Performance evaluations of single ANFIS, single SVM, hybrid Wavelet-ANFIS model and hybrid Wavelet- SVM model

The performance of the single and hybrid models were discussed in the previous sections. In this part a comparison of all the models with all the seven stations has been done to know the best model that could estimate air temperature accurately. The Tables 4.3 to 4.6 give the performance of single models and hybrid models. Using Matlab R2012a software all the models were run in assembled desk top with Intel® core ™ i3-3210 CPU @ 3.20 GHz and 4 GB RAM and 32 bit windows 7 operating system.

In terms of performance, hybrid Wavelet- ANFIS model works better compared to single ANFIS, single SVM and hybrid Wavelet- SVM model where Wavelet analysis for nonlinear data, reduces the noise and makes the periodic information more readily understandable by the model. The wavelet transform has given the strength of generalization to neural network and specialization to Sugeno inference fuzzy logic for training the non-stationary data and predicting the output. Comparing of all models with all the seven stations it is clear from the previous Tables 4.3 to 4.6, that hybrid Wavelet-ANFIS model (Gbell membership) for station Linganamakki showed best performance with CC of 0.954, RMSE and SI 0.21 and 0.027 respectively.

Wavelet analysis helps in extracting more information that can hidden in the frequency component of the signal. So wavelet which represents the data in both frequency and time. ANFIS is a combination of neural network and fuzzy, where optimal fuzzy membership functions are created using neural networks.

For non-stationary data series, pre-processing of data by wavelet transformation is effective by removing noise, spikes, irregularity etc. For nonlinearity of data series, NN is efficient and better than others in input-output mapping. For handling uncertainty in data and processes, fuzzy logic working better than others.

Hence hybridization of W-NN-Fuzzy is providing robust model in these scenarios by eliminating limitation of individual techniques.

4.2.11 Development of Regional model for air temperature

As there are lacks of data or limited meteorological data available, it is proposed to develop regional model which can be used in similar climatic region.

Following are the multiple linear regression equations developed are present below for all the stations based on sensitivity analysis in and around Shimoga district, Karnataka, Shimoga(13° 27' to 14° 39'' North latitude and 74° 38´to 75° 45′ East longitudes).

Davangere:-
$Y=0.004168X_1+1.965408X_2+0.003107X_3+0.005939X_4+24.51542$
Shimoga:-
$Y= 0.005604133X_1+ 4.332163898X_2 - 0.088007766X_3 - 0.00226X_4 + 30.23344$
Honnali:-
$Y= 0.004476 X_1 + 3.212013X_2 - 0.080258X_3 - 0.009377X_4 + 16.52761$
Linganamakki:-
$Y= -0.00056X_1 - 0.01508 X_2 - 0.01392 X_3 + 0.002759X_4 + 26.70307$
B.R.Project:-
$Y= 0.01074X_1 +18.37954X_2 - 0.1248X_3 + 0.004862X_4 + 31.69456$
Raipura:-
$Y= 0.011834X_1 + 5.030336X_2 - 0.03993X_3 + 0.019132X_4 + 20.52125$
Hiriyur:-
$Y= 0.005969X_1 -0.47363X_2 - 0.03993X_3 -0.12399X_4 + 34.20859$
ALL Station:-
Average $Y= 0.011041X_1 + 1.541011X_2 - 0.21976X_3+ 0.00533X_4 + 39.08518$

Where X_1= Rainfall, X_2= Wind speed, X_3=Humidity, X_4=Sunshine hour and Y=Air temperature.

4.2.12 Prediction of air temperature for one season ahead using hybrid models

To examine the potential and applicability of best hybrid model and suggestion to decision maker, the potential of coupled Wavelet-ANFIS models in comparison with Wavelet-SVM models for 3 months ahead (one season) air temperature forecasting has been investigated in this study. The work is an extension of modeling of air temperature. Here we used Wavelet decomposed Db3 level 3 data like rainfall, wind speed, humidity and sunshine as a input for prediction of air temperature for all the seven stations(Sugeno first order with 16 fuzzy rules and Gbell membership function).Performances of the two models are tabulated in the Table 4.7 and Table 4.8. Out of two hybrid model Wavelet – ANFIS model outperformed the hybrid model Wavelet – SVM. Hybrid Wavelet – ANFIS model performance is better in predicting three month ahead (one season) air temperature. Out of Seven stations for Hiriyur station three month ahead (one season) prediction is working better with CC of 0.913 and low RMSE and SI of 1.340 and 0.051 respectively represented in scatter and line diagram(Fig 4.11a and Fig 4.11b), which is acceptable in terms of performance.

Table 4.7 Results of Hybrid model (Wavelet-ANFIS) of seven stations for air temperature Prediction (One season ahead)

Sl. NO	STATIONS	Hybrid (Wavelet-ANFIS)model		
		One season ahead(T+3)		
		CC	RMSE	SI
1	SHIMOGA	0.731	1.498	0.058
2	HONNALI	0.860	1.531	0.059
3	B.R.PROJECT	0.626	2.470	0.102
4	DAVANAGERE	0.802	1.514	0.057
5	LINGANAMAKKI	0.735	1.386	0.054
6	HIRIYUR	0.913	1.342	0.051
7	RAIPURA	0.636	2.570	0.106

Table 4.8 Results of Hybrid model (Wavelet- SVM) of seven stations for air temperature Prediction (One season ahead)

Sl. NO	STATIONS	Hybrid (Wavelet-SVM)model		
		One season ahead(T+3)		
		CC	RMSE	SI
1	SHIMOGA	0.246	2.163	0.060
2	HONNALI	0.264	2.734	0.107
3	B.R.PROJECT	0.059	2.502	0.103
4	DAVANAGERE	0.048	2.805	0.106
5	LINGANAMAKKI	0.208	1.332	0.052
6	HIRIYUR	0.169	3.317	0.126
7	RAIPURA	0.212	2.078	0.086

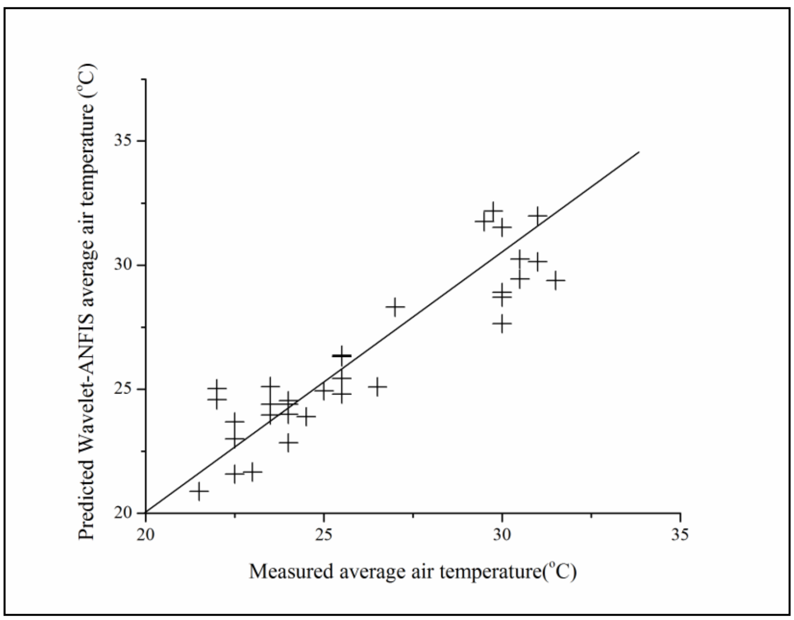


Fig. 4.11a Scatter plot of hybridized Wavelet-ANFIS model performance for Hiriyur station(One season ahead)

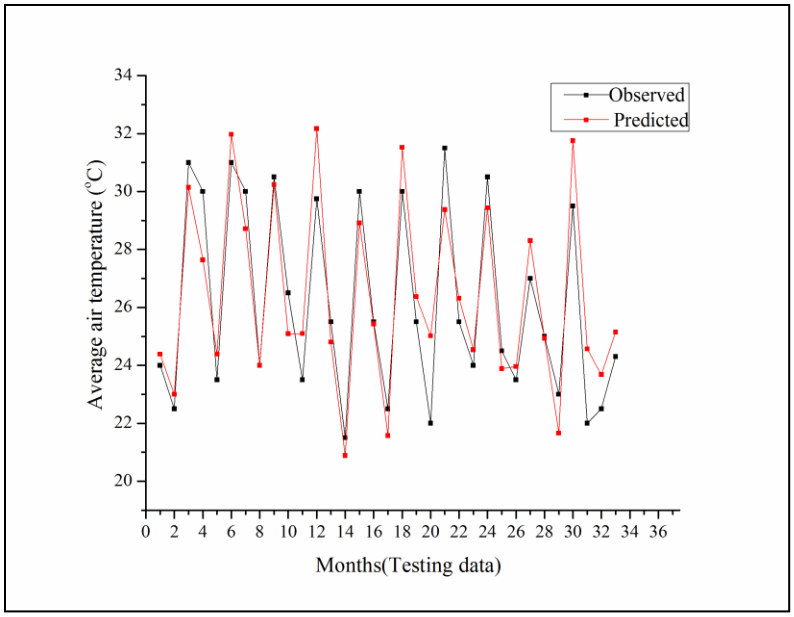


Fig. 4.11b Hybridized Wavelet-ANFIS model performance for Predicted air temperature of test data for Hiriyur station (One season ahead)

4.2.13 Uncertainty analysis in Prediction of air temperature using in Wavelet-Bootstrap-ANFIS (W-B-ANFIS)

Uncertainty analysis in prediction of air temperature has been carried out by ensembled of models using bootstrapped input-out pattern for Hiriyur station with one season ahead (3 month) lead time by W-B-ANFIS model. In the boot strap methods of uncertainty analysis, the total example set is divided into two set; training and testing sets. The testing set is kept aside, and random bootstrapping with replacement is performed on training set in order to evaluate the variation in performance with varying training sets. (Using same initial weights)

A sufficient large number of networks are trained using this procedure (50 in current study).All the networks so developed are evaluated on the testing set kept aside by computing various performance indices. The variation in the weights of the networks and the output of the network over the whole trained networks is a measure of uncertainty in the model parameters and prediction respectively that are coming from the variation in the training dataset. Given a test input vector(X), a pre specified prediction interval for output(Y) is an interval (L, U) such that P(L ≤ Y ≤ U)=C where C is typically taken 0.95, and the probability is computed over repeated random selection of the training set and repeated observation of Y, given the test input X.(Srivastav et al. 2007)

From the total available data for 11 years, 129 patterns (input-output pairs) were identified for the study and were splitting into training (97 sets) and validation (32 sets) data set. The validation datasets are corresponding to continuous time series out of 97 training samples, all were randomly bootstrapped every time. The optimal number of hidden neurons/fuzzy set/rules was found by trial and error. The number of fuzzy sets and rules was varied for only the first model, and for subsequent models, these number was fixed as that were found optimal for first model so as to maintain consistency. Similar strategy has been adopted for hidden neurons also. Earlier studies using bootstrap have revealed that variations in forecast due to changes in structure of architecture are small in comparison with those that arise from sample splitting (Lebarn and weigend, 1998). Thus the uncertainty arising from the architecture has not been considered in this study.

The performance of the models (50) has been evaluated using various statistical indices like RMSE and SI and results are presented in Table 4.9. It is observed that variation of RMSE for most of the models is found insignificant as coefficient of variance was found to be very low as appeared in the Table 4.9. The RMSE has a mean value of 3.29 with standard deviation 0.11. It is noted that more than 90% of the models produced RMSE within a band of ± 5% around mean value, reveals that the impact of training samples does not have a significant effect on model prediction. The mean value of SI was found 0.13 which is satisfactory.

Overall, the results indicate that the variation in training pattern does not have significant effect on the overall model performance. In the current study, the model architecture has been considered to be deterministic.

Table 4.9. **Summary of Statistics of the Performance Measures for 50 Models**

Performance Measures	Mean	Standard deviation	Variance	Coefficient of Skewness
RMSE	3.29	0.110	1.00×10^{-2}	0.92
SI	0.13	0.004	1.79×10^{-5}	0.92

4.3. PERFORMANCE OF MODEL FOR SINGLE BHADRA STATIONS (METEOROLOGICAL PARAMETERS AND AIR POLLUTIONS DATA)

4.3.1. ANFIS model

In this study, monthly averaged eight different parameters (rainfall, wind speed, humidity, sunshine hour, PM_{10}, SO_2, NO_2) of Bhadra station are used in different combination. To begin with, ANFIS model (Sugeno first order with 32 fuzzy rules and Gbell membership or Gauss membership function) was tested for five different parameters combination (Table 4.10) of input with 50 iterations and monthly averaged air temperature as an output parameter.

Table 4.10 Input parameter combination for model

No.	**Combination**
M1	Rainfall, Wind speed, Humidity, Sunshine hour.
M2	Rainfall, Wind speed, Humidity, Sunshine hour, $PM_{10,}$ SO_2, $NO_{2.}$
M3	Rainfall, Wind speed, Humidity, Sunshine hour, $PM_{10.}$
M4	Rainfall, Wind speed, Humidity, Sunshine hour, $SO_{2.}$
M5	Rainfall, Wind speed, Humidity, Sunshine hour, $NO_{2.}$

Table 4.11 represents the model output, which shows that error of the models are more in terms of RMSE (>2°C) and it reveals a poor performance of the model over measured air temperature. When compared to results of Gauss membership and Gbell membership function with different parameter combination (rainfall, wind speed, humidity, sunshine hour, SO_2),Gbell membership works better(CC of 0.62, RMSE of 2.15 and SI of 0.08). For the same combination with Gauss membership function and for different combination with Gbell as well as Gauss membership function model results are beyond acceptable limit as tabulated in Table 4.11.

Table 4.11 ANFIS model performance for different combination of parameters

	M1		**M2**		**M3**		**M4**		**M5**	
	Gbell	**Gauss**	**Gbell**	**Gauss**	**Gbell**	**Gauss**	**Gbell**	**Gauss**	**Gbell**	**Gauss**
CC	0.170	0.580	0.500	0.380	0.570	0.620	0.570	**0.620**	0.670	0.510
RMSE	11.87	4.090	2.850	2.770	2.880	2.760	2.260	**2.150**	2.840	2.970
SI	0.490	0.160	0.110	0.110	0.110	0.110	0.090	**0.080**	0.110	0.120

4.3.2. SVM model:

SVM model is developed for data of Bhadra station (meteorological and air pollution data) for the optimum values of SVM parameters (i.e. for nsv=32; C=50,100; ε = 0.1; and d = 0.5) and Polynomial as kernel function. Performance in terms of CC was less than 0.5, which is very poor in terms of accuracy. In regards with RMSE,values are more than 2.7 and for SI values are more than 0.1 which was highly disagreement with accuracy. (Table 4.12).

Table 4.12 SVM model performance for different combination of parameters (ε= 0.1, d = 0.5)

	M1		M2		M3		M4		M5	
C	**50**	**100**	**50**	**100**	**50**	**100**	**50**	**100**	**50**	**100**
CC	0.260	0.290	0.020	0.060	0.160	0.190	0.190	0.220	0.260	0.280
RMSE	2.660	2.640	2.790	2.770	2.700	2.680	2.700	2.680	2.660	2.640
SI	0.100	0.100	0.110	0.110	0.110	0.110	0.110	0.110	0.100	0.100

4.3.3. Comparison of ANFIS and SVM for Bhadra stations:

The performance of the single ANFIS model with different parameter combination is compared with single SVM model and it is observed that the ANFIS model works better(Table 4.11 and Table 4.12). In the parameter combination (rainfall, wind speed, humidity, sunshine hour, SO_2) for ANFIS model with Gbell membership (CC of 0.67, RMSE of 2.84 and SI of 0.11) which is comparatively better than others parameter combination of ANFIS model and SVM model. For the entire parameter combination SVM model having CC values of less than 0.4 and more than 2.7 which is beyond acceptable limit in terms of accuracy.

4.3.4. Hybrid Wavelet-ANFIS model:

In this part, the Wavelet - ANFIS model was tested for the Bhadra area only. The original data were decomposed by Daubechies mother Wavelet up to order 5 (Db1, Db2, Db3, Db4, Db5) till level 5 (level 1, level 2, level 3, level 4, level 5).Then for the ANFIS model Wavelet decomposed data is used as input and original average air temperature was the output for the ANFIS model.

It was observed that along the 45° line, scatter points are uniformly distributed for Bhadra station. This reveals the capability of the model, in estimation of air temperature. This exhibits that model works better than other models developed in the study for the estimation of air temperature.

In the Table 4.13 to Table 4.17 results of estimation of average air temperature are reported. The results clearly explain the usefulness of hybrid Wavelet-ANFIS model. In fact this hybrid Wavelet- ANFIS outperforms single ANFIS model. The parameter combination (rainfall, wind speed, humidity, sunshine hour) for Db5 with level4 (2MF) and Gauss membership function is having a CC of 0.98 which is best in case of performance. In terms of RMSE value it was 0.7 which is very low and finally SI value is 0.03 which shows better performance in terms of accuracy. High performance of the model is due to the combination of Fuzzy Logic, Neural Network and Wavelet Transformation. This is capable of handling uncertainty, noisy data and somewhat non-stationary data.

Inclusion of PM_{10} with meteorological data, SO_2 with meteorological data and NO_2 with meteorological data separately performances are slightly less compared to meteorological data alone. But Inclusion of all air pollution parameters with meteorological data is less accurate compared to other combination.

Table 4.13. Results of Hybrid model (Wavelet-ANFIS) testing data of parameter combination (Meteorological parameter).

		LEVEL 1			LEVEL 2			LEVEL 3			LEVEL 4			LEVEL 5		
		CC	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**
Db 1	**Gbell**	0.846	1.840	0.080	-0.400	3.280	0.140	-0.160	2.860	0.120	-0.330	2.910	0.120	-0.180	2.840	0.120
	Gauss	0.846	1.840	0.080	-0.400	3.280	0.140	-0.160	2.860	0.120	-0.330	2.910	0.120	-0.180	2.840	0.120
Db 2	**Gbell**	0.280	7.350	0.300	0.630	4.960	0.200	-0.310	9.670	0.400	0.840	1.460	0.060	0.820	1.590	0.070
	Gauss	0.410	7.870	0.330	0.600	3.340	0.140	0.520	3.630	0.150	0.850	1.410	0.060	0.700	1.960	0.080
Db 3	**Gbell**	0.680	6.940	0.290	0.320	11.250	0.460	0.430	4.600	0.190	0.970	0.750	0.030	0.950	1.360	0.060
	Gauss	0.760	6.120	0.250	-0.020	32.080	1.320	0.340	6.040	0.250	0.960	0.800	0.030	0.890	1.230	0.050
Db 4	**Gbell**	0.690	6.440	0.270	0.560	2.820	0.120	0.920	1.060	0.040	0.820	1.710	0.070	0.850	1.420	0.060
	Gauss	0.730	4.710	0.190	0.545	2.390	0.100	0.970	0.900	0.040	0.800	1.620	0.070	0.890	1.180	0.050
Db 5	**Gbell**	0.460	3.340	0.140	0.720	4.740	0.200	0.620	2.270	0.090	**0.960**	**0.680**	**0.030**	0.900	1.240	0.050
	Gauss	0.550	2.600	0.110	0.610	2.220	0.090	0.770	1.720	0.070	**0.980**	**0.700**	**0.030**	0.920	1.210	0.050

Table 4.14. Results of Hybrid model (Wavelet-ANFIS) testing data of parameter combination (Meteorological parameter, PM_{10}, NO_2, and SO_2).

		LEVEL 1			LEVEL 2			LEVEL 3			LEVEL 4			LEVEL 5		
		CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI
DB 1	**Gbell**	0.840	1.840	0.070	-0.400	3.280	0.130	-0.150	2.860	0.110	-0.330	2.410	0.120	-0.180	2.840	0.110
	Gauss	0.840	1.840	0.070	-0.400	3.280	0.130	-0.150	2.860	0.110	-0.330	2.410	0.120	-0.180	2.840	0.110
DB 2	**Gbell**	0.690	2.930	0.120	0.930	0.970	0.040	0.810	2.180	0.090	0.640	2.140	0.080	0.340	8.900	0.360
	Gauss	0.820	1.610	0.060	0.920	1.060	0.040	0.800	1.630	0.060	0.910	1.440	0.050	0.810	1.670	0.600
DB 3	**Gbell**	0.880	2.450	0.100	0.830	2.350	0.090	0.930	1.050	0.040	0.930	1.070	0.040	0.564	2.630	0.100
	Gauss	0.890	2.320	0.090	0.890	1.600	0.060	0.870	1.392	0.050	0.910	1.460	0.060	0.530	2.300	0.090
DB 4	**Gbell**	0.630	3.530	0.140	0.940	0.950	0.030	0.900	1.190	0.040	0.550	3.630	0.150	0.930	1.090	0.040
	Gauss	0.710	3.380	0.130	0.940	1.050	0.040	0.950	0.890	0.030	0.430	4.200	0.170	0.900	1.130	0.040
DB 5	**Gbell**	0.480	2.610	0.110	0.880	1.870	0.080	0.760	1.810	0.070	0.880	1.250	0.050	**0.930**	**0.950**	**0.040**
	Gauss	0.820	1.650	0.070	0.870	1.670	0.070	0.860	1.520	0.060	0.910	1.140	0.050	**0.960**	**0.830**	**0.030**

Table 4.15. Results of Hybrid model (Wavelet-ANFIS) testing data of parameter combination (Meteorological parameter and PM_{10}).

		LEVEL 1			LEVEL 2			LEVEL 3			LEVEL 4			LEVEL 5		
		CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI
Db 1	**Gbell**	0.840	1.840	0.080	-0.400	3.280	0.140	-0.160	2.860	0.120	-0.330	2.910	0.120	-0.180	2.840	0.120
	Gauss	0.840	1.840	0.080	-0.400	3.280	0.140	-0.160	2.860	0.120	-0.330	2.910	0.120	-0.180	2.840	0.120
Db 2	**Gbell**	0.520	4.670	0.190	0.820	1.540	0.060	0.860	1.540	0.060	0.830	1.500	0.060	0.930	1.010	0.040
	Gauss	0.340	6.800	0.280	0.930	1.280	0.050	0.910	1.360	0.060	0.880	1.510	0.060	0.680	2.320	0.100
Db 3	**Gbell**	0.830	3.990	0.160	0.540	3.750	0.150	0.870	1.530	0.060	0.920	1.270	0.050	0.910	1.290	0.050
	Gauss	0.840	2.780	0.110	0.380	5.500	0.230	0.660	2.450	0.100	0.920	1.130	0.050	0.870	1.650	0.070
Db 4	**Gbell**	0.410	3.230	0.130	0.890	1.270	0.050	0.960	0.790	0.030	0.310	4.000	0.160	0.930	1.180	0.050
	Gauss	0.710	2.050	0.080	0.860	1.350	0.060	0.920	1.160	0.050	0.520	2.530	0.100	0.890	1.240	0.050
Db 5	**Gbell**	0.390	2.750	0.110	0.850	1.380	0.060	0.750	1.830	0.080	**0.970**	**0.660**	**0.030**	0.950	0.940	0.040
	Gauss	0.800	1.580	0.070	0.850	1.540	0.060	0.850	1.500	0.060	**0.960**	**0.740**	**0.030**	0.830	1.500	0.060

Table 4.16 Results of Hybrid model (Wavelet-ANFIS) testing data of parameter combination (Meteorological parameter and SO_2).

		LEVEL 1			LEVEL 2			LEVEL 3			LEVEL 4			LEVEL 5		
		CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI
Db 1	Gbell	0.846	1.840	0.080	-0.400	3.280	0.140	-0.160	2.860	0.120	-0.330	2.910	0.120	-0.180	2.840	0.120
	Gauss	0.846	1.840	0.080	-0.400	3.280	0.140	-0.160	2.860	0.120	-0.330	2.910	0.120	-0.180	2.840	0.120
Db 2	Gbell	0.690	2.530	0.100	0.910	1.130	0.050	0.560	2.730	0.110	0.770	1.810	0.070	0.880	1.340	0.060
	Gauss	0.530	4.770	0.200	0.900	1.170	0.050	0.710	2.070	0.090	0.920	1.090	0.050	0.900	1.250	0.050
Db 3	Gbell	0.750	2.660	0.110	0.420	5.070	0.210	0.860	1.420	0.060	0.930	1.120	0.050	0.910	1.440	0.060
	Gauss	0.640	2.480	0.100	0.600	3.750	0.150	0.680	2.100	0.090	0.950	0.950	0.040	0.910	1.260	0.050
Db 4	Gbell	0.750	1.780	0.070	0.830	1.590	0.070	**0.90**	**2.160**	**0.090**	0.740	2.610	0.110	0.880	1.470	0.060
	Gauss	0.670	2.070	0.090	0.780	1.720	0.070	**0.970**	**0.740**	**0.030**	0.490	2.360	0.100	0.890	1.270	0.050
Db 5	Gbell	0.270	2.990	0.120	0.450	4.740	0.200	0.740	1.920	0.080	0.920	1.110	0.050	0.930	0.920	0.040
	Gauss	0.260	3.120	0.130	0.790	2.170	0.090	0.760	1.750	0.070	0.960	0.740	0.030	0.920	1.090	0.050

Table 4.17 Results of Hybrid model (Wavelet-ANFIS) testing data of parameter combination (Meteorological parameter and NO_2).

		LEVEL 1			LEVEL 2			LEVEL 3			LEVEL 4			LEVEL 5		
		CC	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**
Db1	**Gbell**	0.840	1.840	0.080	-0.400	3.280	0.140	-0.160	2.860	0.120	-0.330	2.910	0.120	-0.180	2.840	0.120
	Gauss	0.840	1.840	0.080	-0.400	3.280	0.140	-0.160	2.860	0.120	-0.330	2.910	0.120	-0.180	2.840	0.120
Db 2	**Gbell**	0.650	4.980	0.210	0.930	1.100	0.050	0.840	1.590	0.070	0.860	1.460	0.060	0.770	1.690	0.070
	Gauss	0.600	7.880	0.330	0.920	1.200	0.050	0.870	1.570	0.060	0.880	1.470	0.060	0.870	1.440	0.060
Db 3	**Gbell**	0.750	2.110	0.090	0.510	5.000	0.210	0.390	3.170	0.130	0.950	1.050	0.040	0.930	1.780	0.070
	Gauss	0.830	1.760	0.070	0.790	2.890	0.120	0.760	1.870	0.080	0.890	1.630	0.070	0.860	1.620	0.070
Db 4	**Gbell**	0.810	2.270	0.090	0.800	1.780	0.070	0.950	0.870	0.040	-0.360	15.550	0.640	0.860	1.550	0.060
	Gauss	0.840	1.610	0.070	0.830	1.550	0.060	0.930	1.060	0.040	0.580	2.470	0.100	0.900	1.110	0.050
Db 5	**Gbell**	0.830	1.950	0.080	0.640	3.070	0.130	0.910	1.210	0.050	**0.970**	**0.670**	**0.030**	0.950	0.840	0.030
	Gauss	0.850	2.070	0.090	0.770	1.870	0.080	0.780	1.690	0.070	**0.970**	**0.650**	**0.030**	0.960	0.870	0.040

4.3.5. Comparison of Performance of Hybrid Wavelet-ANFIS model with ANFIS model:

The performance of the hybrid Wavelet-ANFIS model with the combination of meteorological data is high compared to other combination and between hybrid Wavelet-ANFIS model and ANFIS model with 0.98 as CC, 0.7 and 0.03 as RMSE and SI respectively. But for the ANFIS model, only parameter combination (rainfall, wind speed, humidity, sunshine hour, SO_2), Gbell membership works comparatively better in ANFIS model. (CC of 0.62, RMSE of 2.15 and SI of 0.08).

The hybrid Wavelet-ANFIS model with a parameter combination of rainfall, wind speed, humidity, sunshine hour for Db5 with level 4 and Gauss membership function is having a high CC is 0.98 which is finest in terms of performance (Table. 4.13) (Figs 4.12a and 4.12b). The RMSE and SI values are 0.7 and 0.03 respectively. Further, with a parameter combination of rainfall, wind speed, humidity, sunshine hour,PM_{10},SO_2,$NO_2$for Db5 with level 5 and Gauss membership function results in a high CC of 0.96, RMSE and SI of 0.83 and 0.03 respectively (Table. 4.14) (Figs 4.13a and 4.13b). With the parameter combination of rainfall, wind speed, humidity, sunshine hour, PM_{10} for Db5 with level 4 and Gbell membership function is showing a CC of 0.97 which is finest in case of performance (Table. 4.15) (Figs 4.14a and 4.14b). Also, RMSE value is 0.66 which is very low with best accuracy and finally SI value is 0.03 which shows improved results in case of accuracy. Further, for combination of rainfall, wind speed, humidity, sunshine hour, SO_2 along with Db4 with level 3 and Gauss membership function, CC is 0.97, lowest RMSE value with 0.74 and having SI as 0.03 which is better accuracy of the work (Table. 4.16) (Figs.4.15a and 4.15b). The last combination of rainfall, wind speed, humidity, sunshine hour, NO_2 along with Db5 with level 4 and Gauss membership function is having better result in CC of 0.97 and 0.65 of RMSE and in terms of SI having 0.03 lowest values with high performance, which is acceptable in case of accuracy as tabulated in Table. 4.17 (Figs 4.16a and 4.16b).

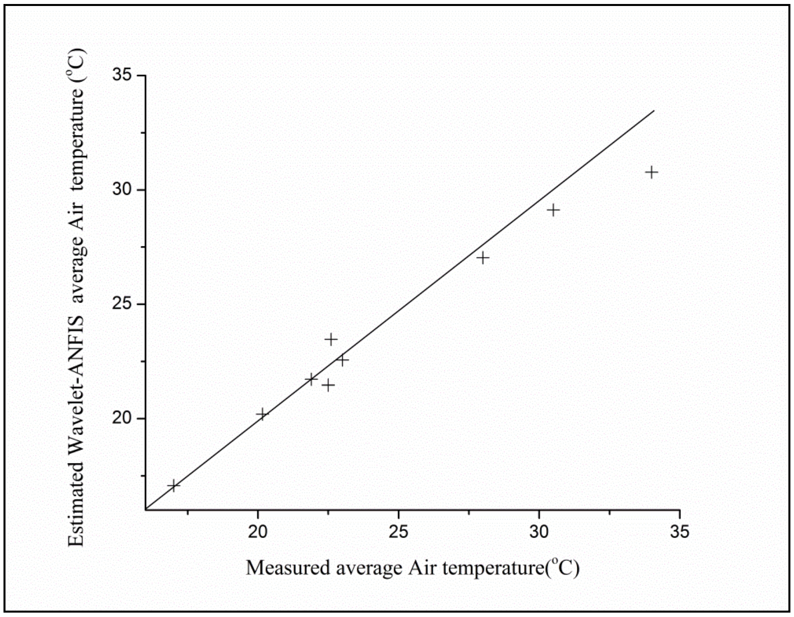


Fig. 4.12a Scatter plot of hybridized Wavelet-ANFIS model performance for Bhadra station (Meteorological parameter) (Db5-L4, Gauss)

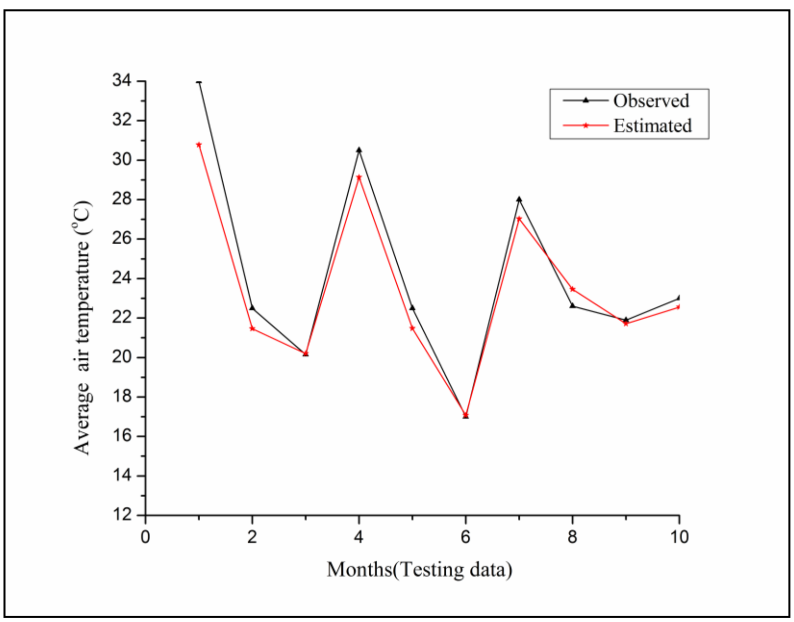


Fig. 4.12b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Bhadra station (Meteorological parameter) (Db5-L4, Gauss)

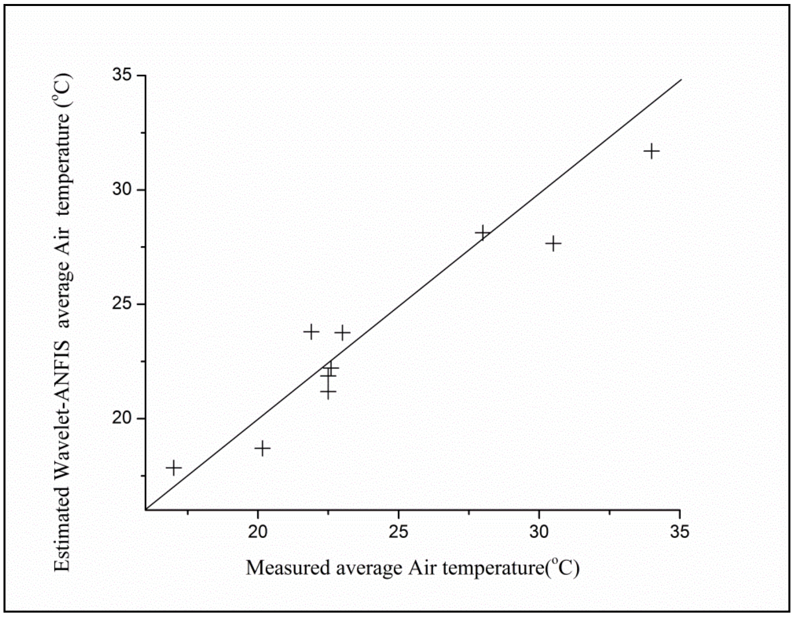


Fig. 4.13a Scatter plot of hybridized Wavelet-ANFIS model performance for Bhadra station (Meteorological parameter, PM_{10}, SO_2 and NO_2) (Db5-L5, Gauss)

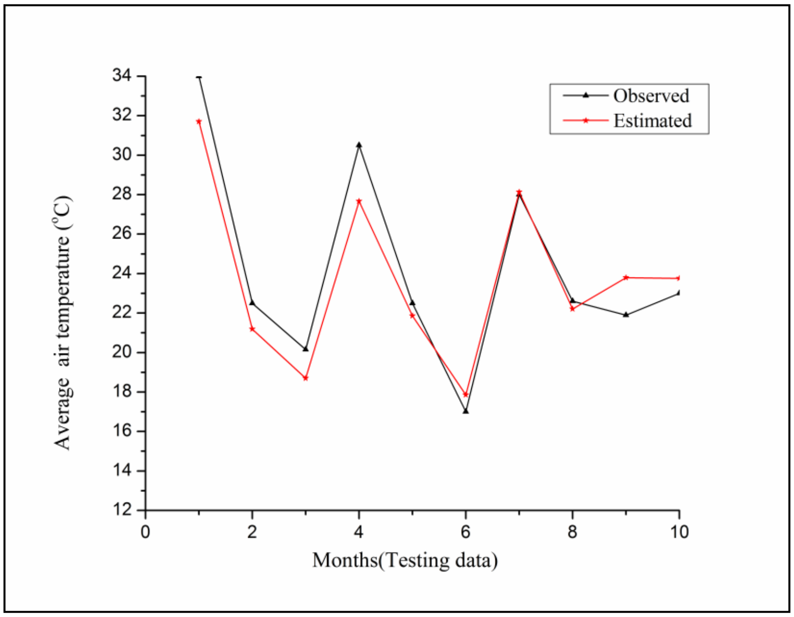


Fig. 4.13b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Bhadra station (Meteorological parameter, PM_{10}, SO_2 and NO_2) (Db5-L5, Gauss)

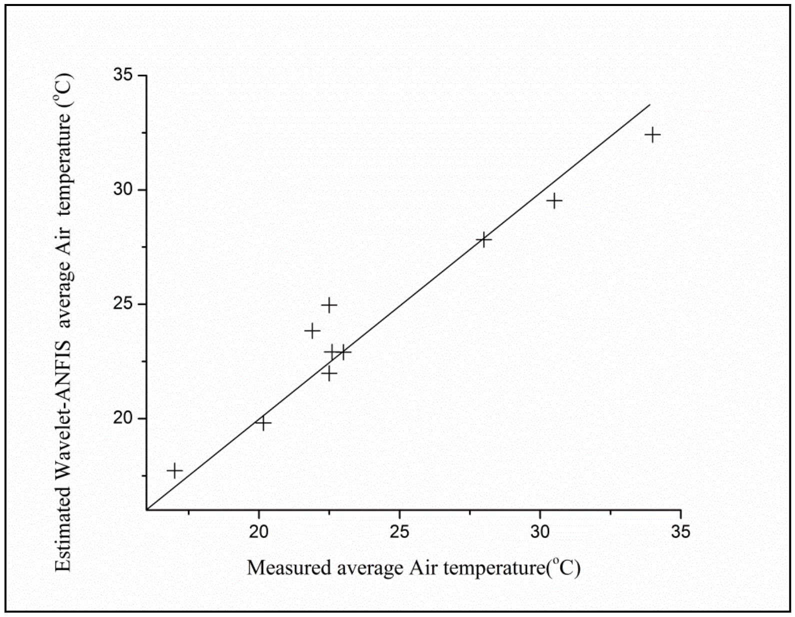


Fig. 4.14a Scatter plot of hybridized Wavelet-ANFIS model performance for Bhadra station (Meteorological parameter and PM_{10}) (Db5-L4, Gbell)

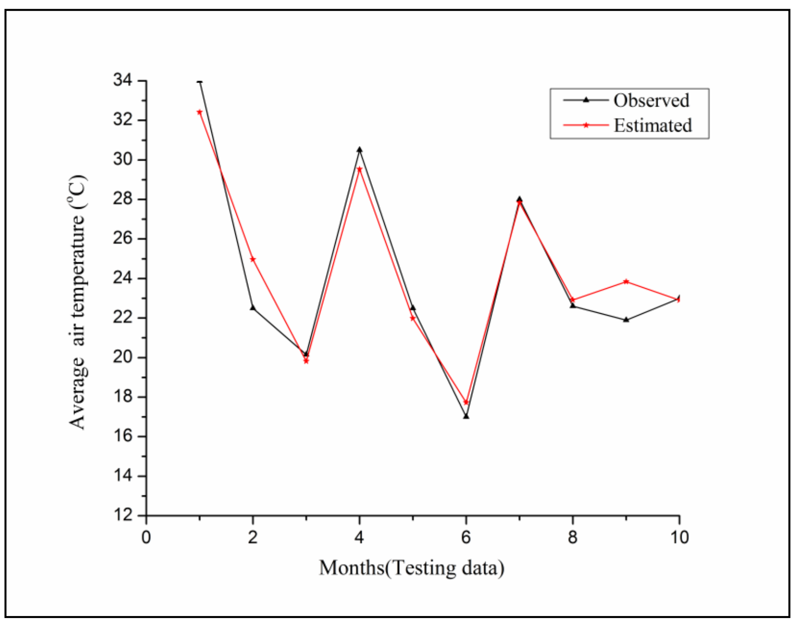


Fig. 4.14b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Bhadra station (Meteorological parameter and PM_{10}) (Db5-L4, Gbell)

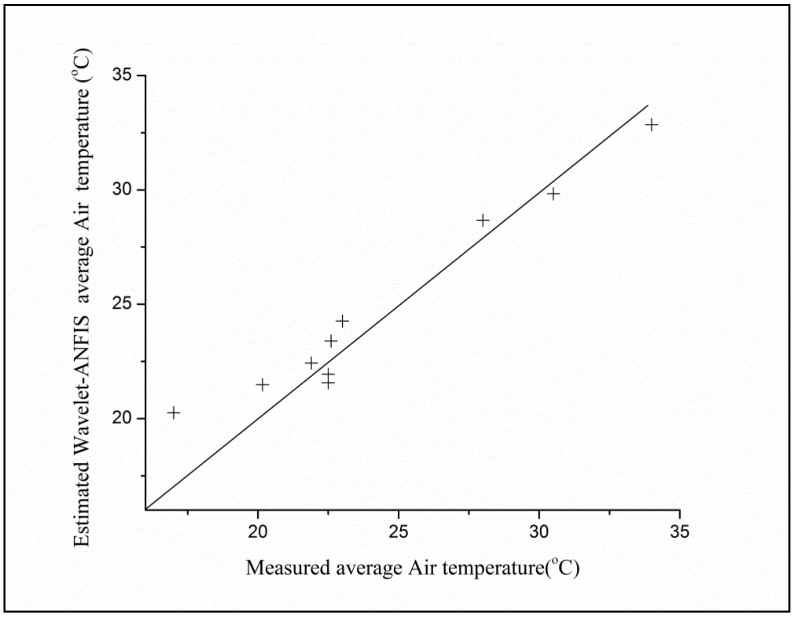


Fig. 4.15a Scatter plot of hybridized Wavelet-ANFIS model performance for Bhadra station (Meteorological parameter and SO_2) (Db4-L3, Gauss)

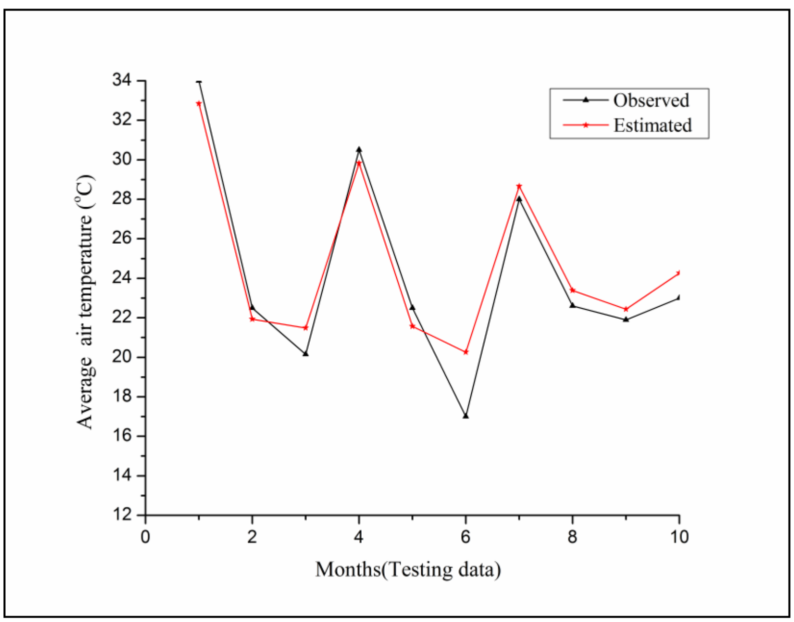


Fig. 4.15b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Bhadra station (Meteorological parameter and SO_2) (Db4-L3, Gauss)

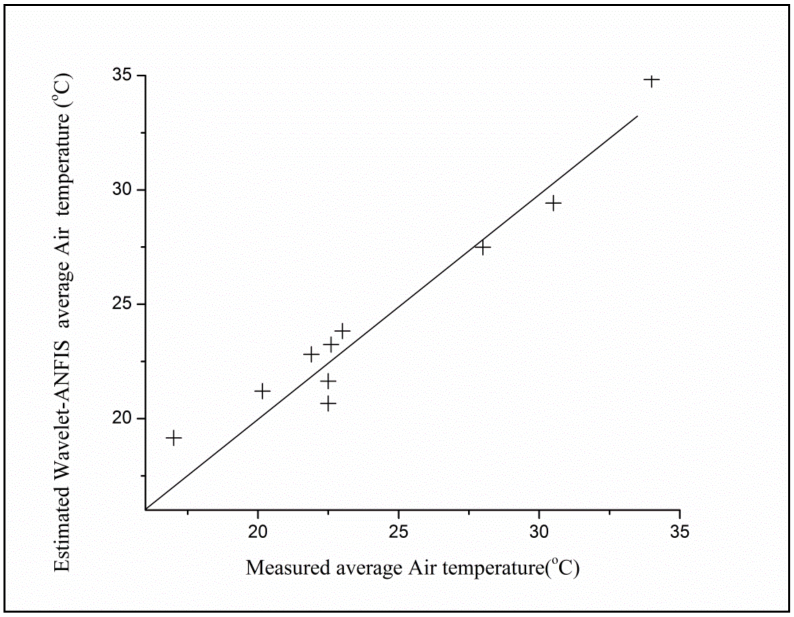


Fig. 4.16a Scatter plot of hybridized Wavelet-ANFIS model performance for Bhadra station (Meteorological parameter and NO_2) (Db5-L4, Gauss)

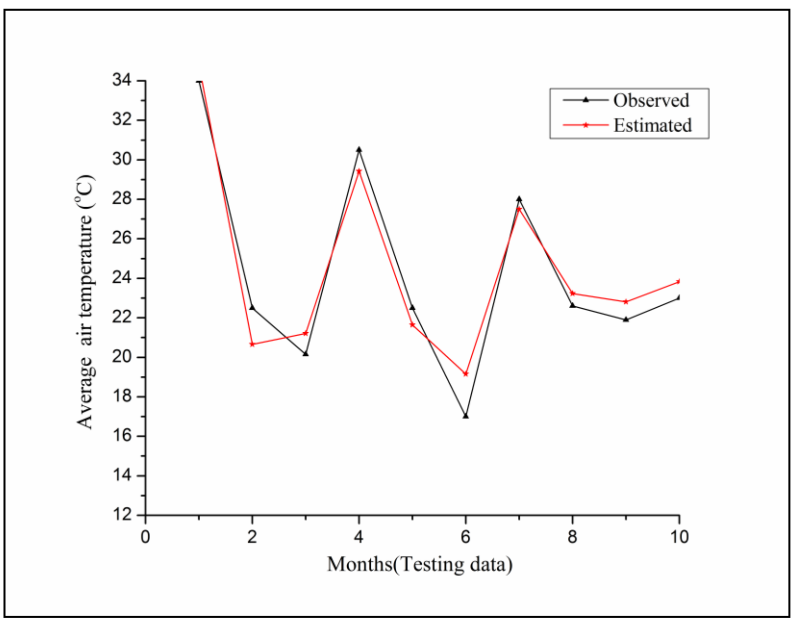


Fig. 4.16b Hybridized Wavelet-ANFIS model performance for modeled air temperature of test data for Bhadra station (Meteorological parameter and NO_2) (Db5-L4, Gauss)

4.3.6. Hybrid Wavelet- SVM model:

As discussed in the above section the Wavelet- SVM model for original data of Bhadra area were decomposed by Daubechies mother Wavelet up to order 5 (Db1, Db2, Db3, Db4, Db5) till level 5(level 1, level 2, level 3, level 4, level 5).Then for the SVM model Wavelet decomposed data is used as input and original average air temperature was the output for the SVM model.

In the Tables 4.18 to 4.22 results of average air temperature using hybrid Wavelet- SVM model are shown. The final optimum values (i.e. for nsv=32; C=50,100; ε = 0.1; and d = 0.5) of SVM and Polynomial as kernel function are used. The number of support vectors used in Wavelet-SVM models is 100% (32), which indicates that every training data set is utilized as support vector. This clearly proves that, there is no noise in the training data set, but there is non-linearity and complexity associated in mapping input and output parameters of average air temperature. The results of testing are less than 0.5 in case of CC, which reveals its poor performance. In regards with RMSE for testing ismore than 2.7 which shows low performance. Also in SI for testing, values are more than 0.1 which is not acceptable in terms of accuracy. But in the parameter combination rainfall, wind speed, humidity, sunshine hour, PM_{10} CC value is 0.57 and in terms of RMSE and SI it is 2.43 and 0.1 which is better compared to all the parameter combination in this model.

Results of the model reveal a poor performance of the model to estimate air temperature. However, far and wide deviated model data were found during lower and higher air temperature. However, sparsely deviated model data were found during lower and higher air temperature. Also it was observed that for middle range of air temperature, model values are relatively closer to 45^{o} line.

Table 4.18. Results of Hybrid model (Wavelet-SVM) testing data of parameter combination (Meteorological parameter).

	Db1			Db2			Db3			Db4			Db5		
	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI
	C=50; ε = 0.1; d = 0.5														
Level 1	-0.210	2.90	0.110	0.200	2.650	0.100	0.480	2.460	0.100	0.520	2.480	0.100	0.450	2.540	0.100
Level 2	-0.640	3.160	0.130	0.350	2.600	0.100	0.490	2.450	0.100	0.150	2.690	0.110	0.160	2.690	0.110
Level 3	-0.280	2.870	0.110	-0.140	2.770	0.140	-0.200	2.860	0.110	0.080	2.720	0.110	0.020	2.750	0.110
Level 4	-0.330	2.910	0.120	-0.390	2.860	0.110	-0.200	2.860	0.120	0.240	2.850	0.110	-0.120	2.830	0.110
Level 5	-0.180	2.840	0.110	-0.410	2.920	0.120	0.030	2.730	0.110	-0.330	2.900	0.120	-0.280	2.890	0.110
	C=100; ε = 0.1; d = 0.5														
Level 1	-0.140	2.880	0.110	0.250	2.620	0.100	0.500	2.420	0.090	0.530	2.430	0.100	0.450	2.540	0.100
Level 2	-0.620	3.150	0.130	0.390	2.570	0.130	0.500	2.400	0.090	0.200	2.670	0.110	0.210	2.660	0.120
Level 3	-0.250	2.870	0.110	-0.060	2.760	0.110	-0.190	2.850	0.120	0.110	2.710	0.110	0.050	2.750	0.110
Level 4	-0.330	2.910	0.120	-0.390	2.860	0.110	-0.180	2.850	0.120	-0.240	2.860	0.110	-0.130	2.830	0.110
Level 5	-0.180	2.840	0.110	-0.410	2.920	0.120	0.030	2.730	0.110	-0.330	2.900	0.120	-0.280	2.890	0.110

Table 4.19. Results of Hybrid model (Wavelet-SVM) testing data of parameter combination (Meteorological parameter, PM_{10}, NO_2, and SO_2).

	Db1			Db2			Db3			Db4			Db5		
	CC	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**
	C=50; ε = 0.1; d = 0.5														
Level 1	-0.140	2.930	0.120	-0.100	2.730	0.110	0.280	2.580	0.100	-0.370	2.900	0.110	0.220	2.640	0.100
Level 2	-0.0340	3.200	0.130	0.280	2.630	0.100	0.340	2.560	0.100	0.180	2.720	0.110	0.220	2.670	0.110
Level 3	-0.150	2.850	0.110	0.060	2.780	0.110	-0.110	2.860	0.110	0.020	2.820	0.110	-0.120	2.860	0.110
Level 4	-0.330	2.910	0.120	-0.450	2.900	0.120	0.350	2.540	0.100	-0.320	2.900	0.120	-0.180	2.870	0.120
Level 5	-0.150	2.910	0.120	0.020	2.690	0.110	0.200	2.640	0.100	-0.330	2.900	0.120	-0.140	2.850	0.110
	C=100; ε = 0.1; d = 0.5														
Level 1	-0.10	2.910	0.100	0.170	0.680	0.110	0.340	2.520	0.100	-0.370	2.900	0.110	0.270	2.600	0.100
Level 2	-0.034	3.200	0.130	0.280	2.620	0.100	0.430	2.430	0.100	0.180	2.720	0.110	0.240	2.660	0.110
Level 3	-0.150	2.860	0.110	0.060	2.780	0.110	-0.110	2.860	0.110	0.020	2.820	0.110	-0.120	2.850	0.110
Level 4	-0.330	2.910	0.120	-0.450	2.900	0.120	0.350	2.540	0.100	-0.310	2.900	0.120	-0.180	2.870	0.120
Level 5	-0.150	2.910	0.120	0.020	2.690	0.110	0.210	2.630	0.100	-0.330	2.900	0.120	-0.110	2.850	0.110

Table 4.20. Results of Hybrid model (Wavelet-SVM) testing data of parameter combination (Meteorological parameter and PM_{10}).

	Db1			**Db2**			**Db3**			**Db4**			**Db5**		
	CC	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**
	C=50; ε = 0.1; d = 0.5														
Level 1	0.180	2.930	0.120	0.170	2.670	0.110	0.480	2.470	0.100	**0. 540**	**2.480**	**0.100**	0.460	2.500	0.100
Level 2	0.500	3.240	0.130	0.220	2.660	0.100	0.420	2.480	0.100	0.070	2.760	0.110	0.070	2.750	0.100
Level 3	0.090	2.800	0.110	0.020	2.780	0.110	0.160	2.800	0.110	0.050	2.780	0.110	-0.050	2.790	0.110
Level 4	0.330	2.930	0.120	0.440	2.910	0.120	0.220	2.910	0.120	0.330	2.910	0.120	0.190	2.890	0.110
Level 5	0.180	2.840	0.110	0.420	2.910	0.120	0.200	2.650	0.100	0.330	2.900	0.120	0.140	2.850	0.110
	C=100; ε = 0.1; d = 0.5														
Level 1	0.130	2.910	0.120	0.230	2.630	0.100	0.510	2.410	0.090	**0.570**	**2.430**	**0.100**	0.500	2.500	0.100
Level 2	0.460	3.230	0.130	0.260	2.630	0.100	0.470	2.410	0.090	0.110	2.740	0.110	0.120	2.730	0.100
Level 3	0.090	2.800	0.110	0.040	2.780	0.110	0.120	2.800	0.110	0.060	2.770	0.110	-0.030	2.790	0.110
Level 4	0.330	2.930	0.120	0.440	2.910	0.120	0.220	2.910	0.120	0.330	2.910	0.120	0.190	2.890	0.110
Level 5	0.180	2.840	0.110	0.430	2.910	0.120	0.210	2.640	0.100	0.330	2.900	0.120	0.120	2.850	0.110

Table 4.21. Results of Hybrid model (Wavelet-SVM) testing data of parameter combination (Meteorological parameter and SO_2).

	Db1			Db2			Db3			Db4			Db5		
	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI	CC	RMSE (°C)	SI
	C=50; ε = 0.1; d = 0.5														
Level 1	0.170	2.910	0.120	0.130	2.720	0.110	0.310	2.560	0.100	0.290	2.580	0.100	0.250	2.620	0.100
Level 2	-0.630	3.230	0.130	0.190	2.680	0.110	0.370	2.530	0.100	0.080	2.740	0.110	0.050	2.750	0.110
Level 3	-0.270	2.910	0.120	0.260	2.840	0.110	0.140	2.850	0.110	0.110	2.810	0.110	0.060	2.800	0.110
Level 4	-0.330	2.910	0.120	-0.410	2.920	0.120	0.210	2.860	0.110	-0.330	2.900	0.120	0.190	2.870	0.110
Level 5	-0.180	2.840	0.110	-0.420	2.920	0.120	0.180	2.660	0.110	0.300	2.900	0.120	0.240	3.000	0.120
	C=100; ε = 0.1; d = 0.5														
Level 1	0.060	2.890	0.110	0.190	2.680	0.110	0.360	2.510	0.100	0.350	2.540	0.100	0.300	2.590	0.100
Level 2	-0.610	3.220	0.130	0.240	2.650	0.100	0.420	2.480	0.100	0.080	2.740	0.110	- 0.050	2.750	0.110
Level 3	-0.270	2.910	0.120	0.190	2.820	0.110	0.110	2.830	0.110	0.090	2.800	0.110	0.040	2.800	0.110
Level 4	-0.330	2.910	0.120	-0.410	2.920	0.120	0.200	2.810	0.110	-0.330	2.900	0.120	0.190	2.870	0.110
Level 5	-0.180	2.840	0.110	-0.420	2.920	0.120	0.190	2.650	0.100	0.300	2.900	0.120	0.120	2.850	0.110

Table 4.22. Results of Hybrid model (Wavelet-SVM) testing data of parameter combination (Meteorological parameter and NO_2).

	Db1			**Db2**			**Db3**			**Db4**			**Db5**		
	CC	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**	**CC**	**RMSE (°C)**	**SI**
	C=50; ε = 0.1; d = 0.5														
Level 1	0.210	2.900	0.120	0.170	2.690	0.110	0.420	2.500	0.100	0.460	2.510	0.100	0.400	2.560	0.100
Level 2	-0.530	3.180	0.130	0.330	2.600	0.100	0.440	2.480	0.100	0.190	2.680	0.110	0.200	2.670	0.110
Level 3	0.180	2.870	0.110	0.210	2.810	0.110	0.180	2.850	0.110	0.080	2.790	0.110	0.010	2.760	0.110
Level 4	0.330	2.910	0.120	0.410	2.920	0.120	0.180	2.850	0.110	-0.330	2.900	0.120	0.200	2.880	0.110
Level 5	0.180	2.840	0.110	0.420	2.920	0.120	0.100	2.700	0.110	-0.270	2.860	0.110	-0.280	2.890	0.110
	C=100; ε = 0.1; d = 0.5														
Level 1	0.140	2.900	0.120	0.230	2.640	0.100	0.460	2.440	0.100	0.490	2.460	0.100	0.440	2.520	0.100
Level 2	-0.520	3.180	0.130	0.360	2.580	0.100	0.490	2.400	0.090	0.220	2.670	0.120	0.200	2.670	0.110
Level 3	0.180	2.870	0.110	0.120	2.790	0.110	0.180	2.850	0.110	0.050	2.780	0.110	0.030	2.750	0.110
Level 4	0.330	2.910	0.120	0.410	2.920	0.120	0.180	2.850	0.110	-0.330	2.900	0.120	0.200	2.880	0.110
Level 5	0.180	2.840	0.110	0.420	2.920	0.120	0.130	2.680	0.110	-0.280	2.860	0.110	-0.280	2.890	0.110

4.3.7. Comparison of SVM model and Hybrid Wavelet-SVM model:

In the parameter combination (rainfall, wind speed, humidity, sunshine hour, PM_{10}) for Db4 with level 1 (i.e. for nsv=32; C=100; ε = 0.1; and d = 0.5) is having the results of CC is 0.57 which is good in case of performance for hybrid Wavelet-SVM model. (Table. 4.20) (Figs 4.17a and 4.17b).But for single SVM model performance in terms of CC was less than 0.5, which is very poor in terms of accuracy. In regards with RMSE, values are more than 2.7 and for SI values are more than 0.1 which was highly disagreement with accuracy.

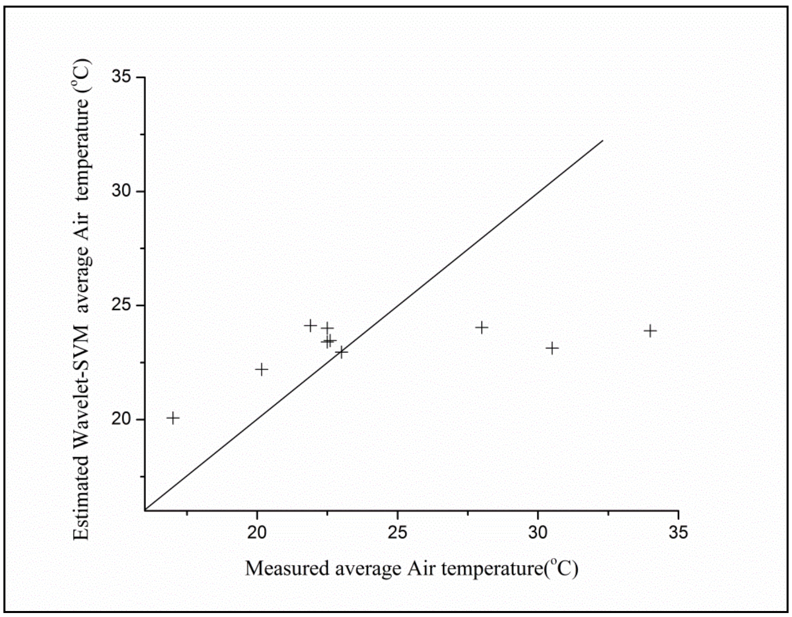


Fig. 4.17a Scatter plot of hybridized Wavelet-SVM model performance for Bhadra station (Meteorological parameter and PM_{10}) (Db4-L1)

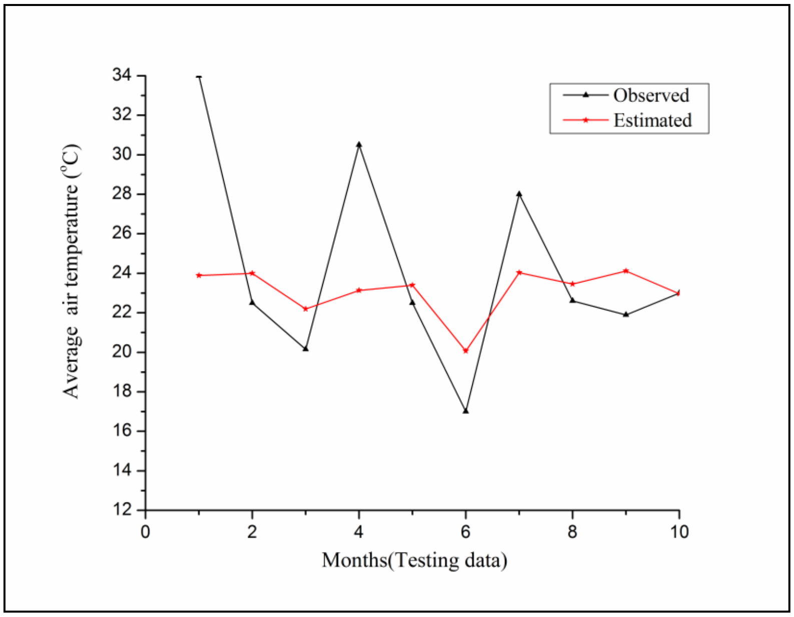


Fig. 4.17b Hybridized Wavelet- SVM model performance for modeled air temperature of test data for Bhadra station (Meteorological parameter and PM_{10}) (Db4-L1)

4.3.8. Comparison of Hybrid model (Wavelet-ANFIS) and Hybrid model (Wavelet-SVM) for Bhadra stations:

The hybrid Wavelet-ANFIS model works better compared to hybrid Wavelet-SVM model (Tables 4.13 to 4.22).Performance of the both hybrid models with different combination of parameters are compared in which in all the parameter combination, for hybrid Wavelet- ANFIS model with Db4 (level3, level4 and level5) and Db5 (level3, level4 and level5) having a better results with more CC values, less RMSE and less SI values which is better in terms of accuracy. But in case of hybrid Wavelet -SVM model, model having CC values less than 0.6 which is less accurate.

4.3.9 Performance evaluations of single ANFIS, single SVM, hybrid Wavelet - ANFIS model and hybrid Wavelet- SVM model

In the previous section the performance of the single and hybrid models were discussed for Bhadra station (B.R.Project) with different parameter combination. In this section for Bhadra station (B.R.Project) all the models with different parameter combinations are compared to give the best model to estimate air temperature accurately and precisely. The Tables 4.11 to 4.22 shows the comparison of single model, comparison of single with hybrid model and comparison of hybrid verses hybrid model.

In terms of performance hybrid Wavelet- ANFIS model works best compared to single ANFIS, single SVM and hybrid Wavelet- SVM model. Hybrid Wavelet - ANFIS model with the parameter combination (rainfall, wind speed, humidity, sunshine hour) for Db5 with level4 (2MF) and Gauss membership function is having the results of CC is 0.98 which is best in case of accuracy. In terms of RMSE value is 0.7 which is very low with high precision and finally in SI value is 0.03 which shows better results in case of performance (Table 4.23).

Table 4.23. Comparison of the best performances of the model

Rank	**Model**	**Parameter combination**	**Level of decomposition**	**CC**	**RMSE (°C)**	**SI**
1	Wavelet-ANFIS model	M1	Db5-Level4(Gauss)	0.980	0.700	0.030
2	Wavelet-ANFIS model	M5	Db5-Level4(Gauss)	0.970	0.650	0.030
3	Wavelet-ANFIS model	M3	Db5-Level4(Gbell)	0.970	0.660	0.030
4	Wavelet-ANFIS model	M5	Db5-Level4(Gbell)	0.970	0.670	0.030
5	Wavelet-ANFIS model	M4	Db4-Level3(Gauss)	0.970	0.740	0.030

4.4 SUMMARY

Initially, for seven stations with four meteorological parameters are used for modeling of air temperature. It is observed that hybrid Wavelet- ANFIS model works better in terms of accuracy. But results of the model show that a room for improvement was available considering air pollution aspect. Hence to check the accuracy, single Bhadra stations with meteorological data and air pollution are used for the modeling of air temperature. Comparing different models with different parameter combination hybrid Wavelet-ANFIS model works better compared to single ANFIS, single SVM and hybrid Wavelet - SVM model.

CHAPTER 5

SUMMARY AND CONCLUSIONS

5.1 SUMMARY

Accurate modeling of average air temperature by considering all the boundary conditions is extremely difficult. Therefore, it needs an approximate analysis system which is capable of handling such boundary conditions. In this circumstance, soft computing models are developed, such as, ANFIS, SVM, hybrid Wavelet-ANFIS and hybrid Wavelet-SVM.

For the development of soft computing models data were collected from meteorological stations in and around Shimoga, Karnataka, India. Dimensional input parameters that control the modeling of average air temperature, such as, rainfall, wind speed, sunshine hours, humidity, suspended particulate matters (SPM), nitrogen dioxide (NO_2) and Sulphur dioxide (SO_2) were used as the inputs for the soft computing techniques.

Initially ANFIS and SVM models are developed to estimate average air temperature for seven stations and then study is extended with inclusion of air pollution data additional to meteorological data for Bhadra station only. SVM showed better performance compared to ANFIS but a room for improvement was available. Hence, to carry over the work hybrid Wavelet-ANFIS and hybrid Wavelet-SVM were developed. ANFIS, SVM, hybrid Wavelet-ANFIS and hybrid Wavelet-SVM models were compared in terms of statistical measures. Based on the present model results, conclusions were drawn and presented here.

5.2 RESEARCH CONTRIBUTION

1. Developed hybrid models to estimate air temperature using meteorological data and air pollution data with higher accuracy.
2. The influence of air pollution in air temperature modeling has been highlighted.

5.3 CONCLUSIONS

Based on the results of the present investigations and discussion thereon, following general conclusions are arrived at:

1. Performance of the proposed hybrid Wavelet-ANFIS model outperforms other models such as ANFIS model, SVM model and Hybrid Wavelet-SVM models in modeling air temperature.
2. Performance of single SVM model is better than single ANFIS model out of seven stations with in same climatic zone.
3. The accuracy of proposed hybrid Wavelet-ANFIS appears to be highly influenced by the choice of membership function such as Gbell and Gauss and the optimal combination of input parameters.
4. Out of seven stations, for station Linganamakki showed best performance with CC of 0.954, RMSE and SI 0.71 and 0.027 respectively with hybrid Wavelet- ANFIS model (Gbell membership).
5. For Bhadra station (B. R. Project), the input parameters combination (rainfall, wind speed, humidity, sunshine hour, SO_2) works better for ANFIS model with Gbell membership (CC of 0.62, RMSE of 2.15 and SI of 0.08) which is found to be superior than other parameter combination of ANFIS model and SVM model.
6. Inclusion of air pollution parameters with meteorological parameters shows encouraging result with significant improvement in modeling accuracy.
7. For Bhadra station (B. R. Project) the parameter combination (rainfall, wind speed, humidity, sunshine hour) for Db5 with Level4 (2MF) and Gauss membership function, the hybrid Wavelet - ANFIS model having higher CC of 0.98 considered best model.

8. For prediction of season ahead air temperature, hybrid Wavelet- ANFIS model shows better performance for Hiriyur station out of seven station with higher CC of 0.913 and low RMSE and SI of 1.340 and 0.051 respectively.
9. The results of uncertainty analysis using Bootstrap resampling indicate that the variation in training pattern does not have significant effect on the overall model performance.

5.4 LIMITATION OF THE WORK

1. Quality of the data is governing factor for this type of analysis, mainly influences the performance of the model.
2. Proposed hybrid model to estimate air temperature is site specific.

5.5 SUGGESTIONS FOR FUTURE WORK

The following suggestions may be considered for further study:

1. Modeling of average air temperature with different parameters like vehicle density, population density, vegetation index, etc., may include in the studies.
2. Space borne observations of atmospheric variables may be incorporated in studies to improve the spatial coverage of air temperature estimations.
3. For modeling of average air temperature by super hybridization like wavelet - ANFIS-SVM or Wavelet - SVM-ANFIS methods may be tried for extension of work.

REFERENCES

- Abrahart, R. J. (2003), Neural network rainfall-runoff forecasting based on continuous resampling, J. Hydroinformatics,5, 51–61.
- Aldrian, E., and Djamil, Y. S. (2008). Application of multivariate ANFIS for daily rainfall prediction: influences of training data size. Makara, Sains, 12(1), 7-14.
- Alvares, C. A., Stape, J. L., Sentelhas, P. C., & de Moraes Gonçalves, J. L. (2013). Modeling monthly mean air temperature for Brazil. Theoretical and applied climatology, 113(3-4), 407-427.
- Ashish, M., and Rashmi, B. (2011). Prediction of daily air pollution using wavelet decomposition and adaptive-network-based fuzzy inference system.International Journal of Environmental Sciences, 2(1), 185-196.
- Aussem, A., Campbell, J., & Murtagh, F. (1998). Wavelet-based feature extraction and decomposition strategies for financial forecasting. Journal of Computational Intelligence in Finance, 6(2), 5-12.
- Baboo, S. S., and Shereef, I. K. (2010). An efficient weather forecasting system using artificial neural network. International journal of environmental science and development, 1(4), 2010-0264.
- Bertini, I., Ceravolo, F., Citterio, M., De Felice, M., Di Pietra, B., Margiotta, F., and Puglisi, G. (2010). Ambient temperature modelling with soft computing techniques. Solar Energy, 84(7), 1264-1272.
- Bilbao, J., De Miguel, A. H., & Kambezidis, H. D. (2002). Air temperature model evaluation in the north Mediterranean belt area. Journal of Applied Meteorology, 41(8), 872-884.
- Bilgili, M., and Sahin, B. (2009). Prediction of long-term monthly temperature and rainfall in Turkey. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 32(1), 60-71.

- Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.
- Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., and Vitabile, S. (2007). Two-days ahead prediction of daily maximum concentrations of SO 2, O 3, PM10, NO 2, CO in the urban area of Palermo, Italy. Atmospheric Environment, 41(14), 2967-2995.
- Catalao, J., Pousinho, H. M. I., and Mendes, V. M. F. (2011). Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal.Sustainable Energy, IEEE Transactions on, 2(1), 50-59.
- Chaturvedi, D. K. (2008). Soft computing: techniques and its applications in electrical engineering (Vol. 103). Springer.
- Chaturvedi, D. K., Premdayal, S. A., and Chandiok, A. (2013). Short Term Load Forecasting using Neuro-fuzzy-Wavelet Approach. International Journal of Computing Academic Research (IJCAR), ISSN, 2305-9184. Vol. 2, pp. 36-48.
- Cherkassky, V., and Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. Neural networks, 17(1), 113-126.
- Christiaens, K., and J. Feyen (2002), Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE SHE model, Water Resour. Res., 38(9), 1169, doi:10.1029/2001WR000478.
- Cobaner, M., Citakoglu, H., Kisi, O., and Haktanir, T. (2014). Estimation of mean monthly air temperatures in Turkey. Computers and Electronics in Agriculture,109, 71-79.
- Dadu, K. S., and Deka, P. C. (2013). Multistep lead time forecasting of hydrologic time series using daubechies wavelet–neural network hybrid model. International Journal of Scientific and Engineering Research, 4(10), 115-124.
- Daneshmand, H., Tavousi, T., Khosravi, M., and Tavakoli, S. (2015). Modeling minimum temperature using adaptive neuro-fuzzy inference system based on

spectral analysis of climate indices: A case study in Iran. Journal of the Saudi Society of Agricultural Sciences, 14(1), 33-40.

- Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. Information Theory, IEEE Transactions on, 36(5), 961-1005.
- De, S. S., and Debnath, A. (2009). Artificial neural network based prediction of maximum and minimum temperature in the summer monsoon months over India. Applied Physics Research, 1(2), p37.
- Deka, P. C., and Prahlada, R. (2012). Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. Ocean Engineering, 43, 32-42.
- Deka, P., andChandramouli, V. (2005). Fuzzy neural network model for hydrologic flow routing. Journal of Hydrologic Engineering, 10(4), 302-314.
- Ding-cheng, W., Chun-xiu, W., Yong-hua, X., and Tian-yi, Z. (2010, December). Air temperature prediction based on EMD and LS-SVM. In Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on (pp. 177-180). IEEE.
- Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability, Vol. 57. New York and London: Chapman and Hall/CRC.
- Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning (Vol. 1). Springer, Berlin: Springer series in statistics.
- Güler, I., and Übeyli, E. D. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. Journal of neuroscience methods, 148(2), 113-121.
- Gunn, S. R. (1998). Support vector machines for classification and regression.ISIS technical report, 14.

- Hatim, A. R.,(2014). A model tool for predicting of outdoor air temperatures on construction materials manufacture performance in Baghdad. ARPN Journal of Engineering and Applied Sciences. VOL. 9, NO. 7, JULY 2014.
- Jang, J. S. R. (1992). Self-learning fuzzy controllers based on temporal backpropagation. Neural Networks, IEEE Transactions on, 3(5), 714-723.
- Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system.Systems, Man and Cybernetics, IEEE Transactions on, 23(3), 665-685.
- Jang, J.S., and Sun, C.T. (1995).Neuro-fuzzy modeling and control. Proceedings of the IEEE, 83(3): 378-406.
- Jarvis, C. H., & Stuart, N. (2001). A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part I: The selection of “guiding” topographic and land cover variables. Journal of Applied Meteorology, 40(6), 1060-1074.
- Jusuf, S. K., & Wong, N. H. (2009, September). Development of empirical models for an estate level air temperature prediction in Singapore. InProceedings of the Second International Conference on Countermeasures to Urban Heat Islands (pp. 21-23).
- Kadu, P., Wagh. K., and Chatur, P. (2012) Analysis and Prediction of Temperature using Statistical Artificial Neural Network. IJCSMS Int J ComputSciManag Stud 12(2):2231–5268.
- Kim, T. W., and Valdés, J. B. (2003). Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering, 8(6), 319-328.
- Kisi, O., Shiri, J., and Nazemi, A. H. (2011). A Wavelet-genetic programming model for predicting shortterm and long-term air temperatures. J. Civil Engineering and Urbanism, 1, 25-37.

- Kumar, P. (2012). Minimum Weekly Temperature Forecasting using ANFIS.Computer Engineering and Intelligent Systems, 3(5), 1-5.
- Liu, X., Yuan, S., and Li, L. (2012). Prediction of temperature time series based on wavelet transform and support vector machine. Journal of computers, 7(8), 1911-1918.
- LeBaron, B., & Weigend, A. S. (1998). A bootstrap evaluation of the effect of data splitting on financial time series. IEEE Transactions on Neural Networks, 9(1), 213-220.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 11(7), 674-693.
- Mattera, D., and Haykin, S. (1999, February). Support vector machines for dynamic reconstruction of a chaotic system. In Advances in kernel methods(pp. 211-241). MIT Press.
- Mellit, A., Khorissi, N., and Salhi, H. (2007). An ANFIS-based forecasting for solar radiation data from sunshine duration and ambient temperature. InPower Engineering Society General Meeting, 2007. IEEE (pp. 1-6). IEEE.
- Min, J. H., and Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert systems with applications, 28(4), 603-614.
- Mohammadi, K., Shamshirband, S., Tong, C. W., Arif, M., Petković, D., &Ch, S. (2015). A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Conversion and Management, 92, 162-171.
- Mohandes, M. A., Halawani, T. O., Rehman, S., and Hussain, A. A. (2004). Support vector machines for wind speed prediction. Renewable Energy, 29(6), 939-947.

- Moosavi, V., Vafakhah, M., Shirmohammadi, B., and Behnia, N. (2013). A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. Water resources management, 27(5), 1301-1321.
- Moser, G., andSerpico, S. B. (2009). Modeling the error statistics in support vector regression of surface temperature from infrared data. Geoscience and Remote Sensing Letters, IEEE, 6(3), 448-452.
- Ninyerola, M., Pons, X., & Roure, J. M. (2000). A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. International Journal of Climatology, 20(14), 1823-1841.
- Nourani, V., Alami, M. T., & Aminfar, M. H. (2009). A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation.Engineering Applications of Artificial Intelligence, 22(3), 466-472.
- Osowski, S., and Garanty, K. (2007). Forecasting of the daily meteorological pollution using wavelets and support vector machine. Engineering Applications of Artificial Intelligence, 20(6), 745-755.
- Pal, N. R., Pal, S., Das, J., and Majumdar, K. (2003). SOFM-MLP: a hybrid neural network for atmospheric temperature prediction. Geoscience and Remote Sensing, IEEE Transactions on, 41(12), 2783-2791.
- Paniagua-Tineo, A., Salcedo-Sanz, S., Casanova-Mateo, C., Ortiz-García, E. G., Cony, M. A., and Hernández-Martín, E. (2011). Prediction of daily maximum temperature using a support vector regression algorithm. Renewable Energy,36(11), 3054-3060.
- Partal, T., and Kişi, Ö. (2007). Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. Journal of Hydrology, 342(1), 199-212.
- Patil, A. P., and Deka, P. C.(2015) Performance evaluation of hybrid Wavelet-ANN and Wavelet-ANFIS models for estimating evapotranspiration in arid regions of India. Neural Computing and Applications, 1-11.

- Pires, J. C. M., and Martins, F. G. (2011). Correction methods for statistical models in tropospheric ozone forecasting. Atmospheric Environment, 45(14), 2413-2417.
- Radhika, Y., and Shashi, M. (2009). Atmospheric temperature prediction using support vector machines. International Journal of Computer Theory and Engineering, 1(1), 1793-8201.
- Rajaee, T., Mirbagheri, S. A., Nourani, V., and Alikhani, A. (2010). Prediction of daily suspended sediment load using wavelet and neurofuzzy combined model.International Journal of Environmental Science and Technology, 7(1), 93-110.
- Rao, T., Rajasekhar, N., and Rajinikanth, T. V. (2012). An efficient approach for weather forecasting using support vector machine. In Proc. Int. Conf. Comput. Technol. Sci.(ICCTS), Int. Proc. Comput. Sci. Inf. Technol.(IPCSIT) (Vol. 47, pp. 208-212).
- Salcedo-Sanz, S., Deo, R. C., Carro-Calvo, L., andSaavedra-Moreno, B. (2015). Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms. Theoretical and Applied Climatology, 1-13.
- Savic, M., Mihajlovic, I., and Zivkovic, Z. (2013). An ANFIS-based air quality model for prediction of SO2 concentration in urban area. Serbian Journal of Management, 8(1), 25-38.
- Schölkopf, B., and Burges, C. J. (1999). Advances in kernel methods: support vector learning. MIT press.
- Shaharuddin, M., Mohd, A., Mohd, J., Othman, N., Karim, A., and Sopian, K. (2008). Application of Wavelet Transform on Airborne Suspended Particulate Matter and Meteorological Temporal Variations. WSEAS Trans. Environ. Dev, 4(2), 89-98.

- Shao, X., Xu, Y., and Yan, C. (2006). Wavelet analysis of rainfall variation in the Yellow River Basin. Actascientiarumnaturalium-universitatispekinensis, 42(4), 503.
- Singh, S., Bhambri, P., and Gill, J. (2011). Time series based temperature prediction using back propagation with genetic algorithm technique.International Journal of Computer Science, 8(5), 28-35.
- Smith, B. A., Hoogenboom, G., and McClendon, R. W. (2009). Artificial neural networks for automated year-round temperature prediction. Computers and Electronics in Agriculture, 68(1), 52-61.
- Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression.Statistics and computing, 14(3), 199-222.
- Smola, A.J. and Schölkopf, B. (1998)."A Tutorial on Support Vector Regression". NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK.
- Srivastav, R. K., Sudheer, K. P., & Chaubey, I. (2007). A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. Water Resources Research, 43(10).
- Stone, M. (1974), Cross validation choice and assessment of statistical predictions, J. R. Stat. Soc., Ser. B, 36, 44–47.
- Strang, G., and Nguyen, T. (1996). Wavelets and filter banks. SIAM.
- SujayRaghavendra, N., and Deka, P. C. (2015). Forecasting monthly groundwater level fluctuations in coastal aquifers using hybrid Wavelet packet–Support vector regression. Cogent Engineering, 2(1), 999414.
- Tasadduq, I., Rehman, S., and Bubshait, K. (2002). Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. Renewable Energy, 25(4), 545-554.

- Vapnik, V. (1995).“The Nature of Statistical Learning Theory”, Springer Verlag, New York.
- Vapnik, V. N., and Vapnik, V. (1998). Statistical learning theory (Vol. 1). New York: Wiley.
- Vong, C. M., Ip, W. F., Wong, P. K., and Yang, J. Y. (2012). Short-term prediction of air pollution in macau using support vector machines. Journal of Control Science and Engineering, pp-11.
- Wagener, T. (2003), Evaluation of catchment models, Hydrol. Processes, 17, 3375–3378.
- Wei, S., Song, J., and Khan, N. I. (2012). Simulating and predicting river discharge time series using a wavelet- neural network hybrid modelling approach. Hydrological Processes, 26(2), 281-296.
- Yarar, A. (2014). A hybrid wavelet and neuro-fuzzy model for forecasting the monthly streamflow data. Water resources management, 28(2), 553-565.
- Yun, Z., Quan, Z., Caixin, S., Shaolan, L., Yuming, L., and Yang, S. (2008). RBF neural network and ANFIS-based short-term load forecasting approach in real-time price environment. Power Systems, IEEE Transactions on, 23(3), 853-858.

PUBLICATIONS BASED ON PRESENT RESEARCH WORK

International Journal

- Karthika, B. S., and Deka, P. C.(2016). "Modeling of Air Temperature using ANFIS by Wavelet Refined Parameters." International Journal of Intelligent Systems and Applications ,1 , 25-34. DOI: 10.5815/ijisa.2016.01.04.
- Karthika, B. S., and Deka, P. C. "Modeling of air temperature by Wavelet-ANFIS hybridization using meteorological and air pollutant parameters." Submitted to Computers and electronics in agriculture. Elsevier Journal. (Under review)

International conferences

- Karthika, B. S., and Deka, P. C. (2015). "Prediction of Air Temperature by Hybridized Model (Wavelet-ANFIS) Using Wavelet Decomposed Data".*AquaticProcedia*,*4*,1155-1161. DOI:10.1016/j.aqpro.2015.02.174(**Elsevier**).

- Karthika, B. S., and Deka, P. C. (2015). "Prediction Of Air Temperature By Conjunctive Method (Wavelet-Anfis) Using Subseries Of Wavelet Decomposed Data", 4th INTERNATIONAL ENGINEERING SYMPOSIUM (IES2015)- held at Kumamoto University, Japan during March 4-6.

RESUME

Name	:	Karthika. B. S
Date of Birth	:	08-04-1982
Address	:	S/o Sridhara.B.N, Hulimande, Bananki, Devangi(post), ThirthahalliTq, ShimogaDist, Karnataka, India-577415
Mobile	:	(+91) 90084-20322
Email(s)	:	bskarthikgowda@gmail.com

Qualification	:	B.E. (Civil), M.Tech. (RS&GIS)
Journal Publications	:	1
No. of Conference Publications	:	2