Forecasting of Monthly Mean of Maximum Surface Air Temperature in India

K. Anitha Kumari[1*], Naveen Kumar Boiroju[2], P. Rajashekara Reddy[1]

[1]Department of Statistics, Sri Venkateswara University, Tirupati, Andhra Pradesh, INDIA.

[2]National Academy of Agricultural Research Management, Hyderabad, Andhra Pradesh, INDIA.

[*]Corresponding Address:

anitha.kumari37@gmail.com

Research Article

Abstract: In this paper, forecasting of monthly mean of maximum surface air temperature of India using seasonal autoregressive integrated moving average (SARIMA) model, feed forward neural networks (FFNN) and hybrid models are discussed. Prediction ability of the models compared with the various statistical methods. The results show that there is an increasing trend in the monthly mean of maximum SAT in India.

Keywords: ARIMA, Neural Networks, Hybrid Models, Surface Air Temperature.

1. Introduction

The surface air temperature (SAT) represents an important element of a regional climate. Therefore maximum and minimum values of SAT are commonly used as an input in various environmental applications, including agriculture, forestry, fisheries and ecological models to predict likely changes at field and landscape level attributes. Changes in SAT during the last century widely discussed among researchers and scientists in the field of climate change. The increase in temperature is global phenomena and many scientists are pondering upon its likely impact on human beings. At national level, over the past century increase of 0.4 ^{0}C surface temperature has been observed (NAPCC, 2008). The likely increase in SAT can have severe effect on crops and livestock if the intensity will be higher. There is speculation that the mean annual temperature would rise up by 1 to 2^o C. The increase in maximum and minimum surface air temperature will lead to increasing evapotranspiration from the plants and eventually the deepening of water table, effect on the physiology of food crops and reduced yield as well as the quality and nutritional values.. Thus our food security may be sooner at risk. The rivers and ponds will get dried due to increase evaporation which may give a call for scarcity of water in the near future. The increase in temperature would also accompany the increasing consumption of energy particularly electricity in urban areas. So the energy industries will be in crisis. Increasing in temperature would also lead to melting of polar ice which causes to increase the sea level and there would be inundation of coastal areas. Also the marine ecosystem may be endangered. Vagaries in monsoon and erratic rainfall could also be noticed. Increase in temperature would also effect on forest and animal ecosystem. There would be migration of birds and animals, severe incidents of forest fire could be seen. SAT prediction is of a concern in environment, industry and agriculture. Thus the study of variability in SAT is very important to deal with the likely impacts. In this context SAT forecasting is useful in determining the probability of tornados and flood occurrence in advance. Fan and Dool (2008) analyzed the monthly land surface air temperature. Mostovoy et al. (2006) discussed on statistical estimation of daily maximum and minimum air temperatures. Modelling of daily temperature extremes and climate changes over Europe was discussed by Kjellstrom et al. (2007). Kitoh and Mukano (2009) present the variability in daily and monthly surface air temperature with the help of multi-model global warming experiments. The fluctuations of surface air temperatures were discussed by Hasanean (2001). Application of neural networks for the prediction of hourly mean surface temperatures was discussed by Tasadduq et al. (2005). Smith (2006), Afzali et al. (2011), Shrivastva et al. (2012) discussed on climate prediction using neural networks. Box-Jenkins methodology for the forecasting of air and water temperatures was discussed by Stein and Lloret (2001). In this paper, forecasting of monthly mean of maximum surface air temperature of India using seasonal autoregressive integrated moving average (SARIMA) model, feed forward neural networks (FFNN) and hybrid models are discussed. Prediction ability of the models compared using the error measures such as mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean squared error (RMSE).

2. Methodology

This section presents the methodology of SARIMA models, neural networks, hybrid models and the testing procedures for equal prediction accuracy of the models. Seasonal autoregressive integrated moving average (SARIMA) model for any variable involves mainly four steps: Identification, Estimation, Diagnostic checking and Forecasting. The basic form of SARIMA model is denoted by $SARIMA(p,d,q)X(P,D,Q)_S$ and the model is given by $\phi_p(B)\Phi_P(B^s)\nabla^d\nabla_s^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t$, where Z_t is the time series value at time t and ϕ, Φ, θ *and* Θ are polynomials of order of p, P, q and Q respectively. B is the backward shift operator, $B^s Z_t = Z_{t-s}$ and $\nabla = (1-B)$. Order of seasonality is represented by s. Non-seasonal and seasonal difference orders are denoted by d and D respectively. White noise process is denoted by a_t. The Box-Jenkins methodology involves four steps (Box et al., 1994): (i) identification (ii) estimation (iii) diagnostic checking and (iv) forecasting. First, the original series must be transformed to become stationary around its mean and its variance. Second, the appropriate order of p and q must be specified using autocorrelation and partial autocorrelation functions. Third, the value of the parameters must be estimated using some non-linear optimization procedure that minimizes the sum of squares of the errors or some other appropriate loss function. Diagnostic checking of the model adequacy is required in the fourth step. This procedure is continued until an adequate model is obtained. Finally, the future forecasts generate using minimum mean square error method (Box et al. 1994). SARIMA models are used as benchmark models to compare the performance of the other models developed on the same data set. The iterative procedure of SARIMA model building was explained by Kumari et al. (2013), Boiroju (2012), Rao (2011) and Box et al. (1994). Recently, ANN models have received increasing attention as decision making tools. Plentiful studies have shown that neural networks can be one of the very useful tools in time series forecasting. Neural networks have general nonlinear function mapping capability which can approximate any continuous function with arbitrarily desired accuracy (Zhang et al., 1998). Neural networks are data driven and data mining techniques provides a single platform for many of the statistical applications (Boiroju, 2012). Feed-forward neural networks (FFNN) is the most popular neural networks models for time series forecasting applications. The input nodes are the previous lagged observations, while the output provides the forecast for the future values. The hidden nodes with appropriate nonlinear transfer functions are used to process the information received by the input nodes. The FFNN model with a single hidden layer can be written as $Z_t = \beta_0 + \sum_{j=1}^{q}\beta_j f\left(\sum_{i=1}^{p}\gamma_{ij}Z_{t-i} + \gamma_{oj}\right) + \varepsilon_t$, where p is the number of input nodes, q is the number of hidden nodes, f is a hyperbolic tangent function. $\{\beta_j, j = 0,1,\cdots,q\}$ is a vector of weights from the hidden to output nodes and $\{\gamma_{ij}, i = 0,1,\cdots,p; j = 1,2,\cdots,q\}$ are weights from the input to hidden nodes. The input and output variables transformed in to the range of the activation functions. Usually, standardization of the input and output variables is sufficient to achieve the acceptable accuracy of the models. Faraway and Chatfield (1998) discovered that FFNN model performs better than the SARIMA model and it also reduces the mean square errors of out-of-sample prediction. A comprehensive review of the current status of research in this area is provided by Kumari et al. (2013), Rao (2011), Boiroju (2012) and Zhang et al. (1998).

The hybrid models have been introduced to overcome the deficiency by using individual models. The hybrid models merge different methods to improve the prediction accuracy (Sallehuddin, 2007). Both the theoretical and empirical findings in the literature show that combining different method can be an effective and efficient way to improve forecasts. Therefore hybrid ARIMA and ANNs methods have been used for modelling both the linear and non-linear patterns equally well (Aladag et al. 2009). A hybrid ARIMA and support vector machines model was discussed by Pai and Lin (2005). A hybrid ARIMA and FFNN model was given by Zhang (2003) and a combined model of SARIMA and FFNN model was presented by Tseng et al. (2002). Aladag et al. (2009) modified the Zhang's hybrid approach by considering Elman's recurrent neural network for the non-linear component in the model. A comparative study of autoregressive neural network hybrids was carried out by Taskaya and Casey (2005). The hybrid SARIMA and FFNN models, composed of seasonal-linear and non-linear components as follows: $\hat{Z}_t = \hat{T}_t + \hat{e}_t$ where Z_t denotes original time series, T_t denotes the seasonal linear component estimated by SARIMA model and e_t is the residual obtained from the SARIMA model, which denotes the non-linear component estimated by FFNN model. FFNN model developed for the residuals generated by the SARIMA model as function of previous residuals to estimate the non-linear component.

3. Results

This section presents the forecasting models for monthly mean maximum surface air temperature using SARIMA, FFNN and hybrid models. The monthly mean of maximum surface air temperature in degrees Celsius data of all India during 1950 to 2007 is collected from Indian Institute of Tropical Meteorology (IITM), Pune, India. From the given data last five years data during 2003-2007 are used as an out-of-sample set to measure the predictability of the selected models using mean absolute error, mean absolute percentage error and root mean squared error. Forecasts are generated for the years from January 2008 to December 2013.

The time plot (Figure 1) of the monthly maximum surface air temperature during 1950-2002 is given below and it clearly indicates that the series is seasonal.

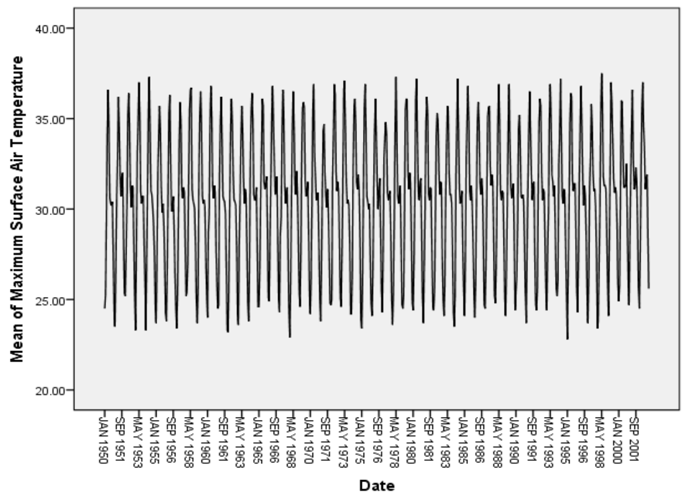

Figure 1: Time plot of monthly mean of maximum SAT in India

The Box-Jenkins methodology is applied to the given data to obtain an adequate model and the model parameters are estimated with the help of SPSS software.

Table 1: SARIMA model parameters

Transformation	**Parameter**		**Estimate**	**SE**	**t-statistic**	**Sig.**
No Transformation	AR	Lag 1	0.717	0.074	9.685	0
	MA	Lag 1	0.459	0.094	4.892	0
	Seasonal Difference		1			
	MA, Seasonal	Lag 1	0.965	0.029	33.372	0

The SARIMA (1, 0, 1) x (0, 1, 1)$_{12}$ forecasting model for the given data is obtained as $(1-0.717B)\nabla^{1}_{12}\tilde{Z}_t = (1-0.459B)(1-0.965B^{12})a_t$. Future forecasts of maximum SAT using this model are presented in the Table 3.

For building the FFNN model, the in-sample data set is partitioned into training set with 70 percent of the observations and the remaining 30 percent of the observations taken under the testing set. The FFNN model consists of an input layer, a hidden layer and an output layer. Input layer consists of 14 units representing the months (numbers from 1 to 12), Z_{t-1} and Z_{t-12} values and the output layer consists of only one neuron, which represents the forecast value ($\hat{Z}_t$) of the series. The optimum number of hidden neurons is four obtained using forward selection procedure. The network is trained with the hyperbolic tangent function as an activation function under batch mode with the scaled conjugate learning method. The network is trained until the number of epochs is equivalent to 10000 as a stopping criterion. SPSS software is used to train the network and with these specifications the following FFNN model is obtained.

$$\hat{Z}_s = -0.334 - 0.708\,H(1:1) - 0.661\,H(1:2) + 0.614\,H(1:3) + 0.865\,H(1:4)$$

$$H(1:1) = Tanh\begin{pmatrix} -0.52 + 0.607I(M=1) - 0.21I(M=2) - 0.922I(M=3) \\ -0.349I(M=4) - 0.630I(M=5) - 0.317I(M=6) \\ +0.618I(M=7) + 0.386I(M=8) - 0.356I(M=9) \\ +0.409I(M=10) - 0.75I(M=11) - 0.590I(M=12) \\ +0.100S(Z_{t-1}) - 0.098S(Z_{t-12}) \end{pmatrix}$$

$$H(1:2) = Tanh\begin{pmatrix} -0.299 + 0.608I(M=1) + 0.425I(M=2) \\ -0.956I(M=3) + 0.008I(M=4) \\ -0.287I(M=5) - 0.469I(M=6) \\ +0.093I(M=7) - 0.236I(M=8) + 0.713I(M=9) \\ +0.068I(M=10) + 0.339I(M=11) + 0.010I(M=12) \\ -0.372S(Z_{t-1}) - 0.616S(Z_{t-12}) \end{pmatrix}$$

$$H(1:3) = Tanh\begin{pmatrix} 0.475 - 0.445I(M=1) + 0.101I(M=2) \\ -0.223I(M=3) + 0.220I(M=4) \\ +0.574I(M=5) - 0.076I(M=6) \\ -0.244I(M=7) - 0.353I(M=8) \\ +0.353I(M=9) - 0.234I(M=10) - 0.381I(M=11) \\ -0.405I(M=12) - 0.089S(Z_{t-1}) - 0.332S(Z_{t-12}) \end{pmatrix}$$

$$H(1:4) = Tanh\begin{pmatrix} 0.060 - 0.309I(M=1) + 0.095I(M=2) \\ -0.411I(M=3) + 0.682I(M=4) + 0.280I(M=5) \\ +0.286I(M=6) - 0.010I(M=7) + 0.028I(M=8) \\ -0.023I(M=9) + 0.298I(M=10) - 0.539I(M=11) \\ -0.617I(M=12) + 0.692S(Z_{t-1}) - 0.134S(Z_{t-12}) \end{pmatrix}$$

M= month, $S(Z_{t-1}) = (Z_{t-1} - \mu(Z_{t-1}))/\sigma(Z_{t-1})$, $S(Z_{t-12}) = (Z_{t-12} - \mu(Z_{t-12}))/\sigma(Z_{t-12})$ and $I(A)$ is an indicator function.

The hybrid forecasting model is a combination of SARIMA and FFNN models. Let the hybrid model is $\hat{Z}_t = \hat{T}_t + \hat{e}_t$ where $\hat{T}_t$ is estimated using $(1-0.717B)\nabla_{12}^{1} T_t = (1-0.459B)(1-0.965B^{12})a_t$ and $\hat{e}_t = 0.061 + 0.70356E_t$ where $E_t = -0.039 - 0.217H(1:1) + 0.401H(1:2)$

$$H(1:1) = Tanh\begin{bmatrix} (-0.155 - 0.029I(M=1) - 0.187I(M=2) \\ +0.416I(M=3) - 0.245I(M=4) + 0.125I(M=5) \\ -0.021I(M=6) - 0.175I(M=7) - 0.007I(M=8) \\ +0.047I(M=9) - 0.322I(M=10) - 0.114I(M=11) \\ +0.329I(M=12) \end{bmatrix}$$

$$H(1:2) = Tanh\begin{bmatrix} (0.068 - 0.341I(M=1) - 0.397I(M=2) \\ +0.175I(M=3) + 0.062I(M=4) \\ -0.059I(M=5) + 0.174I(M=6) \\ -0.328I(M=7) - 0.063I(M=8) \\ +0.073I(M=9) - 0.092I(M=10) \\ +0.375I(M=11) + 0.048I(M=12) \end{bmatrix}$$

Table 2 depicts the forecasting errors of the three models and it is observed that the FFNN model has minimum error compared to other two models in both the samples. The hybrid model has minimum error as compared with the SARIMA model but not to the FFNN model. This model improved the prediction accuracy of the SARIMA model and performed well at the out-of-sample set as compared with other two models. Even though the errors of the three models are approximately same, FFNN model preferred for forecasting the monthly mean of maximum surface air temperature since this model can predict both the linear and nonlinear patterns in the series.

Table 2: Forecasting errors of the models

Sample	Error	SARIMA	FFNN	Hybrid
In-Sample	MAE	0.526	0.508	0.519
	MAPE	1.742	1.696	1.722
	RMSE	0.706	0.681	0.704
Out-of-Sample	MAE	0.512	0.498	0.489
	MAPE	1.677	1.631	1.603
	RMSE	0.820	0.773	0.801

The forecasts of monthly mean of maximum surface air temperature in India for the year 2013 given in the Table 3 manifested that the monthly mean surface air temperature varying from 24^{o}C to 36.5^{o}C during the year 2013. The minimum of maximum surface air temperature observed in the month of January (24.33 ^{0}C) whereas, the maximum of maximum surface air temperature observed in the month of May (36.48 ^{0}C).

Table 3: Forecasts of monthly mean of maximum SAT for the year 2013

Month, 2013	SARIMA	FFNN	Hybrid
Jan	24.19	24.33	24.13
Feb	26.73	26.94	26.68
Mar	31.08	31.28	31.24
Apr	34.84	35.02	34.92
May	36.40	36.48	36.39
Jun	34.56	34.86	34.64
Jul	31.49	31.58	31.45
Aug	30.69	30.82	30.70
Sep	31.15	31.39	31.07
Oct	30.99	31.15	31.04
Nov	28.20	28.29	28.34
Dec	25.09	25.20	25.09

The following table presents the cumulative growth rate (CGR) in the monthly mean of maximum SAT in India during the previous five years. It is observed that, the mean of maximum SAT is increasing at 2.1 per centin the month of January whereas a negative growth was observed in the month of March and the same is decreasing at the rate of -0.64 per centin the previous five years.

Table 4: Compound growth in monthly mean of maximum SAT in India

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
CGR (%)	2.10	1.21	-0.64	0.08	-0.11	-0.54	-0.16	0.23	0.10	0.87	0.17	0.20

With the above compound annual growth rates for each month, we can expect the minimum of maximum temperature in India may touch 30^{o}C by the year 2015 and it will increase to 33.28^{o}C by the year 2020, while the maximum of maximum temperature will reaches 43.10^{o}C by the year 2015 and the same may be expected to 47.83^{o}C by the year 2020.

4. Conclusion

Advances in monthly average of maximum SAT forecasts are important for planning in different economic sectors, such as agriculture, energy industry, food industry, tourism sector etc. Prediction of the solar radiation, energy consumption, ground water level, sea water level, rainfall etc. is related to SAT forecasting. It will give alarm of likely events that are going to happen in near future and help to frame the strategies to deal with. It is observed from the above study, FFNN model performing well at forecasting of monthly mean of maximum SAT in India as compared to the SARIMA and hybrid models. The hybrid model performing better than the SARIMA model and FFNN model is better than the hybrid model at forecasting of maximum SAT. The forecasts reveal that the monthly mean of maximum SAT is around 30^{o}C-36^{o}C during March to October, 2013 and even it is more in the winter season (November to February, 2013) and which varies in between 28^{o}C and 24^{o}C. The monthly mean of maximum SAT is slowly increasing and fluctuating between 23^{o}C and 37^{o}C during 2003-2007 and which can be expected to increase in the near future also. The highest increase in the maximum temperature during the previous five years was observed in the months of January and February and the same was decreasing in the months of March and June, which is evidence that the winter is extending to the month of March and monsoon will arrive in the early of June.

Acknowledgements

The authors are grateful to the Department of Science and Technology (DST) New Delhi for Financial assistance through the award of Inspire Fellowship. We are also indebted to the Department of Statistics, S.V.U College of Sciences, Sri Venkateswara University, Tirupati, Andhra Pradesh, India for providing the space and facilities to complete the above Research work. The authors are grateful in this regard

References

1. Afzali, M., Afzali, A. and Zahedi, G. (2011). Ambient Air Temperature Forecasting Using Artificial Neural Network Approach, 2011 International Conference on Environmental and Computer Science, IPCBEE Vol.19, IACSIT Press, Singapore.
2. Aladag, H.A., Egrioglu, E., Kadilar, C. (2009). Forecasting Nonlinear Time Series with a Hybrid Methodology. Applied Mathematic Letters, 22, pp. 1467-147.
3. Boiroju, N.K. (2012). Forecasting Foreign Exchange Rates using Neural Networks, Unpublished Ph.D. Thesis, Department of Statistics, Osmania University, Hyderabad.
4. Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). Time Series Analysis Forecasting and Control, 3rd ed., Englewood Cliffs, N.J. Prentice Hall.

5. Fan,Y and Dool, H. (2008). A global monthly land surface air temperature analysis for 1948–present. Journal Of Geophysical Research, Vol. 113, D01103, Doi:10.1029/2007jd008470, 2008.
6. Faraway, J. and Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the airline data, Journal of the Royal Statistical Society, Series C, Vol. 47, 2, pp. 231-250.
7. Hasanean. H. M. (2001). Fluctuations of surface air temperature in the Eastern Mediterranean. Theoretical Applied Climatology, 68, pp. 75-87.
8. Kitoh. A, Mukano. T. (2009). Changes in Daily and Monthly Surface Air Temperature Variability by Multi-Model Global Warming Experiments. Journal of the Meteorological Society of Japan, Vol. 87, No. 3, pp. 513--524, 2009.
9. Kjellström, E. Bärring, L., Jacob, D., Jones, R., Lenderink, G. and Schär, C. (2007). Modelling daily temperature extremes: recent climate and future changes over Europe. Climatic Change (2007) 81, pp. 249–265 DOI 10.1007/s10584-006-9220-5.
10. Kumari, A. K., Boiroju, N.K., Ganesh, T. and Reddy, P.R. (2013), Forecasting Surface Air Temperature using Neural Networks, International Journal of Mathematics and Computer Applications Research (IJMCAR), Vol. 3, Issue 2, pp. 65-78.
11. Mostovoy. G. V, King. R. L, Reddy, K.R., Kakani,V.G., Filippova, M.G. (2006). Statistical Estimation of Daily Maximum and Minimum Air Temperatures from MODIS LST Data over the State of Mississippi. GIScience & Remote Sensing, 2006, 43, No. 1, pp. 78-110.
12. NAPCC. (2008). National action plan on climate change, Prime minister's council on climate change, Government of India. (Available online at http://pmindia.gov.in/climate_change_english.pdf accessed on 10.10.2013).
13. Pai, P.F. and Lin, C.S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega - International Journal of Management Science, 33, pp. 497-505.
14. Rao, S.S. (2011). Forecasting of Monthly Rainfall in Andhra Pradesh using Neural Networks, Unpublished Ph.D. Thesis, Department of Statistics, Osmania University, Hyderabad.
15. Sallehuddin, R., Shamsuddin, S. M. H., Hashim, S. Z. M., Abraham, A. (2007). Forecasting time series data using hybrid grey relational artificial neural network and auto regressive integrated moving average model, Neural network world, 6/07, pp. 573-605.
16. Shrivastava, G., Karmakar, S., Kowar, M.K., Guhathakurta, P. (2012). Application of Artificial Neural Networks in Weather Forecasting: A Comprehensive Literature Review, International Journal of Computer Applications, Vol. 51, No.18, August 2012.
17. Smith, B.A. (2006). Air temperature prediction using artificial neural networks, M.Sc. (Thesis), University of Gerogia.
18. Stein, M. and Lloret, J. (2001). Forecasting of Air and Water Temperatures for Fishery Purposes with Selected Examples from Northwest Atlantic, J. Northw. Atl. Fish. Sci., Vol. 29, pp. 23-30.
19. Tasadduq. I., Rehman, S., Bubshait, K. (2005). Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia. Renewable Energy, 25, pp. 545-554.
20. Taskaya, T., and Casey, M. C. (2005). A comparative study of autoregressive neural network hybrids. Neural Networks, 18, pp. 781–789.
21. Tseng, F.M., Yu, H.C., and Tzeng, G.H. (2002). Combining neural network model with seasonal time series ARIMA model. Technological Forecasting and Social Change, 69, pp. 71–87.
22. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, pp. 159–175.
23. Zhang, G., Patuwo, B.E. and Hu, M.Y. (1998). Forecasting with Artificial Neural Networks: The State of the Art, International Journal of Forecasting, 14, pp. 35-62.