Research Journal of Applied Sciences, Engineering and Technology 7(11): 2306-2310, 2014
DOI:10.19026/rjaset.7.530
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: July 19, 2013 Accepted: July 31, 2013 Published: March 20, 2014

Research Article

Linear Regression Based Lead Seven Day Maximum and Minimum Air Temperature Prediction in Chennai, India

[1]K. Ramesh, [2]R. Anitha and [3]P. SelvaGopal
[1]Regional Centre, Anna University, Tirunelveli,
[2] K. S. Rangasamy College of Technology, Tiruchengode, Tamil Nadu, India
[3]Al Musanna College of Technology, Sultanate of Oman

Abstract: The surface temperature is the key determinant for vegetation, animals and human livelihood in a particular location of earth. Timely prediction of minimum and maximum temperature will help in planning and governing very hot and very cold climate. In this study numerical weather parameters based lead seven day minimum and maximum temperature prediction models using multiple linear regression is developed at the location Chennai, India. The result of the analysis states that regression based minimum temperature prediction models provide better accuracy than maximum temperature forecast models with the highest R^2 and lowest MAE, RMSE in independent test dataset. The analysis also emphasizes that the prediction performance is good at smaller lead days and it decreases gradually to higher lead days for both minimum and maximum temperature.

Keywords: Linear regression, temperature forecast

INTRODUCTION

Weather and climate influence the lives of human, animals and vegetation. Particularly temperature has major impact in our lives. Heat exhaustion or heat stroke takes thousands of lives during summer every year all over the world. Over 500 thousand chickens perished in Georgia alone during a two-day period at the peak of the summer heat (Donald, 2011).

Accurate and timely prediction of temperature will help to take precautionary measures. Accurate calculation of what the atmosphere will do in the future is challenging because of the dynamic nature of the atmospheric environment influencing the temperature observed in the earth surface. The objective of this study is to develop single point minimum and maximum temperature prediction models using Multiple Linear Regression (MLR). Scientific Community has recommended many linear and nonlinear temperature prediction techniques, but still MLR is selected since linear models often produce better forecasts than nonlinear models even when the data are nonlinear (Chatfield, 2009) and also statistical schemes require little computation time to make a forecast.

Statistical forecasting techniques, MOS is a powerful tool which generates models based on linear regression which has given significant results in forecasting maximum and minimum temperature (Taylor and Leslie, 2005). Stepwise linear regression approach is used for estimating daily maximum and minimum air temperature with MOD erate-resolution Imaging Spectroradiometer (MODIS) land surface temperature data in east Africa (Shengpan *et al*., 2012). Shengpan *et al*. (2012) in the study correlates surface temperature with air temperature and tries to predict the minimum and maximum air temperature. In short-term (6 to 24 h) single-station forecasts using a multiple regression model for predicting temperature anomalies, the RMSE of the temperature forecasts is of 1.78°C for the 6-h forecast and 2.28°C for the 12-h forecast (Christoph *et al*., 1999). In this study, regression models are created using the predictors listed in Table 1 to forecast the maximum and minimum temperature.

Table 1: Predictors used to formulate prediction models

Predictor variable	Abbreviations	Units
Mean temperature	Tm	°C
Mean dew point	DP	°C
Maximum sea level pressure	SLP	hPa
Mean visibility	Vis	Km
Mean wind speed	WS	Km/h
Maximum wind speed	WSmax	Km/h
Precipitation	P	Mm
Minimum temperature	Tmin	°C
Maximum temperature	Tmax	°C

DATA AND METHODOLOGY

Data: The atmospheric parameters (Table 1) recorded daily in Chennai, India. (Latitude: 13°4'7.3" N, Longitude: 80° 14'48.33" E) are used as predictors to forecast next seven days minimum temperature and maximum temperature. The observed predictor dataset

Corresponding Author: K. Ramesh, Regional Centre, Anna University, Tirunelveli, Tamil Nadu, India

for analysis is obtained from National Data Centre of National Centre for Environmental Prediction (NCEP), USA. (http://www.ncdc.noaa.gov/oa/ncdc.html). For the present study, the overall period used for analysis covers for a duration of nine years (1995-2003). The data from 1996 through 2003 are used for the training purpose and dataset of one year (1995) are used to validate the performance of the derived models.

Methodology:

Multiple Linear Regressions (MLR): Multiple Linear Regression is a classical linear statistical forecasting technique which allows estimation of the accuracy of predictions. MLR models the relationship between multiple variables by fitting a linear equation to observed data. Generally the form of regression model is:

$$y_i = \beta_0 + \beta_1 x_{i1} \ldots\ldots + \beta_p x_{ip} + \varepsilon_i \quad (1)$$

where,

y_i = The predicted variable
β_0 = The intercept
β_1 = Measures the change in y_i with respect to x_{i1}
β_p = Measures the change in y_i with respect to x_{ip}
$x_{i1} \ldots\ldots x_{ip}$ = Predictor variable and ε_i the error

Models: Minimum and maximum forecast models for predicting the minimum temperature (T_{min}) and maximum temperature (T_{max}) that may be felt in next 7 days are devised using the dataset from 1996 to 2003. The data set comprises of nine atmospheric parameters, one dependent and eight independent parameters as shown in Table 1. Fourteen regression models are formulated seven each for T_{min} temperature and T_{max} temperature prediction models are devised. Formulated models fitness is measured by coefficient of determination (R^2), it shows how well the independent variables explain variation in the dependent variable (Wilks, 2006).

The prediction model performance is validated by deploying the models with independent verification dataset and calculating MAE, RMSE and observed vs. predicted correlation coefficient. The MAE is the arithmetic average of the absolute values of the differences between the members of each pair and RMSE is the square root of average squared difference between the forecast and observation pairs. The forecast is perfect if MAE and RMSE are equal to zero. Correlation coefficient between observed and predicted value is another accuracy measure for validating the models.

The absolute error or the residual e_i is obtained by:

$$e_i = |f_i - y_i| \quad (2)$$

where,

f_i = The observed value
y_i = The predicted value

The Mean Absolute Error (MAE) is used to measure how close forecasts are to the eventual outcomes. The MAE is given by:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i| \quad (3)$$

Root Mean Square Error (RMSE) measure of the differences between predicted value and the values actually observed. It is given by:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \quad (4)$$

Daily minimum temperature is used as eighth predictor of minimum temperature prediction models and daily maximum temperature is used as eighth predictor of maximum temperature prediction models.

RESULTS AND DISCUSSION

The performance of the devised forecast models accuracy is assessed by deploying the models with one year data (1995).

Minimum temperature forecast models assessment: The accuracy of prediction for T_{min} temperature are validated and summarized in Table 2 and Fig. 1. The MAE often is used as a verification measure for temperature forecasts in the United States. The MAE for lead day one is 0.59°C and it increases slowly to 1.09°C for lead day seven. The RMSE calculated for these models also gives a least error of less than 1°C for day one and two. It is noted from the analysis the average RMSE for all seven lead days is 1.15°C. The

Table 2: Performance summary of all regression models formulated

	Max. temp			Min. temp		
Lead days	MAE	RMSE	Correlation	MAE	RMSE	Correlation
1	0.86	1.19	0.90	0.59	0.81	0.93
2	1.07	1.44	0.84	0.73	0.92	0.91
3	1.12	1.49	0.85	0.88	1.14	0.85
4	1.24	1.71	0.76	0.98	1.26	0.82
5	1.05	1.37	0.87	0.99	1.31	0.81
6	1.22	1.67	0.76	1.04	1.39	0.80
7	1.22	1.60	0.80	1.09	1.41	0.78

Max.: Maximum; Min.: Minimum

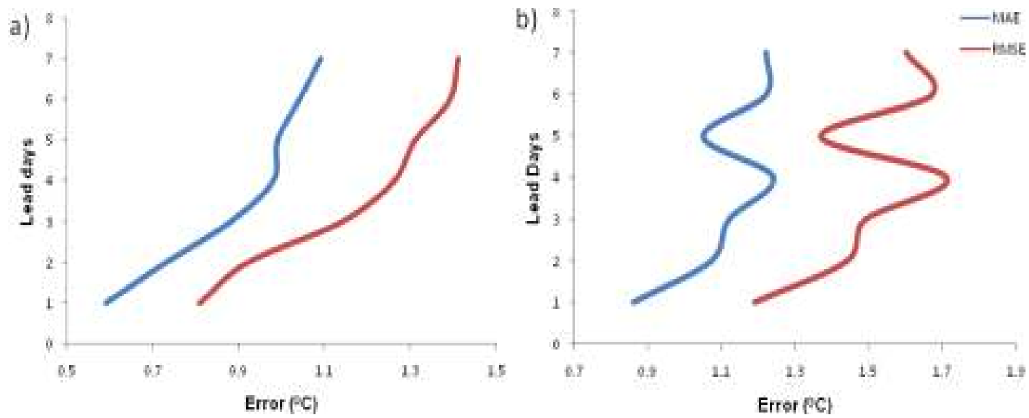

Fig. 1: Performance analysis (MAE and RMSE) of prediction models, a) minimum temperature, b) maximum temperature

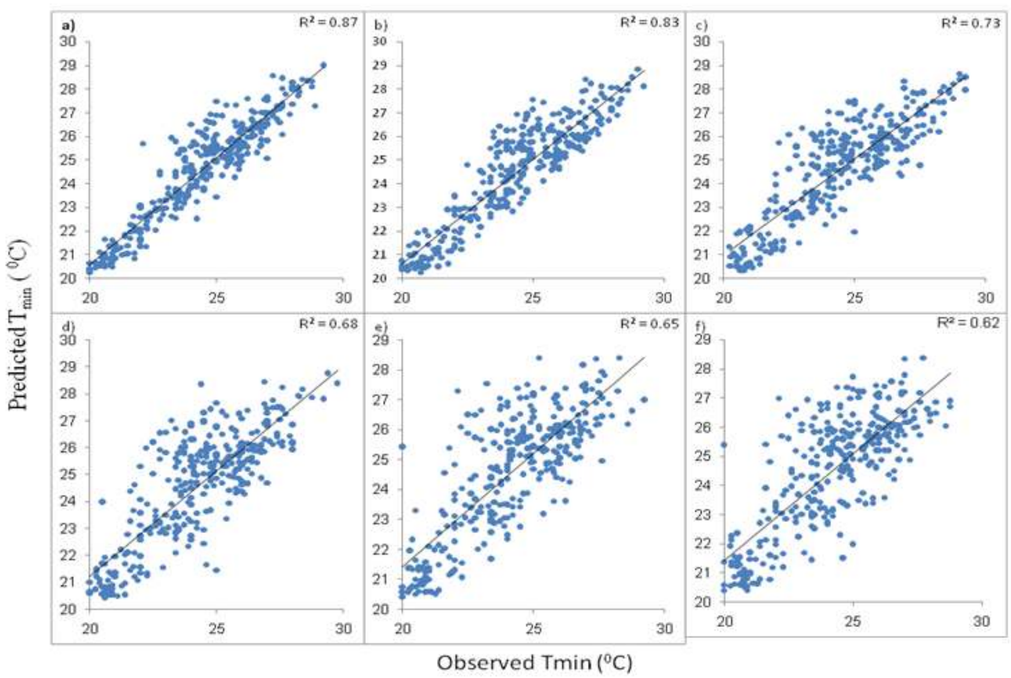

Fig. 2: Association between observed and predicted minimum temperature (°C) (T_{min}), (a-c) lead day one through day three, (d-f) lead day five to seven

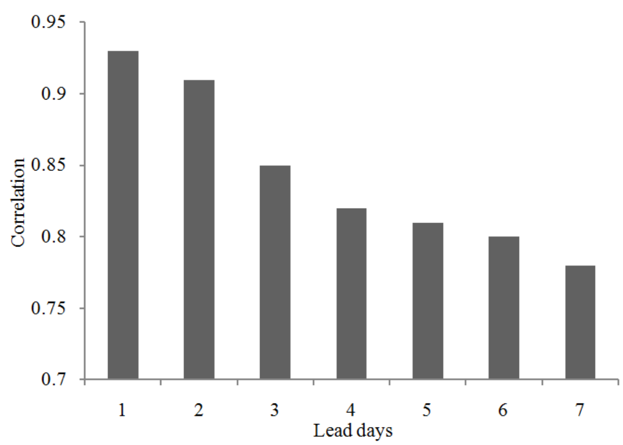

Fig. 3: Correlation of observed verses predicted minimum temperature

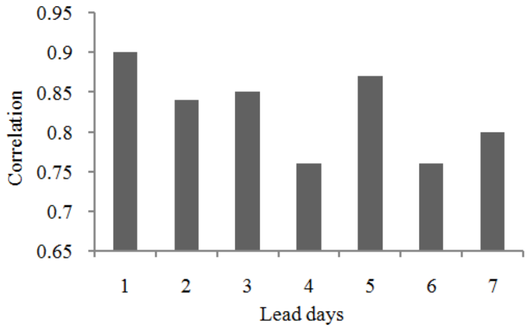

Fig. 4: Correlation of observed verses predicted maximum temperature

correlation coefficient between observed and predicted for day one determines that the predicted values are 93% correlated with the observed. It is also renowned that correlation between observed and predicted is above 80% for all seven lead days (Fig. 2). The results also suggest that the forecast skill is highest in near lead days but decreases and becomes lower as days increases. Investigation on MAE and RMSE confirm that the forecast models provide better accuracy but the forecast skill reduces towards higher lead days. The coefficient of determination (R^2) also emphasizes that model for lead day one has higher R^2 of 0.81 and 0.64 for lead day seven (Fig. 3).

Maximum temperature forecast models assessment: The performance of the maximum temperature forecast models for lead day one and lead day seven for maximum temperature is summarized in Table 2 and Fig. 1. The forecast given by lead day one forecast model is 90% correlated with the observed. The lead days two and three predictions are 85% correlated with the observed. Figure 4 shows that the predictions results are much correlated with the station observations. Figure 1b compares the MAE and RMSE of the models. The analysis on MAE demonstrates that the performance degrades as the lead rises. The RMSE calculated for these models also justify the above point. The coefficient of determination also highlight that the MLR models has better fit for lead day one are 81 and 59% best-fit on seventh day (Fig. 5).

CONCLUSION

The main intension of the study is to propose short term minimum and maximum temperature estimation models in densely populated urban area using multiple linear regression. Foretelling of temperature will help public as well governing authorities to take necessary precaution to handle sour heat and cold weather. Among the models formulated, the models for lesser lead day's produces better accuracy with least MAE, RMSE and higher correlation between observed and predicted. The coefficient of determinant of the models also states the same. The study also suggests that the forecast accuracy is higher for minimum temperature when compared with maximum temperature prediction,

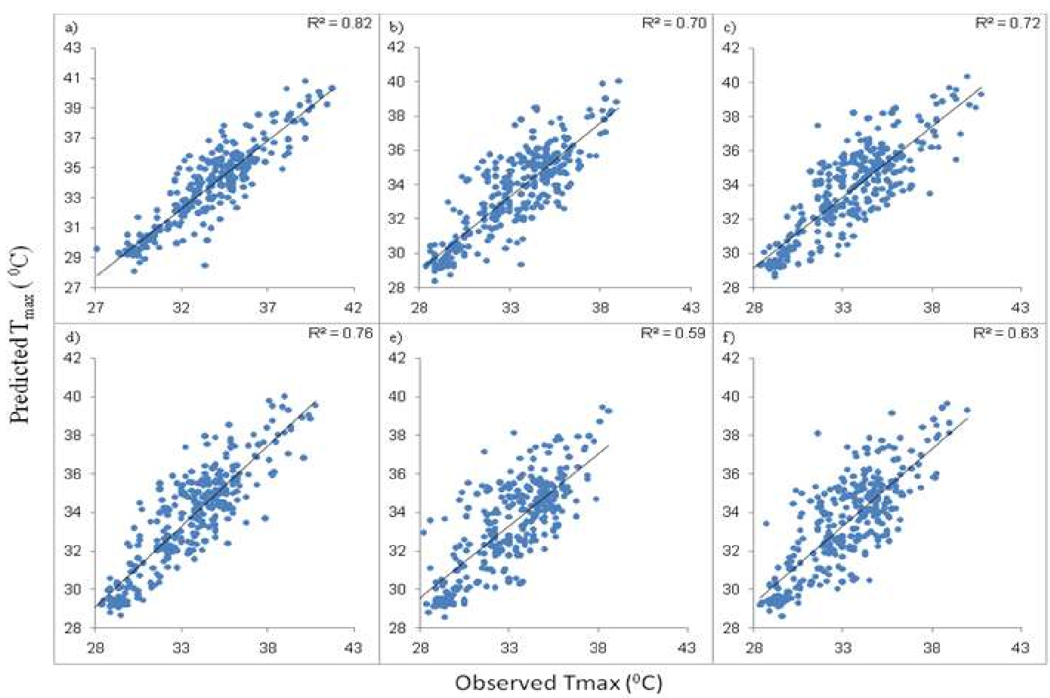

Fig. 5: Association between observed and predicted maximum temperature (°C) (T_{max}), (a-c) lead day one through day three, (d-f) lead day five to seven

the correlation between the temperature and the atmospheric parameters selected for analysis decreases as the lead increases. Although the methodology employed in this study has given significant performance for least lead days, the correlation of the accuracy for greater lead days should be further refined.

REFERENCES

Chatfield, C., 2009. The Analysis of Time Series: An Introduction. Chapman and Hall, CRC, Boca Raton, Florida.

Christoph, C.R., B. Georg, F. Klaus and K. Edilbert, 1999. Statistical single-station short-term forecasting of temperature and probability of precipitation: Area interpolation and NWP combination. Weather Forecast., 14: 203-214.

Donald, A.C., 2011. Essentials of Meteorology: An Invitation to the Atmosphere. 6th Edn., Brooks/Cole, Belmont, CA.

Shengpan, L., J.M. Nathan, P.M. Joseph, H.D. Mark and W. Jiaping, 2012. Evaluation of estimating daily maximum and minimum air temperature with MODIS data in east Africa. Int. J. Appl. Earth Obs., 18: 128-140.

Taylor, A.A. and L.M. Leslie, 2005. A single-station approach to model output statistics temperature forecast error assessment. Weather Forecast., 20: 1006-1019.

Wilks, D.S., 2006. Statistical Methods in the Atmospheric Science. 2nd Edn., Elsevier Inc., pp: 519-522.