Lead seven day maximum and minimum air temperature prediction: Comparative study with Linear and Polynomial Regression.

K.Ramesh, R. Anitha, M. Senthil Kumar

1 Department of Computer Applications, Regional Centre, Anna University, Tirunelveli, Tamil Nadu, India.
2 Department of Computer Applications, K. S. Rangasamy College of Technology, Tiruchengode, Tamil Nadu, India.
3 Department of Computer Applications, SNS College of Technology, Coimbatore, Tamil Nadu, India.

ARTICLE INFO

Article history:
Received 19 November 2013
Received in revised form 20 January 2013
Accepted 23 January 2013
Available online 25 February 2014

Keywords:
Linear regression, Polynomial regression, Temperature forecast.

ABSTRACT

The range of surface air temperature encountered at the earth surface is important as it is favorable for vegetation, animals, human livelihood and the environmental events. Forecast of temperature deviations are challenging due to the dynamic atmospheric parameters involved in the temperature event. In this work a comparative study is done between the statistical forecasting approaches Multiple Linear Regression (MLR) and Multiple Polynomial Regression (MPR) for forecasting the minimum and maximum surface air temperature for lead day one to seven in Chennai, India. Eight atmospheric parameters observed at the ground observing station for nine year (1995 – 2003) is used as predictors in this work. The result of the comparative study highlights that both regression based approaches provide close accuracy with the highest $R^2$ and lowest MAE, RMSE with independent test dataset. The MPR based model has better performance with an overall difference of 0.01 $^\circ$C in minimum temperature prediction and in maximum temperature linear regression performs better. The performance of the statistical analysis reveals that the mean absolute error is less than one degree Celsius upto lead days five and close to 1 $^\circ$C for day six and seven in minimum temperature forecast. Regarding maximum temperature forecast the analysis indicates that the mean absolute error is 1.24 $^\circ$C for day seven. From the statistical analysis it was found that prediction skill degrades as the lead days increases.

© 2014 AENSI Publisher All rights reserved.


INTRODUCTION

Surface air temperature is the degree of hotness or coldness of the air at the earth surface, which is an important weather element in our lives. Those most likely to suffer from overheat and heat exhaustion or heat strokes are the elderly with impaired circulatory systems and infants, whose heat regulatory mechanisms are not yet fully developed. (Donald Ahrens, 2011). Accurate forecast of surface air temperature is challenging because of the dynamic nature of the temperature event. Scientific community has devised many linear and nonlinear temperature prediction methodologies, but still the prediction accuracy for medium and long term forecast need improvement in the prediction skill. Although in recent studies artificial intelligence techniques are used in temperature prediction, in this study MLR is selected since linear models often produce better forecasts than nonlinear models even when the data are nonlinear and also statistical schemes require little computation time to make a forecast (Chatfield, 2009). Statistical forecasting techniques, Model Output Statistics (MOS) is a powerful statistical forecasting technique based on linear regression has given significant results in forecasting maximum and minimum temperature (Taylor et al., 2005) but it is more suitable for short term forecast. Stepwise linear regression approach using MODerate-resolution Imaging Spectroradiometer (MODIS) land surface temperature data in east Africa for estimating daily maximum and minimum air temperature has given promising results (Shenpan Lin et al., 2012). Multiple linear regression based single-station temperature variation forecast model for 6 hour to 24 hour has given RMSE of 1.78 $^\circ$C for the 6-h forecast and 2.28 $^\circ$C for the 12-h forecast (Raible et al., 1999). In this paper, the minimum and maximum temperature observed in Chennai, India for next consecutive seven days is predicted with the daily atmospheric parameters recorded at the station using MLR and MPR techniques. MLR based forecast are well suited for linear data but it is incapable for nonlinear problems like environmental modelling whereas polynomial regression can fit in highly nonlinear data. In computation, the regression coefficient of MLR model grows linearly with the number of predictors but it grows exponentially in MPR based modelling (Caren Marzban, 2003). The objective of this
work is to compare the prediction performance of Multiple Linear Regression based lead seven days minimum and maximum temperature prediction models with Multiple Polynomial Regression.

2. Data and Methodology:

2.1 Data:
The study area selected for forecast analysis is Chennai in India. (Latitude : 13°41’7.3” N, Longitude :80°14’48.33” E). The weather observation station located in Chennai is managed by Indian meteorological Department records various daily surface level atmospheric events. Among the parameters recorded based on previous studies the parameters listed in table 1 are used as predictors to forecast next seven days minimum temperature and maximum temperature. The observed predictor dataset 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 this analysis, the overall period used covers a duration of nine years (1995 – 2003). The data from 1996 through 2003 were used for the designing the models and the performance and accuracy of the derived models are validated by deploying it with one year (1995) of independent data.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Predictor variable</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mean Temperature</td>
<td>°C</td>
</tr>
<tr>
<td>2</td>
<td>Mean Dew Point</td>
<td>°C</td>
</tr>
<tr>
<td>3</td>
<td>Maximum Sea Level Pressure</td>
<td>hPa</td>
</tr>
<tr>
<td>4</td>
<td>Mean Visibility</td>
<td>Km</td>
</tr>
<tr>
<td>5</td>
<td>Mean Wind Speed</td>
<td>km/h</td>
</tr>
<tr>
<td>6</td>
<td>Maximum Wind Speed</td>
<td>km/h</td>
</tr>
<tr>
<td>7</td>
<td>Precipitation</td>
<td>mm</td>
</tr>
<tr>
<td>8</td>
<td>Minimum Temperature</td>
<td>°C</td>
</tr>
<tr>
<td>9</td>
<td>Maximum Temperature</td>
<td>°C</td>
</tr>
</tbody>
</table>

*aDaily minimum temperature is used as the eighth predictor of minimum temperature prediction models and daily maximum temperature is used as the eighth predictor of maximum temperature prediction models.*

2.2 Methodology:

2.2.1 Multiple Linear Regressions:

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

\[ y_i = \beta_0 + \beta_1 x_{i1} \ldots + \beta_p x_{ip} + \epsilon_i \] (1)

Where \( y_i \) is the predicted variable, \( \beta_0 \) is the intercept, \( \beta_1 \) measures the change in \( y_i \) with respect to \( x_{i1} \), \( \beta_p \) measures the change in \( y_i \) with respect to \( x_{ip} \), \( x_{i1} \ldots x_{ip} \) are predictor variable and \( \epsilon_i \) the error.

2.2.1.1 Multiple Polynomial Regression

Polynomial regression used to fit nonlinear (e.g. curvilinear) data into a least squares linear regression model. It is a form of linear regression that allows one to predict \( y \) by decomposing the \( x \) variable into a \( n \)th order polynomial.

Generally has the form:

\[ y = \beta_0 + \beta_1 x^1 + \beta_2 x^2 \ldots + \beta_k x^k \] (2)

where \( \beta_0 \) is the intercept and \( \beta_1 \) to \( \beta_k \) is the regression coefficient.

3. Models:
The dataset is divided into two dataset: training set (1996 to 2003) and validation set (1995). With the training dataset, seven minimum temperature prediction models one each for lead days one to seven and another seven for maximum temperature with eight regressor and one dependent variable are formulated using MPR. The same fourteen models were also developed with eight predictors and minimum or maximum temperature as dependent variable using MLR. The fitness of the formulated models are analyzed by coefficient of
determination ($R^2$), which shows how well the independent variables varies with the dependent variable (Wilks, 2006).

The prediction model performance is validated by deploying the models with validation dataset. The prediction accuracy is accessed by calculating statistical error analysis namely MAE, RMSE and correlation between observed and model predicted temperature. The MAE is the average of the absolute errors used to measure how close forecasts or predictions are to the corresponding observation. RMSE is the square root of average squared difference between the forecast and actual outcome. The MAE and RMSE are used together to diagnose the variation in the error forecast, the forecast is perfect if MAE and RMSE are equal to zero.

Table 2: Performance comparison of forecast models

<table>
<thead>
<tr>
<th>Lead Days</th>
<th>Minimum Temperature</th>
<th>Maximum Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE (MLR)</td>
<td>MAE (MPR)</td>
</tr>
<tr>
<td>1</td>
<td>0.59</td>
<td>0.60</td>
</tr>
<tr>
<td>2</td>
<td>0.73</td>
<td>0.72</td>
</tr>
<tr>
<td>3</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>4</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>5</td>
<td>0.99</td>
<td>0.97</td>
</tr>
<tr>
<td>6</td>
<td>1.04</td>
<td>1.02</td>
</tr>
<tr>
<td>7</td>
<td>1.09</td>
<td>1.06</td>
</tr>
</tbody>
</table>

RESULT AND DISCUSSION

4.1 MLR model performance assessment:

Table 2 shows the model performance measured by statistical error analysis on the deployment output. The MAE on minimum temperature forecast is the least at lead day one with 0.59 °C and the prediction skill shows a solid declaim as the prediction lag increases. The correlation coefficient between observed and predicted for day one has better fit for lead day one with 93% and the prediction models fitness for remaining six lead days are above 80% (figure 2) in minimum temperature forecast. The performance of the MLR based maximum temperature forecast models for lead day one to lead day seven is summarized in Table 2 and figure 1. The forecast error for minimum temperature forecast is higher than minimum temperature forecast. The MAE and RMSE for maximum temperature forecast vary from 0.86°C to 1.22°C and 1.19°C to 1.60°C respectively. Figure 1 also shows that the prediction skill decreases but not in a even form in maximum temperature forecast (in day 4 the skill has a great fall and raises for day 5 and 6). From the correlation analysis it can be interpreted that the forecast skill is highest in near lead days but decreases as the prediction interval increases.

![Fig. 1: Performance comparison (MAE and RMSE) of temperature prediction models.](image-url)
4.2 MPR model performance assessment:

The MPR forecast model deployment results show an average absolute error of 0.60°C and 0.86°C for lead day one minimum and maximum temperature forecast respectively. The RMSE error analysis on the MPR models also clarify that the forecast skill declines as the lead time increases. The average difference between the observed and the predicted temperature is highest at day seven (1.24°C). The correlation between the observed and predicted temperature of both minimum and maximum temperature are shown in figure 2 and figure 3.

![Graphs showing correlation of observed versus predicted temperatures](image)

**Fig. 2:** Correlation of observed versus predicted minimum temperature. (2.a, 2.c, 2.e MPR models and 2.b, 2.d, 2.f MLR models)

4.3 Comparison of MLR and MPR prediction models:

To put forward the better prediction model among the formulated models, the forecast computed with MLR models are compared with the forecast given by MPR approach. The comparison analysis state that polynomial regression based models gives a trivial deviation and betterment in the prediction with an average of 0.01°C degree for higher lead days. The prediction performance is almost same for smaller lead days in minimum temperature forecast for both approaches. In maximum temperature forecast the forecast skill is better for MLR models than MPR models with a deviation of 0.01°C. The difference in MAE with RMSE indicates that very large errors are unlikely to have occurred in the forecast and there are variations in the magnitude of the errors. The regression models reveal that among the predictors, temperature and precipitation has high influence in the temperature event that occurs at the station. As the correlation between the observed and predicted temperature are almost the same, it does not help to compare the performance or to converge to a solution. The residual analysis presented in Figure 4 and figure 5 disclose that the bias is mildly more in MLR approach in minimum temperature forecast than MPR approach and in maximum temperature forecast MLR model has less bias. It is also noted that the models shows warm bias in the winter season (November to march). The comparison analysis helps to conclude that the both approaches (linear and polynomial regression) do not show significant variance in forecast at this station.
Fig. 3: Correlation of observed verses predicted minimum temperature. (3.a, 3.c, 3.e MPR models and 3.b, 3.d, 3.f MLR models)

Fig. 4: Residual analysis of minimum temperature forecast models.
**Conclusion:**

The objective of this study is to compare the performance of linear regression based forecast models with polynomial regression based models in minimum and maximum temperature forecast and to propose the better model in the study in Chennai, India. The analysis of the forecast models illustrates that the forecast skill is statistically good for short term and the prediction skill decreases as the lead increases. From the comparative analysis it can be concluded that the MPR model has a better performance in minimum temperature forecast and in maximum temperature forecast the MLR has better performance. This study also suggests that the models based on atmospheric parameters selected are more suitable for minimum temperature when compared with maximum temperature prediction. The methodologies used in this study has given considerable performance for small lag days, further the prediction accuracy should be refined for long interval as the temperature event variance is an integrated part of lively hood in the earth.

**REFERENCES**


