ISSN: 2278-0181
Vol. 1 Issue 6, August - 2012

Prediction of maximum/minimum temperatures using Holt Winters Method with Excel Spread Sheet for Junagadh Region

Gundalia Manoj j.
Ph. D. Scholar
Gujarat Technological University, Ahmedabad

Dholakia M. B.
Joint Director
Directorate of Technical Education, Gandhinagar

Abstract

Temperature Time Series data are used as major input in many rainfall-runoff models. Temperature prediction is a temporal and time series based process and involve when insufficient data availability. Temperature time series often consist of periodic patterns, i.e. patterns that repeatedly occur in defined periods over time, Holt Winters method is suitable to predict such meteorological process. Microsoft Excel is user friendly and most popular spread sheet program.

In this paper, in order to predict maximum and minimum daily temperature time series of Junagadh region with Holt Winters method using excel spread sheet is proposed. The performance of the method is measured by three standard error measures MSE, MAPE and MAD which are found well within the acceptable limits. Results shows the maximum temperature time series exhibit less fluctuation and gave better results as compared to minimum temperature time series.

Keywords

Maximum/minimum temperatures; Time Series Prediction; Triple Exponential Smoothing (Holt Winters).

1. Introduction

Forecasting is a phenomenon of knowing what may happen to a system in the next coming time periods [22]. The goal is to observe or model the existing data series to enable future unknown data values to be forecasted accurately [10]. As weather is a continuous, data-intensive and dynamic process, the parameters required to predict temperature are enormously complex such that there is uncertainty in prediction even for a short period [1].One way of classifying forecasting problems is to consider the timescale involved in the forecast. Short, medium and long-term are the usual categories but the actual meaning of each will vary according to the situation that is being studied. Exponential smoothing technique is one of the most important quantitative techniques in forecasting. The accuracy of forecasting of this technique depends on exponential smoothing constant.

Once data have been captured for the time series to be forecasted, the analyst's next step is to select a model for forecasting. Various statistical and graphic techniques may be useful to the analyst in the selection process. Three tools for assessing the autocorrelation of a time series are (i) the time series plot, (ii) the lagged scatter plot, and (iii) the autocorrelation function. The best place to start with any time series forecasting analysis is the lagged scatter plots of the time series to be forecasted. A sequence plot is a graph of the data series values, usually on the vertical axis, against time usually on the horizontal axis. The purpose of the lagged scatter plot is to give the analyst a visual impression of the nature of the time series. This visual impression should suggest to the analyst whether there are certain behavioural "components "present within the time series. A lag plot checks whether a data set or time series is random or not. Random data should not exhibit any identifiable structure in the lag plot. Non-random structure in the lag plot indicates that the underlying data are not random. The presence/absence of trend and seasonality components can help the analyst in selecting the model with the potential to produce the best forecasts.

Excel is now the most popular spread sheet program and most people have it on their computers. In Rahmbow and Klimberg's 200 studies of forecasting practices, the leading application software was overwhelmingly Excel (90%). Excel is an electronic spread sheet developed by Microsoft; this computer application program simulates a physical spread sheet by capturing, displaying and manipulating data arranged in rows and columns. The main advantage of Excel is its versatility and functionality when you are doing any type of model. It is a much simpler program than a database program. For these reasons, and to avoid the need to purchase specialized forecasting software, the Excel spread sheet program was selected for use in this paper

The Solver is an add-in for Microsoft Excel which is used for the optimization and simulation of numerical models. It solves complex linear and nonlinear problems and can also be used in conjunction with VBA to automate tasks. It can solve problems by enabling a Target cell to achieve objectives.

2. Data Collection

Maximum/minimum daily temperature data were collected from Agro meteorological Cell, Department of Agronomy, College of Agriculture, Junagadh Agricultural University, Junagadh (Gujarat). Most of the rainfall-runoff models have needed data from month of June to month of October for a year. Maximum daily temperature (^{0}C) data from June to October for period of 1984 to 2010 and Minimum daily temperature (^{0}C) data from June to October for period of 1987 to 2010 were available. In order to complete data set from June to October for period of 1980 to 2010, prediction of maximum temperature time series (MXTTS) for period 1980 to 1983 and minimum temperature time series (MNTTS) for period 1980 to 1986 are needed.

3. Selection of Method

Choosing an appropriate model or class of models is as much an art as a science. There is no single approach that is 'best' for all situations, but it is possible to lay down some general guidelines. The analyst will generally have to (i) get as much background information as is necessary, (ii) assess costs, (iii) clarify objectives, and (iv) have a preliminary look at the data. Many of the above points are amplified in chapter 1of [12], whose main example discusses the difficulties involved when trying to model the dynamics of deforestation in the Amazon region.[11] provided some simple rules based on the variances of differenced time series for choosing an appropriate exponential smoothing method. [27] compared these rules with others proposed by [8] and an approach based on the BIC. [14] also proposed an information criterion approach, but using the underlying state space models.

The choice of suitable forecasting procedure depends on the properties of the data or the objectives of the study [17], [9].The graph should show up important features of the data such as trend, seasonality, outliers, smooth changes in structure, turning points and/or sudden discontinuities, and is vital, both in describing the data, in helping to formulate a sensible model and in choosing an appropriate forecasting method. A lag plot of mean minimum and mean maximum temperatures of available data series are shown in **Fig.1** and **Fig.2** respectively. Plots show tight clustering of points along the diagonal. This is the lag plot signature of a process with strong positive autocorrelation. Such processes are highly non-random and there is strong association between an observation and a succeeding observation. In short, if you know X_{i-1} you can make a strong guess as to what X_i will be.

The introduction of a class of state-space models underlying exponential smoothing methods enabled them to enjoy the advantages that forecasting procedures based on a proper statistical model have [16], [15], [24], [2].Therefore, Holt winters method is appropriate and selected for prediction of these time series.

4. Holt Winters Method

Twenty-five years ago, exponential smoothing methods were often considered a collection of ad hoc techniques for extrapolating various types of univariate time series. Although exponential smoothing methods were widely used in business and industry, they had received little attention from statisticians and did not have a well-developed statistical foundation. These methods originated in the 1950s and 1960s with the work of [5], [6], [13], and [28]. [23] Provided a simple but useful classification of the trend and the seasonal patterns depending on whether they are additive (linear) or multiplicative (nonlinear). [21] was the first to suggest a statistical foundation for simple exponential smoothing (SES)by demonstrating that it provided the optimal forecasts for a random walk plus noise. Further steps towards putting exponential smoothing within a statistical framework are provided by [5], [25], and [3], [4], who showed that some linear exponential smoothing forecasts arise as special cases of ARIMA models. However, these results did not extend to any nonlinear exponential smoothing methods. Exponential smoothing methods received a boost from two papers published in 1985, which laid the foundation for much of the subsequent work in this area. First, Gardner (1985) provided a thorough review and synthesis of work in exponential smoothing to that date and extended Pegel's classification to include damped trend. This paper brought together a lot of existing work which stimulated the use of these methods and prompted a substantial amount of additional research. Later in the same year, [26] showed that SES could be considered as arising from an innovation state space model (i.e., a model with a single source of error). Although this insight went largely unnoticed at the time, in recent years it has provided the basis for a large amount of work on state space models underlying exponential smoothing methods.

Triple Exponential Smoothing (Holt Winters Method) method is appropriate when trend and seasonality are present in the time series. It decomposes the times series down into three components: base, trend and seasonal components.

$$F_t = \alpha\, (Y_t / S_{t-p}) + (1-\alpha)\, (F_{t-1} + T_{t-1}) \qquad (01)$$

$$S_t = \gamma\, (Y_t / F_t) + (1-\gamma)\, S_{t-p} \qquad (02)$$

$$T_t = \beta\, (F_t - F_{t-1}) + (1-\beta)\, T_{t-1} \qquad (03)$$

$$W_{t+m} = (F_t + mT_t)\, S_{t+m-p} \qquad (04)$$

Where,

F_t= smoothed value of the level of series for period t

F_{t-1} = smoothed value for period t-1

Y_t = actual value in period t

T_t = trend estimate
S_t = seasonality estimate
α = Smoothing constant for the data ($0< \alpha<1$)
β = smoothing constant for the trend estimate ($0<\beta <1$)
γ = smoothing constant for seasonality estimate ($0<\gamma <1$)
p = number of periods in seasonal cycle
m = number of periods ahead to be forecast
W_{t+m}=Winters' forecast for m periods into the future

5. Methodology

The main aim of the present study is to predict maximum and minimum temperature time series for Junagadh district (Gujarat).To start Holt Winters method, there is need F1, T1, and a seasonal factor S1for each period in the cycle. Cyclic period of seven year have been considered for minimum temperature time series and that of four years for maximum temperature time series based on time plot graphs observation and periods to be forecast into the future. For initialization, collect p (Cyclic period) observations and estimates the seasonal factor using equation [05].

$$S_i = y_i \, [\, (1/p)(y_1 + y_2 + y_3 + \ldots + y_p)] \qquad (05)$$

Enter the maximum temperature of day 01of June from year 2010 to 1984 in column Yi of excel spread sheet. Assume $F_p = y_p / S_p$ and $T_p = 0$ and initialized values of Fi and Si as shown in **Fig. 3 and 4.**The equations (01), (02) and (03) are formulated in cells C7, D7 and E7as shown **in Fig. 5, 6 and 7** respectively. Cell G7, H7 and I7 show the MAD, MSE and MAPE respectively. To optimize values of smoothing constant α, β and γ, mean value of MAD should be minimized. Cell J5 is the Invoke Excel Solver to minimize the MAD. Enter locations of parameters α, β and γ with constraints in solver dialogue box **Fig. 8**. Click the Solve button. Solver will start to optimize the parameters α, β and γ and Keep the Solver solution. The solution is instantly found and the optimized values of α, β and γ appear in cell J2, K2 and L2 respectively **Fig. 9**. To predict values of maximum temperatures for year 1983 to 1980, simply fill down the formula in cell F30 to cell F33**Fig. 9**. The graph shows the nature of existing and predicted values **Fig. 10**. Similar methodology is applied for prediction of minimum temperature time series.

6. Goodness of Fit

Accuracy can be defined as "goodness of fit" or how well the forecasting model is able to reproduce data that is already known [20]. This study used three standard error measures: mean squared error (MSE), mean absolute percentage error (MAPE) and mean absolute deviation (MAD).

MAD is the average of the absolute value of the error without regard to whether the error was an overestimate or underestimate [18], equation (06) illustrate the MAD formula.

$$\text{MAD} = \frac{\sum_{t=1}^{n} |X_t - \bar{X}_t|}{n} \qquad [06]$$

MSE is a measure of dispersion of forecast errors; statisticians have taken the average of the squared individual errors. Interpreting the MSE value can be misleading; for the mean squared error will accentuate large error terms. Equation (07) describes the MSE measurement.

$$\text{MSE} = \frac{\sum_{t=1}^{n} (X_t - \widehat{X_t})^2}{n} \qquad (07)$$

MAPE is regarded as a better error measurement than MSE because it does not accentuate large errors. Equation (08) illustrates the MAPE formula. MAPE is not, dependent on the unit of measurement.

$$MAPE = \frac{\sum_{t=1}^{n} \left(\frac{|X_t - X_t|}{X_t} \right)}{n} \times 100\% \qquad (08)$$

For all three measures, the smaller the value, the better the fit of the model. MAD, MSE and MAPE are calculated as an average of the values computed in column G, H and I respectively **Fig. 3**. A scale to judge the accuracy of the model based on the MAPE measure, developed by [19].MAD and MSE measure is that there is no context to indicate whether the model is good or not. Using MAPE, and applying Lewis's scale, provides some framework as shown in **Table 1** to judge the model.

7. Results and Discussion

The main aim of the present study is to predict maximum and Variations of the triple exponential smoothing techniques were applied to maximum and minimum temperature time series data, the results of which are discussed in this section. The tests were conducted for 153 days on (MXTTS) from year 1984 to 2010 and prediction period were from year 1980 to 1983 while on MNTTS from year 1987 to 2010 and prediction period were from year 1980 to 1986. The mean value of smoothing constants α, β and γ for maximum/minimum daily temperature series are obtained as shown in **Table 2**. The difference in mean values for both the series are found more in γ and less in α. Series are more diverge each other in trend and cyclic patterns. Maximum, minimum and mean values of all three measures MAD, MSE and MAPE are computed and presented in **Table3**.Mean values of

MAD, MSE and MAPE of MNTTS are less than that of MXTTS. Standard deviation of MAD, MSE and MAPE of MXTTS are less than that of MNTTS.

8. Conclusions

The Triple Exponential smoothing (Holt-Winters Method) is widely used technique to predict the time series. It is used when the data exhibits both trend and seasonality. In this study Triple exponential technique is applied on maximum/minimum temperature time series using excel spread sheet. Values of smoothing constants α, β and γ are optimized by minimizing MAD using solver tool of Microsoft excel.

Mean values of smoothing constants α, β and γ for (MXTTS) are found less than that of (MNTTS). This suggests that Prediction of (MXTTS) more depends on previous observations than current observations. Mean MAPE value of (MXTTS) is found greater than that of (MNTTS), but standard deviation is low. Out of 153 observations of MAPE only 3 observations are found with MAPE greater than 10 in (MXTTS) while 19 observations are found that of (MNTTS). This indicates that (MXTTS) exhibit less fluctuations and gave better results as compared to (MNTTS).

9. Acknowledgement

We express sincere thanks to Agro meteorological Cell, Department of Agronomy, College of Agriculture, Junagadh Agricultural University, Junagadh (Gujarat) for providing maximum/minimum temperature data.

10. References

[1] Abraham et. al., "Soft Computing Models for Weather Forecasting" J. of Applied Sciences and Computations, USA, vol. 11, no. 3, pp. 106-117, 2004.
[2] Corber´an-Vallet, J.D. Berm´udez, J.V. Segura JV, and E. Vercher,"A Forecasting Support System based on Exponential Smoothing", in Handbook on Decision Making (Chapter 8), Edited by L. C. Jain and C. P. Lim, Berlin: Springer-Verlag, 2010.
[3] Abraham, B., & Ledolter, J. (1983). Statistical methods for forecasting. New York7 John Wiley and Sons.
[4] Abraham, B., & Ledolter, J. (1986). Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. International Statistical Review, 54, 51–66.
[5] Box, G. E. P., & Jenkins, G. M. (1970). Time series analysis: Forecasting and control. San Francisco7 Holden Day (revised. 1976).
[6] Brown, R. G. (1959). Statistical forecasting for inventory control. New York7 McGraw-Hill.
[7] Brown, R. G. (1963). Smoothing, forecasting and prediction of discrete time series. Englewood Cliffs, NJ7 Prentice-Hall.
[8] Collopy, F., & Armstrong, J. S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. Management Science, 38, 1394–1414.
[9] D.C. Montgomery, C.L. Jenkins, and M. Kulahci, Introduction to TimeSeries Analysis and Forecasting, New Jersey: Wiley, 2008.
[10] E. A. Plummer, "Time Series Forecasting with Feed-forward Neural Networks: Guidelines and Limitations", M.S. thesis, Department of Computer Science, The Graduate School of The University of Wyoming, Laramie, USA, 2000.
[11] Gardner Jr., E. S., & McKenzie, E. (1988). Model identification in exponential smoothing. Journal of the Operational Research Society, 39, 863–867.
[12] Granger, C.W.J. (1999) Empirical Modeling in Economics. Am bridge: Cambridge Univ.Press.[3.5]
[13] Holt, C. C. (1957). Forecasting seasonal and trends by exponentially weighted averages. O.N.R. Memorandum 52/1957, Carnegie Institute of Technology. Reprinted with discussion in2004. International Journal of Forecasting, 20, 5 – 13.
[14] Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2005).Prediction intervals for exponential smoothing state space models. Journal of Forecasting, 24, 17– 37.
[15] J.D. Berm´udez, J.V. Segura JV, and E. Vercher, "Holt-Winters forecasting: an alternative formulation applied to UK air passenger data", Journal of Applied Statistics, vol. 34, pp. 1075–1090, 2007.
[16] J.K. Ord, A.B. Koehler, and R.D. Snyder, "Estimation and prediction for a class of dynamic nonlinear statistical models", Journal of American Statistical Association, vol. 92, pp. 1621–1629, 1997.
[17] J.R.T. Arnold, S.N. Chapman, and L.M. Clive, Introduction to Materials Management, 6th edn.,New Jersey: Pearson Prentice Hall, 2008.
[18] Krajewski, L. J., Ritzman, L. R., (1993). Operations Management, Strategy and Analysis, Addison-Wesley Publishing Company, Inc., p. 436.
[19] Lewis, C. D. "Industrial and business forecasting methods: A Radical guide to exponential smoothing and curve fitting." Butterworth Scientific, London; Boston: 1982.
[20] Makridakis, S., Wheelwright, S. C., (1989).Forecasting Methods for Management, John Wiley & Sons, New York, P. 3.
[21] Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. Journal of the American Statistical Association, 55,299– 306.
[22] P. Sarangi and et al., "Short Term Load Forecasting using Artificial Neural Network: A Comparison with Genetic Algorithm Implementation", J. of ARPN Engineering and Applied Sciences, vol. 9, 2009.
[23] Pegels, C. C. (1969). Exponential smoothing: Some new variations. Management Science, 12, 311 –315.
[24] R.J. Hyndman, A.B. Koehler, J.K. Ord, and R.D. Snyder, Forecasting with Exponential Smoothing. The State Space Approach, Berlin: Springer, 2008.
[25] Roberts, S. A. (1982). A general class of Holt–Winters type forecasting models. Management Science, 28, 808–820.

[26] Snyder, R. D. (1985). Recursive estimation of dynamic linear statistical models. Journal of the Royal Statistical Society (B), 47, 272–276.
[27] Tashman, L., & Kruk, J. M. (1996). The use of protocols to select exponential smoothing procedures: A reconsideration of forecasting competitions. International Journal of Forecasting, 12,235– 253.
[28] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6, 324–342.

GUNDALIA MANOJ J. earned his B.E. in Civil Engineering from M.S. University and did M.E. in Civil (Water Resources Management) Engineering from Gujarat University. He has presented and published 3 research papers in national conference and 2 published in international journal. He has 14 years field experience, 4 years teaching experience and currently he is Assistant Professor in Dr. Subhash Technical Campus, Junagadh, (Gujarat), INDIA.

Table 1. A scale of Judgment of Forecast Accuracy (Lewis)

MAPE	Judgment of Forecast Accuracy
Less than 10%	Highly Accurate
11% to 20%	Good Forecast
21% to 50%	Reasonable Forecast
51% or more	Inaccurat Forecast

Table 2. Values of smoothing constants

Parameters		Max. Temp Time Series	Min. Temp. Time Series
α	Mean	0.0270	0.0303
β	Mean	0.0723	0.1238
γ	Mean	0.2812	0.4568

Table 3. Measures of errors

Parameters		Max. Temp Time Series	Min. Temp. Time Series
MAD	Max	3.4303	3.5446
	Min	1.2912	0.5059
	Mean	2.0914	1.2796
	Stdev	0.4195	0.6192
MSE	Max	16.8096	17.3642
	Min	2.9674	0.4435
	Mean	7.6313	3.3867
	Stdev	2.8467	3.3492
MAPE	Max	10.4488	18.9432
	Min	3.5394	2.1047
	Mean	6.3902	5.6224
	Stdev	1.3088	3.5366

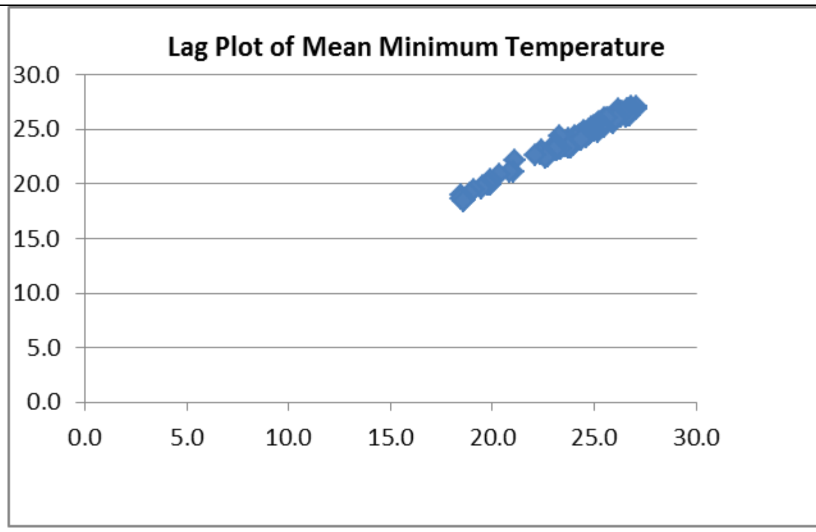


Fig. 1 Lag plot of Mean Minimum Temperature for period 1987 to 2010

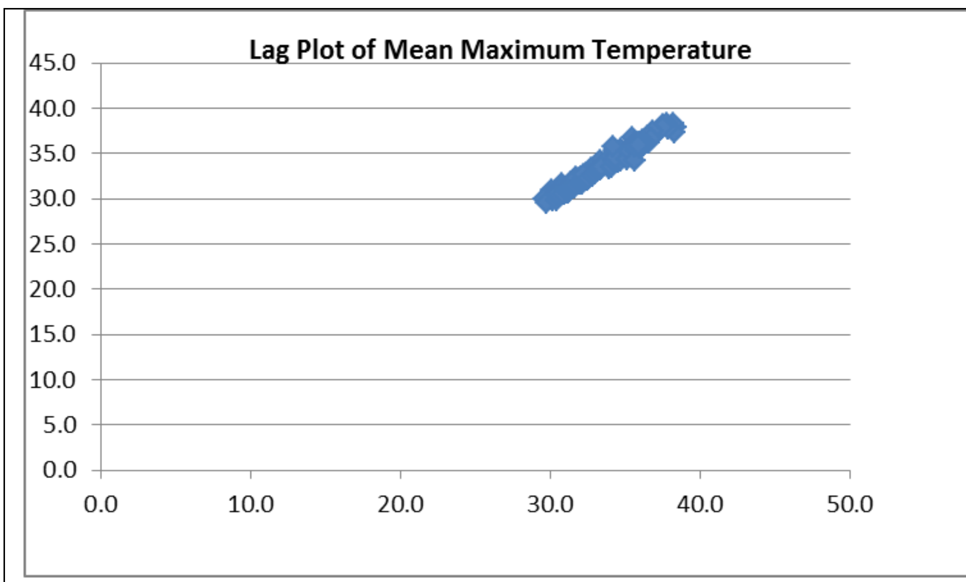


Fig. 2 Lag plot of Mean Maximum Temperature for period 1984 to 2010

C6 =AVERAGE(B3:B6)

	A	B	C	D	E	F	G
1		Y_i	F_i	T_i	S_i	F	e
2	T_{Max}						
3	01-Jun-10	40			1.03		
4	01-Jun-09	36.7			0.95		
5	01-Jun-08	37			0.96		
6	01-Jun-07	41	38.73	0.00	1.06		MAD
7	01-Jun-06	30	38.61	0.00	1.00	40.00	10.00

Fig. 3 Initialization of F_i

E6 | fx =B6/AVERAGE($B$3:$B$6)

	A	B	C	D	E	F	G	H	I
1		Y_i	F_i	T_i	S_i	F	e	e^2	
2	T_{Max}								
3	01-Jun-10	40			1.03				
4	01-Jun-09	36.7			0.95				
5	01-Jun-08	37			0.96				
6	01-Jun-07	41.2	38.73	0.00	1.06		MAD	MSE	MAPE
7	01-Jun-06	30	38.61	0.00	1.00	40.00	10.00	100.00	33.33

Fig. 4 Initialization of S_i

C7 | fx =$J$2*(B7/E3)+(1-$J$2)*(C6+D6)

	A	B	C	D	E	F	G	H	I
1		Y_i	F_i	T_i	S_i	F	e	e^2	
2	T_{Max}								
3	01-Jun-10	40			1.03				
4	01-Jun-09	36.7			0.95				
5	01-Jun-08	37			0.96				
6	01-Jun-07	41.2	38.73	0.00	1.06		MAD	MSE	MAPE
7	01-Jun-06	30	38.61	0.00	1.00	40.00	10.00	100.00	33.33

Fig. 5 Formulation of equation [01]

D7 | fx =$K$2*(C7-C6)+(1-$K$2)*D6

	A	B	C	D	E	F	G	H	I
1		Y_i	F_i	T_i	S_i	F	e	e^2	
2	T_{Max}								
3	01-Jun-10	40			1.03				
4	01-Jun-09	36.7			0.95				
5	01-Jun-08	37			0.96				
6	01-Jun-07	41.2	38.73	0.00	1.06		MAD	MSE	MAPE
7	01-Jun-06	30	38.61	0.00	1.00	40.00	10.00	100.00	33.33

Fig. 6 Formulation of equation [02]

E7 | fx =$L$2*(B7/C7)+(1-$L$2)*E3

	A	B	C	D	E	F	G	H	I
1		Y_i	F_i	T_i	S_i	F	e	e^2	
2	T_{Max}								
3	01-Jun-10	40			1.03				
4	01-Jun-09	36.7			0.95				
5	01-Jun-08	37			0.96				
6	01-Jun-07	41.2	38.73	0.00	1.06		MAD	MSE	MAPE
7	01-Jun-06	30	38.61	0.00	1.00	40.00	10.00	100.00	33.33

Fig. 7 Formulation of equation [03]

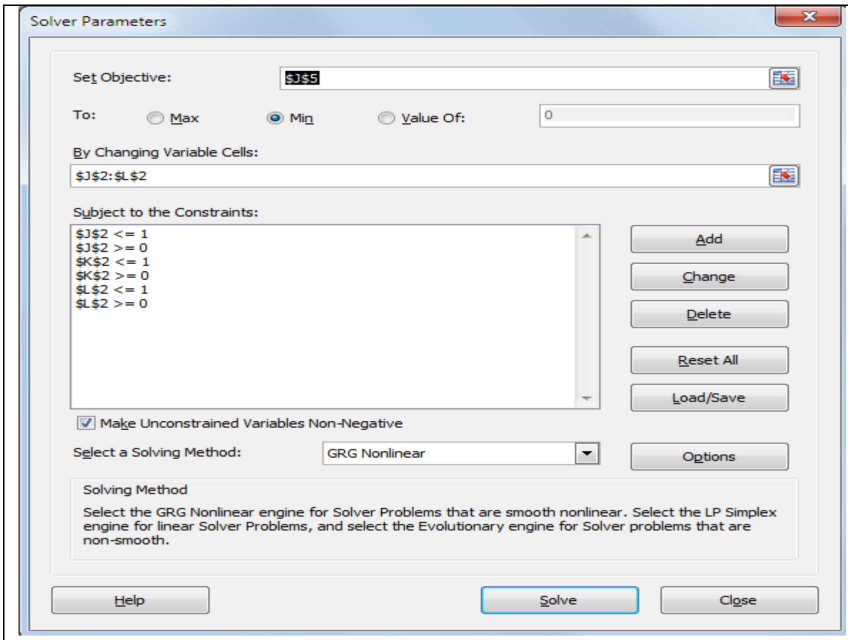
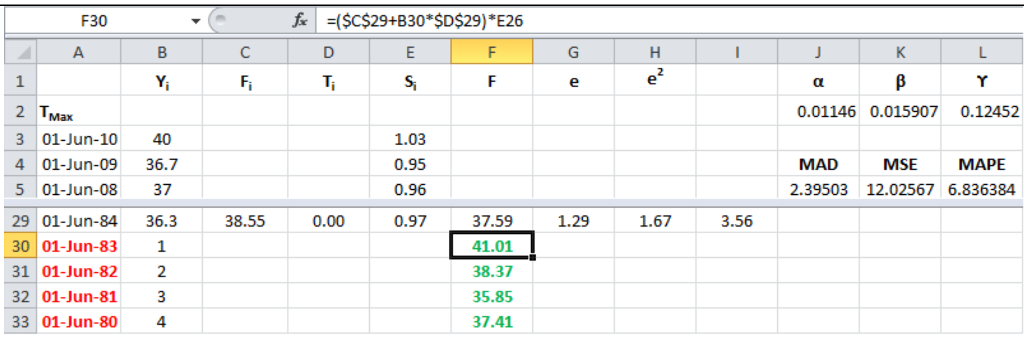


Fig. 8 Formulation of equation [01]



F30 =($C$29+B30*$D$29)*E26

	A	B	C	D	E	F	G	H	I	J	K	L
1		Y_i	F_i	T_i	S_i	F	e	e^2		α	β	γ
2	T_{Max}									0.01146	0.015907	0.12452
3	01-Jun-10	40			1.03							
4	01-Jun-09	36.7			0.95					MAD	MSE	MAPE
5	01-Jun-08	37			0.96					2.39503	12.02567	6.836384
29	01-Jun-84	36.3	38.55	0.00	0.97	37.59	1.29	1.67	3.56			
30	01-Jun-83	1				41.01						
31	01-Jun-82	2				38.37						
32	01-Jun-81	3				35.85						
33	01-Jun-80	4				37.41						

Fig. 9 Optimized parameters α, β and γ and predicted values of maximum temperature

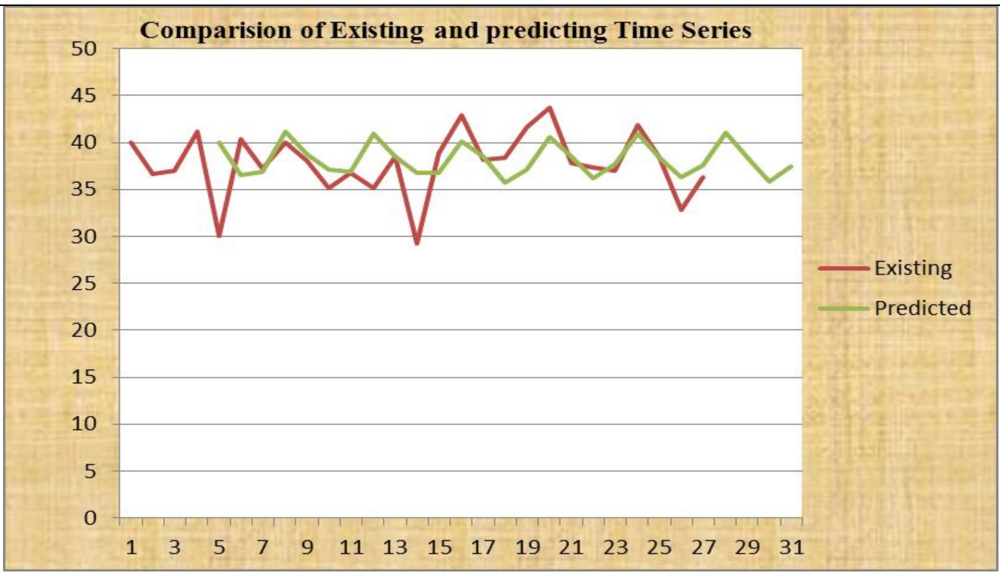


Fig. 10 Existing and predicted values of maximum temperature time series for day June 01