

## A STATISTICAL ANALYSIS OF THE MONTHLY MEAN MAXIMUM AIR TEMPERATURE IN COLOMBO, SRI LANKA

T.K. HEWAPATHIRANA,

Department of Statistics & Computer Science, University of Kelaniya, Kelaniya, Sri Lanka.

E-mail: hewapath@kln.ac.lk

### ABSTRACT

The present study was carried out to fit a mathematical model to describe the variation pattern of monthly mean maximum air temperature in Colombo in order to predict the future values. Monthly mean maximum temperature values for a period of over 35 years were used for the analysis. Time series statistical methods were considered to study the trend and seasonal, cyclic and irregular components. The long term pattern in the variation of monthly mean maximum temperature in Colombo appears to be dominated by a pronounced seasonal effect. The highest seasonal effect was found to be in March. It was found that the temperature of a particular month depends on the lagged temperature values of the two preceding months.

The model

$$\hat{x}_t = \hat{s}_t + 12.8 + 0.367(x_{t-1} - \hat{s}_{t-1}) + 0.214(x_{t-2} - \hat{s}_{t-2}) + 0.000567 * t$$

where  $\hat{x}_t$  = the estimated temperature at time t,  
 $\hat{s}_{t-k}$  = the estimated seasonal factor at time t-k for  
 $k = 0, 1, 2$  and  
 $x_{t-k}$  = observed mean maximum air temperature value at  
time t-k for  $k = 1, 2$

could be effectively used to predict the monthly mean maximum air temperature in Colombo.

The present analysis indicates that the mean maximum temperature in Colombo increases at a rate of 0.175 °C per decade, which is less than the rate of increase in the mean global temperature.

**Key words:** Time series analysis, trend, seasonal variation, cyclic variation, residuals, decomposition method.

## INTRODUCTION

Analysis of air temperature data at a particular region for long term trends is useful to determine whether there are localized factors which affect the average rate of increase in global temperature. By the end of the 21<sup>st</sup> century, the global temperature is predicted to increase by 2 °C – 5 °C (Veteli *et al.*, 2002) with an average rate of around 0.3 °C per decade (Anon., 2002). Earlier studies on temperature variations have shown spatial differences in the rate of change in temperature in different regions. For example, the variation of sea surface temperature of the eastern tropical Pacific has shown no correlation with the variation of global average temperature or with that of tropical sea surface temperature (Singer, 1999).

The mean air temperature in some areas is reported to increase not only due to enhanced greenhouse effect but also due to urbanization (Fernando and Basnayaka, 2002). Further, the nature and extent of clouds, and the vertical distribution of water vapour are also reported to have an effect on greenhouse effect (Singer, 1999) in addition to factors such as increased emission of carbon dioxide and methane due to anthropogenic activities. It is also hypothesized that carbon dioxide and other waste gases generate only a small amount of warming compared to water vapour (Stevens, 2002).

The present study was carried out to fit a mathematical model to explain the trend in the monthly mean maximum air temperature in Colombo during the past 35 years in order to forecast the future trend with an objective of determining whether it differs from the global trend.

## MATERIALS AND METHODS

Monthly mean maximum air temperature data from 1967 to 2002 obtained from MOFE (2000) and the Department of Meteorology were analysed using a traditional model building approach of time series analysis. Traditional methods of time series analysis are devoted to an analysis of the data to identify a trend or long term movement, oscillations about the trend of greater or less regularity, a seasonal effect, and random, unsystematic or irregular components, in order to isolate them for separate studies (Kendall and Stuart, 1968). As a matter of mathematical description, it is convenient to represent the series as one of these constituents or the sum of several of them (Kendall, 1973).

As the magnitude of the seasonal swing did not change with the average level, the following simple additive model was used in this analysis.

$$x_t = T_t + S_t + CL_t + Ir_t$$

Where

$x_t$	=	the observation at time $t$ ,
$T_t$	=	the trend component,
$S_t$	=	the seasonal component,
$CL_t$	=	the cyclic component and
$Ir_t$	=	the irregular or error component, which is usually assumed to be uncorrelated.



As the observations were recorded monthly, a centred moving average of twelve observations was used in order to eliminate the seasonal effect.

The centred moving average  $[sm(x_t)]$  used in order to smooth out local fluctuations could be written as follows:

$$sm(x_t) = \frac{\frac{1}{2}x_{t-6} + x_{t-5} + \dots + x_{t+5} + \frac{1}{2}x_{t+6}}{12}$$

According to Chatfield (1980), this simple moving average is not generally recommended by itself for measuring trend but it can be used to remove seasonal variation. Since only one observation from each month was used in calculating the moving averages  $[sm(x_t)]$ , this method removes the seasonal variation from the data. It removes the irregular variations too, leaving a combination of the cyclic variation and the general long term movement embedded in the data.

The residuals obtained after subtracting the centred moving averages from the appropriate observations were grouped according to the months and averaged in order to remove the influence of irregular variation. Finally these averages were normalized so that the seasonal factors obtained add to zero. This was done by subtracting the mean of the averaged residuals, so that any excess of the sum of seasonal effects could be absorbed into the long term movement as suggested by Kendall (1973).

Once the seasonal factors have been estimated, a seasonally adjusted series was obtained by subtracting the appropriate seasonal factor from the monthly observation in order to get a more accurate estimate of the trend component. Traditionally, the trend component  $T_t$  is modelled by a low order polynomial of time  $t$  in the form:

$$T_t = \sum_{i=0}^k \beta_i \frac{t^i}{i!}$$

where  $\beta_i$  = the model parameters for  $i = 0, 1, \dots, k$ .

It was assumed that the order of the polynomial will not exceed 2 as suggested by Abraham and Ledolter (1983) for similar analyses. The method of least squares was used to fit the polynomial model in order to estimate the trend embedded in the data.

However, as this simple polynomial model was found to be inadequate, the trend embedded in the data was estimated using a non-linear least squares estimation procedure after including the lagged variables in the linear polynomial model.

As a well defined repeating cycle with a reasonably constant duration could not be identified after the removal of the seasonal and trend components from the original data, forecasts were computed using the estimated trend component and the seasonal component of the appropriate month.

## RESULTS

The first step in time series analysis is to obtain the time plot of the observations in order to visualize the important features such as the trend, seasonality, discontinuities and outliers (Chatfield, 1980). The time plot of  $x_t$  is shown in Figure 1. It gives evidence for a long term movement together with a periodic pattern. As the variation in the time series is dominated by a long term movement and oscillations around the long term movement, a simple model which attempts to describe the above features was used to analyse the data.

As the variance of the data did not appear to change with the mean level, a transformation was not carried out.

The residuals obtained after smoothening out the local fluctuation by the method of moving averages represent a combination of the seasonal variation and the irregular variations (Abraham and Ledolter, 1983). The seasonal factors were calculated by normalizing these residuals. The seasonal factor for March was found to be the highest followed by that of April while the seasonal factor for July was found to be the lowest (Figure 2).

After the removal of seasonal variation from the observed data, as the deseasonalized data were found to increase in an approximately linear pattern (Figure 3), a linear trend was fitted to the deseasonalized data. The trend was found to be

$$\hat{T}_t = 30.5 + 0.00013 * \text{time} \quad (1)$$

However, the coefficient of determination was found to be quite low (Table 1) casting doubt on the fitted model.

After the removal of the estimated trend and the estimated seasonal effects from the observations, the adequacy of the fitted model was checked by analyzing the residuals. But however, the correlogram of the residuals displayed serial correlations exhibiting non randomness. The first few autocorrelation coefficients of the residuals were well outside the 95% confidence limits (Figure 4).

When the true errors of a model are serially correlated, according to Abraham and Ledolter (1983), the usual least squares procedure will no longer be suitable. As the correlogram of the residuals exhibited a dying down pattern (Figure 4) and the partial correlogram exhibited a cut off after lag 2 (Figure 5), the true errors of the trend model could be considered to follow an autoregressive model of order 2.

Hence the simple linear regression model which could be used to explain the variation embedded in the trend was generalized using a non-linear least squares procedure. The new trend model was found to be

$$\hat{T}_t = 12.8 + 0.367 * y_{t-1} + 0.214 * y_{t-2} + 0.000567 * t \quad (2)$$

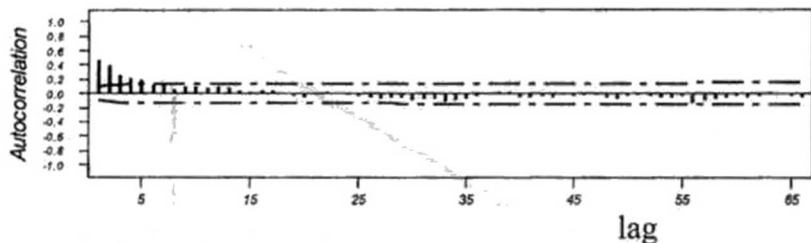
where  $y_{t-1}$  and  $y_{t-2}$  are the deseasonalized observations at time  $t-1$  and time  $t-2$ .

This model was found to be more suitable in explaining the variation in the trend (Table 2). The coefficient of determination was found to increase from 9% to 32.1% and according to the correlogram of the residuals of the new trend model (Figure 6), the errors of the model were found to be uncorrelated. The errors were also found to be normally distributed (Figure 7), with constant variance (Figure 8).

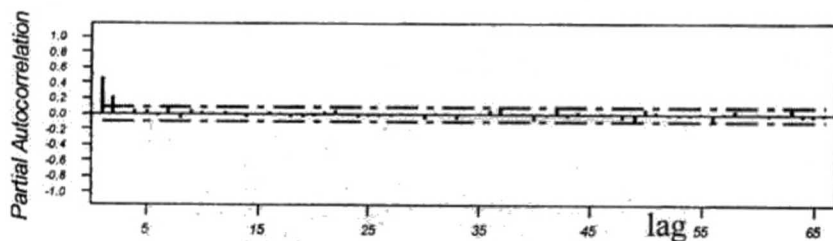
However, twelve different regression models were also fitted to the twelve monthly observations collected over the years from 1967 to 2002 in order to examine any relationship between the observed values of different months. These are given in Table 3. But for all twelve models, the coefficients of determination were found to be rather low (Table 3). Hence, in order to give a simple description of the underlying trend in the monthly mean maximum temperature values, a linear curve was fitted to the successive yearly averages of the monthly mean maximum temperature values using the method of least squares. According to the Chatfield (1980), with seasonal data showing a fairly small trend such a method will be perfectly adequate to provide a

simple description of the underlying trend. The underlying trend in the yearly averages of the monthly mean maximum temperature values was found to be:

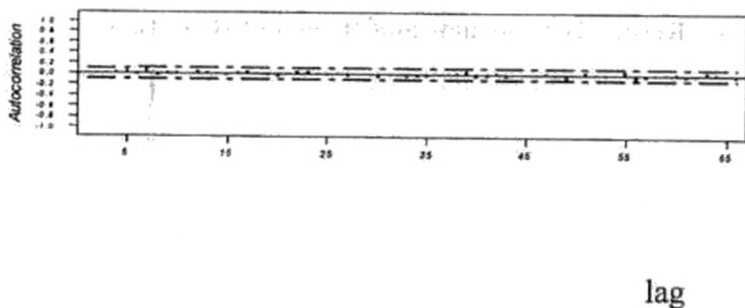
$$\text{trend} = -3.82 + 0.0175 * \text{time} \quad (3)$$



**Figure 4 :** Correlogram of the residuals. (The broken horizontal lines represent the 95 % confidence limits)

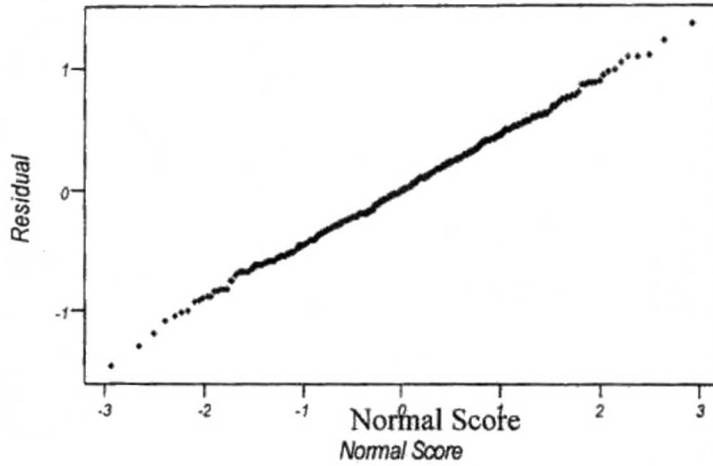


**Figure 5 :** Partial correlogram of the residuals. (The broken horizontal lines represent the 95% confidence limits)

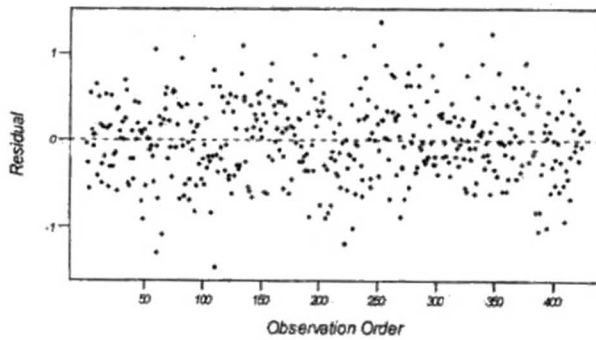


**Figure 6** –Correlogram of the residuals of the new model. (The broken horizontal lines represent the 95% confidence limits)





**Figure 7 – Normal probability plot of the residuals of the new model.**



**Figure 8 – Residuals of the new model verses fitted order**

**Table 1 : Regression Analysis for the deseasonalized observations****Regression Analysis**

The regression equation is  
 $\text{trend} = 30.5 + 0.00133 * \text{time}$

Predictor	Coef	S.D.	T	P
Constant	30.545	0.0504	605.49	0.000
Time	0.0013	0.0002	6.50	0.000

S = 0.5203    R-Sq = 9.0%    R-Sq(adj) = 8.8%

**Analysis of Variance**

Source	DF	SS	MS	F	P
Regression	1	11.430	11.43	42.22	0.000
Error	425	115.06	0.271		
Total	426	126.49			

**Table 2 : Regression Analysis for the deseasonalized observations after including the lagged variables**

The regression equation is

$\text{Trend} = 12.8 + 0.367*y_{t-1} + 0.214*y_{t-2} + 0.000567*t$

Predictor	Coef	S.D.	T	P
Constant	12.798	1.5030	8.52	0.000
$y_{t-1}$	0.367	0.0476	7.72	0.000
$y_{t-2}$	0.214	0.0476	4.49	0.000
time	0.001	0.0002	2.99	0.003

S = 0.4514    R-Sq = 32.1%    R-Sq(adj) = 31.6%

**Analysis of Variance**

Source	DF	SS	MS	F	P
Regression	3	40.490	13.497	66.25	0.000
Error	421	85.772	0.204		
Total	424	126.26			

$y_{t-k}$  = deseasonalized observation at time t-k for k=1,2.

Table 3 – Monthly regression models and the  $R^2$  values.

Month	Regression model	$R^2$
January	$Y = 30.8 + 0.0209t$	9.4%
February	$Y = 31.2 + 0.0139t$	5.3%
March	$Y = 31.5 + 0.0225t$	13.7%
April	$Y = 31.7 + 0.00912t$	2.7%
May	$Y = 30.9 + 0.0206t$	17.4%
June	$Y = 30.4 + 0.00412t$	0.8%
July	$Y = 29.9 + 0.0153t$	10%
August	$Y = 29.9 + 0.014t$	14.1%
September	$Y = 30.2 + 0.00493t$	1.1%
October	$Y = 29.9 + 0.0158t$	12.5%
November	$Y = 30.0 + 0.0194t$	20%
December	$Y = 30.1 + 0.0333t$	24.9%

Table 4 – Observed and the forecasted values of the last fifteen observations

Observed values ( $^{\circ}\text{C}$ )	Forecasted values ( $^{\circ}\text{C}$ )
31.7	32.0
31.7	32.1
31.8	31.3
30.5	30.7
29.9	30.6
30.8	30.5
30.8	30.7
30.5	30.6
30.8	30.8
30.6	31.0
31.3	31.5
31.9	31.7
32.2	32.1
32.0	32.1
31.6	31.5

## DISCUSSION

The seasonality in most meteorological variables can be explained on the basis of physical reasons. The earth's rotation around the sun introduces an yearly seasonal pattern into many meteorological variables (Abraham and Ledolter, 1983). Although the seasonal pattern in many meteorological variables are usually quite stable, there is a possibility of the seasonal factors for time periods in the same season to change over time, resulting a shifting seasonal pattern. However, evidence for a changing seasonal



pattern was not found in the present study. Hence the seasonal pattern in the mean maximum temperature in Colombo could be considered as approximately constant.

Most common use of the information provided by decomposition methods involve generating forecasts using the trend and seasonal variation only. Irregular movements are unpredictable by their very nature and do not reveal any pattern of the repetitive tendency. Usually irregular components are not considered when predicting the future values (Pankratz, 1983). Mostly these complex fluctuations remain unanalysed. Cycles generally exhibit semi regular periodicity as they are neither as regular as the seasonal variations nor as accidental as the random fluctuations. If a well defined repeating cycle with a reasonably constant duration could have been identified, cyclic factors could have been estimated too. However, this was not possible in the present study. Nevertheless, as the cyclic behaviour in the time series is reflected in the estimated trend component, it does not mean that the cyclic behaviour of the time series has been completely ignored.

Finding a suitable model for forecasting future values depends on a number of factors such as availability of money, time, data, expected lead time, accuracy of the forecast and expected usage of the forecasting procedure. As there is no universally applicable forecasting procedure, forecaster should select the procedure that is most appropriate for the given set of conditions. When fitting a model to a data set one should always try to fit a parsimonious model as the estimation of each unnecessary parameter introduces additional uncertainty into forecasts. Before a fitted model is used for forecasting, investigations should be made to check whether the model provides an adequate description of the historical data. A limitation of the present model is the low  $R^2$  value. This may have resulted by the outlier values produced due to El-niño and La-niña effects.

When the forecasted values of the last fifteen months were examined (Table 4), the additive seasonal model with the estimated trend and seasonal components, i.e.

$$\hat{x}_t = \hat{T}_t + \hat{S}_t$$

where  $\hat{x}_t$  = the estimated temperature at time  $t$ ,

$\hat{T}_t$  = the estimated trend at time  $t$  as described in equation (2)

and  $\hat{S}_t$  = the estimated seasonal factor at time  $t$ ,

was found to fit the past data fairly well and this model could be effectively used to predict the future monthly mean maximum temperature values in Colombo.

The present study indicates that the mean maximum temperature in Colombo increases at a rate of 0.0175 °C per annum. This rate of increase is smaller than those recorded for some other regions of Sri Lanka such as Puttalam which was 0.0325 °C per annum. However, the rate of increase in mean maximum air temperature in Colombo is greater than that recorded for Nuwara Eliya which was 0.0013 °C per annum (Basnayaka *et al.*, 2002). The rate of increase in mean maximum air temperature in Colombo is smaller than the mean global rate of increase in temperature, which is 0.02°C – 0.05°C (Singer, 1998, Veteli *et al.*, 2002, Anon., 2002). It has been suggested that urbanization also contribute to increase in mean temperature (Fernando and Basnayaka, 2002). However, since the rate of increase of temperature in Puttalam is higher than that of Colombo, it appears that a complex of factors affects the

rate of increase in mean air temperature. Those factors may include the amount of cloud cover in the sky and emission of smoke from vehicle exhaust in addition to other factors such as urbanization, amount of greenhouse gases emitted and vertical distribution of water vapour in the atmosphere.

The present analysis of the monthly mean maximum temperature in Colombo indicates that short term correlation in a month which is warmer/colder than the average for that month will tend to be followed by one or two further months which is/are warmer/colder than the average.

### CONCLUSIONS

As the variation in the mean maximum temperature in Colombo is dominated by a pronounced seasonal effect and a slight trend, the model,

$$\hat{x}_t = \hat{s}_t + 12.8 + 0.367(x_{t-1} - \hat{s}_{t-1}) + 0.214(x_{t-2} - \hat{s}_{t-2}) + 0.000567 * t$$

where

$\hat{x}_t$  = the estimated temperature at time t,

$\hat{s}_{t-k}$  = the estimated seasonal factor at time t-k for k = 0, 1, 2

and  $x_{t-k}$  = observed mean maximum temperature value at time t-k for k = 1, 2

could be effectively used to predict the future monthly mean maximum temperature values in Colombo. It appears that there is a tendency for the monthly mean maximum temperature in Colombo to increase at a rate of 0.0175 °C per annum, which is less than the rate of increase in the mean global temperature.

### REFERENCES

- Abraham, B. & Ledolter, J. (1983). *Statistical Methods for Forecasting*. John Wiley and Sons, New York.
- Anonymous (2002). NAS report confuses public. [www.globalwarming.org](http://www.globalwarming.org).
- Basnayaka, B. R. S. B., Fernando, T.K. & Vithanage, J. C. (2002). Variation of air temperature and rainfall during Yala and Maha agricultural seasons, p 212. *Proceeding of the Sri Lanka Association for the Advancement of Science, Part 1 (Abstracts)*, 58<sup>th</sup> Annual session, 02 – 07 December 2002, Colombo.
- Chatfield, C. (1980). *The Analysis of Time Series: An Introduction*. Chapman and Hall, London and New York.
- Fernando, T.K. & Basnayaka, B. R. S. B. (2002). A comparative study of the ambient air temperature at Anuradhapura and Mahailuppalama. p 213, *Proceedings of the Sri Lanka Association for the Advancement of Science, Part 1 (Abstracts)*. 58<sup>th</sup> Annual Session, 02 – 07 December, 2002, Colombo.
- Kendall, M.G. (1973). *Time Series*. Griffin, London.

- Kendall, M.G. & Stuart, A. (1968). *The Advanced Theory of Statistics: Vol 3*. Griffin, London.
- Pankratz, A. (1983). *Forecasting with Univariate Box-Jenkins Models*. John Wiley and Sons, New York.
- MOFE (2000). *Statistical Compendium on Natural Resources Management For Sustainable Development*. Ministry of Forestry and Environment, Colombo, Sri Lanka.
- Singer, S.F. (1998). Why are we not seeing global warming?. Paper presented at the 9<sup>th</sup> symposium on global change studies. American Meteorological Society. January 11 – 16, 1998, Phoenix, Arizona.  
[www.sepp.org/scirsrch/amsglwarm.html](http://www.sepp.org/scirsrch/amsglwarm.html).
- Singer, S.F. (1999). Is there a connection between El-niño and global temperature?  
[www.sepp.org/scirsrch/elnino.html](http://www.sepp.org/scirsrch/elnino.html).
- Veteli, T.O., Kuokkanen, K., Julkunen - Titto, R., Roninen, H. & Tahvaninen, J. (2002). Effects of elevated CO<sub>2</sub> and temperature on plant growth and herbivore defensive chemistry. *Global Change Biology*, 8 (12) : 1240 – 1252.
- Stevens, W.K. (2002). Skeptic asks, is it really warmer?  
[www.globalwarming.org/news/lindzen.html](http://www.globalwarming.org/news/lindzen.html).



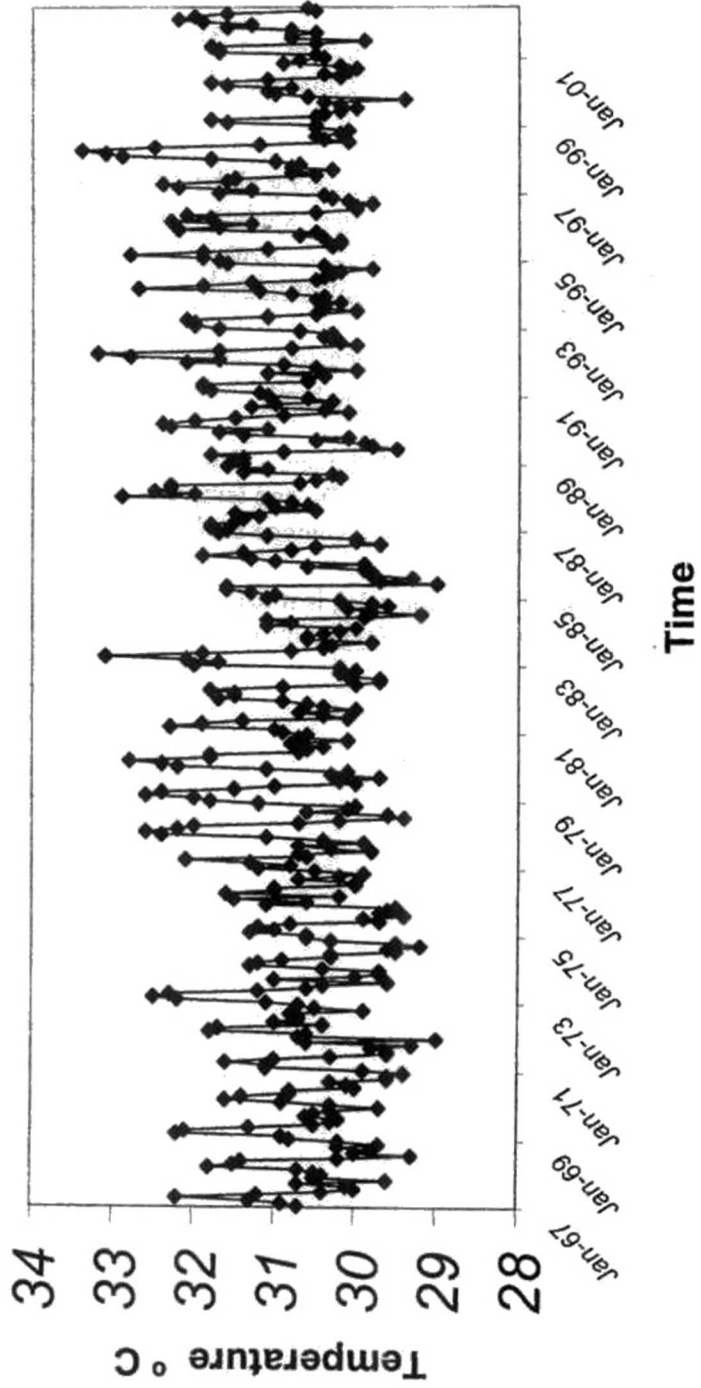


Figure 1 - Time Plot of the observations

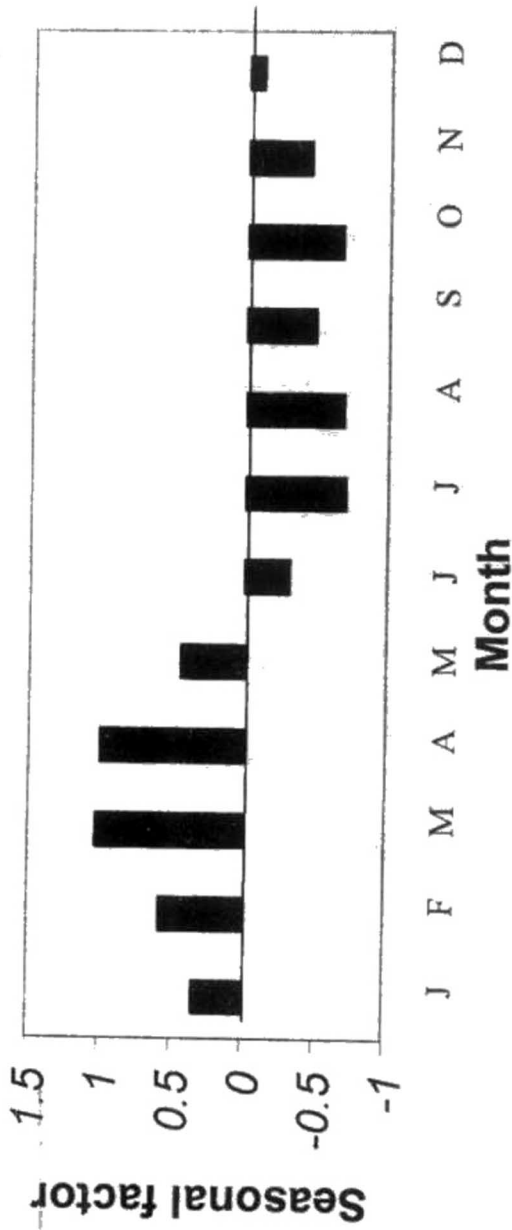


Figure 2 – Monthly seasonal factors

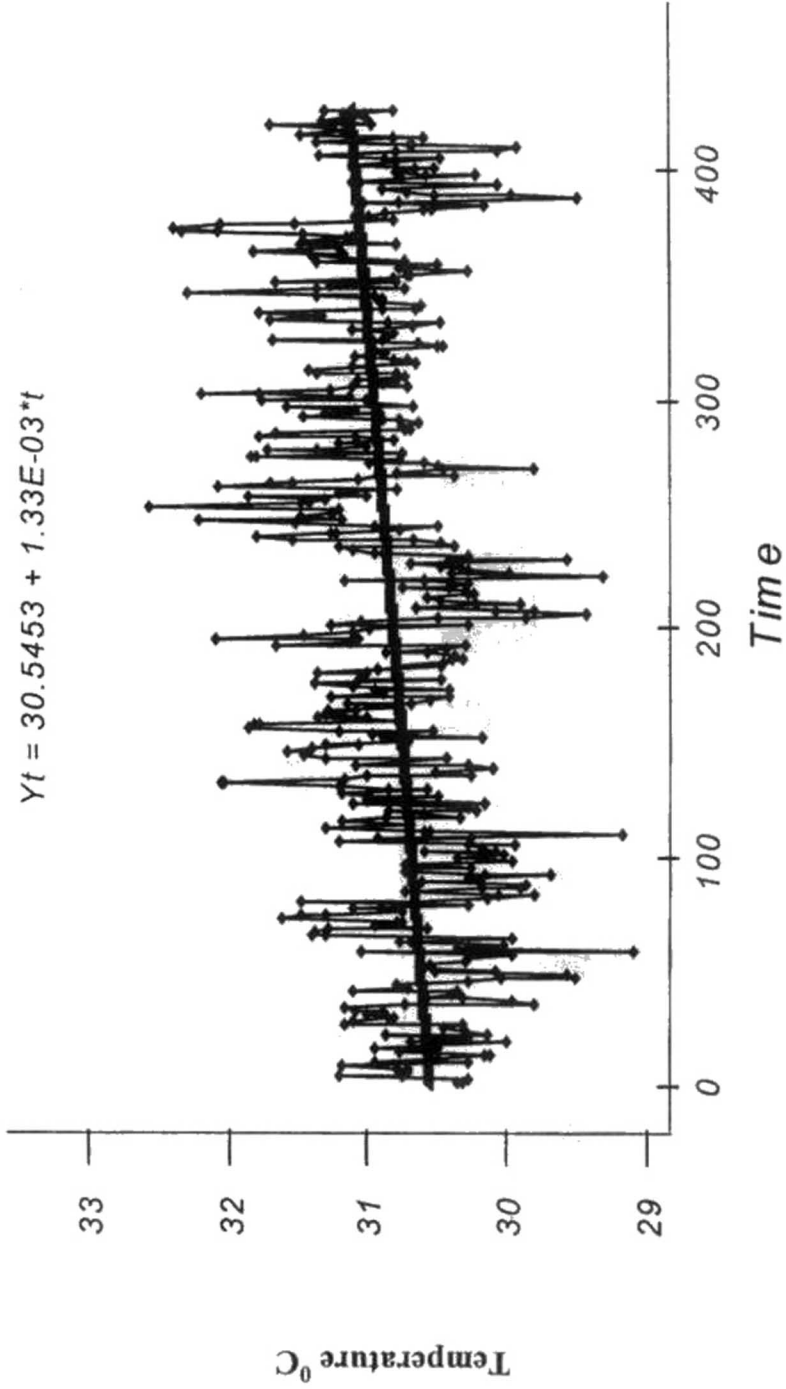


Figure 3 –Linear model fitted to the deseasonalized observations