Temperature Trend Analysis Using Non-Linear Regression of Kohat, Northwestern Pakistan

Muhammad, J.[1], M. Adnan[2, 3] I. Din[1], A. A. Khan[2], W. Ali[4], S. Jehan[1]

ABSTRACT

Accurate climatic prediction is a need of the day and it is possible if the factors affecting climate are thoroughly investigated. To know the annual temperature trend the data of past 50 years (1961-2010) of Kohat station, Pakistan was studied. Short-term changes (10 years) and long-term changes (50 years) are analyzed by using regression models. Tenth order polynomial is the best-fitted model for short-term changes in temperature. The cyclical variation of minimum temperature showing a long-term decline and the cycle time span varies from 10 to 15 years. The maximum temperature recorded at the station exhibit cyclic variations with cycles ranging over a decade. Overall trend of the maximum temperature is showing positive trend. The annual mean temperature pattern indicates an overall increasing trend. The trend of fitted models indicate 2.08 °C rise in the maximum temperature, 0.36 °C fall in minimum temperature and 0.86°C rise in the annual temperature during the study period.

Key Words: Temperature, Trend analyses, Polynomial Equation, Climate Change, Pakistan

Introduction

The issues relating to climate change got strong attention on a global scale during last decades. Gradually rising temperatures and change in rainfall pattern are evident in many regions around the world. From the scientific evidence, it has been concluded that the current rate of global warming is 2°C per century (Salinger 2005), which goes to increase global-average surface temperatures by 2°C~4.5°C by the end of this century. Other studies showed that there was an increase of 0.3–0.6 °C in the19th century and it was raised 0.2–0.3 °C only in last 40 years(Jones & Briffa 1992), while the1990s decade was considered as the warmest decade. Another study (Hingane and Kumar 1985) reported 0.4°C rise per 100 years during the 20th century until the 1980s. While other scientist (Kothawale and Kumar 2002) reported a rise of 0.5°C in last hundred years. Accelerated melting glaciers, intense and extreme climatic events like drought and floods are the signs of climate variation in Pakistan (Harasawa et al., 2007). A number of studies with different approaches have been conducted in Pakistan to study climate variation as well as the warming trend in the country (Chaudhry 2002; Chaudhry and Rasul 2007; Afzaal and Haroon 2009).

These studies indicate that Pakistan's climate is experiencing some prominent changes including temperature rise. This increasing trend in temperature affects water resources and agriculture sector among others. Most of the water resources of Pakistan depends on snow melting. The increase in snowmelt due to the rise in annual maximum temperature may result in flooding in lower plains. Glacial melts serve as an important source of surface water in Pakistan. The rise in temperature in winter may also cause a significant decrease in snowfall, hence, the existing surface water resources of Pakistan are under severe threat. On the other hand, increase in temperature directly increases crop water requirements thus increases demand for water resources. Such a situation may lead to decrease in the crop yield. Among the other environmental factors, minimum temperature is the most important factor which affects plant leaf growth (Lyons 1973).

Pakistan is an agricultural country, while agriculture is totally dependent on the temperature of the growing season. So, projections and prediction of temperatures in the climate change era are very important. For developing good relation in hindcasting and forecasting of temperature, a modelling approach is required

[1] The University of Agriculture, Peshawar- Pakistan.
[2] Global Change Impact Studies Centre, Ministry of Climate Change, Islamabad, Pakistan
[3] scarpion982@gmail.com
[4] National Centre of Excellence in Geology, University of Peshawar- Pakistan.

(Houghton et al. 1995). Method of nonlinear forecasting was used for predicting the behavior of cloud coverage time series several hours in advance (Muñnuzuri & Gelpi 2000). Linear regression is one of the most commonly used methods for identifying trends in climate data (Sajjad et al. 2009; Singh et al. 2013). In this study, a new approach was used and changes in temperature trends were extracted with nonlinear regression models and the best-fitted model has been selected via linear regression graphs with R^2 . At the end, best fit models have been selected and their variables were defined. The analyses can be used as a baseline study and as an input for the upcoming planning and projects.

This study is limited to meteorological data of Kohat station, situated in the province of Khyber Pakhtunkhwa in North West-Pakistan. This station was selected because this is near to the snow-covered mountains of Hangu in the west, while this station also experiences warm blows from the plains situated in its south-east. So, the station is vulnerable to the small changes in either minimum temperature and in maximum temperature. The climate of Kohat and surrounding areas is hot from May to September having June as the hottest month. The mean maximum and mean minimum temperature recorded during June is about 40 °C and 27 °C respectively. The winter is cold and severe. In the winter winds from west known as "Hangu Breeze" often blows down the Miranzai valley towards Kohat for weeks and resulting in significant decrease in temperature. The mean maximum and mean minimum temperature, recorded during the month of January, is about 18 °C and 6 °C respectively. Annual rainfall is about 638 mm and most of the rainfall is received in the monsoon season. The monsoon rain is received from June to September, while August is the wettest month with an average of about 114 mm. The winter rain occurs from December to March and the highest winter rainfall is received in the month of March. The time series analyses of temperature was conducted for Kohat station, southern zone of Khyber Pakhtunkhwa (KP), Pakistan.

Data and Methodology

Monthly data of temperature during 1961 to 2010 were collected from Pakistan Meteorological Department (PMD), Government of Pakistan, which is the only reliable, competent and official source for meteorological data in Pakistan. Mean minimum temperature ($M_{min}T$), mean maximum temperature ($M_{max}T$) and mean annual temperature ($M_{ann}T$) from 1961 to 2010 were used to find out the possible changes in temperature of Kohat station. The DataFit software was used to analyze the collected data of temperature for the mentioned period. DataFit is a statistical tool that simplifies the tasks of data plotting, regression analysis (curve fitting) and statistical analysis. The best fit option in the software provides linear and nonlinear regression models which are ranked according to the goodness of the fit.

Linear regression requires a linear model where each term is either a constant or the product of a parameter and a predictor variable has a power or index of one. This constrains the equation to just one basic form:

Response = constant + parameter * predictor

$$Y = b_o + b_1 X \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

In equation (1), Y is response/dependent variable, X is independent variable, b_o is the model intercept, and b_1 is the regression coefficient. While analyzing the data, Y is replaced by annual values of MminT, MmaxT and MannT as a dependent variable while X is the time in years is used as the independent variable.

On the other hand, nonlinear equations (Bates and Watts 2007; Rhinehart 2016) can take many different forms. The easiest way to determine whether an equation is nonlinear is to focus on the term "nonlinear" itself. If the equation doesn't meet the criteria for a linear equation, then it is called nonlinear. Non-linear equations provide the most flexible curve-fitting functionality. In the present study, regression models, non-linear in variable(s) were applied to achieve the required objectives. Generally, non-linear equation of order k, which is not linear in variables, is defined in equation (2).

$$Y = a + b_1X + b_2X^2 + \ldots + b_kX^k \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

In equation (2), all the terms have the same meanings as described in equation (1). By analyzing the data, Y is replaced by annual values of MminT, MmaxT and MannT as a dependent variable while X is the time in years is used as the independent variable. It is pre-requisite to mention that equation (2) is linear in parameters but not linear in variable as entering in different powers instead of power 1.

In order to solve more than one regression models at a time on DataFit, the summary results provides a quick way to view the basic goodness of fit statistics for all the models at the same time. Best model Data Fit is based on Standard Error, Residual Sum, Residual Average, Residual Sum of Squares and R^2. The goal of nonlinear regression is to determine the best-fit parameters for a model by minimizing a chosen merit good fit criteria. The non-linear regression differs in that the model has a nonlinear dependence on the unknown parameters, and the process of merit function minimization is an iterative approach. The process is to start with some initial estimates and incorporates algorithms to improve the estimates iteratively. The new estimates then become a starting point for the next iteration. These iterations continue until the merit function effectively stops decreasing. The nonlinear model to be fitted can be represented by:

$$Y = Y(X; a) \ldots \ldots \ldots \ldots (3)$$

The merit function minimized in performing nonlinear regression is given in equation (4):

$$x^2(a) = \sum_{i=1}^{N} \{\frac{y_1 - y(x_i; a)}{\sigma_i}\}^2 \ldots \ldots \ldots \ldots (4)$$

Where σ_i is the measurement error, or standard deviation of the ith data point. As with nonlinear regression, we are minimizing the sum of the squares of the distances between the actual data points and the regression line. To check the statistical significance, the null hypothesis Ho: $\beta_i = 0$ is tested by using t-ratio for each of the parameters involved in the model at 5 % level of significance with the following equation (5).

$$t_{cal} = \frac{\hat{\beta i}}{S.E(\hat{\beta i})} \ldots \ldots \ldots \ldots (5)$$

Where, $\hat{\beta i}$ is the regression coefficient and S.E is the standard error of $\hat{\beta i}$ which is given in equation (6):

$$S.E_{yx} = \sqrt{\frac{\sum(y-\bar{y})^2 \frac{[\sum(x-\bar{x})(y-\bar{y})]^2}{\sum(x-\bar{x})^2}}{n-k}} \ldots \ldots \ldots \ldots (6)$$

Where n is the sample size.

Furthermore, the best fit model was selected on the basis of relatively higher coefficient of determination (R^2), which shows the percent of variation in temperature explained by time in years.

Results and Discussions

Figure 1 shows mean minimum, mean maximum and mean annual temperature recorded, with their respective years. The simple linear regression model is fitted to temperature data to know long-term (50 years) changes at Kohat station. The parameters calculated for linear equation are shown in Table 1. It shows the trend line of the linear regression equation and coefficient of determination (R^2). In Table 1, regression equation of mean maximum and mean annual temperature's beta values are positive, which means that they have positive relationship. In other words, there is an increase in the long-term mean maximum and mean annual temperature with increase in time. On the other hand, a decreasing trend has been identified in the mean minimum temperature data. This could be attributed to the overall rise in surface temperature which causes an increase in snowmelt and cold breeze movement from the hills of Wadi Terrah towards Kohat and make its winter cooler and causing a decrease in mean minimum temperature. The long-term analysis shows that the climate of Kohat is becoming intense with the summers becoming hotter and winter becoming cooler with time.

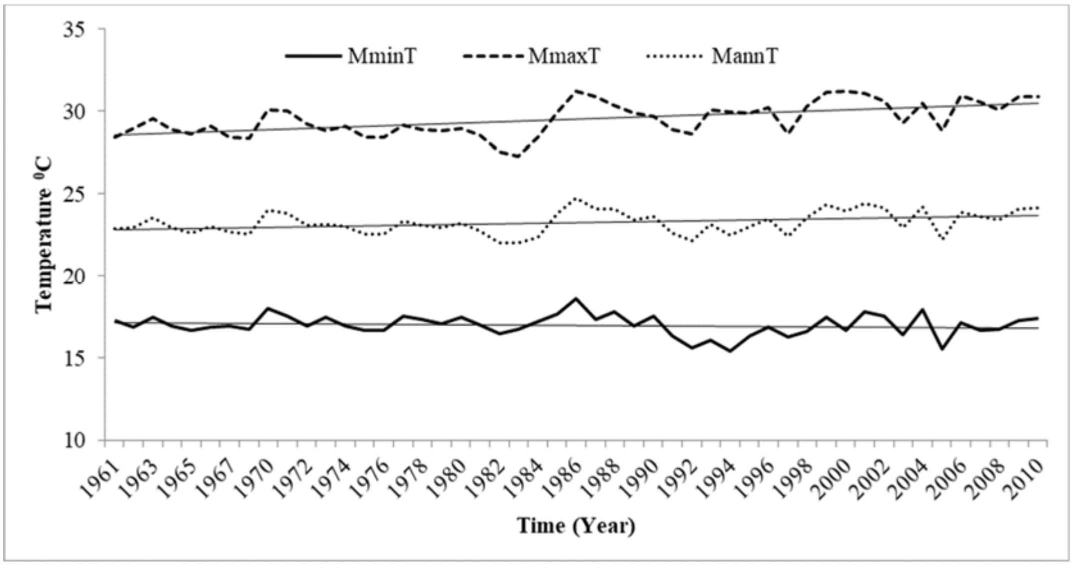

Figure 1: Mean minimum, Mean maximum and Mean annual temperature trend

Table 1: Linear equations for long-term temperature projection

Temperature	Alpha (α)	Beta (β)	R^2	Projected 50 years Temperature (^{0}C)
Mean Maximum ($M_{max}T$)	28.480	0.0416	0.35	30.560
Mean Annual ($M_{ann}T$)	22.811	0.0173	0.13	23.676
Mean Minimum ($M_{min}T$)	17.175	-0.0072	0.03	16.815

Mean Minimum Temperature($M_{min}T$) Trend

Figure 2 shows the observed mean annual minimum temperature and the best fit model that explains its variation and changes. It is clear from the figure that at a regular interval of ten years there is cyclic variation (rise and fall) in the mean annual minimum temperature. It is concluded that the average temperature of the region rise and fall gradually in a decades time period as observed in trend analysis (Ahmad et al. 2009). But overall there is a decreasing trend in the mean annual minimum temperature in the study area.

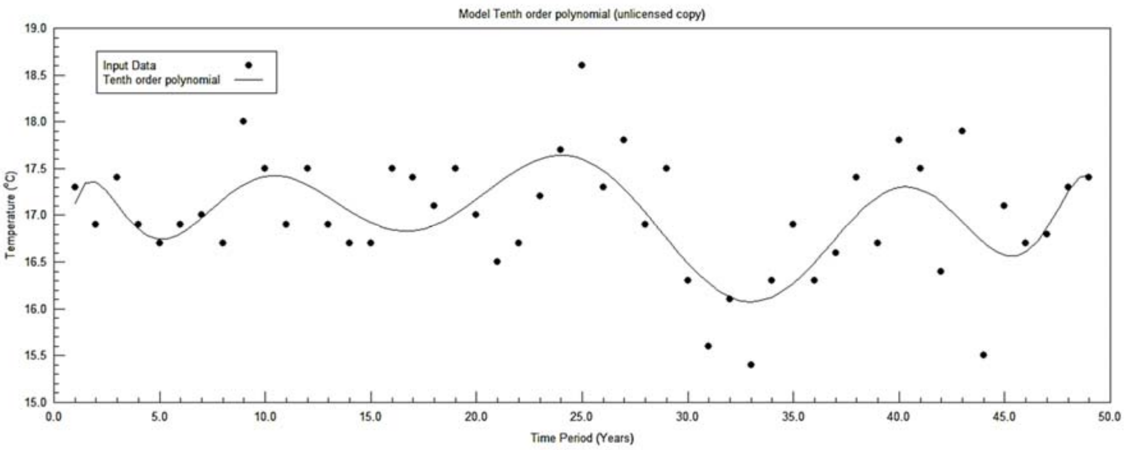

Figure 2: Observed Trend in Mean minimum temperature (MminT).

Table 2 shows the model fitness and comparison of the best-fit models. Based on the defined criteria of the high value of R^2 it was found that Tenth order polynomial was the best predicted model for the Mean minimum temperature ($M_{min}T$) in the study area followed by Ninth (The results of 9^{th} order polynomial are missing in Table-2) and then Eight order polynomial model.

Table 2: Model fitness Comparison for Mean minimum temperature (MminT).

Rank	Model	Std. Error	RSS	R^2
1	Tenth order polynomial	0.54	11.37	0.40
2	Eighth order polynomial	0.58	13.70	0.28
3	$ax^4+bx^3+cx^2+dx$ (fourth order polynomial without intercept)	0.62	17.43	0.09
4	Fifth order logarithm	0.64	17.61	0.08
5	Fifth order inverse polynomial	0.64	18.04	0.05

Std. Error = standard error of estimate; RSS = residual sum of squares; R2 = coefficient of determination; x =independent variable.

Table 3 shows the results of analysis of variance (ANOVA) for the Tenth order polynomial, which indicates that the model is overall significant ($P < 0.05$). While Table4 shows all the regression models of various order of polynomial which include standard error, t-statistic and its p-value. While "x" in the equation is time period in a number of years from 1961 to 2010. The p-value for the t-statistics of all the regression coefficient shows that all of them are significantly affecting temperature trend.

[These results are not shown in Table4], actually it shows the summary of goodness of fit of the models of various orders of polynomial.

Table 3: Analysis of Variance (ANOVA) for Tenth order polynomial

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Regression	10	7.80	0.78	2.61	0.02
Error	38	11.37	0.30		
Total	48	19.17			

Where DF, F-ratio and P-value denote degrees of freedom, F-statistic and probability value, respectively.

Mean Maximum Temperature (MmaxT) Trend

The natural trend followed by annual Mean maximum temperature is shown in Figure 3 with the best-fit polynomial model. It is evident that mean maximum temperature follow a decade wise cyclic and inclining nature as followed by mean minimum temperature. While the overall trend shows slightly upward trend. Based on the results, it is inferred that mean maximum temperature is cyclic with increasing trend.

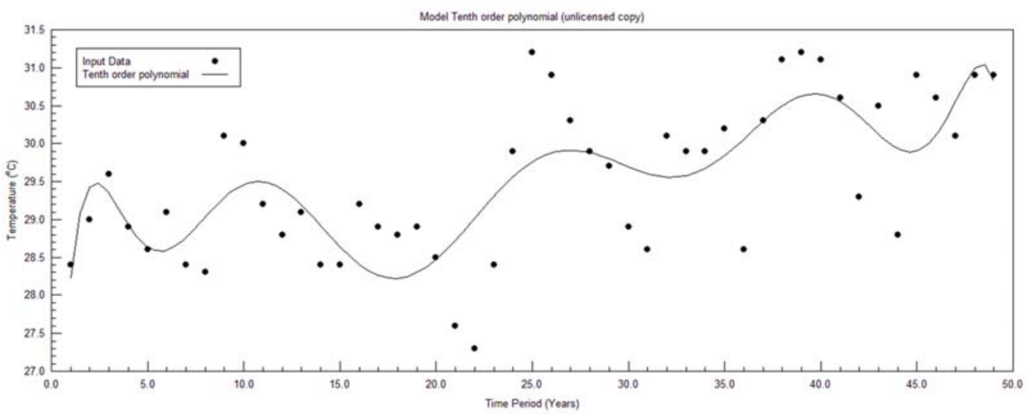

Figure 3: Observed Trend in Mean Maximum Temperature (MmaxT)

Table 4 shows model fitness comparison of top five models with different orders of polynomial. It is clear that the highest R^2 was produced by Tenth order polynomial model. In other words, 54 % of the variation in mean maximum temperature is explained by the Tenth order polynomial model while this value is less for the others fitted model. Therefore, Tenth order polynomial model was considered as the best fitted model for the trend forecasting of Mean Maximum Temperature (MmaxT) in the study area.

Table 4: Model fitness Comparison for Mean Maximum Temperature (MmaxT).

Rank	Model	Std. Error	RSS	R^2
1	Tenth order polynomial	0.76	22.33	0.54
2	Seventh order polynomial	0.83	28.53	0.41
3	Fifth order logarithm	0.83	29.73	0.39
4	Fourth order logarithm	0.82	29.92	0.38
5	Third order logarithm	0.81	29.93	0.38

The results of ANOVA for the Tenth order polynomial model (MmaxT) are displayed in Table 5. It depicts that the P-value of F-ratio is less than 5 % level of probability suggesting that the fitted model is overall significant and best fitted for MmaxT data of Kohat district.

Table 5: Variance Analysis for Mean Maximum Temperature (MmaxT)

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Regression	10	26.67	2.66	4.53	0.0003
Error	38	22.33	0.58		
Total	48	49.00			

Where DF, F-ratio and P-value denote degrees of freedom, F-statistic and probability value, respectively.

Mean Annual Temperature (MannT) Trend

Mean annual temperature short-term trend is shown in Figure 4. As it is the average of the mean maximum and mean minimum temperature so it also follows the same cyclic nature but with a little increasing trend as compared to the trend shown in Figure 2 and 3. The increasing trend in the mean annual temperature is due to relatively increase in mean maximum temperature which overshadowed the decreasing effect of mean minimum temperature

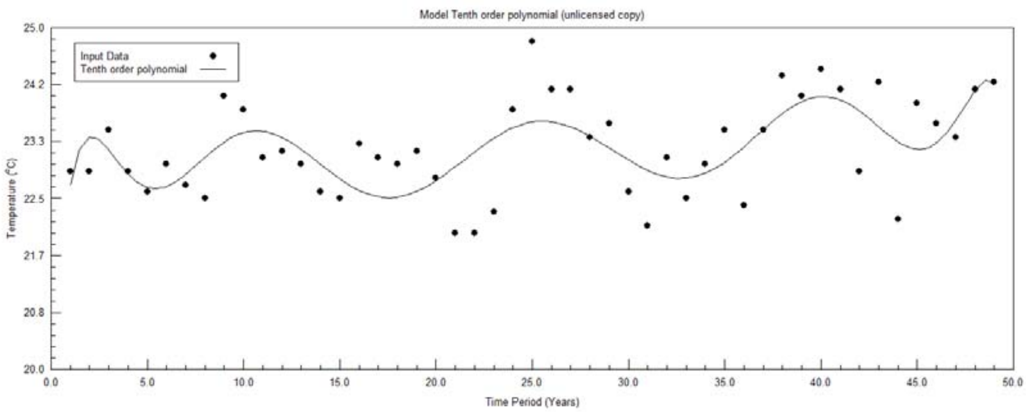

Figure 4: Observed Trend in Mean Annual Temperature (MannT)

Based on the best fit model, Mean Annual Temperature (MannT) was also following Tenth order polynomial model as shown in Table 6. The relevant P-value (0.0268) of the F-ratio indicates that significant variation in the temperature data is explained by tenth order polynomial (Table 7). **Table 6:** Model fitness Comparison for Mean Annual Temperature (MannT)

Rank	Model	Std. Error	RSS	R^2
1	Tenth order polynomial	0.61	14.52	0.38
2	Eighth order polynomial	0.67	18.21	0.22
3	$a+bx^2\ln(x)+c.\exp(x)$	0.65	20.03	0.15
4	$a+bx^2+c\exp(x)$	0.66	20.04	0.15
5	$a+bx^3+c.\exp(x)$	0.66	20.06	0.14

Table 7: Variance Analysis for Mean Annual Temperature (MannT)

Source	DF	Sum of Squares	Mean Square	F-ratio	P-value
Regression	10	9.07	0.90	2.37	0.02681
Error	38	14.52	0.38		
Total	48	23.59			

Where DF, F-ratio and P-value denote degrees of freedom, F-statistic and probability value, respectively.

Discussion

Simple linear regression method has been used by various authors for identifying trends in climatic data owing to its simplicity and ease of use (Sajjad et al. 2009; Singh et al. 2013). However, the procedure has a few weaknesses; (i) Non-normal distribution of climate data and the presence of outliers may limit its usefulness in climate studies (ii). It also does not account for the high amount of variability found in climate data variables. In order to tackle these problems in many instances, climate data is studied using non-parametric statistics like Mann Kendall test. However, in this study, nonlinear regression models were compared with linear regression methods for the identification of trends in temperature data sets. The dataset were acquired from a station subjected to high temporal and seasonal variations. The analysis shows an expected rise of 0.86 °C in mean annual temperature and 2.08 °C increase in mean maximum temperature, it also shows a decline of 0.36 °C in mean minimum temperature, suggesting a trend towards severity in climate. For Pakistan as a whole, a 0.57 °C of increase in mean annual temperature has been reported during the period of 1901 till 2000 and an increase of 0.87 °C in mean maximum temperature during the period of 1960-2007 by others (Chaudhry and Rasul 2007). Pakistan's climate is highly variable and therefore changes are not linear (Chaudhry 2002). Other studies conducted with different methods and at different scales also support the warming tendency of the Pakistani climate (Afzaal and Haroon 2009; Ahmad et al. 2014). Being an agriculture country, climate change is an important concern for Pakistan's economy and society. Kohat is by and large a rain-fed area; a study conducted on the impacts of Temperature rise in Punjab, Pakistan has shown that a reduction in per capita wheat production is expected with the rising temperature scenario. In addition, the impact of rise in temperature on wheat production is expected to be more profound in rainfed areas as compared to the irrigated areas. The rise in temperature may also lead to increase in vector-borne diseases. A comprehensive study conducted by Bouma et al. (1996) in Khyber Pakhtunkhwa province (Peshawar City) reported an increase in November and December temperatures by 2 °C and 1.5 °C respectively since 1876, which, they suggested has led to the increase of incidences of Falciparum Malaria in the region.

Conclusions

From the results, it is concluded that temperature trend in the study area is best explained by the nonlinear regression as compared to linear regression analysis. The decade wise rise and fall were impossible to be covered by linear regression as a linear trend only shows the linear relationship between temperature and time. However, the non-linear regression was unable to capture the 100 % decadal variation but it captures some portion of the decadal variations as non-linear regression has the component of the higher degree polynomial which has the capability of capturing the portion of variation other than the trend. The Mean minimum temperature ($M_{min}T$), mean maximum temperature ($M_{max}T$) and mean annual temperature ($M_{ann}T$) follows Tenth order polynomial model and suggested to be adopted for prediction/forecasting purpose. A significant increase in mean maximum and mean annual temperature and a decreasing trend in mean minimum temperature was identified using the linear and non-linear regression models in the current study. It is recommended that for stations subject to high variations in climate data nonlinear models might be an appropriate tool to identify trends in temperature. The cyclical variations of minimum temperature show long-term decline and the cycle time span vary from 10 to 15 years. The maximum temperature recorded at the station also exhibit cyclic variations (the cycles duration ranging over a decade). The upward trend of annual mean temperature again shows cyclic variation. The linear fit models indicate 2.08 ºC rise in the mean maximum temperature, 0.36 ºC fall in mean minimum temperature and 0.86ºC rise in the mean annual temperature during the study period.

References

Afzaal, M., & M. A. Haroon, 2009: Interdecadal Oscillations and the Warming Trend in the Area-Weighted Annual Mean Temperature of Pakistan. *Pakistan Journal of Meteorology*, 6(11), pp.13–19.

Ahmad, I., et al., 2009: Trend analysis of january temperature in pakistan over the period of 1961-2006 : geographical perspective. *Pakistan Journal of Meteorology*, 7(13), pp.11–22.

Ahmad, I., et al., 2014: The analysis of temperature based on interdecadal variability of regional temperature in pakistan : case of february over the period of 1950s-2000s. , 69(1), pp.1–6.

Bates, D. M., and D. G. Watts. 2007: Nonlinear Regression Analysis and Its Applications. *Willy Series in Probability and Statistics.*

Bouma, M. J., C. Dye, & H. J. V. D. Kaay, 1996: Falciparum malaria and climate change in the Northwest Frontier Province of Pakistan. *American Journal of Tropical Medicine and Hygiene*, 55(2), pp.131–137.

Chaudhry, Q. Z., and G. Rasul, 2007: Global Warming and Expected snowline shift along Northern Mountains of Pakistan. Proceedings of 1stAsiaclic symposium.

Chaudhry, Q. Z., et al., 2002: Climate Change and its Impact on the Water Resources of Mountain Regions of Pakistan. *Pakistan Journal of Geography*, 11.

Harasawa, H., et al., 2007: Asia. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. , pp.469–506.

Hingane, L. S., K. R. Kumar, B. V. R. Murty, 1985: Long term trends of surface air temperature in India. *Journal of climatology*, 5, pp.521–528.

Houghton J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. V. D. Linden, X. Dai, K. Maskell, C. A. Johnson, 1995: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press: Cambridge*.

Jones, P. D., & K. R. Briffa, 1992: Global surface air temperature variations during the twentieth century: part 1, spatial, temporal and seasonal details. *Holocene*, 2(2), pp.165–179.

Kothawale, D. R., and K. R. Kumar, 2002: Tropospheric temperature variation over India and links with the Indian summer monsoon : 1971-2000. *Mausam*, 3(July), pp.289–308.

Lyons, J. M., 1973: Chilling Injury in Plants. *Annual Review of Plant Physiology*, 24(1), pp.445–466.

Pérez-Muñnuzuri, V., & I. R. Gelpi, 2000: Application of nonlinear forecasting techniques for meteorological modeling, Annales Geophysicae, 18: 1349. https://doi.org/10.1007/s00585-000-1349-3

Rhinehart, R. R., 2016: Nonlinear Regression Models for Engineering Applications, Modeling, Model Validation, and Enabling Design of Experiments. 2016. *John Wiley & Sons, Ltd*.

Sajjad, S.H., et al., 2009: On rising temperature trends of Karachi in Pakistan. *Climatic Change*, 96(4), pp.539–547.

Salinger, M. J., 2005: Guest editorial, increasing climate variability and change: reducing the vulnerability. *Clim Change*, 70, pp.1–3.

Singh, O., P. Arya, & B. S. Chaudhary, 2013: On rising temperature trends at Dehradun in Doon valley of Uttarakhand, India. *Journal of Earth System Science*, 122(3), pp.613–622.