International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 5 Issue 4, May-June 2021 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Surface Air Temperature of Kolkata District of West Bengal, India - A Statistical Study

Amartya Bhattacharya

National Atlas & Thematic Mapping Organisation (NATMO), Department of Science & Technology, Government of India, Salt Lake, Kolkata-700064, West Bengal, India

ABSTRACT

The impact of steady increase in surface air temperature on climate change is a serious topic of today's discussion. This study deals with the trend analysis and time series analysis and its future forecasts of surface air temperature of Kolkata district of West Bengal state, India during 1901-2002. The mean of annual average, maximum and minimum surface air temperature (^{0}C) of last 102 years are respectively 26.73, 31.26 and 22.23 with 1.27 %, 1.28 % and 1.73 % coefficients of variation. A non-parametric trend test namely the Mann-Kendall (MK) Trend Test along with the Sen's Slope estimator has been used to determine the monotonic trend of temperature and the magnitude of trend respectively. On the other hand, the Innovative Trend Test is also being performed for pattern determination. Both of the Mann-Kendall (MK) Trend Test and Innovative Trend Test show that the trend of average, minimum and maximum surface air temperatures of Kolkata are increasing (i.e. upwards) in nature. The time series forecasts of surface air temperature are done by Autoregressive Integrated Moving Average (ARIMA) model. This upward trend of surface air temperature has an impact on human life and urbanization as well. The impact of climate change and related socio-economic development will affect the ability of a nation to achieve sustainable development goals. The statistical analysis is carried out by using "R" (Version 3.6.1) statistical software.

KEYWORDS: *Mann-Kendall Trend Test, Innovative Trend Test, Sen's slope, Time Series, ARIMA*

How to cite this paper: Amartya Bhattacharya "Surface Air Temperature of Kolkata District of West Bengal, India - A Statistical Study" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-4, June 2021, pp.1620-1624, URL: www.ijtsrd.com/papers/ijtsrd43684.pdf

IJTSRD43684

INTRODUCTION

Climate change is a burning issue of recent days. The earth's climate is continuously changing its behavior over the past century due to variation of surface air temperature, precipitation and other climatic conditions. The changing of surface air temperature pattern is one of the major impacts of climate change. Changes in behavior of surface air temperature influences the hydrological cycle too. Moreover, the Climatic conditions are responsible for the economic condition, livelihood as well as overall sustainable development of a region. Many researchers have studied on different themes of climatic conditions. *Pattanaik et.al (2007) [1]* investigated the analysis of rainfall over different homogeneous regions of India. The study was done in relation to variability in westward movement frequency of monsoon depressions. *Mondal et.al. (2012) [2]* showed the trend analysis of rainfall of north-eastern part of Cuttack district of Odisha, India. *Anie et.al.(2018) [3]*have studied the pattern of rainfall in Vamanapuram River Basin of Kerala, India by using non-parametric statistical trend analysis. The rainfall pattern of the districts of West Bengal has been studied by *Mukherjee (2017) [4].Girma et. Al (2020) [5] has studied on recent trends on climatic variability using Innovative Trend analysison upper Huai River Basin.*

In view of the above, study of surface air temperature trend is a significant and interesting area of research today. Changing pattern of surface air temperature is not only the important global and national climatic scenario, but also an important scenario for a district too.

The Kolkata district is located on the east bank of the Hooghly river, it is the major commercial, cultural and educational center of East India. The district has a Tropical wet-and-dry climate. Huge population as well as pollution are the major concerns in Kolkata. The Suspended Particulate Matter (SPM) level is high as compared to other major districts and cities of India, leading to regular smog and haze. Severe air pollution in the district has caused rise in pollution-related respiratory ailments.

The main objective of the present study is to detect the long-term monotonic significant trend of mean annual surface air temperature (average, minimum and maximum) of Kolkata district during 1901-2002 and its related magnitude by the Mann-Kendall (MK) Trend Test Method and the Innovative Trend (IT) Test Method. The Box-Jenkins autoregressive integrated moving average (ARIMA) method is performed to show appropriate time series model and future forecasts (2003 to 2050) of the mean annual surface air temperature (average) of the district.

Study Area

The Capital of West Bengal is Kolkata which is the only cent percent urban district of West Bengal, India. Kolkata district lies between 22^0 37' and 22^0 30' North Latitude and 88^0 23' and 88^0 18' East Longitude. As per Census 2011, the district covers total 185 sq. Km area with population density 24306 person/sq. Km. The total population of Kolkata district is

4496694 with sex ratio 908, child sex ratio 933, literacy rate 86.30%, work participation rate 39.90 %.

Materials

The datasets for mean annual surface air temperature (average, minimum and maximum) (^{0}C) of Kolkata district of West Bengal state, India during 1901-2002 have been collected from India Water Portal, The Indian Meteorological Department (IMD). The demographic data are collected from the Census of India, 2011. The Statistical analysis is done by using "R" (Version 3.6.1) Statistical software.

Methodology

Statistical Inference has basically two parts, one is statistical estimation of parameter(s) and other is statistical testing of hypothesis. Parametric testing and Non-parametric are the two commonly used testing procedures. A parametric statistical test is a test where the model specifies certain conditions about the parameters of the population. In this case there must be a parent probability distribution form which the samples are to be drawn. But in reality, many situations arise where the assumptions about the parent probability distribution is known to us. Then we have to introduce non-parametric approach. A non-parametric test is a test where the model does not specify conditions about the parameters of the population or the parent probability distribution. But the most interesting fact is, in most of the non-parametric model, the test statistic follows a Normal Distribution asymptotically.

The mean annual surface air temperature (average, minimum and maximum) (^{0}C) of Kolkata district of West Bengal state of India during 1901-2002 have been taken into consideration separately for analysis purpose. The time series analysis is done only on the mean annual surface air temperature (average) of the district during 1901-2002 and its future prediction have been forecasted for next 48 years, i.e. up to year 2050.

Mann-Kendall (MK) Trend Test Method

The Mann-Kendall test is a non-parametric test which is widely used to detect monotonic trends in series of environmental data, climate data or hydrological data. This test is the result of the development of the test first proposed by Mann (1945) which is further studied by Kendall (1975).

In Mann-Kendall test we are to test the null hypothesis (H_0): There is no significant trend in surface air temperature against the alternative hypothesis (H_1): There is a specific significant trend in surface air temperature at 5% level of significance.

The Mann-Kendall (MK) test statistic is denoted as S and is given by

$$S=\sum_{k=1}^{n-1}\sum_{j=k+n}^{n}\operatorname{sgn}(X_j-X_k) \qquad (1)$$

where, n is the number of data points; X_j, X_k are the data values for the time period j and k respectively; $\operatorname{sgn}(X_j-X_k)$ is signum function and is equal to 1, 0 or -1 if (X_j-X_k) is positive, zero or negative respectively.

Under the null hypothesis, the mean of S is zero and the variance of S is denoted by Var (S)and is given by

$$Var(S)=\{n(n-1)(2n+5)-\sum_{j=1}^{m}t_j(t_j-1)(2t_j+5)\}/18 \qquad (2)$$

where m is the number of the tied groups in the data set and t_j is the number of data points in thej^{th}tied group. For $n\geq 10$ the statistic S is approximately normally distributed with mean 0 and variance Var(S). The normally distributed test statistic is defined by

$$Z=\begin{cases}\frac{S-1}{\sqrt{Var(S)}}, & if\ S>0\\ 0, & if\ S=0\\ \frac{S+1}{\sqrt{Var(S)}}, & if\ S<0\end{cases} \qquad (3)$$

The Kendall's tau is denoted by τ and is given by

$$\tau=\frac{S}{[\frac{1}{2}n(n-1)-\frac{1}{2}\sum_{j=1}^{p}t_j(t_j-1)]^{1/2}[\frac{1}{2}n(n-1)]^{1/2}} \qquad (4)$$

In situations where there is no tie(s) occur, then Var (S) and τ reduce to

$$Var(S)=\{n(n-1)(2n+5)\}/18 \qquad (5)$$

and

$$\tau=\frac{S}{[\frac{1}{2}n(n-1)]} \qquad (6)$$

respectively.

Kendall's tau (τ) has been also calculated to show the nature of trend. The positive and negative τ value define the increasing and decreasing trends respectively.

Now, for determining the hydro-meteorological magnitude of trend the Sen's slope estimator is used.

A set of linear slopes is calculated as follows

$$d_i=\frac{X_j-X_k}{j-k} \qquad (7)$$

for $(1\leq k\leq j\leq n)$, d is the slope and X_j, X_k are the data values for the time period j and k respectively.

The Sen's slope(β) is calculated as the median from all slopes. The percentage change (%) is defined by the following formula.

Percentage change = β × Time Length × 100 / Mean (8)

The positive β value indicates upward (i.e. increasing) trend and the negative β value indicates decreasing (i.e. downward) trend. The test also derives the p-values which is the probability of obtaining the observed results of a test assuming the null hypothesis is true. This p-value and Z value play the significant role for taking decision about accepting or rejecting the null hypothesis.

Innovative Trend (IT) Test Method

Innovative Trend (IT) Testis widely used technique for detection of monotonic pattern of hydro meteorological data. In Innovative Trend method, the dataset is classified into two classes and then the data points arranged independently in increasing order.

Let X_i, i=1,2, 3,…,n/2 and X_j, j= (n/2)+1,…,n be placed on X-axis and Y-axis on a coordinate system respectively. . If the time series data on a scattered plot are collected on the 1:1 (45°) straight line, it indicates no trend. However, the trend is increasing when data points accumulate above the 1:1 straight line (i.e. upper triangle) and decreasing trend when data points accumulate below the 1:1 straight line (i.e. lower triangle).

The trend indicator (φ) is defined by

$$\varphi = \frac{1}{n} \sum \frac{10 \left(X_i - X_j \right)}{\mu} \quad (9)$$

where n is the total number of data points, μ is the mean of the data series in the first half subseries class.

The positive φ values defines the increasing trend and negative φ value defines the decreasing trend. When the scattered data points closest around 1: 1 straight line, it indicates non-existence of trend.

Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) Method:

A time series is defined as an arrangement of data over specific time period [6]. Box-Jenkins autoregressive integrated moving average (ARIMA) method due to Box and Jenkins. ARIMA is a very popular method used to produce accurate future forecasts. This ARIMA (p, d, q) method is a combination of autoregressive (AR) of order p and moving average (MA) of order q with degree of differencing involved (d). The differencing (d) is used to make the data stationary to non-stationary. Therefore, the model is called as Integrated ARMA, i.e. ARIMA.

Let us assume Y_t, t=1(1) n be a time series then the general form of difference taking time series is given by

$$\nabla^d Y_t = \nabla^{d-1} Y_t - \nabla^{d-1} Y_{t-1} \quad (10)$$

where d is the degree of differencing involved.

An autoregressive moving average (ARMA) model with order of autoregressive part (p) and order of moving average part (q) is defined by ARMA (p,q) model and the model is expressed by the following model expression

$$Y_t = \mu + \sum_{i=1}^{p} \theta_i Y_{t-i} + \sum_{i=1}^{q} \varphi_i \varepsilon_{t-i} + \varepsilon_t \quad (11)$$

where μ is the mean value and $\theta \,\&\, \varphi$ are the parameters of autoregressive and moving average part respectively which are to be estimated. It is to be noted that the error term (ε_t) is the white noise which is i.i.d $(0, \sigma^2)$. A time series is said to ARIMA (p, d, q) if $\nabla^d Y_t$ is ARMA (p, q).

The Augmented Dickey Fuller (ADF) test is used to check if the time series is stationary or not. The p-value of the ADF test defines whether the time series is stationary or not where the null hypothesis is taken as the time series is not stationary against the alternative hypothesis is that the time series is stationary at a pre-assigned level of significance. For making the time series stationary, the method of differencing is needed.

Autocorrelation function (ACF) graph and partial autocorrelation (PACF) graph are used to find the initial degree of qand p in ARIMA (p,q) model. ARIMA model deals with stationary data.

Based on the values of p, d and q, some special cases are discussed as follow:

White noise model: ARIMA (0, 0, 0) with p=0, d=0, q=0

Random walk: ARIMA (0, 1, 0) with no constant with p=0, d=1, q=0

Random walk with drift: ARIMA (0, 1, 0) with constant with p=0, d=1, q=0

AR (p): ARIMA (p, 0, 0) with d=0, q=0

MA (q): ARIMA (0, 0, q) with p=0, d=0

ARMA (p, q): ARIMA (p, 0, q) with d=0

Akaike's information criteria (AIC) isapplied for selecting the appropriate ARIMA model. Usually the model with lowest value of AIC among the all possible models is to be selected as the best fitted model.

Based on the best fitted ARIMA (p, d, q) model, forecasts of the time series isobtained with 80 % and 95 % (low and high) prediction intervals.

Results and Discussion

The mean annual surface air temperature (average, minimum and maximum) during 1901-2002 have been taken into consideration.

Table 1: Descriptive Statistics Table for Mean Annual Surface Air Temperature of Kolkata District

Temperature	Minimum	Maximum	Mean	Median	Coefficient of variation (%)	Skewness	Kurtosis
Average Temperature	25.62	27.76	26.73	26.73	1.27	- 0.11	0.69
Maximum Temperature	29.92	32.42	31.26	31.24	1.28	-0.08	0.87
Minimum Temperature	21.34	23.17	22.23	22.18	1.73	0.33	-0.15

***N.B.:** The table is tabulated by the author*

Table 1 shows that the mean of annual average, maximum and minimum surface air temperature (^{0}C) of last 102 years are respectively 26.73, 31.26 and 22.23 with 1.27 %, 1.28 % and 1.73 % coefficients of variation and median values 26.73, 31.24 and 22.18. The skewness values define that the annual average (-0.11) and maximum (-0.08) temperature are skewed left in nature where minimum temperature is skewed right (0.33). The kurtosis values for annual average (0.69) and maximum (0.87) temperature are positive and the for the case of minimum (-0.15) temperature the kurtosis value is negative during 1901-2002.

Table 2: Trend Test Table for Mean Annual Surface Air Temperature of Kolkata District

Temperature	Mann-Kendall Trend Test						Innovative Trend Test		Significant Trend (at 95% Confidence Interval)
	Z value	Mann-Kendall Statistic (S)	Kendall's Tau (τ)	Sen's Slope (β)	p -value	% Change	Innovative Trend indicator (φ)	Innovative Trend slope (D)	
Average Temperature	4.85	1.67×10^3	0.33	0.0051	1.24×10^{-6}	1.95	0.1065	0.0055	Increasing
Maximum Temperature	3.39	1.17×10^3	0.23	0.0042	7×10^{-4}	1.37	0.1020	0.0062	Increasing
Minimum Temperature	4.40	1.52×10^3	0.29	0.0056	1.09×10^{-5}	2.57	0.1074	0.0047	Increasing

***N.B.:** The table is tabulated by the author*

From Table 2, it is clear that for mean annual surface air temperature (average, maximum and minimum), the Z-values,Mann-Kendall Statistic (S) and Kendall's tau (τ)values are all positive. The Sen's slop estimator (β) values are also positive for all the above three cases. The MK test is performed at 5 % level of significance. The p-values are all less than 0.05. Hence it is clear that the trend of meanannual surface air temperature (average, maximum and minimum) are all significantly increasing (i.e. upward) at 5% level of significance or 95% confidence interval with percentage change 1.95 %, 1.37% and 2.57% during 1901-2002. In case of Innovative Trend test, the innovative trend indictors (φ) are all positive and hence it may be found that the trend for all the above three cases are significantly increasing. From this it may be found that for both trend test viz. Mann-Kendall Test and Innovative Test, the mean annual surface air temperature (average, maximum and minimum) of Kolkata district are all significantly increasing in nature during 1901-2002.

A time series is defined as an arrangement of data over specific time period. ARIMA model is widely used reputed time series model for forecasts or future prediction.

The mean annual surface air temperature (average) during 1901-2002 have been taken into consideration for time series analysis. The data is transformed to a time series data to carry out the ARIMA model analysis. Then we are to check whether the data is stationary or not. The stationary is checked by the Augmented Dickey Fuller (ADF) test at 5% level of significance. The p-value plays a significant role for taking decision about accepting or rejecting null hypothesis. The test result is tabulated as follows:

Table 3: Augmented Dickey Fuller (ADF) test

Test	Order of differencing (d) involved	p-value	Remarks
ADF	d=0	0.64	Non-Stationary
ADF	d=1	0.01	Stationary

***N.B.:** The table is tabulated by the author*

From Table 3, it is observed that the 1st order differencing of the time series is stationary with p-value 0.01 (< 0.05) as the test is performed at 5 % level of significance.

Now by using "R" Statistical software (Version 3.6.1), the ARIMA model is constructed for 1st order differenced (d=1) data series.

Table 4: ARIMA Model Summary Table

Model	μ	AR (1)	AR (2)	AR (3)	AR (4)	MA (1)	MA (2)	MA (3)	AIC
ARIMA (4,1,3)	0	-0.6560	-1.2055	-0.1343	-0.2979	0.0184	0.5385	-0.6702	43.49

***N.B.:** The table is tabulated by the author*

From Table 4, it is found that for ARIMA (4,1,3) with zero mean, the calculated value of Akaike's information criteria (AIC) (43.29) is the smallest among the possible models. Hence the model "ARIMA (4,1,3) with zero mean" is selected as the best fitted ARIMA model for mean annual surface air temperature (average) of Kolkata district during 1901-2002.

The model expression of ARIMA (4,1,3) with zero mean is given by the following equation:

$$Y_t^{/} = -0.6560Y_{t-1}^{/} - 1.2055Y_{t-2}^{/} - 0.1343Y_{t-3}^{/} - 0.2979Y_{t-4}^{/} + 0.0184\varepsilon_{t-1} + 0.5385\varepsilon_{t-2} - 0.6702\varepsilon_{t-3} + \varepsilon_t \quad (12)$$

where $Y^{/}_t$ is the 1st order difference of the time series Y_t.

Based on the above ARIMA (4,1,3) with zero mean, the future forecasts for mean annual surface air temperature (average) of Kolkata district from 2003 to 2050 is done with 80% and 95% (low and high) prediction intervals.

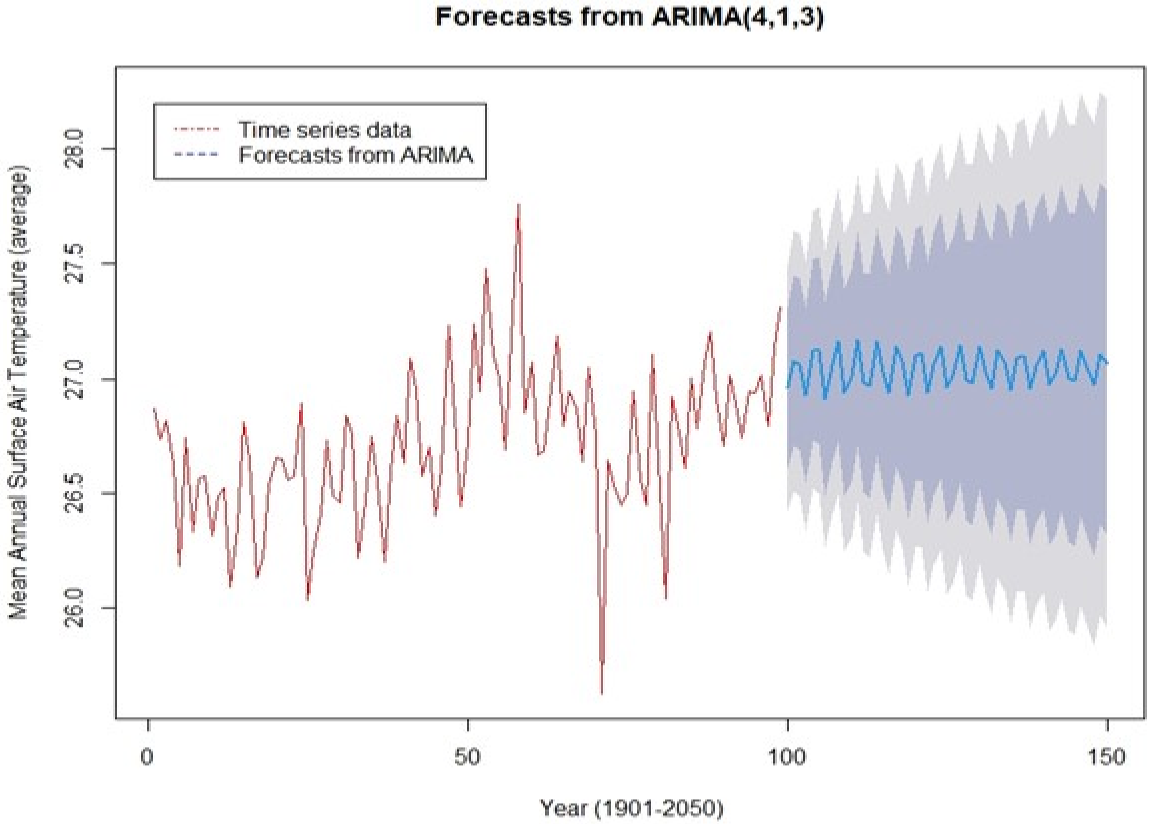

Figure 1: Forecasts from ARIMA (4, 1, 3)

Figure 1 depicts the actual time series data on mean annual surface air temperature (average) of Kolkata district during 1901 to 2002 and its forecasts from ARIMA from2003 to 2050 at 80% and 95% (low and high) prediction intervals.

Concluding Remarks

The impact of climate change on annual surface air temperature is a serious issue now a days. The mean annual surface air temperature (average, minimum and maximum) are all significantly increasing in nature. The time series forecasts from ARIMA of mean annual surface air temperature (average) is not also decreasing in pattern. For a district with huge population and pollution like Kolkata, the increasing trend of surface air temperature is a very alarming agenda today. Rapid growth of population, concretization, urbanization, factories & industries, emission of greenhouse gases, destroying trees and green places and other alike activities are mainly responsible for long-term upward trend of surface air temperature of Kolkata. Monotonic upward pattern of surface air temperature has a deep impact on environment, livelihood, health, socio-economic conditions and overall sustainable development of a region. This study will give the overall information to the Government, community, planners, researchers, policy makers about the long-term increasing pattern of surface air temperature of the district and help them to make necessary action plan and framework to save our environment and planet from the climatic disasters in near future.

References

[1] Pattanaik, D. R. (2007). Analysis of rainfall over different homogeneous regions of India in relation to variability in westward movement frequency of monsoon depressions. *Nat. Hazards*, 40(3), 635–646.

[2] Mondal, A., Mokhopadhyay A., Kundu, S. (2012). Rainfall Trend Analysis by Mann-Kendall Test: A Case Study of North-Eastern Part of Cuttack District, Orissa. *IJGEVS,* 2(1), 70-78.

[3] John, A., Brema, J. (2018). Rainfall Trend Analysis by Mann-Kendall Test for Vamanapuram River Basin, Kerala. *IJCIET,* 9 (13), 1549-1556.

[4] Mukherjee, K. (2017). Trend Analysis of Rainfall in The Districts of West Bengal, India A Study for The Last Century. *JECAS*, 6 (1).

[5] Girma, A., Qin, T., Wang, H., Yan, D., Gedefaw, M., Abiyu, A., Batsuren, D. (2020). Study on Recent Trends of Climate Variability Using Innovative Trend Analysis: The Case of the Upper Huai River Basin. *Pol. J. Environ. Stud.,* Vol. 29, No. 3 (2020), 2199-2210.

[6] Bhattacharya, A., Banerjee, T. (2020). COVID -19 Pandemic in India: A Statistical Approach. *International Research Journal of Modernization in Engineering Technology and Science*, Vol. 2, Issue 07.