RESEARCH ARTICLE OPEN ACCESS

Regression And Augmentation Analytics on Earth's Surface Temperature

Siddharth Banga [1], Saksham Mongia [2],VaibhavTiwari [3]
Mrs. Sunita Dhotre [4]
Student [1], [2] & [3], Associate Professor [4]
Department of Computer Engineering
Bharati Vidyapeeth University
College Of Engineering, Pune
Maharashtra, India

ABSTRACT
Big Data Analysis is the process of handling the huge amount of data that overcomes hidden patterns, market trends and other useful information that can help organization to make better decision [2]. GISS temperature scheme was defined in 1970's by James Hansen when method of global temperature was needed. The scheme was based on finding the co-relation of temperature change between the stations separated by 1200km [9].These facts were sufficient to obtain useful estimate for global mean temperature change .Our study is necessary to define the co-relation between Land's Average Temperature and Uncertainty temperature, so as to show the increasing temperature year after year which is the major cause for Global Warming [3].

Keywords:- Uncertainty, Linear Regression, Anomaly, precision.

I. INTRODUCTION

Earth's temperature observation plays an important role in Big Data Analysis. This provides useful information of Climatic change through which detection of the future problem like Global Warming, melting of Glaciers and many more can be overcome for the sake of future generations. This study is based upon the large data sets starting from year 1901 to 2015.Evaluation of monthly temperature of every year and then plotting the graph of temperature change that has happened after every 10 years using Linear Regression [8]. This Earth's temperature program will help the future Earth's program with an important role in academic's and decision making for sustainable environment.

Year	*Land Average Temperature	*Land Average Uncertainty
01-01-1901	8.224833	0.275606
01-01-1911	8.29442	0.25853
01-01-1921	8.52015	0.249766
01-01-1931	8.65478	0.239741
01-01-1941	8.68547	0.217308
01-01-1951	8.64268	0.15295
01-01-1961	8.64484	0.097191
01-01-1971	8.68634	0.097775
01-01-1981	8.936875	0.08843
01-01-1991	9.151833	0.078408
01-01-2001	9.56482	0.085483

Table 1 (Values in Degree Celsius)

II. METHODS/APPROACH

Temperature Anomaly with regression analysis

Temperature anomaly can be defined as a divergence from a reference value or long-term average [6]. It can be classified into positive and negative anomalies where from a positive anomaly demonstrates that temperature was warmer than the reference value, while a negative anomaly demonstrates that the observed temperature was cooler than the reference value [10]. It is a diagnostic tool for global scale climate which provides a big picture overview of average global temperature with a reference value [7].

Regression analysis: **regression analysis** is a statistical process for estimating the relationships among variables [5].

Regression line formula x on y-axis is defined as:

$$x - \bar{x} = b_{xy}\,(y - \bar{y}) \ldots\ldots\ldots (1)$$

Regression line formula x on y-axis is defined as:

$$y - \bar{y} = b_{yx}\,(x - \bar{x}) \ldots\ldots\ldots (2)$$

Estimation of global average temperature is difficult that is why elevation of temperature of region is also considered. For its representation over years process of Normalization and computation of reference values a base line is established on which temperature anomalies is processed [13]. In Table 1, average land temperature and average land uncertainty of time span of 10 years is calculated for over a period of 100 years i.e. 1901-2001.

This process of estimation and evaluation of entire planet inherits some uncertainty level [14]. These values can be either positive or negative for e.g. as "*+0.54°C +/- 0.08°C"*. This concept of uncertainty range can be written as "Precision" or "Margin of error".

Regression Statistics	
Multiple R	0.756190283
R Square	0.571823744
Adjusted R Square	0.524248604
Standard Error	0.055768676
Observations	11

Table 2

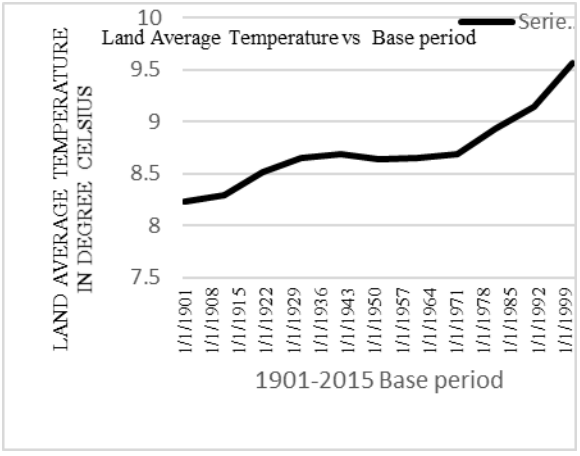
	df	SS	MS	F
Regression	1	0.03738202	12.01938132	0.007083325
Residual	9	0.02799130	0.003110145	
Total	10	0.06537332		

Table 3

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.57803868 3	0.40724070 1	3.87495326 3	0.00376023 8	0.65679621 5	2.49928115 2
X Variable 1	-0.16162603 3	0.04661978 4	-3.46689793 9	0.00708332 5	-0.26708731 2	-0.05616475 5

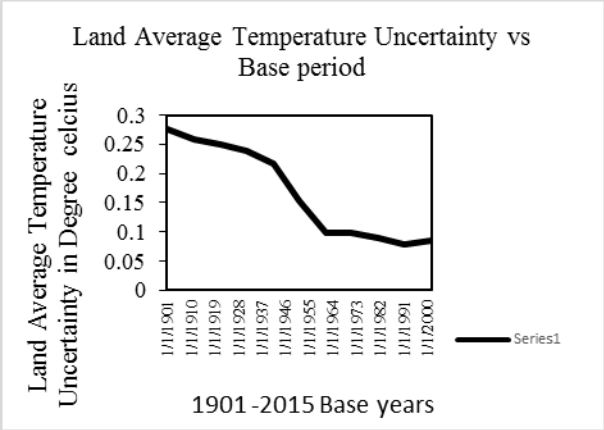
Table 4

III. RESULTS



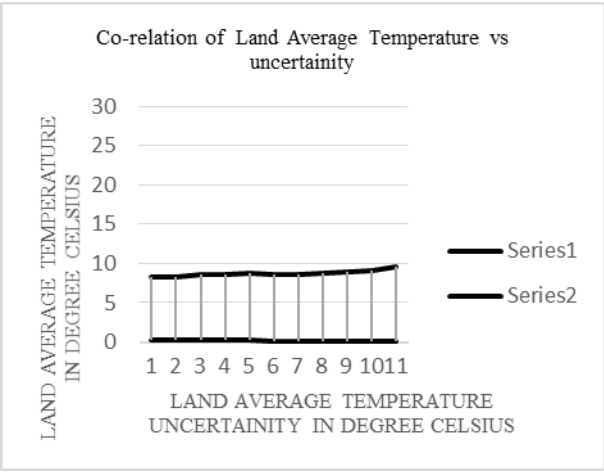
Graph 1: Land Average Temperature vs Base period

Graph 1 can be represented with the help of function y= f(x), this function represents the relation for all the values of 'y' with respect to function 'x'. Let a, b be any intercepts on the curve, for the values of a and b the f(x) will be positive from which it can be inferred from that the function or graph is strictly increasing [21]. Strictly increasing function means the increase in value of y leads to increase in x. Slope of any tangent line drawn to the curve is positive, i.e. positive rate of change with respect to x which is maximum after year 1991 and almost no rate of change from year 1931-1970.



Graph 2: Land Average Temperature Uncertainty vs Base period

Graph 2 can be represented with the help of function y= f(x), this function represents the relation for all the values of 'y' with respect to function 'x'. Let a, b be any intercepts on the curve before 1991, for the values of a and b the f(x) will be positive from which it can be inferred from that the function or graph is decreasing while after 1991 the graph is increasing. The graph is both decreasing as well as increasing. Slope of any tangent line drawn to the curve after 1991 is positive and the slope of tangent before 1991 is negative. Negative rate of change with respect to x is maximum from period 1940-1960 and rate of change is positive after year 1990.



Graph 3: Co-relation of Land Average Temperature vs uncertainity

In graph 3—relation of land average Temperature vs land average uncertainty on x-axis and y-axis respectively. From the graph it can be interpreted that the land average temperature is increasing every year while the land average uncertainty temperature is decreasing over a time period from 1901-2015 [20].

IV. FUTURE SCOPE

On the basis of our Analysis on the Land's Average temperature since years , we have seen the climatic change in successive years ,due to which there is increase in Greenhouse gas concentration, melting of the Snowpack's ,change in the Sea Level, future Precipitation and Strom Event [17]. Thus our project is useful so as to overcome these problems in future by analysing it before in advance and to take measures so as to have a sustainable lifestyle for our future generations [8].

REFERENCES

[1] Eckerson, W. (2011) "BigDataAnalytics: Profiling the Use of Analytical Platforms in User Organizations," TDWI, September.

[2] "Research in Big Data and Analytics: An Overview" International Journal of Computer Applications (0975 – 8887) Volume 108 –No 14, December 2014

[3] Douglas, Laney. "The Importance of 'Big Data': A Definition". Gartner. Retrieved 21 June 2012.

[4] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," interactions, vol. 19, no. 3, pp. 50–59, May 2012

[5] Sen, P. K., Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc., 1968, 63, 1379–1389.

[6] Fyfe, J. C., Gillett N. P. & Zwiers F. W. Overestimated global warming over the past 20 years. Nature Climate Change 3, 767–769 (2013)

[7] Cayan, D.R.& Douglas, A.V., 1984. Urban influences on surface temperatures in the southwestern UnitedStates during recent decades. Journal of Climate and Applied Meteorology, 23, 1520-1530.

[8] Sinha Ray, K.C., Mukhopadhayay, R.K. & Choudhury, S.K., 1997. Trends in maximum, minimum temperatures and sea level pressure over India. Paper presented in INTROMET-97 held at IIT, New Delhi during 2-5 December 1997.

[9] GISTEMP Team, 2017: *GISS Surface Temperature Analysis (GISTEMP)*. NASA Goddard Institute for Space Studies. Dataset accessed 20YY-MM-DD at https://data.giss.nasa.gov/gistemp/.

[10] Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change, *Rev. Geophys.*, **48**, RG4004, doi:10.1029/2010RG000345.

[11] USGCRP (2014) Melillo, Jerry M., Terese (T.C.) Richmond, and Gary W. Yohe, Eds., 2014: *Climate Change Impacts in the United States: The Third National Climate Assessment*. U.S. Global Change Research Program.

[12] IPCC (2013). Climate Change 2013: The Physical Science Basis EXIT. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[13] NRC (2011). Climate Stabilization Targets: Emissions, Concentrations, and Impacts over Decades to Millennia EXIT. National Research Council. The National Academies Press, Washington, DC, USA.

[14] USGCRP (2009). *Global Climate Change Impacts in the United States*. Thomas R. Karl, Jerry M. Melillo, and Thomas C. Peterson (eds.). United States Global Change Research Program. Cambridge University Press, New York, NY, USA.

[15] IPCC (2014). Climate Change 2014: Impacts, Adaptation, and Vulnerability.

[16] https://www.ncdc.noaa.gov/sotc/global/201603

[17] https://www.epa.gov/climate-change-science/future-climate-change

[18] https://data.giss.nasa.gov/gistemp/

[19] https://www.ncdc.noaa.gov/monitoring-references/faq/global-precision.php

[20] https://wattsupwiththat.com/2010/07/13/calculating-global-temperature/

[21] https://www.khanacademy.org/math/algebra/algebra-functions/positive-negative-increasing-decreasing-intervals/v/increasing-decreasing-positive-and-negative-intervals