

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/344052862>

# Time Series Analysis and Forecasting of Temperatures Records of Aurangabad District in Maharashtra

Article in *International Journal for Modern Trends in Science and Technology* · August 2020

DOI: 10.46501/IJMTST060847

---

CITATIONS

0

READS

120

3 authors, including:



[Manoj Mhaske](#)

Dr. Babasaheb Ambedkar Marathwada University

7 PUBLICATIONS 3 CITATIONS

SEE PROFILE

# Time Series Analysis and Forecasting of Temperatures Records of Aurangabad District in Maharashtra

Sagar S. Vakhare<sup>1</sup> | Ramesh R. Manza<sup>1</sup> | Manoj M. Mhaske<sup>1</sup>

<sup>1</sup>Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India.

## To Cite this Article

Sagar S. Vakhare, Ramesh R. Manza and Manoj M. Mhaske, "Time Series Analysis and Forecasting of Temperatures Records of Aurangabad District in Maharashtra", *International Journal for Modern Trends in Science and Technology*, 6(8): 291-295, 2020.

## Article Info

Received on 24-July-2020, Revised on 12-August-2020, Accepted on 18-August-2020, Published on 25-August-2020.

## ABSTRACT

*Time series analysis and forecasting has become a significant tool in various applications in meteorology and other environmental areas. Temperature prediction is viewed as a difficult analysis activity. Numerous techniques used to predict temperature records in which fundamental and technical analysis are one among them. In this paper latest time series forecasting feature of Weka is used. Results have been checked using three different Regression algorithm that are SMO Regression, Linear Regression and Gaussian Regression. The result shows that these Regression algorithms provide better ability to predict time series temperature records of Aurangabad district in Maharashtra.*

**KEYWORDS:** *Temperature forecasting, SMO Regression, Linear Regression and Gaussian Regression.*

## I. INTRODUCTION

Weather forecast has been a tricky issue in meteorological department considering that an extended term. Even in current date, this domain remains as a studies topic in which scientists and mathematicians are working to provide a model or an algorithm as a way to appropriately expect weather. On this analysis we choose Aurangabad district for time series forecasting of temperature data. Aurangabad District stated as one of the thirty six districts of Maharashtra state in western India. Time series forecasting is commonly known as a process of utilizing models in order to obtain the prediction of the output result for future trends and event. The future data result is generated according to the previous variable data event of temperature information. It is verified that this

forecast method has abilities to give a higher result based on many applications along with commercial enterprise, finance, clinical and others. Consequently, that allows you to find the most accurate prediction of temperature records; variety models approach and algorithms were proposed in the preceding studies. This paper discusses the temperature prediction by the use of Time series forecasting which used to analyze and observe the final result based on the one year data forecast with algorithms.

## II. LITERATURE REVIEW

Climate change is for many countries within the world one of the biggest environmental threats to food production, water availability, forest

biodiversity and livelihoods [1]. Furthermore, it is broadly believed that developing countries in tropical areas of the world may be impacted extra critically than developed ones [2] [3]. Understanding the character and scale of viable climate adjustments in north-eastern Bangladesh is of importance to the coverage makers and people who're working there because it gives them a chance to be organized for higher mitigation and edition measures. In fact, in keeping with data from the reinsurance industry, the quantity of climate-associated disasters has increased drastically the 1970's [4].

Time series evaluation and forecasting has become a primary tool in numerous hydro-meteorological applications to take a look at traits and variations in variables like rainfall, humidity, temperature, stream flow and plenty of other environmental parameters[2],[3],[4],[5]. A number of classical time series studies have been conducted in recent years to assess the nature of the climate change [1],[2],[3], as it has occurred over the world as well as in Bangladesh [4] in the recent past and as it will more in likely do so in the future. This study in [6] used the Least Squares Support Vector algorithms as a method to forecast the future stream flow discharge with using the data set of past stream flow and gage height. Qingwei et al [7] utilize the techniques of genetic programming, use in the Evolutionary Modelling to forecast the minimal data sets. This time series forecasting uses the GP to obtain quick and flexible means that create model between input and output to predict and estimate the fluctuation China stream flow with easy use and cost effective. Chiew et al [8], has been used an approach of statistical seasonal to improve stream flow forecasting in multiple sites. It helps to maximize available data extraction and provide good parametric structure to facilitate stream flow data learning.

### III. OBJECTIVE OF THE RESEARCH

The primary objective of this paper is to apply Weka 3.7.8 tool to gain extra accurate temperature prediction records using time series forecasting package of Weka. Our objective is to investigate prediction of various techniques on 102 years record of Temperature in Aurangabad District. Examine and evaluate results with actual data by using Relative percent errors. We will forecast 1 year result using all 3 techniques.

## IV. RESEARCH METHODOLOGY

### A. Research Framework

The studies started with the data series, then observed by using of data pre-processing. After that, the raw data are transformed into CSV format.

### A. Data Set

On these studies of temperature prediction, we will use data set with the attribute of Date, maximum Temperature, average Temperature and minimum Temperature in Aurangabad District of Maharashtra. All of the datasets used inside the present research had been sourced from the publicly available records of the Indian government from the year 1901 to 2002.

### B. Data pre-processing

The data pre-processing is the subsequent step of data mining after collection of records. Demanding situations in maximum temperature, average Temperature and minimum Temperature data; knowledge discovery process is facing bad data quality. For this reason, the data is pre-processing is used to remove noise and undesirable records.

## V. PROPOSED APPROACH

### A. Gaussian Regression

This classifier function implements Gaussian process for regression without hyper-parameter tuning. Here the missing values are replaced with global mean/mode values [9].

### B. Linear Regression

This function uses linear regression for predication. Weighted instances are possible to be used with this approach [10].

### C. SMO Regression

Support vector classifier will be trained using polynomial or RBF kernels where John C. Platt's sequential minimal optimization algorithm is implemented here [11].

### D. Relative Error as a Measure of Accuracy

The relative accuracy of a measurement can be expressed as percentage and these percentages can easy to calculate. Subtract the difference between the actual value and measurement from actual value and divide the result by the actual value to obtain the accuracy of the measurement. Multiply the result by 100 percent to convert the accuracy to a percentage.



### E. Forecasting Execution Environment

On this research, Waikato environment for knowledge analysis-version 3.7.6 (Weka 3.7.6) tool and a package referred as "Time series forecasting environment" are getting used for appearing the predication modeling and prediction. The input data set preprocessing and attribute-Relation file format (ARFF) generation steps are achieved manually and "Time series forecasting environment" is used for forecasting.

### VI. FORECAST RESULTS

The result of the forecasted records is shown in the Table 1 and Table 2. These results include values returned by applying each of the predication methods and actual value. Hence predicted temperature information can be compared between each prediction methods; Gaussian, Linear regression, and SMO regression.

**Table 1. Result of Linear Regression with Actual Values**

Actual Values				Predicted Values			
Month	Max Temp	Avg Temp	Min Temp	Month	Max Temp	Avg Temp	Min Temp
Jan-2002	29.6	20.9	12.2	Jan-2002	30.6	22	13.1
Feb-2002	34.1	25.0	15.9	Feb-2002	33.2	23.94	14.9
Mar-2002	36.7	27.8	18.9	Mar-2002	36.6	27.61	18.7
Apr-2002	40.0	31.4	23.0	Apr-2002	39.1	30.57	22.2
May-2002	40.1	32.1	24.1	May-2002	39.9	31.81	23.9

**Table 2. Result of Gaussian Regression with Actual Values**

Actual Values				Predicted Values			
Month	Max Temp	Avg Temp	Min Temp	Month	Max Temp	Avg Temp	Min Temp
Jan-2002	29.6	20.9	12.2	Jan-2002	30.9	22.2	13.6
Feb-2002	34.1	25.0	15.9	Feb-2002	33.0	23.8	14.7
Mar-2002	36.7	27.8	18.9	Mar-2002	36.7	27.7	18.8
Apr-2002	40.0	31.4	23.0	Apr-2002	39.3	30.8	22.3
May-2002	40.1	32.1	24.1	May-2002	39.8	31.9	23.9

**Table 3. Result of SMO Regression with Actual Values**

Actual Values				Predicted Values			
Month	Max Temp	Avg Temp	Min Temp	Month	Max Temp	Avg Temp	Min Temp
Jan-2002	29.6	20.9	12.2	Jan-2002	30.7	22.0	13.4
Feb-2002	34.1	25.0	15.9	Feb-2002	32.9	23.8	14.6
Mar-2002	36.7	27.8	18.9	Mar-2002	36.5	27.6	18.7
Apr-2002	40.0	31.4	23.0	Apr-2002	39.2	30.7	22.2
May-2002	40.1	32.1	24.1	May-2002	39.8	31.8	23.8

**Table 4. Relative Percentage Error of Linear Regression with Actual Values**

Month	Max Temp	Avg Temp	Min Temp
Jan-2002	3.35	5.24	7.87
Feb-2002	2.88	4.38	6.36
Mar-2002	0.51	0.80	0.92
Apr-2002	2.25	2.93	3.83
May-2002	0.52	0.95	0.84

**Table 5. Relative Percentage Error of Gaussian Regression with Actual Values**

Month	Max Temp	Avg Temp	Min Temp
Jan-2002	4.57	6.53	11.39
Feb-2002	3.44	4.62	7.24
Mar-2002	0.24	0.26	0.39
Apr-2002	1.75	2.20	2.87
May-2002	0.59	0.67	0.72

**Table 6. Relative Percentage Error of SMO Regression with Actual Values**

Month	Max Temp	Avg Temp	Min Temp
Jan-2002	3.96	5.53	9.92
Feb-2002	3.53	4.94	7.93
Mar-2002	0.62	0.77	1.24
Apr-2002	1.82	2.42	3.13
May-2002	0.67	0.73	1.05

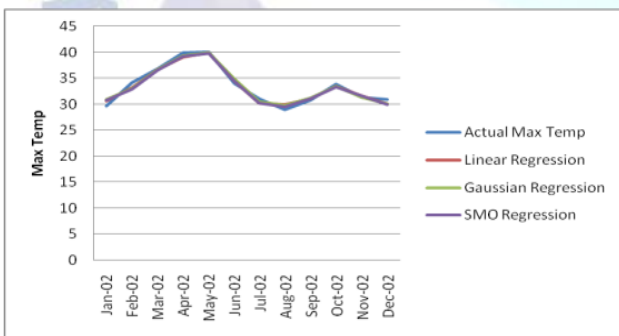


Figure 1

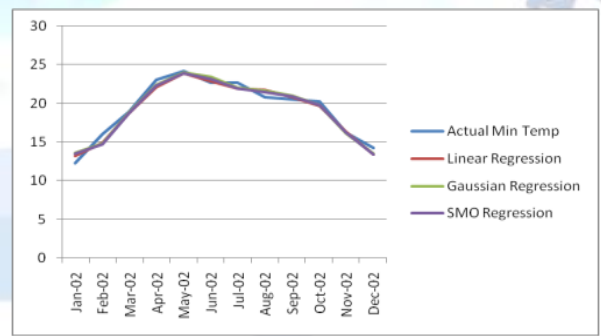


Figure 3

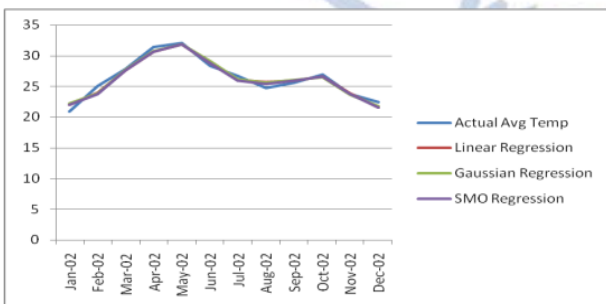


Figure 2

Figure 1, Figure 2 and Figure 3 describes a line graph representation of forecast result of Maximum Temperature, Average Temperature and Minimum Temperature with actual values respectively.

## VII. CONCLUSION

The accuracy of the above forecast results has been analyzed and evaluated based on Relative Percentage Error. Table 4, Table 5 and Table 6 are shown with results of accuracy for each of the four forecasting methods used. Actual value and forecasted graph representation shown in Figure 3, Figure 4 and Figure 5. Table 4 shows that the SMO regression and Linear regression algorithm have the smallest relative percentage error for Maximum Temperature data. Based on the result, it is apparent that the Gaussian regression algorithm produces the lowest difference values between the actual and predicted for average temperature and it have lowest relative percentage error. From Table 6 and Table 4, Linear regression and SMO Regression algorithm have less error percentage for Minimum Temperature data. This performance comparison leads us to our conclusion that all these three regression algorithms are the most suitable algorithm for time series forecasting of Temperature records in Aurangabad district.

## REFERENCES

- [1] E.S. Chung, K. Park and K.S. Lee. 2011. The relative impacts of climate change and urbanization on the hydrological response of a Korean urban watershed. *Hydrological Processes*, 25:544-560.
- [2] D. Machiwal and M. K. Jha. 2006. Time Series Analysis of Hydrologic Data For Water Resources Planning and Management: A Review. *J. Hydrol. Hydromech.*, 54(3): 237-257.
- [3] S.A. Shamsnia, N. Shahidi, A. Liaghat, A. Sarraf and S.F. Vahdat. 2011. Modeling Of Weather Parameters (Temperature, Rainfall And Humidity) Using Stochastic Methods. *Internat. Conference on Environment and Industrial Innovation, IPCBEE, Singapore*, 282-285.
- [4] S.B. Cheema, G. Rasul, G. Ali and D.H. Kazmi. 2011. A comparison of minimum temperature trends with model projections. *Pakistan Journal of Meteorology*, 8:39-52.
- [5] S. Soltani, R. Modarres and S. S. Eslamian. 2007. The use of time series modeling for the determination of rainfall climates of Iran. *International Journal of Climatology*, 27: 819- 829.
- [6] Zhang, N., Williams, C., Ososanya, E., & Mahmoud, W. (2013, October). Streamflow Prediction Based on Least Squares Support Vector Machines. In *ASEE-2013 Mid-Atlantic Fall Conference*, Washington, DC.
- [7] Ni, Q., Wang, L., Ye, R., Yang, F., & Sivakumar, M. (2010). Evolutionary modeling for streamflow forecasting with minimal datasets: a case study in the West Malian River, China. *Environmental Engineering Science*, 27(5), 377-385.
- [8] Francis Chiew, Q. W. (2008). Statistical Seasonal Streamflow Forecasting Model. *South Eastern Australian Climate Initiative*
- [9] K. Driessens, "GaussianProcesses." [Online]. Available:<http://weka.sourceforge.net/doc.dev/index.html?weka/classifiers/functions/GaussianProcesses.html>. [Accessed: 6-Aug-2012].
- [10] E. Frank and L. Trigg, "LinearRegression." [Online]. Available:<http://weka.sourceforge.net/doc/weka/classifiers/functions/LinearRegression.html>. [Accessed: 10Aug-2012].
- [11] E. Frank, L. Shane, and S. Inglis, "SMO." [Online]. Available:<http://weka.sourceforge.net/doc/weka/classifiers/functions/SMO.html>. [Accessed:7-Aug-2012].