DEVELOPMENT OF MULTIPLE LINEAR REGRESSION MODEL FOR PREDICTION OF TEMPERATURE LEVELS IN OGUN BASIN, ABEOKUTA, NIGERIA

JIBRIL, Ibrahim MUSA, John Jiya and ADEOYE, Peter Aderemi
Department of Agricultural and Bioresources Engineering, Federal University of Technology,
P M B 65, Minna, Nigeria. Tel: +234-8094-747-246
Corresponding Email: johnmusa@futminna.edu.ng

ABSTRACT

The study centered on developing empirical equations to predict maximum and minimum temperatures, using multiple linear regressions. The technique utilizes previous data in the study area in order to effectively predict the future outcome. To facilitate the research work, 29 years information of the observed minimum and maximum temperatures of Ogun basin were obtained from the Nigeria Meteorological Agency, Federal Ministry of Aviation, Abeokuta, Nigeria. The data collected covers the periods between 1982 and 2009 and were preprocessed to extract some features or statistical indicators embedded in the time series. Features extracted are; the moving averages, exponential smoothening, oscillators and rate of change. The multiple linear regressions for each climatic element under consideration were generated using first part of the extracted features, and the remaining was used for validation of the models. R – Square values obtained are 0.98 and 0.92 for maximum temperature and minimum temperature respectively, which signifies a good correlation between the actual and predicted values of the weather parameters. It was therefore concluded that applying soft computing approach for weather modeling to forecast Minimum and Maximum Temperature is most feasible rather than any other short term weather forecasting approach.

Keywords: Features, Forecast, MLR Model, Ogun Basin, Weather Parameters

INTRODUCTION

Since the beginning of industrial revolution the influence of human activity has begun to extend to a global scale. Today, environmental issue becomes the biggest concern of mankind because of the increasing concentration of greenhouse gases in the atmosphere and the changing climate of the earth and recently at global scale temperature is increasing and the amount and distribution of rainfall is being altered (Cubasch, *et al*. 2001). A general consensus is that the average global surface temperature has risen by about 4.6 °c under high emission scenario by the end of last century (IPCC, 2013) with the doubling of the CO_2 concentration in the atmosphere. Sea level rise, change in precipitation pattern (up to ±20%), and change in other local climate conditions are expected to occur as a consequence of rising global temperature (Cubasch, *et al*., 2001). The IPCC findings indicate that developing countries will be more vulnerable to climate change and may have far reaching implications for various reasons, since the economy largely depends on agriculture, large part is highly prone to desertification and drought (NMSA, 2007). Hence, evaluation of climate change is important to develop alternative strategies and policies to mitigate the impacts of global warming (IPCC, 2007). Climate change is considered as one of the biggest challenges of 21st century to the whole world. It is now widely accepted that climate change is already happening and further change is inevitable; over the last century (between 1906 and 2005), the average global temperature rose by about 0.74 0C. This has occurred in two phases, from 1910s to 1940s and more strongly from the 1970s to the present (IPCC. 2007). IPPC also states that the projected global surface warming lies within the range 0.6 to 4.0 ^{0}C at the end of next century (IPCC. 2007) and other projections of IPPC show that global average temperature will increase by 1.9 - 4.6 °C at the end 2100 and by 0.2°C per decade with the doubling of the CO_2 concentration in the atmosphere.

Temperature is the key determinant of global weather change and one of the main meteorological variables measured by meteorological service networks. According to (Deshmukh and Lunge, 2013), the Intergovernmental Panel on Climate Change (IPCC) reports maintained that the surface temperature of the earth has risen by 0.6 $\pm$ 0.2 ^{0}C over the 20th century. Further. surface air temperature could rise by between 1.1 ^{0}C to 6.4 ^{0}C over 21st Century due to global warming. Precipitation amount, type and timing are changing or are expected to change because of increased evaporation, especially in the tropics (Ritter, 2006).

Although greenhouse gas concentrations, especially carbon dioxide, have risen steadily, the climate we experience is affected by a number of other factors. These factors can be internal to the climate system (e.g. interactions occurring between the atmosphere and ocean, often referred to as natural internal variability) or external (e.g. occurring outside of the climate system. such as a change in the amount of energy being output by the Sun). This is clearly mentioned in (Lean and Rind, 2009; Kaufmann *et al.*. 2011 and Folland *et al.*, 2011).

In view of this effect of global warming, it is imperative that an effective method of predicting climatic parameters is develop, most especially for temperature in order to take proper precaution against adverse weather effect. Thus, the main objective of this study is to develop a Multiple Linear Regression (MLR) model to predict future maximum and minimum temperature for Ogun basin.

Materials and Method

The name 'Ogun basin' is derived from two major rivers that drains within; Rivers Ogun and Osun, though they have smaller tributaries like; Sasa, Ona, Ibu, Ofiki, Yewa rivers etc. The basin under consideration is located in South Western Nigeria. The entire basin is bounded by Oyo state in the north, Osun and Ondo States in the east and Lagos State in the South as shown in Figure 1.

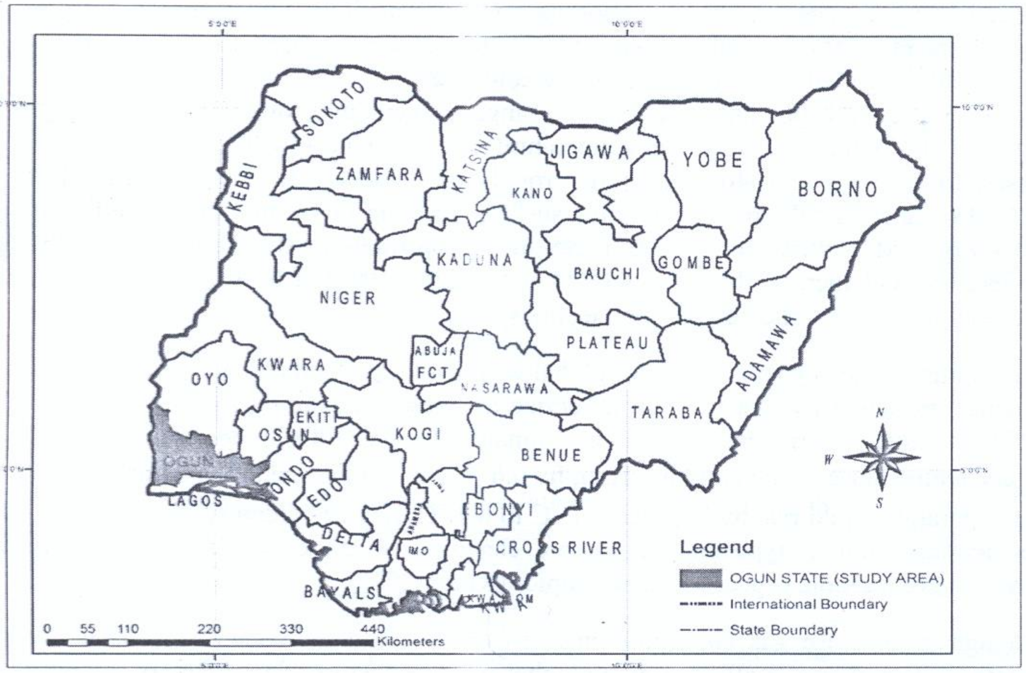

Fig 1: Map of Nigeria showing the study Area.

The Ogun basin covers the whole of Ogun State, located in southern Nigeria, bordered geographically by latitudes 6^0 26^I N and 9^0 10^I N and longitudes 2^0 28^I E and 4^0 8^I E. About 2% of the basin area falls outside Nigeria in the Benin Republic. The land area is about 23,000km^2. The relief is generally low, with the gradient in the North-South direction.

The two major vegetation zones that can be identified the area are the high forest vegetation in the north and central parts, and the swamp/mangrove forests that cover the southern coastal and floodplains, next to the lagoon. It has two distinct seasons throughout the year. The monthly rainfall distribution in the study area shows a distinct dry season extending from November through March and a rainy season divided into two periods: April – July and September – October. The mean annual rainfall data for 30 years showed a variation from about 1,150mm in the northern part to around 2,285mm in the southern extremity. The estimates of total annual potential evapotranspiration have been put between 1600 and 1900mm. (Ewomoje and Ewomooje, 2011).

Data Collection and Preprocessing

The data which includes minimum and maximum temperatures, rainfall and relative humidity used for this study were obtained from the Nigeria meteorological Agency, Abeokuta, Nigeria. The data collected covered a period of twenty nine years (1982-2009). These values were obtained by the use of GPS (general position satellite) equipment. Data preprocessing is an important task in almost all modeling techniques. The data obtained are the time series types which are collected monthly for a period of 29 years. For the purpose of this study, the mean annual values of the data were first determined before use.

Determination and Selection of Predictors

Predictors are characteristics of time series data embedded in them. Therefore, for an effective prediction of any time series, the predictors or features should be determined and used to obtain the model equation. The models are more unique and appear to perform better when some hidden features are presented which enhances its adaptability. The predictors in this research work include some statistical indicators, and the ones with least errors were selected for the model. The methodologies employed in the determination of the predictors are according to (Para and Sanjay, 2012), and are as below:

- Moving Average (MA): It is calculated progressively as an average of N number data values over certain period. The term moving is used because every time a new observation becomes available for the time series, it replaces the oldest observation in the equation and a new average is computed. As a result. the average will change, or move, as new observations become available. Data set is represented by d_t, d_{t-1}, d_{t-2},………………, d_0, where d_t is present and d_0 is the first data value, the moving average with a sliding window of period N is given by:

$$MA = \frac{(dt+dt-1\ dt-2 \ldots dt-N)}{N} \qquad (1)$$

- Exponential Smoothening (ESM): It also uses a weighted average of past time series values as a forecast; it is a special case of the weighted moving averages method in which we select only one weight, the weight for the most recent observation. The weights for the other data values are computed automatically and become smaller as the observations move farther into the past. The exponential smoothing equation is given as:

$$ESM = F_{t+1} = \alpha Y_t + (1 - \alpha)F_t \qquad (2)$$

Where;

F_{t+1} = forecast of the time series for period t+1

Y_t = actual value of the time series in period t

F_t = forecast of the time series for period t

α = is called the smoothing constant having value ($0 \leq \alpha \leq 1$).

- Oscillator (OSC): Oscillator is used to indicate the rising or falling trend present in the time series when the values are plotted against time. It is defined as difference of moving averages or exponential smoothening of two different periods.

$$OSC = MAN_1 - MAN_2 \quad (3)$$

Or

$$OSC = ESMN_1 - ESMN_2 \quad (4)$$

Where, $N_1$and N_2 are different periods and $N_1 > N_2$.

- Rate of Change (ROC): It indicates the rate of change of the variable at present, as compared to the value of the variable at certain period back. Thus percentage ROC at 'a' times back is given by:

$$ROC = \left(1 - \frac{d_t}{d_{t-a}}\right) 100 \quad (5)$$

Where, d_t = the value of the time series at present time t

d_{t-a} = the value of the time series at time t – a back.

Building Model Using Regression

Using the features or predictors determined, the multiple linear regressions are build. For this research, the whole dataset is divided in two parts, the first part is used to develop the regression equations; and the remaining is used for testing their validity. Microsoft excel and Minitab are the major tools used to process the data.

Results and Discussion

Predictors Determined for Maximum and Minimum Temperatures

The monthly data during the period of 1983-2009 were used to calibrate and validate the model for Ogun basin. Table 1 presents the developed MLR model for the maximum and minimum temperature with their R-square values. This is similar to the works of Dimri and Mohanty (2007); Mitra and Nath (2007) and Karimi, Karimi, Yavari, and Niksokhan (2015). The R square values for the maximum and minimum temperature were 0.98 and 0.92 respectively. The regression equations generated were used to predict values of the maximum and minimum temperatures shown in the last column of the tables. It was observed fro Figures 1 and 2 that both the observed and forecasted values for both the maximum and minimum temperature values were observed and synthetic to the actual data as they are seen to have almost the same frequency distribution. The scatter diagrams show the linear relationship between the actual weather parameter and the output of the MLR model.

Table 1: Regression equation and R- squared values obtained for maximum and minimum temperature

Experiment	Regression equation	R- squared value
Maximum temperature estimation using features of maximum temperature	$Y_1 = -1.01 - 0.448 X_1 + 1.47 X_2 + 0.514 X_3 - 0.414 X_4$	0.98
Minimum temperature estimation using features of minimum temperature	$Y_2 = -2.92 - 0.352 X_1 + 1.46 X_2 + 0.795 X_3 - 0.222 X_4$	0.92

From the regression equation shown in the tables above, Y_1 and Y_2 represent the maximum and minimum temperatures while, X_1, X_2, X_3 and X_4 are the moving average, exponential smoothening, oscillator and rate of change respectively.

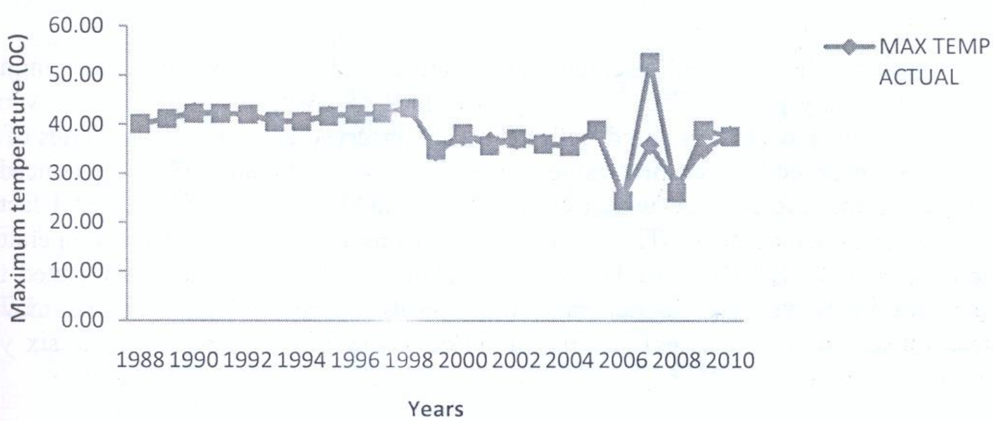

Fig. 1: Actual and predicted maximum temperatures

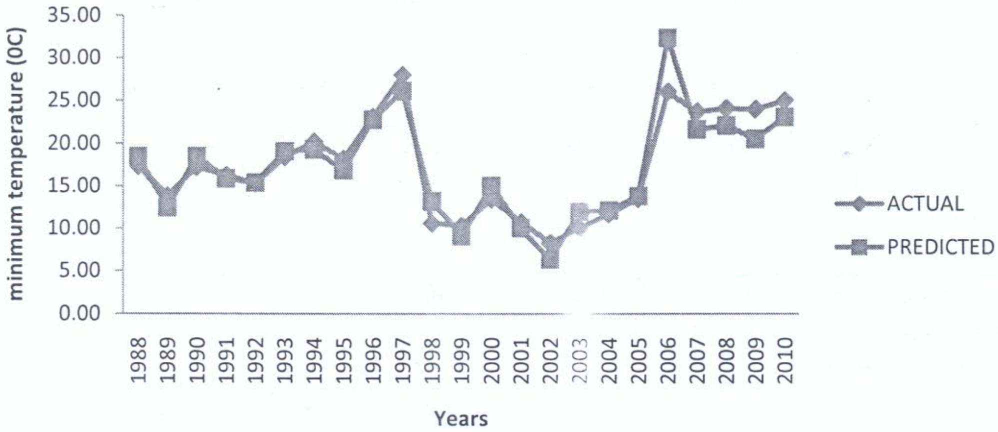

Fig. 2: Actual and predicted minimum temperatures

Tables 2 and 3 presents the observed, predicted and feature values of maximum and minimum temperatures for the study period. The first ten year values of the three year moving average were observed to be lower when compared with the initial observed average yearly values of the maximum temperature except for the first value of the year 1985. 1996 and 1997 experienced an increase in the three moving average which could be attributed to other environmental factors affecting the change in temperature. This trend was also observed for minimum temperature. This finding is similar to the works of Dimri and Mohanty (2007) in which they stated that maximum and minimum temperature depends considerably upon local features and moving synoptic systems (advective processes). A similar process was also observed for the six year moving average.

Table 1: Observed, predicted and feature values of maximum temperature

Year	Observed Max. Temp (^{0}C)	3 Years MA	6 Years MA	ESM 0.8	OSC	ROC	Predicted Max. Temp (^{0}C)
1982	42.60						
1983	37.00			42.60		13.15	
1984	35.50			38.12		4.05	
1985	35.50	38.37		36.02		0.00	
1986	38.05	36.00		35.60		-7.18	
1987	40.60	36.35		37.56		-6.70	
1988	40.00	38.05	38.21	39.99	0.16	1.48	40.13

1989	41.20	39.55	37.78	40.00	-1.78	-3.00	41.19
1990	42.60	40.60	38.48	40.96	-2.13	-3.40	42.28
1991	42.35	41.27	39.66	42.27	-1.61	0.59	42.29
1992	42.10	42.05	40.80	42.33	-1.25	0.59	42.06
1993	40.70	42.35	41.48	42.15	-0.88	3.33	40.54
1994	40.80	41.72	41.49	40.99	-0.23	-0.25	40.64
1995	41.90	41.20	41.63	40.84	0.42	-2.70	41.71
1996	42.10	41.13	41.74	41.69	0.61	-0.48	42.08
1997	42.30	41.60	41.66	42.02	0.06	-0.48	42.32
1998	43.50	42.10	41.65	42.24	-0.45	-2.84	43.37
1999	34.50	42.63	41.88	43.25	-0.75	20.69	34.85
2000	37.66	40.10	40.85	36.25	0.75	-9.16	38.15
2001	36.70	38.55	40.33	37.38	1.77	2.55	35.73
2002	36.80	36.29	39.46	36.84	3.17	-0.27	37.20
2003	36.30	37.05	38.58	36.81	1.52	1.36	36.03
2004	35.80	36.60	37.58	36.40	0.98	1.38	35.60
2005	38.80	36.30	36.29	35.92	-0.01	-8.38	39.00
2006	25.50	36.97	37.01	38.22	0.04	34.28	24.43
2007	35.90	35.80	36.40	35.92	0.60	-40.78	52.68
2008	27.70	35.90	34.85	35.90	-1.05	22.84	26.16
2009	35.00	31.80	33.33	29.34	1.53	-26.35	38.89
2010	38.00	32.87	33.12	33.87	0.25	-8.57	37.62

Table 2: Observed, predicted and feature values of minimum temperature.

Year	Observed Min. Temp.	3 Years MA	6 Years MA	ESM 0.80	OSC	ROC	Predicted Min. Temp.
1982	20.80						
1983	17.20			20.80		17.31	
1984	29.50			17.92		-71.51	
1985	18.20	22.50		27.18		38.31	
1986	21.85	21.63		20.00		-20.05	
1987	25.50	23.18		21.48		-16.70	
1988	17.30	21.85	22.18	24.70	0.33	32.16	18.45
1989	13.80	21.55	21.59	18.78	0.04	20.23	12.44
1990	17.20	18.87	21.03	14.80	2.16	-24.64	18.47
1991	16.25	16.10	18.98	16.72	2.88	5.52	15.87

1992	15.30	15.75	18.65	16.34	2.90	5.85	15.38
1993	18.40	16.25	17.56	15.51	1.31	-20.26	19.08
1994	20.20	16.65	16.38	17.82	-0.28	-9.78	19.29
1995	18.20	17.97	16.86	19.72	-1.11	9.90	16.86
1996	23.10	18.93	17.59	18.50	-1.34	-26.92	22.82
1997	28.00	20.50	18.58	22.18	-1.93	-21.21	26.10
1998	10.60	23.10	20.53	26.84	-2.57	62.14	13.20
1999	10.30	20.57	19.75	13.85	-0.82	2.83	9.07
2000	13.40	16.30	18.40	11.01	2.10	-30.10	15.03
2001	10.70	11.43	17.27	12.92	5.83	20.15	10.03
2002	8.30	11.47	16.02	11.14	4.55	22.43	6.35
2003	10.00	10.80	13.55	8.87	2.75	-20.48	11.99
2004	11.70	9.67	10.55	9.77	0.88	-17.00	12.11
2005	13.50	10.00	10.73	11.31	0.73	-15.38	13.82
2006	26.00	11.73	11.27	13.06	-0.47	-92.59	32.37
2007	23.70	17.07	13.37	23.41	-3.70	8.85	21.65
2008	24.10	21.07	15.53	23.64	-5.53	-1.69	22.11
2009	24.00	24.60	18.17	24.01	-6.43	0.41	20.53
2010	25.00	23.93	20.50	24.00	-3.43	-4.17	23.10

Conclusion

From the above study and discussion we see that applying soft computing approach for weather modeling to forecast Minimum and Maximum Temperature is most feasible rather than any other short term weather forecasting approach. The study also says that the valid and authentic dataset selection, input variable selection, the proper training set and the proper MLR are most vital for the best prediction results. In conclusion, this study shows a good R square value for both the maximum and minimum temperature for the study area. This also shows that the model produces excellent performance in downscaling Tmax and Tmin in the study region when considering the moving average for three and six years respectively. The predicted results showed that the average annual maximum and minimum temperature increases as the respective actual temperature increases and vice versa which supports the works of Mathur and Mathur (2012) and Karimi, *et al.*, (2015).

References

Deshmukh, D. T., and Lunge, H. S., (2013).A Study of Temperature and Rainfall Trends in Buldana District of Vidarbha, India.International Journal of Scientific and Technology Research. Vol. 2, Issue 2.

Dimri, A. P. and Mohanty, U. C. (2007): Location-specific prediction of maximum and minimum temperature over the western Himalayas. *Meteorol. Appl.* 14: 79–93. DOI: 10.1002/met.8

Ewemoje, T. A., and Ewemooje, O. S., (2011). Best Distribution and Fitting Positions of Daily Maximum Flood Estimation at Ona River in Ogun-Osun River Basin, Nigeria. Agricultural Engineering International: CIGR Journal. Vol. 13, No. 3.Manuscript No. 1380.

Folland, C., Colman, A., Kennedy, J., Knight, J., Parker, D., Stott, P., Smith, D., and Boucher, O., (2011). High Predictive Skill of Global Surface Temperature a Year ahead, in AGU Fall Meeting Abstracts, vol. 1, p. 01.

IPPC, (2013). Fifth Assessment Report on Climate Change: The Physical Science Basis. Stockholm, Sweden. Available at: http://www.ipcc.ch/.

IPCC, (2007). Climate Change: The Physical Science Basis. Summary for Policy Makers. Contribution of WG1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Available at: http://www.ipcc.ch/ipccreports/ar4-wg1.htm

Karimi, S., Karimi, S., Yavari, A. R. and Niksokhan, M. H. (2015): Prediction of Temperature and Precipitation in Damavand Catchment in Iran by Using LARS-WG in Future. Earth Science; 4(3): 95-100. Doi:10.11648/j.earth.20150403.12

Kaufmann, R., Kauppi, H., Mann, M., and Stock, J., (2011). Reconciling Anthropogenic Climate Change with Observed Temperature 1998–2008, Proceedings of the National Academy of Sciences, 108 (29), 11,790.

Lean, J., and D. H. Rind (2009). "How will earth's Surface Temperature Change in Future Decades?" Geophysical Research Letters, 36, L15,708.

Mathur, P., and Mathur, S. (2012): A Simple Weather Forecasting Model Using Mathematical Regression. Indian Research Journal of Extension Education Special Issue (Volume I), 161-168

Mitra, A. K., and Nath, S. (2007): Forecasting Maximum Temperatures Via Fuzzy Nearest Neighbour Model over Delhi. Appl. Comput. Math. 6; No. 2, pp288-294

NMSA, (2007).Climate Change National Adaptation Program of Action.Ethiopian National Meteorological Service Agency.Edited by Abebe, Teshome. Addis Ababa, Ethiopia.

Ritter, M. E., (2006). "The Physical Environment: An Introduction to Physical Geography," Available online at: http://www.uwsp.edu/geo/faculty/ritter/geog101/textbook/title_page.html.

Para, S., and Sanjay, M., (2012). A Simple Weather Forecasting Model Using Mathematical Regression. Indian Research Journal of Extension Education Special Issue, Vol. I