Mathematical Regression Modeling for Smart Environmental Weather Forecasting

Haftamu Menker GebreYohannes[#1]

[#]Lecturer Middle East College
Middle East College, Knowledge Oasis Muscat Campus, Muscat, Sultanate of Oman

Abstract; *Environmental weather forecasts have been increase the city's ability to be self-sustaining and lowering environmental impact as well as metrological support for critical decisions as climate change brings more extreme weather events. To predict weather meteorologists form models based on the land's geography and starting weather conditions and they can calculate future forecast by entering it into a computer to be processed. In this paper we will describe a survey of Mathematical techniques and Practices used to model weather forecasting today.*

Keywords — *Weather forecasting; Multiple Linear Regression Model; Time Series Data; Moving Average Model; Exponential Smoothing Model.*

I. Introduction

Weather forecasting is a kind of scientific and technological activity, which contributes to social and economic welfare in many sections of the community to-day. In this context, the purpose of weather forecasting is to provide information on the expected weather with forecast projection times ranging from a few hours to a few months. Weather conditions are required to be predicted not only for future planning in agriculture and industries but also in many other fields like defense, mountaineering, shipping and aerospace navigation etc. It is often used to warn about natural disasters which are caused by abrupt change in climatic conditions. At macro level, weather forecasting is usually done using the data gathered by remote sensing satellites. Weather parameters like maximum temperature, minimum temperature, extent of rainfall, cloud conditions, wind streams and their directions, are estimated using images and data taken by these meteorological satellites to access future trends. Those variables defining weather conditions vary continuously with time, forming time series of each parameter and can be used to develop a forecasting model either mathematically or using some other means that uses time series data.

Mathematics in weather forecasting has been starting with the father of Numerical Weather Prediction Vilhelm Bjerkness and Lewis Fry Richardson. In this approach we will provide illustrative challenges of Multiple Regression Modeling in smart weather forecasting. The Proposed Model is capable of forecasting weather conditions for a particular place using data collected locally. The whole data set will be classified in to two parts; the first is used for experimental work to obtain a 3-month Moving average Model forecast for the recent daily temperature, and to find the exponential Smoothing model through assumed smoothing constant and finally to find the Multiple Linear Regression (MLR) equations models and the second to test the validity of the model. The weather data for this paper is obtained from data collected by empirical and inferential interpretations (EMINF) in Muscat, Oman.

The command of the MS-Excel Tool Pack is used to analyze the data.

Multiple Linear Regression Modeling

Regression is a statistical approach to forecasting change in a dependent variable on the basis of change in one or more independent variable.

The general regression equation of Y on X is

$Y = \delta + \beta X + c$ ………………………….. (1)

This equations is Known as the mathematical modeling for linear regression. More specifically the equation $Y - c + \beta X$ is call deterministic model. Multiple Linear Regression equation is a linear model that describes how a Y variable relates to two or more X variables. The General Structure of Multiple Linear Regression Model is given by

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$…… (2)

Where Y is the predictant or the variable to be predicted and $X_1, X_2, X_3, \ldots, X_K$ is the predictors, β_0 is Y intercept and ε is the error which is distributed normally with zero mean and variance 2.

Basic Features Enhancing Weather Forcasting

There are basic features that enhance our weather forecasting capability. They are called statistical indicators.

A. Moving Average Model (MA)

A moving average is a time series data constructed by taking averages of several sequential values over another time series. It is a type of mathematical convolution.

If we represent the original time series by $x_1, x_2 \ldots x_n$ the a two sided moving average of the time series is given by

MA=$\frac{1}{2k+1}\sum_{j=-k}^{k} x_{t+j}$ where t=k+1,K+2,n-k… (3)

Where, K is the number of terms in the moving average.

The Moving average Model doesn't handle trends or seasonality very well although it can do better than the total mean. Let us have a look at the following data.

The following data shows recent daily temperature of Muscat (Oman) from January to December for two weeks.

II. TABLE

Average Temperature

Average Daily Temperatures (C) for Muscat, Oman												
Minimum Average Maximum												
	January	February	March	April	May	June	July	August	September	October	November	December
1	20 22 24	21 22 24	21 24 26	24 28 33	27 32 37	30 34 40	31 34 41	28 31 34	27 29 31	29 30 34	27 29 31	24 25 26
2	20 22 24	19 22 24	22 24 25	25 29 32	28 32 37	29 34 40	31 34 37	27 31 34	26 29 31	28 30 33	27 29 30	24 25 26
3	20 22 24	19 22 24	21 24 30	25 29 33	30 33 37	28 33 37	30 34 39	26 31 34	27 29 31	29 31 33	27 28 31	22 25 26
4	20 22 24	19 23 26	21 24 26	24 28 30	30 33 37	25 33 39	30 34 39	26 31 37	28 30 33	29 31 34	27 28 29	21 24 26
5	19 23 24	16 23 27	22 24 26	25 28 30	30 34 37	26 33 39	29 34 39	26 30 39	28 31 34	29 31 34	27 28 30	21 24 26
6	20 23 24	16 22 25	22 24 27	26 29 32	31 34 38	29 34 40	29 33 37	25 30 38	28 30 36	27 31 34	26 28 30	21 24 26
7	20 23 24	19 23 25	22 25 29	26 29 32	29 33 37	30 34 41	30 32 36	28 31 37	27 30 33	29 31 34	26 28 30	22 25 26
8	19 22 24	19 22 26	23 25 29	25 28 32	28 33 37	29 33 38	30 32 37	26 31 39	27 30 34	30 31 35	26 28 30	22 25 27
9	21 23 25	20 22 24	24 25 27	26 29 33	30 32 34	29 32 37	30 32 39	28 31 36	28 30 33	30 31 35	25 28 29	23 25 28
10	20 23 25	21 22 24	24 25 28	26 29 35	30 32 34	30 32 35	28 31 34	29 30 33	27 30 32	29 31 34	26 28 30	23 25 27
11	21 23 25	21 22 24	23 26 28	26 29 34	31 32 36	30 33 36	28 31 32	28 30 32	27 30 32	28 31 34	26 27 29	23 25 26
12	22 23 26	21 23 24	23 25 29	26 29 32	31 34 38	30 33 35	28 31 36	27 30 33	29 30 32	28 30 32	26 27 29	23 25 28

The first value of each day is the minimum daily temperature and the second value of each day is the average daily temperature and the last value indicates the maximum daily temperature for each month.

To make our life easy, we can start our work by Using 3- month Moving Average Models (MA (3)) to forecast the Minimum, Average and Maximum recent daily temperature in (℃) of Muscat Oman.

The Moving average Model uses the last t periods in order to predict the temperature in the last t+1.

III. Table

Average Temperature

Month	Actual Min Daily temp(℃)	Forecasted Min. Daily temp(℃)	Actual Average Daily temp(℃)	Forecasted Average Daily temp(℃)	Actual Max. Daily temp(℃)	Forecasted Max Daily temp(℃)
Jan	18		22		25	
Feb	16		23		29	
Mar	21		25		33	
Apr	24	18.33	29	23.33	35	29
May	27	20.33	32	25.66	41	32.33
June	25	24	33	28.66	41	36.33
July	28	25.33	31	32.33	41	39
Aug	25	26.66	30	32	39	41
Sept	26	26	30	31.33	35	40.33
Oct	26	26.33	29	30.33	34	38.33
Nov	25	25.66	27	29.66	31	36
Dec	21	25.66	24	28.66	28	33.33

B. Exponential Smoothing Models

This is a common scheme to produce a smoothed time series.

And is given by

$F_{t+1} = F_t + \alpha(A_t - F_t)$……………...……….. (4)

Where; F_{t+1} is the forecast for the period t+1.

A_t; Is the actual forecast for period t.

α;is the smoothing constant. 0 < α ≤ 1, t > 0.

The smoothing constant α expresses how much our forecast will react to observed difference.

Assuming the smoothing constant α=0.1 the exponential smoothing forecast for the minimum, average and maximum recent daily temperature of Muscat (Oman), given the above Table (Table I) is calculated as follow.

To find the exponential smoothing forecast for the month of January for the above weather data we need to assume that the actual daily temperature is the same as that of the forecasted daily temperature unless it given. and hence

$F_{Feb}(\min) = F_{Jan} + 0.1(A_{Jan} - F_{Jan})$

= 18+0.1*(18-18)

=18°C

$F_{Mar}(\min) = F_{Feb} + 0.1 * (A_{Feb} - F_{Feb})$

=18+0.1*(16-18)

=15.8°C

Suppose given the forecasted Exponential Smoothing daily average temperature for the month of November is 26.63℃, then here we can forecast therefore for the month of December by

$$F_{Dec}\,(\text{Average}) = F_{Nov} + 0.1(A_{Nov} - F_{Nov})$$

=26.59+0.1*(24-26.59)

=26.63℃

The daily maximum temperature for the month of December can also be forecasted as

$$F_{Dec}\,(\text{Max}) = F_{Nov} + 0.1(A_{Nov} - F_{Nov})$$

=33.9+0.1*(31-33.9)

=33.61℃,

Similarly the exponential smoothing forecast of the daily minimum, average and maximum temperature for the rest of months is given the table as summary.

The table below shows the complete output predicted by the Exponential Smoothing forecast for the recent daily Minimum, Average and maximum temperature of Muscat (Oman).

IV. Table

Month	Actual Min temp(°C)	Forecasted Min. temp(°C)	Actual Average temp(°C)	Forecasted Average temp(°C)	Actual Max. temp(°C)	Forecasted Max temp(°C)
Jan	18	18(Assumed value)	22	22(Assumed value)	25	25(Assumed Value)
Feb	16	18	23	22	29	29
Mar	21	15.8	25	22.1	33	29.4
Apr	24	16.32	29	22.39	35	29.76
May	27	17.088	32	23.051	41	30.284
June	25	18.0792	33	23.9459	41	31.3556
July	28	18.7713	31	24.8513	41	32.32
Aug	25	19.6942	30	25.4662	39	33.188
Sept	26	20.2247	30	25.9196	35	33.7692
Oct	26	20.8023	29	26.3276	34	33.8923
Nov	25	21.322	27	26.5948	31	33.9031
Dec	21	21.6898	24	26.6354	28	33.6128

C. Implementation of MLR Model

The estimation of the maximum and minimum daily temperature by Multiple Linear Regression model for the recent weather data is based on the Multiple Linear Regression Excel Analysis ToolPak. Sometimes the weather parameters to be forecasted can be estimated based on the same basic features of the weather data set.

But there may be some cases that the parameter to be forecasted shows strong dependence on other parameter.

In this case the model will include some basic futures from other parameter. And hence to forecast the minimum, average and maximum daily temperature, independent variables should include all these basic inputs which have been strong part of the weather data set throughout the observation. so we will use our MA, and EMA, as independent variable to predict the temperature.

The Experiment has two phases, the first phase is to use some of the existing weather data for creating the model, and the second phase is to check the validity of the model.

The output Multiple Linear Regression Model for the Daily minimum Temperature by the Excel TOOLPAK data analysis is given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

$$Y = 20.95 + 1.08X_1 + (-1.179X_2) \ldots (5)$$

The validity of the model can be seen from the table attached below.

V. Table

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.78979784
R Square	0.62378062
Adjusted R Square	0.54017632
Standard Error	2.5123943
Observations	12

ANOVA

	df	SS	MS	F	Significance F
Regression	2	94.19087386	47.0954369	7.46110636	0.012288114
Residual	9	56.80912614	6.31212513		
Total	11	151			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	20.9545417	7.486142309	2.79911078	0.02074275	4.019711222	37.8893721	4.01971122	37.8893721
X1	1.08633349	0.290829176	3.73529749	0.00465919	0.428432187	1.7442348	0.42843219	1.7442348
x2	-1.17978601	0.572464798	-2.0608883	0.0693803	-2.474791352	0.11521934	-2.47479135	0.11521934

Fig.1 Minimum Temperature

As we can see from the above table the multiple regression is 0.78 which fairly acceptable and the significance level from the ANOVA analysis is 0.02**.**This proves that the model is valid. Similarly the predictive values are less the value of 0.05 and hence the Model for the daily minimum temperature is valid.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.666639
R Square	0.444407
Adjusted R Square	0.320942
Standard Error	4.476884
Observations	12

ANOVA

	df	SS	MS	F	Significance F
Regression	2	144.2842	72.14212	3.599459	0.071023867
Residual	9	180.3824	20.04249		
Total	11	324.6667			

	Coefficientstandard Err		t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	Ipper 95.0%
Intercept	29.16052	17.41701	1.674255	0.128405	-10.23950189	68.56054	-10.2395	68.56054
X1	-1.1274	0.962352	-1.1715	0.271476	-3.304390091	1.049592	-3.30439	1.049592
x2	1.176241	0.523331	2.247603	0.051203	-0.007616575	2.360098	-0.00762	2.360098

Fig 2. Average Temperature

The above figure indicates the Output for the multiple linear Regression model for the Daily average temperature. And hence the MLR Model for daily average Temperature is given by;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n.$$

$$Y = 29.16 + 1.34X_1 \text{...} \quad (6)$$

We ignore X_1 as its corresponding P-Value is 0.27 which is greater than the predictive value and hence we can say that X_1 doesn't have significance effect in predicting the average daily temperature. To check the validity; multiple regression square and significance level are all are suitable values of the analysis for predicting. Hence it is a valid Model.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.679062159
R Square	0.461125416
Adjusted R	0.341375508
Standard E	2.965911096
Observatio	12

ANOVA

	df	SS	MS	F	Significance F
Regression	2	67.74700901	33.87350451	3.850737133	0.061900571
Residual	9	79.16965765	8.796628628		
Total	11	146.9166667			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	32.3160905	14.77467522	2.187262327	0.05650182	-1.106546876	65.73872789	-1.106546876	65.73872789
X1	-1.720216352	1.112548541	-1.546194425	0.15645986	-4.236976002	0.796543299	-4.236976002	0.796543299
x2	1.349269481	0.560393979	2.407715878	0.039396301	0.081570227	2.616968735	0.081570227	2.616968735

Fig 3. Maximum Temperature

The above table indicates the MLR Model summary Output for the daily maximum temperature. As we can see from the result the ANOVA results it is very significance and helpful in predicting our maximum daily temperature. More over the P-value for the first independent variable (X_1) is valuable. Therefore the corresponding MLR Model for maximum daily temperature is given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n.$$

$$Y = 32.31 + 1.34X_2 \text{...........................} \quad (7)$$

Regarding the second variable X_2 it doesn't have significance effect in predicting our temperature since the predictive value is more than the required. (P-value must be less than 0.05).

To sum up the first variable X_1 doesn't have significance effect in predicting our temperature since the predictive value is more than the required.(P-value must be less than 0.05).

GRAPHIC ANALYSIS OF THE FORECASTED AND ACTUAL TEMPERATURES.

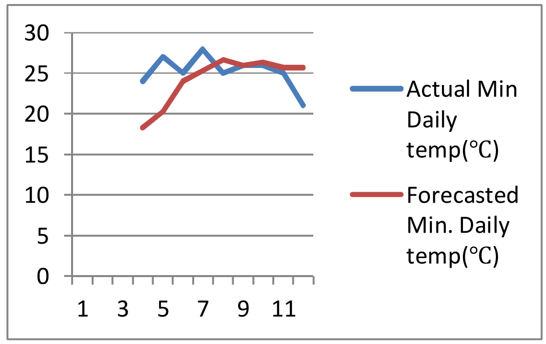

Fig.4 Shows the 3 Months Moving Averages for the Actual and Forecasted Daily Minimum Temperature.

From the above data we can see that the actual data and the forecasted temperature have a flat or smoothed time series

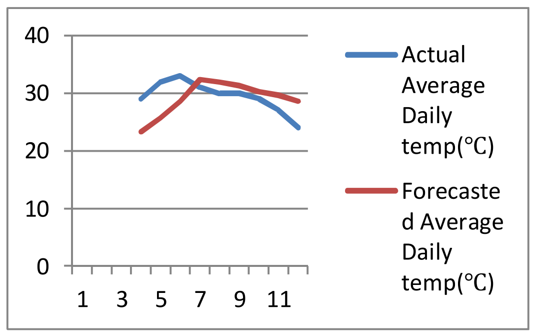

Fig 5.Show Three Month Moving Average for the Actual and Forecasted Daily Average Temperature.

From the above figure it can be inferred that the actual daily average temperature and the forecasted daily average temperature follows smoothed time series.

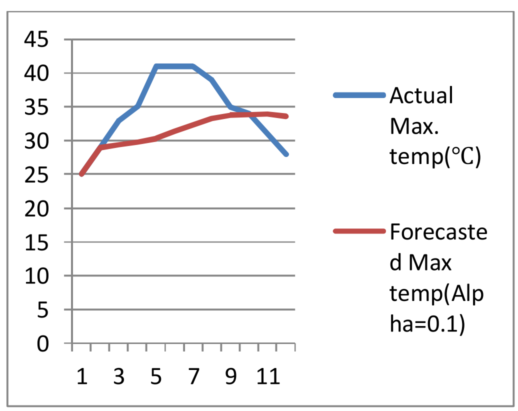

Fig 6. Shows the Exponential smoothing Model between the Actual and Forecasted daily Maximum Temperature.

As we can see from the above line diagram the actual daily maximum temperature and the forecasted daily maximum temperature that was computed though EMA model follows the smoothed time series. The less the smoothing constant $\propto$ the more flat or smooth will be the time series.

VI. CONCLUSIONS

Mathematical and Statistical modeling can be used for smart environmental weather forecasting, specifically for predicting the maximum minimum and average daily temperature by using the Moving average and exponential smoothing as main input parameter for forecasting using Multiple Linear Regression Model (MLR).The M-excel Tool Pack package is the best and the most simple way to analyze and to check the validity of the Model. Further environmental weather forecasting for humidity can also be predicted using the MLR Model.

ACKNOWLEDGMENT

The author would like to thank his colleagues for their valuable comments and input.

REFERENCES

[1] E. Lorenz.(1977)..*An Experiment in Nonlinear Statistical Weather Forecasting.* Mon. Wea. Rev., vol. 105, no. 5, pp. 590-602.

[2] P. Alaton, B. Djehiche and D.Stillberger. (2002). *'On modelling and pricing weather derivatives'.* Applied Mathematical Finance, vol. 9, no. 1, pp. 1-20..

[3] P. Lynch.(2008).*The origins of computer weather prediction and climate modeling*. Journal of Computational Physics, vol. 227, no. 7, pp. 3431-3444.

[4] P. Sanjay Mathur.(2015).*A Simple Weather Forecasting Model Using Mathematical Regression.* Indian Research Journal of Extension Education, vol. , 2012, no. 12432, p. 161.

[5] R. Adhikari and R. Agrawal. (2013). *An Introductory Study on Time Series Modeling and Forecasting*. Ratnadip Adhikari R.K. LAP Lambert Academic Publishing, Germany.

[6] S. Marras, J. Kelly, M. Moragues, A. Müller, M. Kopera, M. Vázquez, F. Giraldo, G. Houzeaux and O. Jorba, (2015). *'A Review of Element-Based Galerkin Methods for Numerical Weather Prediction*. Finite Elements, Spectral Elements, and Discontinuous Galerkin', Arch Computat Methods Eng.

[7] S. N, R. S, P. Ray, K. Chen, A. Lassman and J. Brownlee. (2013).*'Grids in Numerical Weather and Climate Models*. Climate Change and Regional/Local Responses.

[8] S.L.Belousov, L.V Berkovich and I.G Sitnikov (2015).M*athematical modeling in Meteorology and Weather forecasting*. Moscow: Hydro meteorological Research Center of Russia, Moscow, p. 10.