An Application of ARIMA Models in Weather Forecasting: A Case Study of Heipang Airport – Jos Plateau, Nigeria

1Datong, Godwin Monday & 2Golton, Emmanuel Nengak

1Department of Mathematics
University of Jos
Jos, Nigeria.
Email of Correspondence Author: mallong2007@yahoo.com

2Department of Mathematics,
Ahmadu Bello University
Zaria, Nigeria.
Email: spomariae@gmail.com.

ABSTRACT
This paper presents an application of Autoregressive Integrated Moving Average (ARIMA) models for weather forecasting. The data used for developing the models consisted of a five-year (2011 – 2016) monthly records of average temperature, rainfall and wind velocity of Heipang airport - Jos Plateau, Nigeria.

In the model estimation section, we were able to fit seasonal autoregressive and integrated moving average (SARIMA) models for temperature and rainfall. On the other hand, no ARIMA or SARIMA model was fitted to the wind velocity data because the data resembles white noise as close as possible. This means that the wind velocity data was pattern less.

Secondly, the selected model for temperature had a relatively small variance of 1.87. This means that forecasts from such a model will have some degree of accuracy as manifested in the small standard errors associated with the individual forecasts from the model. On the other hand, the model for the rainfall data has a very large variance of 2230 which implies that forecasts from such a model have questionable accuracies as can be verified by the large standard errors of the individual forecasts. Lastly, parameters of both models where statistically significant and values predicted by the models were lying between 95% confidence interval of the forecasts in each case.

Keywords: Weather Forecasting, ARIMA, SARIMA, variogram.

1. INTRODUCTION
Every now and then, forecasts are being made concerning future weather conditions. Closest to us are the daily forecasts of temperature, rainfall and how bright or gloomy the next day will be on our local television channels.

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. Human beings have attempted to predict the weather informally for millennia, and formally since the nineteenth century. Weather forecasting is made by collecting quantitative data about the current state of the atmosphere on a given place and using scientific understanding of atmospheric processes, to project how the atmosphere will evolve on that place. Several weather forecasting methods
are available, some of which are ensemble probabilistic, persistence (today equals tomorrow), climatology and analogue method. There are however highly complex computerized mathematical models that are now available for weather forecasting.

Data recorded by the meteorological centre of Heipang airport, Jos-Plateau have been used to develop the ARIMA models. These models were then used for further analysis and forecasting.

2. LITERATURE REVIEW

Scientific researchers and meteorologists have seen weather forecasting as one of the most challenging problems because of its practical importance in agriculture, air transport other spheres of human life. It is evident that there is a need for accurate estimates of the amount of rainfall, temperature and wind speed on a variety of spatio-temporal scales. Some of the research works on weather forecasting includes:

Edward Lorenz, proposed in 1963 that it is impossible for long-range forecasts – those made more than two weeks in advance – to predict the state of the atmosphere with any degree of skill, owing to the chaotic nature of the fluid dynamics equations involved.

Edward Epstein recognized in 1969 that the atmosphere could not be completely described with a single forecast run due to inherent uncertainty, and proposed a stochastic dynamic model that produced means and variances for the state of the atmosphere.

It has been discovered by Bilham (1934) and Hawke (1934) that monthly rainfall persistence is generally less than temperature persistence. Baur (1958, 1963) recognized many sequential and recurrence tendencies in weather lore in Germany in particular.

Ratcliffe (1978) reported some success with climatic forecasting in both the northern and southern hemispheres using knowledge of the phases of the QBO and saw prospects for increasingly reliable forecasts a year or so ahead.

Ozelkan and Duckstein (1996) compared the performance of regression analysis and fuzzy logic in studying the relationship between monthly atmospheric circulation patterns and precipitations.

Mehmet Tektas (2010) studied weather forecasting of Istanbul using ANFIS and ARIMA models. He discovered that the performance comparisons of ANFIS and ARIMA models due to moving average error (MAE), root mean square error (RMSE) indicate that ANFIS yields better results.

3. THEORETICAL FRAMEWORK

The ARIMA Model

Autoregressive integrated moving average (ARIMA), sometimes called Box-Jenkins models result from differencing a nonstationary time series before fitting an autoregressive moving average (ARMA) model to the differenced series.

Consider the time series $y_t$, the first difference is given by

$$y_t - y_{t-1} = (1 - B)y_t$$

where $B$ is the backshift operator defined as $B^j = y_{t-j}$. Hence, the dth difference is therefore given as $(1 - B)^d y_t = \nabla^d y_t$ and if the original series is differenced $d$ times before fitting an ARMA($p$, $q$) process, then the model for the original undifferenced series is said to be an ARIMA($p,d,q$) generally defined as

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where $w_t = \nabla^d y_t$, the dth differences of the original series. $\phi_1, \ldots, \phi_p$ are
the AR parameters to be estimated and $\theta_1, \ldots, \theta_q$ are the MA parameters to be estimated. As can be observed, an ARIMA (p,d,q) process is the dth differences of the original nonstationary series having p autoregressive terms and q moving average terms.

The least squares regression is applied in the estimation of $\phi$ and $\theta$ by first replacing the unobserved quantities $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ by estimated values, $\hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-q}$. The parameters $\phi$ and $\theta$ are then estimated by regressing $y_t$ on $y_{t-1}, \ldots, y_{t-p}, \hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-q}$.

One of the objectives of time series analysis is forecasting. After selecting an appropriate model based on the data at hand, the desire is to make forecasts about future values.

Our forecast model for $y_{t+k}$ is as follows:

$$\hat{y}_t(k) = \sum_{i=1}^{p+d} \varphi_i \hat{y}_{t-i}(k) + \hat{\varepsilon}_t(k) + \sum_{i=1}^{q} \theta_i \hat{y}_{t-i}(k)$$

The actual calculation of the forecasts is done based on the following rules:

i) The conditional expectations of the present and past observations are actually themselves. That is, $\hat{y}_t(-j) = y_{t-j}, for j = 0,1, \ldots$

ii) The conditional expectations of the future observations are replaced by their future forecasts. That is,

$$E_t[y_{t+j} | y_t, y_{t-1}, \ldots] = \hat{y}_t(j), for j = 1,2, \ldots$$

iii) The conditional expectations of the present and past shocks are actually themselves, but since these random shocks are not observed, we replace them by the one-step-ahead forecast errors. That is,

$$[\varepsilon_{t-j}] = \varepsilon_{t-j}(1) = \varepsilon_{t-j} - \hat{\varepsilon}_{t-j-1}(1), for j = 0,1,2,\ldots$$

iv) The conditional expectations of the future shocks are replaced by zero as earlier assumed. That is,

$$[\varepsilon_{t+j}] = 0, for j = 1,2, \ldots$$

The R programming software will be used the analysis and forecast of the data.

4. DATA ANALYSIS AND CASE STUDY CHARACTERISTICS

Data recorded by the meteorological centre of Heipang airport, Jos-Plateau are used for analysis and forecast in this study. The variables considered for the study are: temperature, rainfall and wind speed. These variables have strong relationship with other weather parameters. To develop the ARIMA models for the variables, the first step in the analysis of data is exploratory data analysis where a plot of the data is examined.
4.1. Temperature
Temperature can be measured more precisely than any other weather variables. Since establishment of the airport, temperature data are recorded and used for different purposes. Figure 1a shows the time series plot of the mean temperature.

![Plot of Temperature](image1.png)

Figure 1a

To have a better understanding about the stationarity or otherwise of the process above, we make use of the correlogram and partial correlogram plots.

![ACF of Temperature](image2.png)
![PACF of Temperature](image3.png)

Figure 1b
From the correlogram in figure 1b, it can be observed that the ACF does not die off fairly quickly. Consequently, we shall use the variogram to determine the stationarity or otherwise of the temperature data.

![Variogram of temp](image)

Fig. 1c

Similarly, the variogram in Fig. 1c indicates that the temperature data is not stationary as indicated earlier by the correlogram. So, we difference the series and examine its correlogram and partial correlogram again.
Figure 1d

The plots in figure 1d were obtained by differencing the original temperature data at lag 12 before evaluating the ACF and PACF. Examining the two plots together, we are left with the option of fitting an ARMA model to the differenced series since the two plots do not satisfy the AR nor the MA model. Using trial and error, we shall select the model with a relatively small variance. A sample (of 60 observations) of the original temperature data shall be used to fit the desired model.

4.2 Rainfall

The amount of rainfall is measured during the wet season (March to October) at the airport’s meteorology centre. A time series plot of the average amount of the rainfall data as displayed in fig.2a

Figure 2a
Looking at the ACF in figure 2b, we observe that the ACF does not die out even after 30 lags with consistent larger autocorrelations after six lags. This is an indication that there is seasonality in the data and needs to be differenced at lag twelve.

![ACF of Rainfall](image1)

**Figure 2b**

Consequently, we shall examine the variogram of the rainfall data displayed in figure 2c. An examination of the variogram reveals that the rainfall data is not stationary since it doesn’t seem to reach any asymptote. The option now is to difference the data and see if we can attain some stationarity.
Figure 2d below displays the correlogram and partial correlogram of the rainfall data that was differenced twice at lag 12.
As is the practice, we use 62 out of the 69 observations to fit an ARIMA model to the rainfall data.

4.3 Wind velocity
Wind is one of the most difficult meteorological parameters to forecast. At any temperature, wind velocity is significant and is a major factor determining the cooling rate. Fig. 3a presents the wind velocity recorded at Heipang airport.

Figure 2d
From the correlogram and partial correlogram in figure 3b, the wind velocity data seems to resemble white noise. To further confirm our suspicion, we use the Ljung-Box test to establish the presence or otherwise of significant autocorrelations in the wind velocity data.
5. RESULTS AND DISCUSSIONS

In this study last section, we were able to fit an ARIMA model for the temperature data and the rainfall data. We may wish to compare the original time series plot with the plot that contains the forecasts generated by the selected model to have a visual judgment of the model’s performance. The two graphs are placed side by side in figures. The red line represents the forecasts generated by the estimated model while the dotted lines denote 95% confidence interval about the forecasts.
A summary of the findings are presented in table 4 below.

Table 4 Summary of Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Temperature Model</th>
<th>Rainfall Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(1,0,0)(0,1,1)[12]</td>
<td>$y_t = 0.6054y_{t-1} + \varepsilon_t - 1.4975\varepsilon_{t-1} + 0.4975\varepsilon_{t-13}$</td>
<td>ARIMA(0,0,0)(2,2,0)[12]</td>
</tr>
<tr>
<td>Variance</td>
<td>1.87</td>
<td>2230</td>
</tr>
<tr>
<td>AIC</td>
<td>176.13</td>
<td>428.38</td>
</tr>
<tr>
<td>AICC</td>
<td>176.67</td>
<td>429.09</td>
</tr>
<tr>
<td>BIC</td>
<td>181.74</td>
<td>433.3</td>
</tr>
</tbody>
</table>
6. CONCLUSION
The data collected from the meteorological centre of Heipang airport was used to fit an ARIMA model to the temperature data (a Seasonal ARIMA model to be precise). The variance of the model was 1.87 which is relatively small. This means that forecasts from this model can be considered reliable to a large extent. Also, the forecasts were seen to be close to the corresponding actual values which were lying between the forecasts plus or minus the standard errors of the forecasts. Thus, the selected model is considered to be useful for forecasting monthly temperatures of Jos-Plateau.

In the same vein, the researcher was also able to fit an ARIMA (a Seasonal ARIMA) model to the rainfall data. In this case, the model had a variance of 2230 which is very large. This implies that the precision in forecasting by this model is questionable even though the forecasts generated from the model for seven months ahead seem to be close to the actual values. Therefore, despite the large variance of the model, it can be considered as reliable when used to forecast future rainfall data.

Lastly, the wind velocity data proved to be white noise therefore, no ARIMA model was fit to the data as that will be inappropriate.

7. REFERENCES
Avril Coghlan (2013), A Little Book of R for Time Series, Release 2.0