A time-series forecasting model for Windhoek Rainfall, Namibia.

by

***Pazvakawambwa G.T. and Ogunmokun A. A.**

University of Namibia, Faculty of Engineering and IT

*corresponding author email address: PazvakawambwaG@namwater.com.na

Abstract

The objective of the study was to use time series forecasting techniques to model Windhoek rainfall based on secondary monthly data from 1891 to 2011. Descriptive summary statistics in the form of measures of centrality and dispersion, time series plots, and autocorrelation functions were generated using R time series statistical software. The Box Jenkin's ARIMA modelling procedure (model identification, model estimation, model validation) was used to determine the best models for the data. Model diagnostics based on residual analysis were performed to assess the adequacy of the identified models. The final model was then used to forecast monthly rainfall for Windhoek up to year 2047. The forecast values suggest that for instance in the 2046/47 , the winter season monthly rainfall point estimates are around 15mm (June 14.5mm; July 14.5 mm; and August 14.3mm) which can technically be higher than expected. However, the lower 95% confidence limits for the same winter months are zero highlighting the possibility of no rainfall during those periods. Based on the ARIMA modelling of the Windhoek rainfall, despite the seasonal and irregular fluctuations, the mean monthly rainfall levels did not suggest an upward or downward trend over the century. Even though the results indicate constant mean monthly rainfall, the limitation of the results is that the analysis is based on a small spatial area to completely rule out climate change effects. Therefore, more adaptive governance initiatives should be explored on the available secondary sources for water security and the sustainable development of the USB.

1. Introduction

Times series analysis plays a significant role in modelling meteorological data such as humidity, temperature, rainfall and other environmental variables (Collischonn et al., 2005; Hung et al., 2009; Meher & Jha, 2013; Kanna et al., 2010; Mahsin et al., 2012; Ansari, 2013; Htike & Khalifa, 2010). Rainfall forecasting is crucial for making important decisions and performing strategic planning. The ability to predict and forecast rainfall quantitatively guides the management of water related problems such as extreme rainfall conditions like floods and droughts among other issues (Htike & Khalifa, 2010; Ansari, 2013; Kanna et al., 2010; Meher & Jha, 2013).

Therefore, predicting hydrological variables like rainfall, floodstream and run-off flow as probabilistic events is a key subject in water resources planning. These hydrological variables are usually measured longitudinally across time. This makes times series analysis of their occurrences in discrete time appropriate for monitoring and simulating their hydrological behaviours (Ansari, 2013). Rainfall is among the sophisticated and challenging components of the hydrological cycle to model and forecast because of various dynamic and environmental factors and random variations both spatially and temporally (Htike & Khalifa., 2010).

Time Series Analysis

A time series analysis often exhibits four main components such as trends, seasonality, cycles and irregular fluctuations. Trends are the long term underlying movements representing growth or decline in a time series over an extended period of time due to natural, human, economic and other processes. They are usually described by a smooth, continuous curve or a straight line. Trends can be captured using times series regression, double exponential smoothing and moving average methods. Seasonal variations are fluctuations in a time series that are repeated at regular short intervals usually within a year (e.g. daily, weekly, monthly, quarterly). These variations are usually due to recurring environmental influences such as climatic conditions (seasons and climate change, Wegner, 2010). Seasonal patterns are usually measured using index numbers, which can be very useful in the estimation of short term forecasts. Cyclical variations are wavelike fluctuations around a trend which can vary substantially in duration and amplitude, suggesting possible existence of periodicity with longer intervals. To a limited extent, index numbers can also be used to measure cyclical fluctuations in terms of the phase of the cycle through which a particular time series is moving and this can be used to adjust forecasts and to account for the likely influence of cyclical forces. Irregular fluctuations are the residuals after other components of the time series have been removed. Causes of irregular components are generally unpredictable once-off events such as natural disasters (floods, droughts, fires) or man-made disasters (strikes, boycotts, accidents, and acts of violence, Wegner, 2010).

Time series decomposition facilitates the separation of the components either additively or multiplicatively. Additive time series decomposition is according to the model given in Equation 1, while multiplicative decomposition is according to the model given in Equation 2

$$\boldsymbol{Y_t = T_t + S_t + C_t + I_t} \qquad \textbf{Equation 1}$$

$$\boldsymbol{Y_t = T_t . S_t . C_t . I_t} \qquad \textbf{Equation 2}$$

where Y_t is the observed time series, T_t is the trend component, S_t is the seasonal component, C_t is the cyclical component, and I_t is the irregular component.

Among the popular statistical techniques used for rainfall forecasting are the Box-Jenkins Auto-regressive Moving Average (ARIMA) models ARIMA models possess many desirable features. They allow the analyst who has observations only on past years (e.g. historical data on rainfall) to forecast future events without having to search for other related time series data (e.g. temperature). ARIMA models have been widely applied in a variety of water and environmental management applications (Otok & Suhartono, 2009; Rabenja et al., 2009; Mauludiyanto et al., 2010; Abudu et al., 2010; Turalam & Ilahee, 2010; Chattopadhyay & Chattopadyay, 2010; Shamsinia et al., 2011; Babu et al., 2011; Mahsin et al., 2012).

If historical data on the dependent variable (e.g. water demand) is available, ARIMA models can be employed for analysis and forecasting as they account for autocorrelation in the water demand time series and use the previous period's value (day, week, month or year) as an independent variable.

An autoregressive model of order *p* is conventionally classified as AR (p) and a moving average model with *q* terms is known as an MA (q). A combined model that contains *p* autoregressive terms and *q* moving average terms is called ARMA (p,q). If the object series is differenced d times to achieve stationarity, the model is classified as ARIMA (p,d,q) where the letter "I" stands for "Integrated". This means that an ARIMA model is a combination of an autoregressive (AR) process and a moving average (MA) process applied to a non-stationary data series (Mahsin et al., 2012; Jakasa et al., 2011; Meher & Jha, 2013). Thus the general non-seasonal ARIMA (p,d,q) model as given in Equation 3 is composed of :

AR:p=order of the autoregressive part,

I: d=degree of differencing involved and,

MA: q=order of the moving average part.

$$\boldsymbol{Y_t = c + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_t - \vartheta_1 e_{t-1} - \vartheta_2 e_{t-2} - \cdots - \vartheta_q e_{t-q}}$$

Equation 3

The same equation can also be presented using the backshift notation as given in Equation 4:

$$\boldsymbol{\left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right) Y_t = c + \left(1 - \vartheta_1 B - \vartheta_2 B^2 - \cdots - \vartheta_q B^q\right)}$$

Equation 4

Where for both equations, c= constant term, $\emptyset_i$ =i^{th} autoregressive parameter, ϑ_j =j^{th} moving average parameter, e_t is the error term at time *t*, and B^k is the k^{th} order backward shift operator.

Further to the non-seasonal ARIMA (p,d,q), one can identify Seasonal ARIMA(P,D,Q) parameters for the time series data known as SARIMA. These parameters are seasonal autoregressive (P), seasonal differencing (D) and seasonal moving average (Q). The general form of the SARIMA(p,d,q)(P,D,Q)$_s$ model using the backshift notation is given by equation 5.

$$\boldsymbol{\emptyset_{AR}(B_{sAR})(B^s)(1-B)^d(1-B^s)^D Y_t = \vartheta_{MA}(B)\vartheta_{sMA}(B^s)e_t} \quad \textbf{Equation 5}$$

where s = number of periods per season;

$\emptyset_{AR}$ = non-seasonal autoregressive parameter

ϑ_{MA} = non-seasonal moving average parameter

ϑ_{sMA} = seasonal moving average parameter

The Box-Jenkins methodology applies ARMA, ARIMA or SARIMA to establish the best fit of a time series historical values to make forecasts. The methodology consists of four stages namely model identification, estimation of model parameters, diagnostic checking for the identified model appropriateness for modelling, and application of the model (i.e. forecasting). Model identification involves testing whether the time series is stationary and if there is significant seasonality to be modelled. The data can be examined to check for the most appropriate class of ARIMA processes by selecting the order of the successive seasonal differencing required to make the series stationary as well as the specification of the order or regular and Seasonal Autoregressive Integrated Moving Average (SARIMA) polynomial required to sufficiently represent the time series model. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are the most important tools for time series analysis and forecasting. The ACF quantifies the extent of linear dependence between time series observations separated by lag *k*. The PACF plot is used to determine how many auto regressive terms are necessary to reveal one or more of the following characteristics: time lags where high correlations appear; seasonality of the series; and the trends either in the means or variances of the series. Stationarity of data can be assessed by the Ljung-Box test with test statistic as given by Equation 6.

$$\boldsymbol{Q = n(n+2)\sum_{k=1}^{n}\frac{r_k^2}{n-k}} \quad \sim \quad \chi^2_{h-m}$$ **Equation 6**

where
r_k is the autocorrelation at lag k
h is the maximum lag being considered
n is the sample size
m is the number of parameters in the model which has been fitted to the data.

The Akaike Information Criterion (AIC) is mostly used to choose the best model from among competing models and is given by Equation 7.

$$\boldsymbol{AIC = -2\log(L) + 2(p+q+P+Q+k)}$$ **Equation 7**

The model which has the minimum AIC is considered the best model (Mahsin et al., 2012). When the most appropriate model has been chosen, the model parameters can be estimated using the least squares method. The values of the parameters can be chosen such that they minimize the Sum of Squared Residuals (SSR) between the actual data and the estimated values. Alternatively, non-linear estimation can be used to estimate the identified parameters using the generally preferred maximum likelihood estimation techniques. The next stage of the process is the diagnostic checking where the residuals from the fitted model are examined, usually by correlation analysis with the aid of ACF plots. If the residuals are correlated, the model will need to be revisited. Otherwise the correlations are Gaussian white noise (i.e. normal with zero mean and constant variance) and the model is adequate to represent the time series.

2. Data and Methods

Daily rainfall data were measured and collected from Namibia Meteorological Service, who used standard rain gauges in millimetres according the World Meteorological Organization (WMO) standards (WMO, 2015). Descriptive summary statistics in the form of measures of centrality and dispersion, time series plots, and autocorrelation functions were generated using R time series statistical software. The Box Jenkin's ARIMA modelling procedure (model identification, model estimation, model validation) was used to determine the best models for the data. Model diagnostics based on residual analysis were performed to assess the adequacy of the identified models. The final models were then used to forecast monthly rainfall for Windhoek up to year 2050.

3. Results and Discussion

The time series plot of the monthly rainfall is presented in Figure 1. Monthly rainfall was very varied with a minimum of 0 mm and a maximum of 321 mm. The average monthly rainfall was 31.21 mm (95% confidence interval: 28.21 mm, 33.77 mm) with a standard deviation of 48.81 mm. There were notably high rainfall figures in February 1923 (303.00 mm), January 1893 (308.00 mm), March 1954 (312.20 mm), January 2011 (320.90 mm) and January 2006 (321.30 mm).

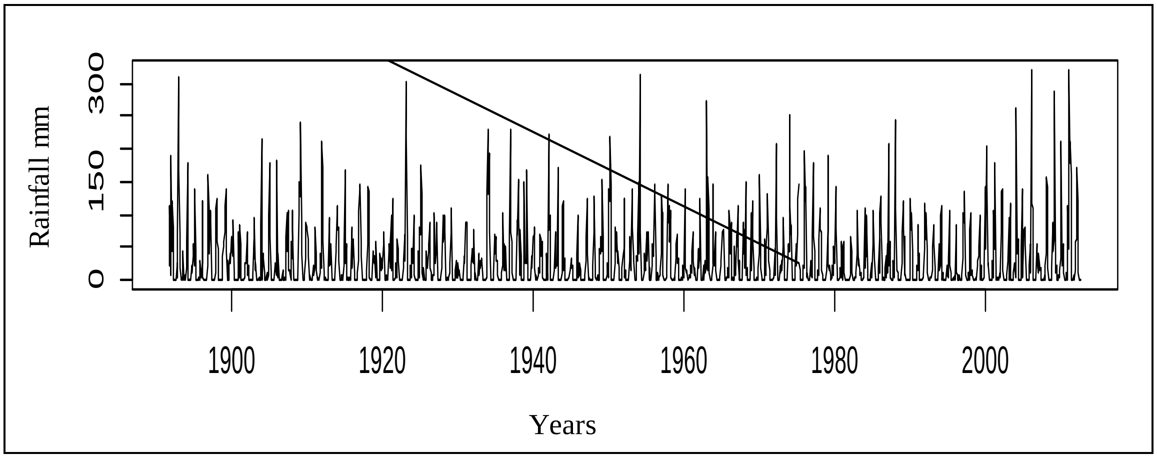

Figure 1 Time series plot of Windhoek monthly rainfall from 1891 to 2012

The time series plot suggests that the time series is stationary since the mean and the variance of the series do not seem to vary with the level of the series, making additive decomposition ideal for the separation of the time series components. The results of the additive time series decomposition are presented in Figure 2**.**

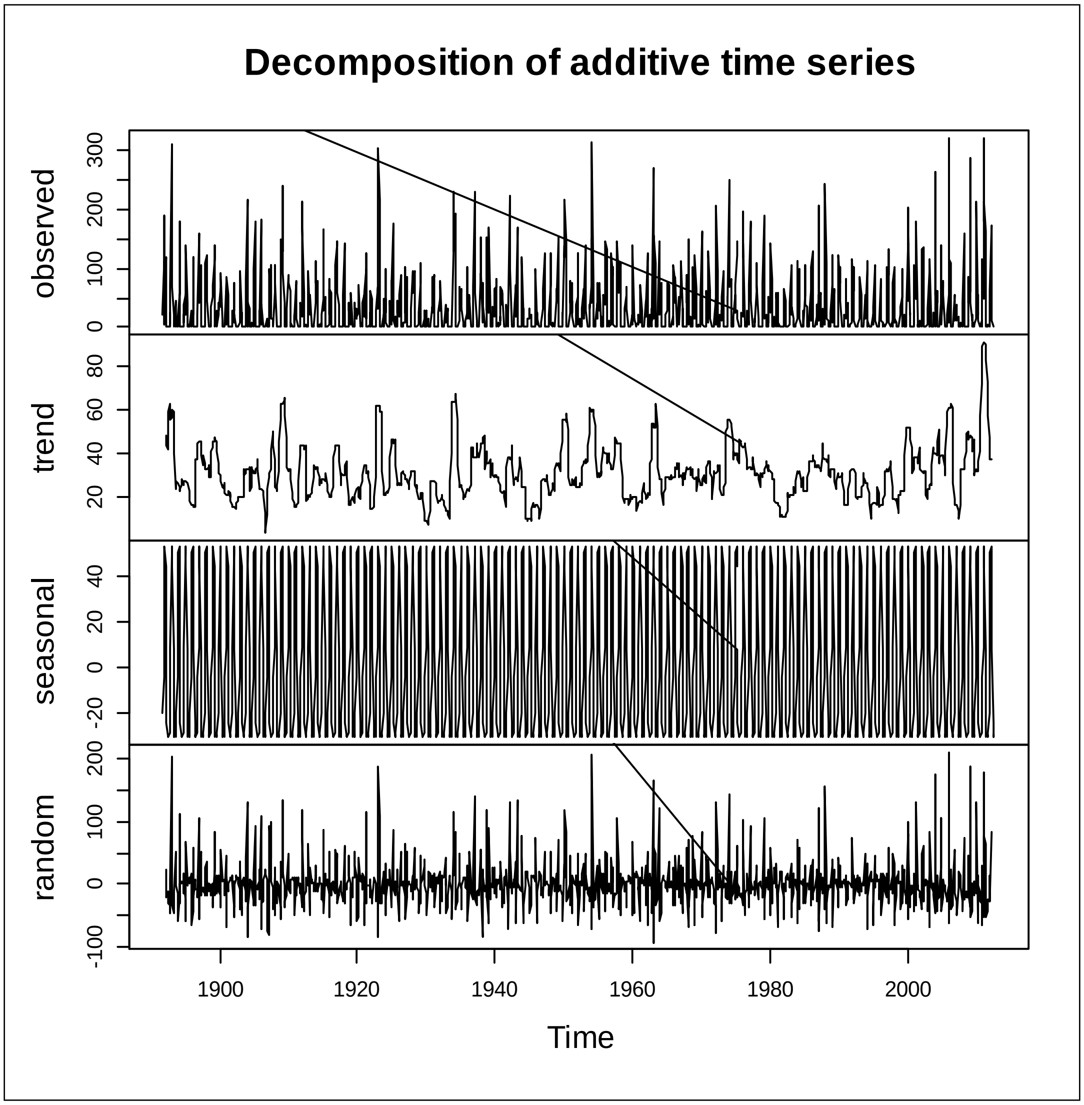

Figure 2 **Decomposition of the Windhoek Rainfall Time Series**

The ACF and the PACF of the rainfall series are presented in Figure 3 and they show seasonality in the data. Therefore, a general ARIMA $(p,0,q)(P,0,Q)_{12}$ was proposed for the rainfall data. After model identification, the p, q, P, and Q parameters were estimated.

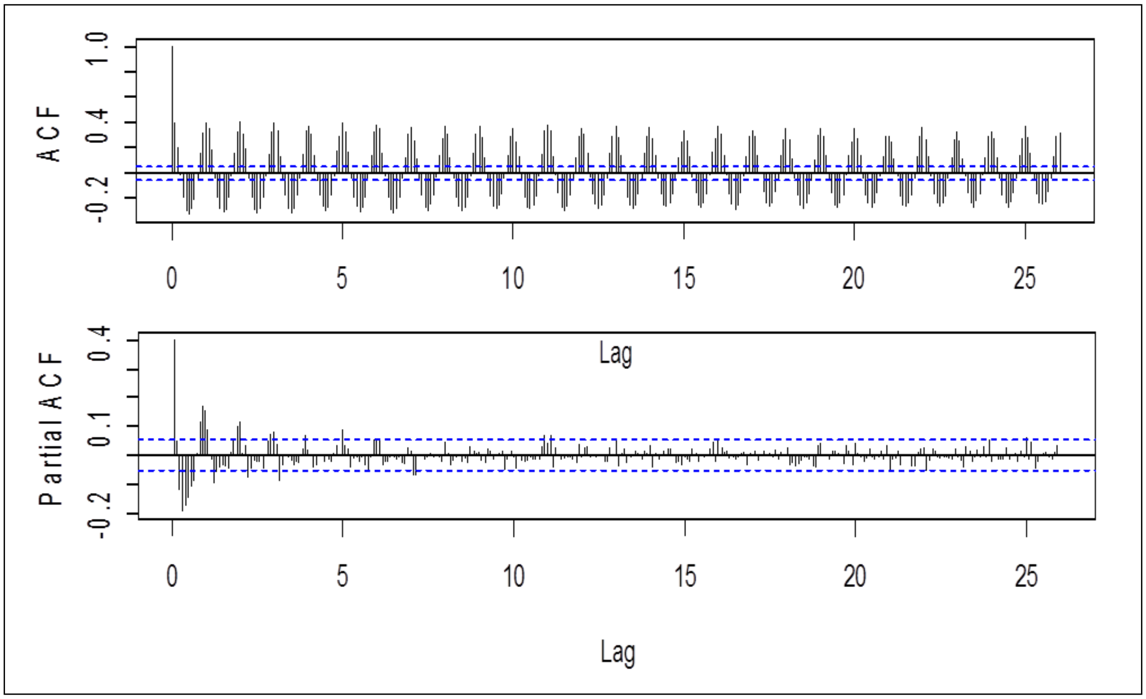

Figure 3 ACF and the PACF of the Windhoek rainfall series

On the basis of automatic ARIMA forecasting, the selected model was a seasonal ARIMA(1,0,1)(1,0,2)$_{12}$ i.e. SARIMA(1,0,1)(1,0,2)$_{12}$. This model was adequate to represent the data and could be used to forecast future rainfall. The maximum likelihood estimates of the SARIMA(1,0,1)(1,0,2)$_{12}$ model and their standard errors are presented in Table 3. The forecast residual ACF, time plot, Normal Q-Q plot and histogram are displayed in Figure 4.

Table 2 Maximum likelihood estimates of the SARIMA(1,0,1)(1,0,2)$_{12}$ model and their standard errors

Coefficients	Estimate	Standard Error
Constant	31.1735	12.1868
Ar1	0.6135	0.1168
Ma1	-0.5122	0.1265
Sar1	0.9852	0.0006
Sma1	-0.9154	0.0270
Sma2	0.0502	0.0264

$\hat{\sigma}_e^2 = 1505$, log likelihood=-7371.98, AIC=14758.13

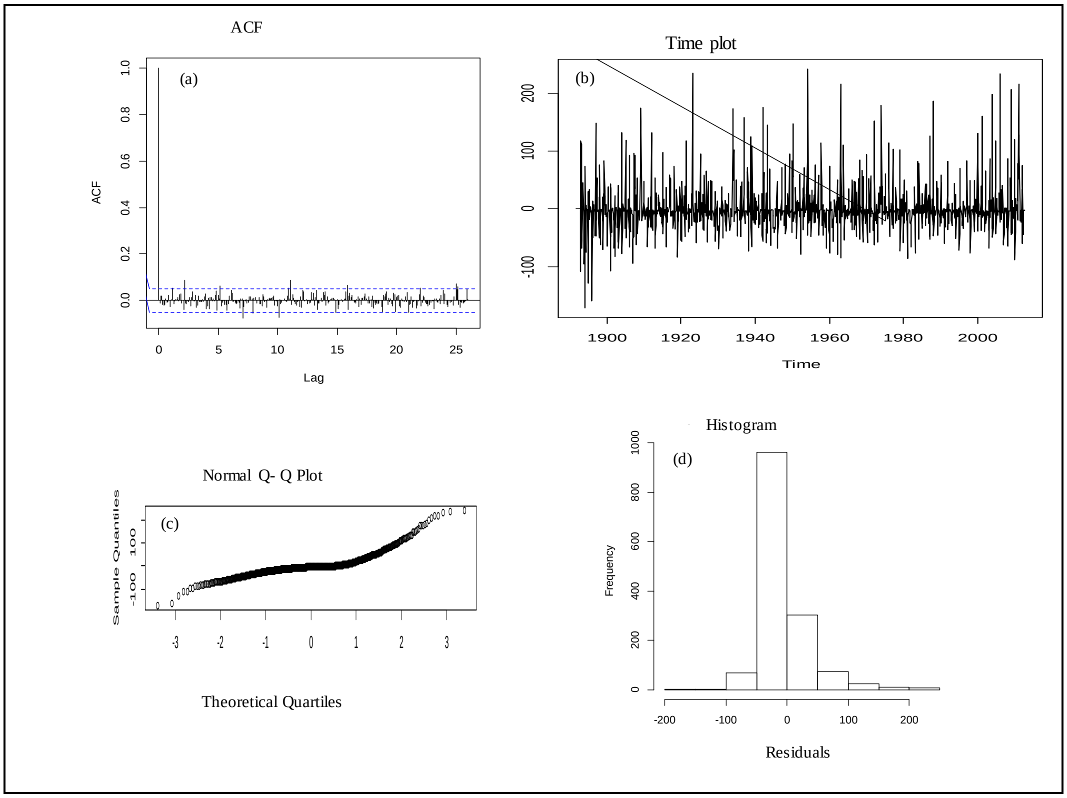

Figure 4 **Diagnostics for the SARIMA (1,0,1)(1,0,2)$_{12}$ fit on the rainfall data a) ACF of forecast residuals; b) Forecast residual time plot; c) Normal Q- Q plot of forecast residuals; d) Histogram of forecast residuals**

The ACF of the residuals shows no serious violations of model assumption suggesting that this model is good (Ljung-Box Chi-squared statistic = 0.1877, p-value = 0.6648). On the basis of the developed model, the forecasted monthly rainfall along with the 95% confidence intervals is as shown in Figure 5.

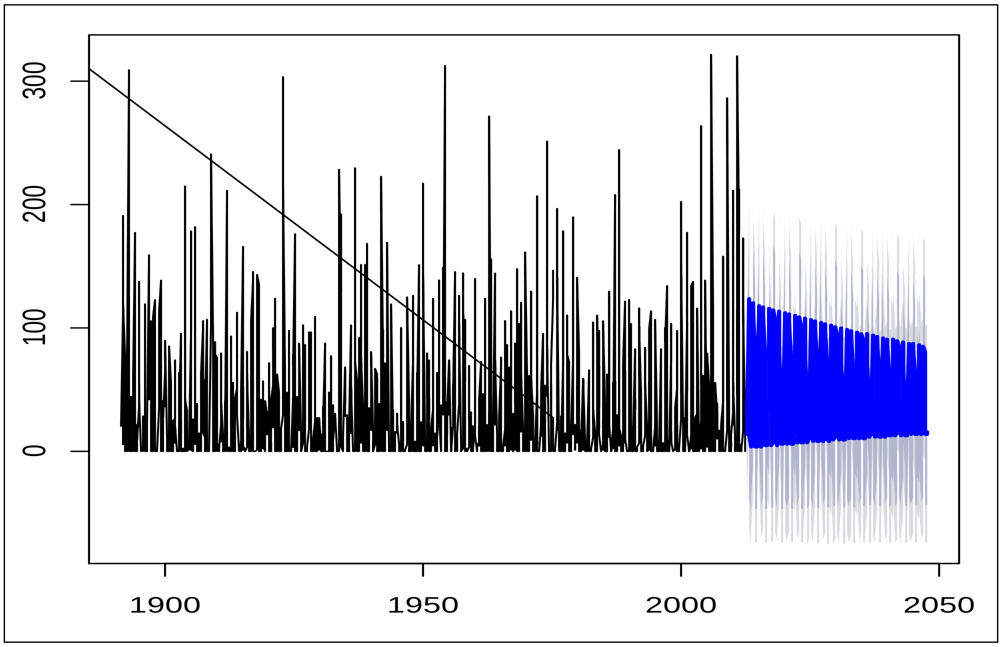

Figure 5 Windhoek Monthly Rainfall forecast with 80% and 95% confidence Interval

An extract of the actual forecast values for together with their corresponding confidence intervals are listed in Table 3.

Table 3 Forecasted Monthly Rainfall with 95% Confidence Intervals

		95% Confidence Interval	
Year	**Point Forecast**	**Lower**	**Upper**
Sep 2046	16.013	0	103.447
Oct 2046	20.880	0	108.494
Nov 2046	37.344	0	124.960
Dec 2046	32.284	0	119.901
Jan 2047	85.604	0	173.221
Feb 2047	81.319	0	168.936
Mar 2047	62.818	0	150.434
Apr 2047	43.441	0	131.056
May 2047	20.196	0	107.056
Jun 2047	14.472	0	102.088
July 2047	14.470	0	102.095
Aug 2047	14.372	0	101.988
Sept 2047	16.238	0	103.855
.			
.			
.			

The forecast values suggest that for instance in the 2046/47 , the winter season monthly rainfall point estimates are around 15mm (June 14.5mm, 95% Confidence Interval:0 - 102.09; July 14.5, 95% Confidence Interval:0 - 102.10 mm; and August 14.3mm, 95% Confidence Interval:0 – 101.99) which can technically be higher than expected. However, the lower 95% confidence limits for the same winter months are zero highlighting the possibility of no rainfall during those periods. The confidence intervals are symmetrical and mathematically include negative values which do not make sense for physical quantities like rainfall which is always greater than or equal to zero. In this case the confidence bands were adjusted to zero in the event of a negative value.

4. Conclusions and Recommendations

Based on the ARIMA modelling of the Windhoek rainfall, despite the seasonal and irregular fluctuations, the mean monthly rainfall levels did not suggest an upward or downward trend over the century. Even though the results indicate constant mean monthly rainfall, the limitation of the results is that the analysis is based on a small spatial area to completely rule out climate change effects.

5. Acknowledgements

The authors acknowledge Namibia Water Corporation (NamWater) and the Namibia Meteorological Services for providing rainfall data and the University of Namibia for funding this research.

6. References

Ansari, H. (2013). Forecasting seasonal and annual rainfall based on non-linear modelling with Gamma test in North of Iran, International Journal of Engineering Practical Research, 2 (1) 16-29.

Abudu, S., Cui, C.L., King, J.P., Abudukadeer, K. (2010). Comparison of performance of statistical models in forecasting monthly stream flow of Kizil River, China. Water Science and Engineering, 3(3) 269–281.

Babu, S.K.K., Karthikeyan, K., Ramanaiah, M.V., Ramanah, D. (2011). Prediction of rain-fall flow time series using Auto-Regressive Models. Advances in Applied Science Research, 2(2) 128–133.

Chattopadhyay, S., Chattopadhyay, G. (2010). Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ARNN. Comptes Rendus Geoscience, 342(2) 100–107.

Collischonn, W., Haas, R., Andreolli, I., and Tucci, C.E.M. (2005). Forecasting river Uruguay flow using rainfall forecasting from a reagional weather predicted model, Journal of hydrology, 305 (1-4) 87-98.

Htike, K. K., & Khalifa, O. O. (2010, May). Rainfall forecasting models using focused time-delay neural networks. In Computer and Communication Engineering (ICCCE), 2010 International Conference on (pp. 1-6). IEEE.

Hung, N.Q.,Babel, M.S.,Weesakul, S. & Tripathi, N.K. (2009). An artificial neural network model for rainfall forecasting in Bangkok, Thailand, Hydrological. Earth. System. Science, 13, 1413-1425.

Jakasa, T., Androsec, I. & Sprcic, P. (2011). Electricity price forecasting –ARIMA model approach, Proceedings of the International Conference on the European Energy market (EEM) 25-27 May 2011, Zagreb, Croatia page 222-225.

Kanna, M., Prabakharan, S., &Ramachandran, P. (2010). Rainfall forecasting using Data Mining Technique, International Journal of Engineering Technology, 2 (6) 397-401.

Mahsin, M., Akhter, Y., & Begum, M. (2012). Modelling rainfall in Dhaka Division of Bangadesh using time series analysis, Journal of mathematical modelling and application, 1 (5) 67-73.

Mauludiyanto, A., Hendrantoro, G., Purnomo, M.H., Ramadhany, T., & Matsushima, A. (2010). ARIMA modeling of tropical rain attenuation on a short28-GHz terrestrial link. IEEE Antennas and Wireless Propagation Letters,9 223–227.

Meher, J., & Jha, R. (2013). Time-series analysis of monthly rainfall data for the Mahanadi River Basin, India. Sciences in Cold and Arid Regions (SCAR), 5(1), 73-84.

Otok, B.W., & Suhartono, F.R. (2009). Development of rainfall forecasting model in Indonesia by using ASTAR, transfer function, and ARIMA methods. European Journal of Scientific Research, 38(3) 386–395.

Rabenja, A. T., Ratiarison, A., & Rabeharisoa, J. M. (2009). Forecasting of the Rainfall and the Discharge of the Namorona River in Vohiparara and FFT Analyses of These Data. In Proceedings, 4th International Conference in High-Energy Physics, Antananarivo, Madagascar (pp. 1-12).

Shamsnia, S.A., Shahidi, N., Liaghat, A., Sarraf, A., & Vahdat, S.F. (2011). Modeling of weather parameters using stochastic methods (ARIMA Model) (Casestudy: Abadeh Region, Iran). International Conference on Environment and Industrial Innovation, 12 282–285.

Turalam, G.A. & Ilalee, M. (2010). Time series analysis of rainfall and temperature interaction in coastal catchments, Journal of mathematics and statistics, 6 (3) 372-380.

Wegner, T. (2010). Applied business statistics: Methods and Excel-based applications. Juta and Company Ltd, South Africa.

World Meteorological Organization (wmo) Extranet (2015) accessed 20 August 2015,| www.wmo.int.