

Seasonal Modelling of Fourier Series with Linear Trend

Iberedem A. Iwok¹

¹ Department of Mathematics/Statistics, University of Port-Harcourt, P.M.B.5323, Port-Harcourt, Rivers State, Nigeria.

Correspondence: Iberedem A. Iwok, Department of Mathematics/Statistics, University of Port-Harcourt, P.M.B.5323, Port-Harcourt, Rivers State, Nigeria. E-mail: ibywok@yahoo.com

Received: July 28, 2016 Accepted: September 6, 2016 Online Published: October 26, 2016

doi:10.5539/ijsp.v5n6p65

URL: <http://dx.doi.org/10.5539/ijsp.v5n6p65>

Abstract

This work was motivated by the need to model a periodic time series function with linear trend. A Fourier series representation with detrended linear function was proposed. In this representation, the time series X_t is expressed as a combination of the linear trend component and a linear combination of s orthogonal trigonometric functions; where s is the number of seasons. The method was applied to a rainfall data and the proposed model was found to give a good fit. Comparative study was carried out with the complete Fourier representation. Diagnostic checks revealed that the proposed method performs better the pure Fourier approach.

Keywords: Fourier series, seasonal model, linear trend, periodogram, spectral density and white noise process.

1. Introduction

The usual autoregressive integrated moving average (ARIMA) models developed by Box and Jenkins (1970) has been extensively used in modelling linear time series. The ARIMA models assume that the current observation depends on weighted previous observations, weighted previous random shocks and the current shock. However, most time series arising in nature do not assume linearity but rather, periodic or seasonal with linear trend. Seasonal time series contain a seasonal phenomenon that repeats itself after a regular period of time. Such phenomena stem from factors such as weather, which affects many business and economic activities, cultural events and graduation ceremonies. Series with seasonal pattern cannot be adequately represented by ARIMA models. To analyze such series, Wold (1974) arranged the series in a two dimensional table according to the season; and the totals and averages were computed. In Wold (1974) representation, a time series is thought to consist of trend-cycle, seasonal and irregular components. To estimate these components, several decompositions are usually involved. Box, Jenkins and Reinsel (2008) made an extension of the Box and Jenkins (1970) ARIMA models to include the seasonal part and is called the seasonal autoregressive integrated moving average (SARIMA) models. Despite these efforts, the models do not adequately represent most periodic series.

A better procedure extensively used for modelling periodic time series is the Fourier analysis. This method represent the time series by a set of elementary functions called basis such that all functions under study can be written as linear combinations of the elementary functions in the basis. These elementary functions involve the sine and cosine functions or complex exponentials. The Fourier series approach describes the fluctuation of time series in terms of sinusoidal behaviour at various frequencies. Despite the wider acceptability of the method, however, Fourier approach still suffers some set backs. One major problem associated with it is the cumbersomeness in Fourier representation and non inclusion of trend component. As will be seen in the methodology, the inconveniences in representing the time series is enormous if we are to include all the terms required in Fourier series. This cannot go well with a series of large sample size because representing all the terms will consume several pages and can be boring to both the researcher and the reader. Hence, there is need to shorten the number of terms in the Fourier expression and give a summarized representation that adequately describe the time series. This is the intent of this work. As earlier stated, seasonal variations in time series can be caused by climatic factor and we are going to use rainfall data in our illustration.

2. Literature Review

The need for accurate rainfall prediction is necessary when considering the importance in which such information would give for river control, reservoir operation, forestry interests, flood mitigation, etc. Due to the numerous benefits of rainfall modelling and prediction, studies on rainfall analysis have been on the fore front in the research world.

Afshar, Joshua, Buckman, and Samuel (2014) modelled rainfall data using ARIMA model and artificial neural networks (ANNs). The ARIMA was found to give discouraging result. However, with application of artificial neural network to the

ARIMA component, the combined model was found to be adequate and forecast were generated.

Cenis (1989) studied temperature in solarized soil using Fourier analysis. He obtained the daily maximum and minimum temperatures at two dept on daily basis for three summer periods. He used the values to fit sinusoidal equations which accounted for 93% variation. The variation and the hourly mean differences between the measured temperatures were calculated. The analysis gave an overall encouraging result.

Serangelo, Ferrari and De Luca (2011) applied non-homogeneous Poisson process to examine the seasonal effects of daily rainfall. The modelling process involved the partitioning of observed daily rainfall data into calibration periods for the estimation of parameters. Though the validation period for checking the occurrence process changed; the model which was applied to the set of rain gauges placed at different geographical areas was shown to provide good fit.

Falahah and Suorapto (2010) carried out research on rainfall data using analytic factor method. The data was obtained from 50 weather stations for a period of 30 years. The result was plotted on pattern factors to reveal dominant factor for each region and inspection period. The method explained factors that influence rainfall in Indonesia and the reasons for having relatively high humidity in one area than the other.

Necholas, Mahmood and Hazan (2013) modelled rainfall data amounts for agriculture planning using gamma distribution models. Daily rainfall data of two stations having two different mean annual rainfalls were analyzed. Generalized linear models were used to fit smooth regression curves. The mean amount of rain per rainy day was computed using the estimates of parameters of the model for each day of the season. The adequacy of the fitted model was check by the analysis of deviance residuals and was found to be satisfactory. Fourier approach was employed for comparative study. It was discovered that though reasonable results were obtained, Fourier analysis was time consuming and boring. However, Fourier series was found suitable in fitting gamma distribution for the determination of mean rain per rainy day.

Zakaria (2013) conducted a study on periodic and stochastic modelling of monthly rainfall and the periodicities were determined. Stochastic components were estimated using the auto-regressive model approach. Residuals obtained from the model were shown to follow a white noise process; thus indicating the adequacy of the fitted model.

Beatrice, Nasser, Afshar, Selaman and Fahmi (2014) analyzed data from eight rain gauge stations. Annual rainfall data for 27 years were computed with the Fourier series equation. The result was compared with that obtained from harmonic series models. It was discovered that both models were capable of describing rainfall pattern and were able to provide reasonable relationship between the simulated and the observed data.

Akpanta, Okorie and Okoye (2015) adopted SARIMA modelling of the frequency approach in analyzing monthly rainfall data in Umuahia. Probability time series approach was considered. The original data plotted showed seasonality which was removed by differencing. After subjecting the model to diagnostic checks, SARIMA (0,0,0)(0,1,1)₁₂ was found to fit the data well and was used for prediction.

3. Methodology

In this method, a periodic time series is first observed whether it contains a linear trend or not. Visual inspection of the raw data plot can reveal this pattern. Assuming a linear trend is detected, a linear regression model of the form

$$Y_t = \beta_0 + \beta_1 t + e_t \quad (1)$$

is first fitted to the data;

where Y_t is the observed time series, t is the time points ($t = 1, 2, \dots, n$), n is the number of observations, β_0 and β_1 are the regression parameters, and e_t is the error component.

Fitting the above model (1) to the data Y_t , we can obtain the estimate of the error component

$$\hat{e}_t = \hat{Y}_t - \hat{\beta}_0 - \hat{\beta}_1 t$$

which can be tested for randomness or white noise.

After obtaining the trend equation (i.e. $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 t$), the main series Y_t is detrended by the expression

$$y_t = Y_t - \hat{Y}_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 t \quad (2)$$

The resulting series y_t is then used to fit seasonal model using Fourier representation.

3.1 Fourier Series Representation of the Time Series y_t

Given a time series of n observations, the Fourier representation is the set of q orthogonal trigonometric functions shown below:

$$y_t = \sum_{i=1}^q (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + e_t \quad (3)$$

estimated by

$$\hat{y}_t = \sum_{i=1}^q (a_i \cos 2\pi f_i t + b_i \sin 2\pi f_i t) \tag{4}$$

where $q = n/2$, $a_i = \frac{2}{n} \sum_{t=1}^n y_t \cos 2\pi f_i t$, $b_i = \frac{2}{n} \sum_{t=1}^n y_t \sin 2\pi f_i t$,

$e_t \sim NIID(0, \sigma^2)$; period = $p_i = n/i$ and $f_i = i/n$ is the i^{th} harmonic of the fundamental frequency $1/n$.

3.2 The Periodogram

The periodogram is defined as the function of intensities $I(f_i)$ at frequency $f_i = i/n$ and is given as

$$I(f_i) = \frac{n}{2} (a_i^2 + b_i^2) \quad ; \quad i = 1, 2, \dots, q.$$

Periodogram is the plot of the intensities against the frequencies or periods. The periodogram $I(f_i)$ is simply the sum of squares associated with the pair of coefficients (a_i, b_i) and hence with the frequency f_i or period p_i . That is,

$$\sum_{t=1}^n (y_t - \bar{y})^2 = \sum_{i=1}^{n/2} I(f_i).$$

In the context at hand, the periodogram is used to determine the seasonality or periodicity of a time series. This is usually indicated by the largest peak in the periodogram plot.

3.3 The Spectrum

The sample spectrum is obtained by allowing the frequency f to vary continuously in the range 0 to 0.5 cycle so that the periodogram can be re-defined as

$$I(f) = \frac{n}{2} (a_j^2 + b_j^2) \quad ; \quad 0 \leq f \leq 0.5.$$

The function $I(f)$ is called the spectrum.

3.4 Autocorrelation Function

This is the plot of autocorrelation at lag k (ρ_k) versus k .

3.5 Spectral Density Function

Spectral density is the Fourier transform of the auto-correlation function and is estimated by

$$g(f) = 2[1 + \sum_{k=1}^{\infty} \rho_k \cos(2\pi f k)] \quad ; \quad 0 \leq f \leq 0.5$$

where ρ_k is the autocorrelation at lag k . The spectral density performs the same function as the periodogram. The period or seasonality of a time series is obtained at where the spectral density is maximum.

3.6 White Noise Process

A process $\{\varepsilon_t\}$ is said to be a white noise process with mean 0 and variance σ_ε^2 written $\{\varepsilon_t\} \sim WN(0, \sigma_\varepsilon^2)$, if it is a sequence of uncorrelated random variables from a fixed distribution.

3.7 The Seasonal Fourier Representation

Rather than fitting the entire Fourier series expression in equation (3), we fit only the Fourier terms up to the season detected by the periodogram. That is, suppose the season determined by the periodogram in the detrended series y_t is s , then equation (3) reduces to

$$y_t = \sum_{i=1}^s (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + \varepsilon_t \tag{5}$$

$$\Rightarrow \hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \sum_{i=1}^s (a_i \cos 2\pi f_i t + b_i \sin 2\pi f_i t) \tag{6}$$

and

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t \tag{7}$$

Comparatively, the expression (5) is less cumbersome in carrying out analysis than the complete Fourier form expressed in equation (3). The model (5) can be fitted to any periodic or seasonal data and the estimated residuals $\hat{\varepsilon}_t$ obtained from (7) can be tested for white noise to determine whether the model is adequate or not.

4. Data Analysis and Result

The data used for this work is the average monthly rainfall data (Y_t) in Calabar, Nigeria between 2005-2015 (Source: www.cbn.gov.ng); and the analysis is carried out using Minitab and gretl software.

4.1 Complete Fourier Series Model

Fitting the full Fourier series in equation (3) where $q = \frac{120}{2} = 60$ result in a residual variance of 11.23 and the Fourier coefficients are displayed in Appendix C. The residual autocorrelation function is displayed in figure 5. Clearly, there is a significant spike at lag 12 ($\rho_k = -0.36$). This shows that the residuals are correlated (at lag 12) and hence do not follow a white noise process. The actual and estimate values plots displayed in figure 6 shows a low correlation between these values. Thus, the full Fourier series, despite it cumbersome nature does not fit adequately to the data.

4.2 The Proposed Approach

4.2.1 Seasonality and the Estimated Trend

The raw data plot in figure 1 clearly shows the existence of seasonality and trend. This is indicated by the periodic pattern and upward movement of the graph. Fitting the trend equation gives the Minitab output in table 1 below.

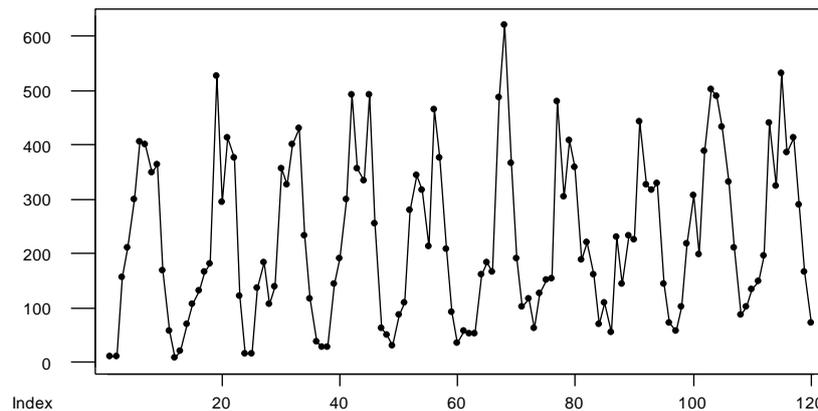


Figure 1. Raw data plot of the series Y_t

Table 1. Minitab Output for the Trend equation

Predictor	Coef	StDev	T	P
Constant	185.28	26.96	6.87	0.000
t	0.6639	0.3898	1.70	0.031

The regression equation is $Y_t = 185 + 0.6639 t$

The p -values in table 1 shows that both the constant term and the coefficient of t are significant. Thus, the estimated trend equation is

$$\widehat{Y}_t = 185.28 + 0.6639t$$

Next, we obtain the new series from equation (2) as

$$y_t = Y_t - \widehat{Y}_t = Y_t - 185.28 - 0.6639t$$

The new series y_t now becomes our working data.

4.2.2 The Peridogram and Spectral Density of y_t

The periodogram analysis was conducted using the gretl software. The periodogram is displayed in figure 2. The values for the spectral densities, periods and frequencies are equally displayed in appendix A. From the table of appendix A and figure 2; it is observed that the periodogram is dominated by a very large peak at scaled frequency, $f_i^* = i = 10$ ($f_i = i/n = 10/120 = 0.0833$). This is indicated by the largest spectral density of 16.397 shown in bold figures. This density and frequency correspond to a *period* or *season*, $s = n/i = 120/10 = 12$ months. This indicates a 12 – month cycle.

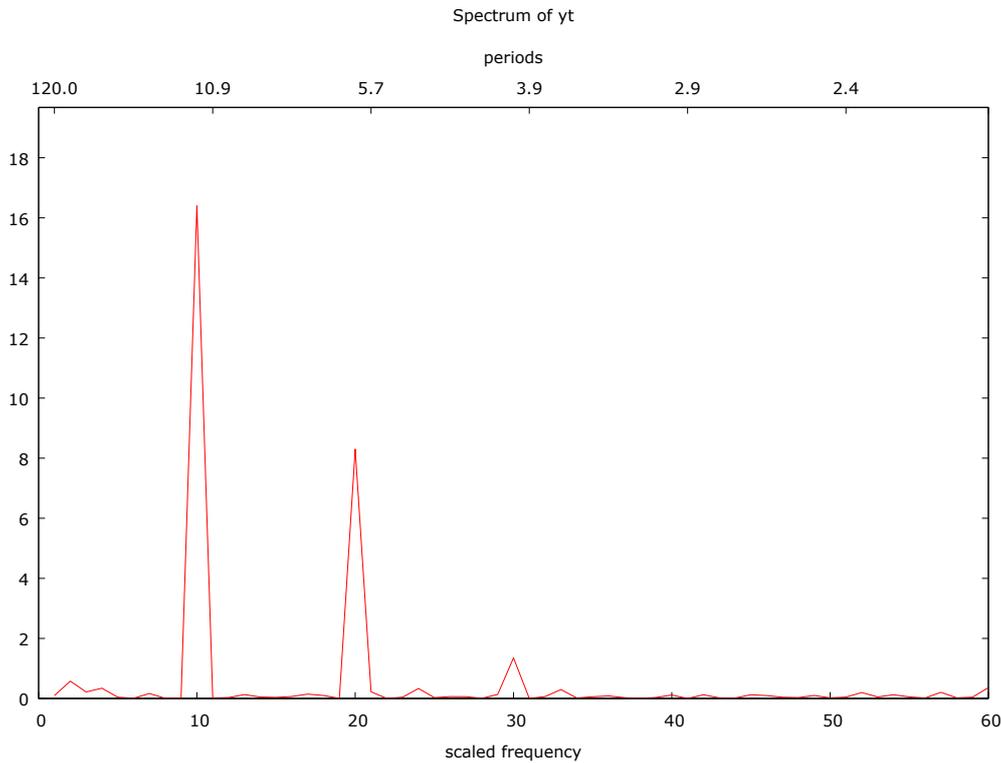


Figure 2. Periodogram plot of y_t

4.2.3 The Seasonal Fourier Representation of y_t

Rather than use $q = 60$ in equation (3), we reduce the Fourier components to the number of seasons, $s = q = 12$ and apply equation (5). That is,

$$y_t = \sum_{i=1}^{12} (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + \varepsilon_t \tag{8}$$

Equivalently, equation (8) can be expressed as

$$y_t = \sum_{i=1}^{12} (\alpha_i \cos \omega_i t + \beta_i \sin \omega_i t) + \varepsilon_t \tag{9}$$

where, $\omega_i = 2\pi f_i$.

Subjecting equation (9) to regression analysis give the parameter estimates displayed in Appendix B. In appendix B, some coefficients of the variables are not statistically significant (i.e. their p values are less than 0.05) and are therefore excluded in the overall equation.

Hence, from (6), the resulting estimated model is

$$\begin{aligned} \hat{Y}_t = & 185.28 + 0.6639t - 0.621 \sin \omega_1 t - 0.314 \sin \omega_2 t - 0.129 \sin \omega_3 t - 0.105 \sin \omega_4 t \\ & - 0.085 \sin \omega_5 t - 0.081 \cos \omega_8 t - 0.155 \cos \omega_9 t - 0.074 \sin \omega_{10} t \\ & + 0.142 \cos \omega_{11} t - 0.206 \cos \omega_{11} t + 0.094 \cos \omega_{12} t \end{aligned} \tag{10}$$

The residual $\hat{\varepsilon}_t$ of the fitted is obtained from

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t$$

5. Diagnosis

We present here two diagnostic checks to ensure that the proposed model (10) fitted to the data is adequate.

5.1 Actual and Estimate Plots

The overlaid plots of the actual values (Y_t) and the estimated values (\hat{Y}_t) is displayed in figure 3. The two superimposed plots move together in the same direction indicating closeness and strong correlation between the values of the two variables. This shows that the model is adequate.

Time Series Plot of Y_t and \hat{Y}_t

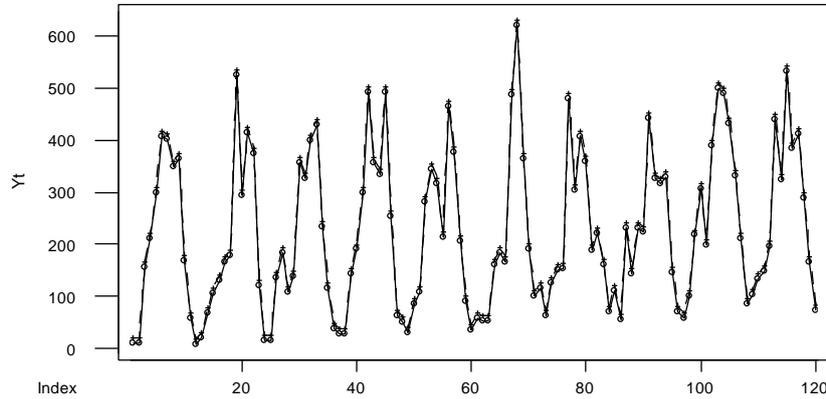


Figure 3. Actual and estimate plot of Y_t

5.2 Residual Variance, Autocorrelation and White Noise

The residual variance of the fitted model is 4.78. This is significantly smaller than the 11.23 obtained from fitting the full Fourier series. The residual autocorrelation function is displayed in figure 4 and it shows that there is no significant autocorrelation of the residuals. That is, all autocorrelations at all lags are within the range $\pm 2/\sqrt{n}$ (as indicated by the two red lines in figure 4). This means the residuals of the fitted model are not serially correlated. In more precise terms, the residuals follow a white noise process. Hence, the model is adequate.

Autocorrelation Function for Res(yt)

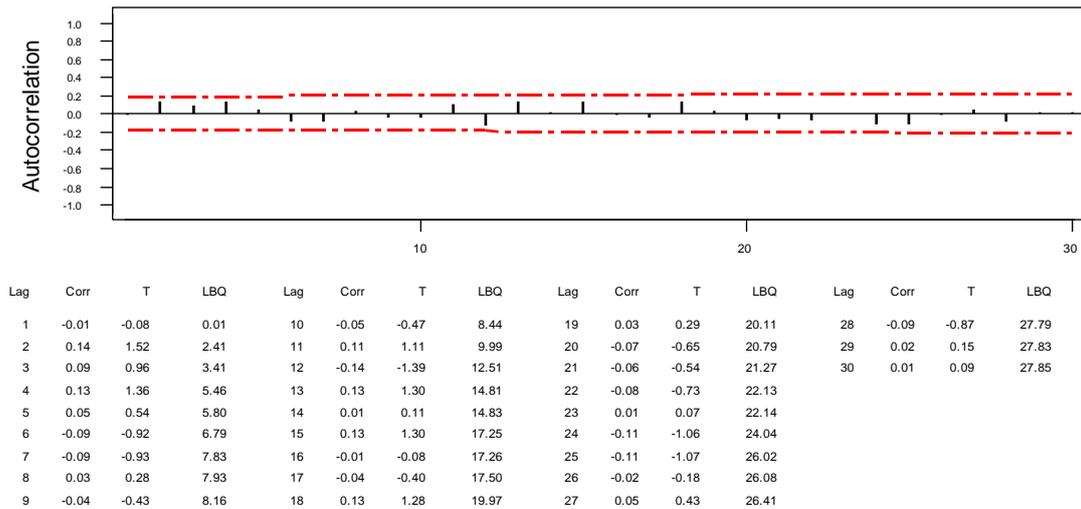


Figure 4. Residual Autocorrelation function plot of $\hat{\epsilon}_t$

6. Discussion and Conclusion

It has been noted that the Fourier series model can only be applied to periodic series that are stationary in mean. If the series contain trend, however, special technique is required for the modelling process. Perhaps, this constituted the problem of Afshar *et al* (2014) that made them to obtain a discouraging result by applying ARIMA model to a periodic data without considering the trend. Besides, as noted by Necholas *et al* (2013), Fourier series modelling is time consuming and boring because of the large number of Fourier coefficients involved. In this work, however, these problems have been addressed. As clearly demonstrated in this method, the trend component of a periodic series is taken care of. Also, setting $q = s$ has reduced the computational burden by 80% and has given adequate Fourier representation as confirmed by the diagnostic checks. It is believed that this work has opened another possibility of addressing periodic functions.

Autocorrelation Function for RES(Yt*)

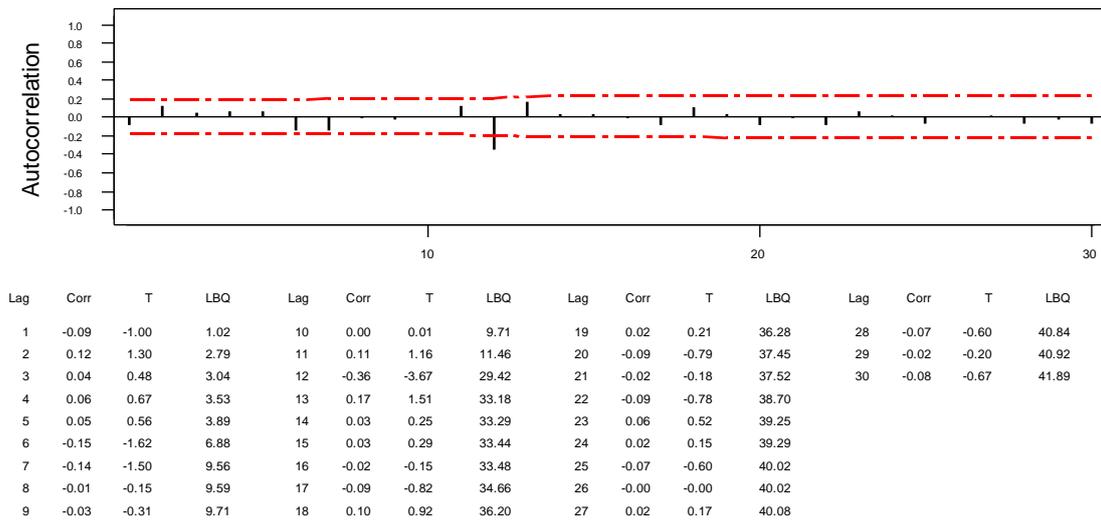


Figure 5. Residual Autocorrelation function plot of $\hat{\epsilon}_t$

Time Series Plot of Yt and Fit(Yt*)

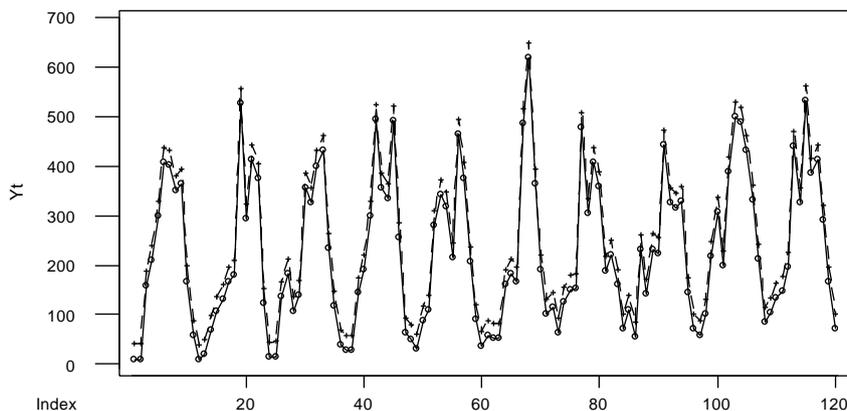


Figure 6. Actual and estimate plot of Y_t^*

References

Afshar, B. N., Joshua, M. A., Buckman, A., & Samuel, T. (2014). Time series modeling of rainfall in new Juaben municipality of the Eastern region of Ghana. *Special issue on contemporary research in business and social science*, 4(7), 21-28.

Akpanta, C. A., Okorie, I. E., & Okoye, N. N. (2015). SARIMA modelling of the frequency of monthly rainfall in Umuahia, Abia state of Nigeria. *American journal of mathematics and statistics*, 5, 82-87.

Beatrice, C. B., Nasser, R., Afshar, K., Selaman, O., & Fahmi, H. (2014). Application of mathematical modelling in rainfall forecast: A case study in SGS. *Sarawak basin. International Journal of Research in Engineering and Technology*, 3, 2321-7308.

Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis; forecasting and control*. Holden Day. San Francisco, California.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). *Time Series Analysis; Forecasting and Control*. John Willey & Sons.

Cenis, J. L. (1989). Temperature evaluation in solarized soils by Fourier analysis. *Photopathology*, 79, 506-510. <http://dx.doi.org/10.1094/Phyto-79-506>

Falahah, M., & Suprpto, S. (2010). Interpretation of rainfall data using analysis factor method. *Proceeding of third*

international conference on mathematics and natural sciences, Indonesia (pp. 1288-1293).

Necholas, Z., Mahmood, Z., & Hazan, Y. (2013). Modeling the daily rainfall amounts of north-west Pakistan for agricultural planning. *Sarhad J. Agric*, 30, 5-23.

Serangelo, B. A., Ferrari, H. K., & DeLuca, T. (2011), Occurrence analysis of daily rainfall through non-homogeneous Poisson process. *Nat. Hazards Earth Syst. Sci.*, 11,1657-2011. <http://dx.doi.org/10.5194/nhess-11-1657-2011>

Wold, H. O. A. (1974). *A study in the analysis of stationary time series*. Almqvist and Wiksell, Uppsala.

Zakaria, A. (2013). A study of periodic and stochastic modelling of monthly rainfall from Purajaya station. *Asian Transactions on Engineering*, 1(3), 1-7.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).