العدد الحادي عشر
تاريخ الإصدار: 2 – 9 – 2019
www.ajsp.net
ISSN: 2663-5798

"Appling Autoregressive, Fractionally-Integrated, Moving Average Models of ARFIMA (p, d, q) Order for Daily Minimum Electric Load at West Tripoli Electricity Station in Libya"

Researchers Name:

Maryouma E Enaami
Ma.enaami@uot.edu.ly

Rida M khaga
r.khaga@uot.edu.ly

Mustafa A Almahmodi
Mustafa.almahmoudi@gmail.com

Fauzia A taweab
fauziataweab@yahoo.com

Department Of Statistics, Faculty of Science, Tripoli University, Libya

العدد الحادي عشر
تاريخ الإصدار: 2 – 9 – 2019
ISSN: 2663-5798
www.ajsp.net

Abstract

In this paper, we report on our investigation of the long memory of the Daily Minimum Electric Load (DMEL) at West Tripoli Electricity Station as recorded by the Electricity Company in Libya. We fitted an Autoregressive, Fractionally-Integrated, Moving-Average (ARFIMA) Model to the measured loads using 361 daily records covering the period of almost one year, extending from 5 January 2008 to 31 December 2008. The results show that the time series is of the long memory type and that it can become stationary with fractional differencing. After performing fractional differencing and determining the number of lags of the autoregressive and moving average (ARMA) components, the long memory ARFIMA (3, 0.499, 3) model was used to fit the data. Even though this model fit the data well, its forecasts were infected by the swing in the data. We estimated the parameters of the model and used those estimates to forecast 20 out-of-sample data points. In light of the forecasting results of the model, we concluded that the ARFIMA is a great model in this regard.

Keywords: Time Series, ARFIMA, Short-Range Dependence (SRD), Long-Range Dependence (LRD).

1. Introduction

In time series modeling, the analysis usually comes to the classic Box and Jekin's method for time series models, that was developed in the 1970s, which can only capture short-range dependence (SRD). The SRD can be characterized by an exponential decay of the autocorrelation function, $p(k)$, or described as the case when the sum of the values which $p(k)$ assumes over all lags is finite (Brockwell and Davis, 1991). On the other hand, long-range dependence (LRD) can be considered as the case when the current observations correlate significantly with the observations that are farther away in time (Baillie, 1996).

Classic models describing SRD, such as the Autoregressive, Integrated, Moving Average (ARIMA) model, can not accurately describe the LRD. As a result, vrious models have been suggested to overcome this limitation of these models. One example is the model proposed by Erfani and Samimi (2009). Another example is the well-known autoregressive, fractionally-integrated, moving average (ARFIMA) model. In this respect, this paper investigated appropriateness of the ARFIMA model as a solution to the persistent non-stationarity of time series in the long-term data.

The ARFIMA model was introduced by Granger and Joyeux (1980) and Hosking (1981). This model has proved to be more successful than the autoregressive, integrated, moving average (ARIMA) model. Within this context, there is bulky literature so far that deliberates on the contradictions between the ARFIMA and ARIMA models in time series forecasting. For instance, Baillie and Chung (2002) found that the ARFIMA model is superior to the Autoregressive Moving Average (ARMA) model and remarkably more successful than it in predicting time series data.

In view of the foregoing discussions, this study examined viability of both the ARIMA and AFRIMA models for forecasting the Daily Minimum Electric Load (DMEL) at West Tripoli Electricity Station in Libya and compared the levels of performance of these two models to determine the one of which that can more accurately predict the DMEL at this station.

The rest of the article is organized as follows. Section Two provides a brief review and description of LRD and SRD. Section Three introduces the AFRIMA model. Then, Section Four discusses model performance evaluation and the performance diagnostics. Section Five describes the study data and illustrates the data modeling process. Thereafter, the forecasting results are presented and discussed in Section Six. Then, Section Seven concludes this paper with conclusions drawn from this study.

2. Long-Range versus Short-Range Dependence

Long-range dependence (LRD) can be described as the case when the current observations correlate significantly with the observations that are farther away in time. One formal definition of the LRD stationary process states that the sum of the values of $p(k)$ of the stationary process is infinite (i.e., $\sum_{k=0}^{\infty} |\rho(k)| = \infty$) over all lags. This implies that $p(k)$ slowly decays to zero at a very low speed of convergence. The SRD, on the other hand, can be characterized by an exponential decay of $p(k)$ or described as the case when the sum of the values of $p(k)$ of the stationary process is finite (i.e. $\sum_{k=0}^{\infty} |\rho(k)| < \infty$) over all lags.

3. The Autoregressive, Fractionally-Integrated, Moving Average (ARFIMA) Model

A stochastic process, x_τ, is called an ARIMA process of order **(p, d, q),** $\boldsymbol{\phi_p(B)\, \Delta^d x_t} = \boldsymbol{\theta_q(B)\, a_t}$, $\boldsymbol{d}$ = 0, 1, 2, …., where $\boldsymbol{\Delta} = (\mathbf{1} - \boldsymbol{B})$ is the difference operator, and $\boldsymbol{\phi_p(B)}$, and $\boldsymbol{\theta_q(B)}$ are the p^{th} and q^{th} degree polynomials, respectively, where **d** ∈ (-0.5, 0.5).

The ARFIMA(p, d, q) processes are widely used in modeling the LRD time series, where *p* is the autoregressive order, *q* is the moving average order, and *d* is the level of differencing (Liu, Chen and Zhang, 2017). The larger the value of *d*, the more closely it approximates a simple integrated series, and the more likely it is to better approximate a general integrated series than a mixed-fractional difference and an ARMA model. Therefore, Whittle's (1953) method has good performance in estimation of the fractional differencing parameter, *d*, according to Velasco and Robinson (2000) and Shimotsu and Phillips (2006).

4. Model Performance Evaluation and Perormance Diagnostics

Before interpretation and use of an ARFIMA model, we have to check whether the model is specified correctly or not. In the present study, the following tests were applied to the model residuals:

(i)- Test of the Mean of the Residuals

This test is a modified, two-sided test. Its null and alternative hypotheses are:

$$H_0 = E(a_t) = 0$$
$$H_A : E(a_t) \neq 0$$

(ii)- *Autocorrelation and Partial Autocorrelation of the Residuals*

In testing autocorrelation and partial autocorrelation of the residuals, the null and alternative hypotheses are the following:

$$H_0 : p_k = 0, k=1$$
$$H_A : p_k \neq 0$$

(iii)- *Test for Normality of the Residuals*

The most common approach to determining whether or not distribution of a variable does, or does not, follow the normal distribution is to draw and check the normal quantile-quantile plot (the Q-Q plot) or the scatter plot of the standardized empirical quantiles of x_t against the quantiles of a standard normal random variable. Another option is to run the diagnostic tests of serial correlation (the Box-Pierce test), residual autocorrelation (the Ljung-Box test), normality (the Kolmogorov-Smirnov test), and stationarity (Kwiatkowski et al.'s (1992) test (KPSS test)).

5. Data Analysis

Data for this study were DMEL time series data for West Tripoli Electricity Station in Libya that were compiled by the Electricity Company. The dataset comprised 361 measurements of DMEL for a period of nearly one year, extending from 5 January 2008 to 31 December 2008. A plot of these load data as a function of time is provided by Figure 1.

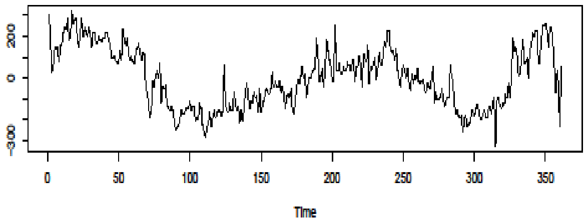


Figure 1: Daily minimum electric load (DMEL during the study period (5 January 2008 - 31 December 2008)

5.1 The DMEL0M Dataset and its Graphical Properties

Figures 2 and 3 present the results obtained from the autocorrelation function (ACF) and the partial autocorrelation function (PACF) in the preliminary analysis of the DMEL data with zero mean, that is, DMEL0M. The ACF in in Figure 2 indicates that the DMELs at West Tripoli Electricity Station decay at a hyperbolic rate, i.e., slower than the short memory time series data. Accordingly, the researchers conclude that the time series demonstrates evidence of long memory.

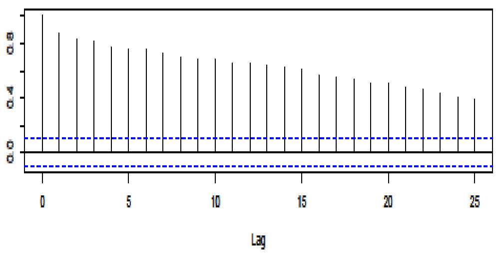


Figure 2: The autocorrelation function (ACF) of the DMEL data series with zero mean (DMEL0M)

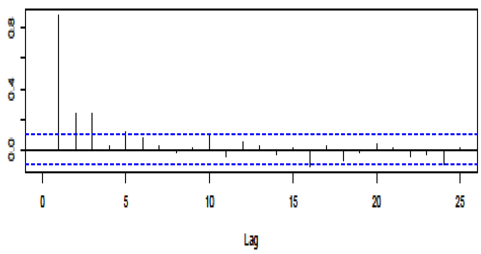


Figure 3: The partial autocorrelation function (PACF) of the DMEL data series with zero mean (DMEL0M)

5.2 The Unit Root Tests of the DMEL0M Series

We tested the unit root of the DMEL0M data based on Phillips and Perron's (1988) test (PP), the Augmented Dickey and Fuller's (1981) test (ADF), and Kwiatkowski et al.'s (1992) test (KPSS). The results of these three tests are listed in Table 1. Since the stationarity assumption was not met we considered applying necessary procedures such as differencing and fractional integration to the data.

Table 1: Results of the PP, ADF, and KPSS tests of the DMEL0M time series

Test	Value of Statistic	Truncation lag parameter	*p*	Decision
PP (1)	-29.4927	5	0.01	Stationary
ADF (2)	-2.2397	7	0.4757	Non-stationary
KPSS (3)	0.7541	4	0.01	Non-stationary

(1) PP: Phillips and Perron's (1988) test.
(2) ADF: The Augmented Dickey and Fuller's (1981) test.
(3) KPSS: Kwiatkowski et al.'s (1992) test.

5.3 Establishment of the ARFIMA model

The ACF of the DMEL0M series introduced by Figure 2 indicates that this time series demonstrates evidence of long memory. Therefore, we used the fractionally-integrated parts according to Velasco and Robinson (2000) as shown in in Figure 4, with a *d* value of 0.4999 (Whittle estimator).

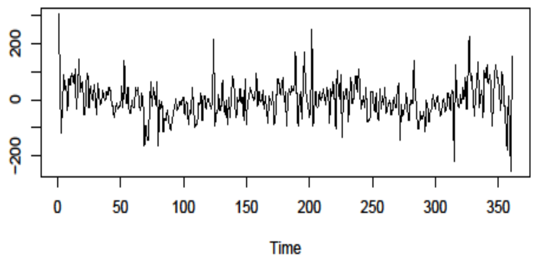


Figure 4: Time plot of the fractional-differencing time series

The results of the ACF and PACF of the fractionally-differenced time series are depicted in figures 5 and 6. Similar unit root tests were performed on these series. The results produced by the three aforementioned tests (Table 2) pointed out that this series is stationary.

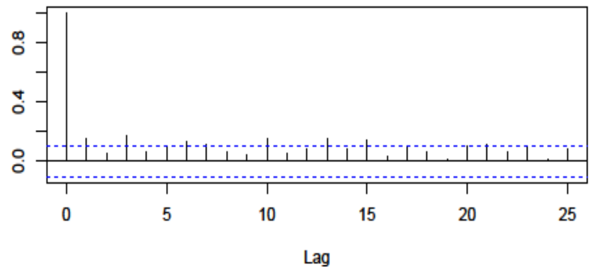


Figure 5: The ACF of the fractionally-differenced time series

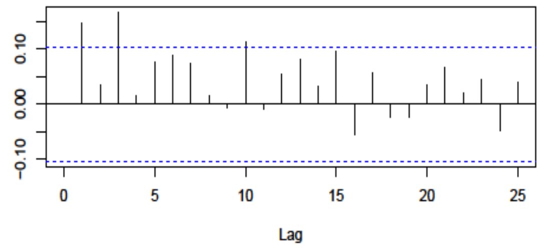


Figure 6: The PACF of the fractionally-differenced time series

Table 2: Results of the PP, ADF, and KPSS tests for the fractionally-differenced time series

Test	Value of Statistic	Truncation lag parameter	*p*	Decision
PP (1)	-353.2034	5	0.01	Stationary
ADF (2)	- 4.5057	7	0.01	Stationary
KPSS (3)	0.2284	4	0.1	Stationary

(1) PP: Phillips and Perron's (1988) test.
(2) ADF: The Augmented Dickey and Fuller's (1981) test.
(3) KPSS: Kwiatkowski et al.'s (1992) test.

5.3.1 Identification of the Optimum ARFIMA Model

The optimization routine of the log-likelihood function shows that the optimal ARFIMA model for the fractionally-differenced time series is ARFIMA (3, 0.4999, 3) since it has the lowest value of the Akaike Information Criterion (AIC) (Akaike (1973)). Accordingly, the results (Table 3) show that the ARFIMA (3, 0.4999, 3) model is the best model.

Table 3: The AIC and σ^2 values for the different ARFIMA models

Model	AIC (1)	Estimated σ^2
***ARFIMA*(1, 0.4999, 1)**	**4021.72**	**3964**
***ARFIMA*(1, 0.4999, 3)**	**4022.03**	**3923**
***ARFIMA*(2, 0.4999, 2)**	**4022.03**	**3924**
***ARFIMA*(3, 0.4999, 3)**	**4019.72**	**3850**

AIC: The Akaike Information Criterion.

Before we could draw any conclusion, it was necessary for us to examine the residuals of all models to see if they did, or did not, pass the four aforementioned diagnostic tests. The outcomes of the diagnostic tests (Table 4) disclose that all the studied ARFIMA models passed the diagnostic tests of serial correlation, residual autocorrelation, normality, and stationarity. We, therefore, proceeded to estimate the parameters of the selected model, namely, the ARFIMA (3, 0.4999, 3) model.

Table 4: Results of the diagnostic tests for the different ARFIMA models

Model \ Test	Box-Pierce	Ljung-Box	Kolmogorov-Smirnov	KPSS (1)
	P			
***ARFIMA*(1, 0.4999, 1)**	**0.7524**	**0.4839**	**0.9890**	**0.1**
***ARFIMA*(1, 0.4999, 3)**	**0.7691**	**0.7579**	**0.9779**	**0.1**
***ARFIMA*(2, 0.4999, 2)**	**0.4335**	**0.7668**	**0.9945**	**0.1**
***ARFIMA*(3, 0.4999, 3)**	**0.9967**	**0.9079**	**0.9779**	**0.1**

(1) KPSS: Kwiatkowski et al.'s (1992) test.

5.3.2 Estimating the Parameters of the ARFIMA (3, 0.499, 3) Model

The estimated values of the parameters of the ARFIMA (3,0.4999,3) model are displayed in Table 5. Having fitted a model to the fractionally-differenced DMEL0M time series, we checked the model for adequacy. The results (Table 5) indicate that all the model parameters are significant.

Table 5: The estimated values of the ARFIMA (3, 0.4999, 3) model parameters

Parameter	Coefficient	S.E (1)	Student's *t*	*p*
ϕ_1	- 0.6572	0.1007	-20.46	0.000
ϕ_2	0.6377	0.0430	15.49	0.000
ϕ_3	0.9024	0.0968	31.39	0.000
θ_1	0.7745	0.0876	-29.28	0.000
θ_2	- 0.5457	0.0726	9.37	0.000
θ_3	- 0.8421	0.0950	129.83	0.000

(1) S.E: Standard error of the estimate (ε).

5.3.3 Checking the ARFIMA (3, 0.499, 3) Model

Figures 7, 8, and 9 depict the residual time, the ACF, and the PACF of the fitted ARFIMA (3, 0.4999, 3) model, respectively, of the fractionally-differenced DMEL0M time series. A study

ISSN: 2663-5798

of the ACF and the PACF with 95% confidence intervals for the residuals points out that almost all correlations fall inside the corresponding 95% confidence intervals. The plots show that there is no serial correlations among the residuals of the series, hence suggesting stationarity and independence of the data. Consequently, the ARFIMA (3, 0.4999, 3) model is adequate and good.

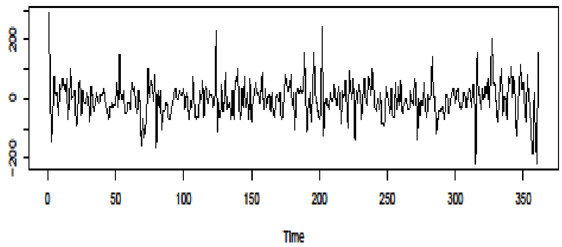
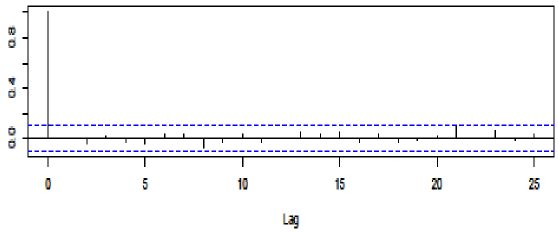
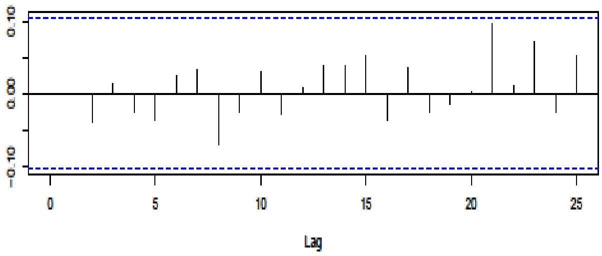


Figure 7: Time plot of the residuals for the ARFIMA (3, 0.4999, 3) model





Figures 8: The ACF of the residuals of the ARFIMA (3, 0.4999, 3) model

Figures 9: The PACF of the residuals of the ARFIMA (3, 0.4999, 3) model

Based on the results of the diagnostic tests shown in Table 6, all the tests agree on that the residuals are normally distributed. There also exist alternative ways for graphically investigating the normality of distribution of variables. For example, the QQ-plot (e.g., Figure 10) is a plot of the quantiles of the variable under consideration against the quantiles of a normal distribution. Another graphical way serving the same purpose is the histogram plot of the residual values against the normal distribution (e.g., Figure 11). The histogram of a normally-distributed variable should be symmetric.

ISSN: 2663-5798

Table 6: Diagnostic tests of residuals for the ARFIMA (3, 0.4999, 3) model

Test	Aspect tested	Test statistic	*p*	Decision
Box-Pierce	Serial Correlation	0.0000	0.9967	Reject
Ljung-Box	Resid Correlation	4.7385	0.9079	Reject
Kolmogorov-Smirnov	Normality	0.5125	0.9779	Accept
KPSS (1)	Stationarity	0.0647	0.1	Accept
t	Zero mean	- 0.1656	0.8686	Accept

(1) KPSS: Kwiatkowski et al.'s (1992) test.

Three different plots (Figure 12) were drawn to analyze the behavior of the residuals of the fitted ARFIMA (3, 0.4999, 3) model and check if the residuals are white noise or not. The results of the tests support that the model is adequate at the level of significance (α) of 0.05 as all the associated *p* values are lower than 0.05. The model was then used to forecast future values of the DMEL series.

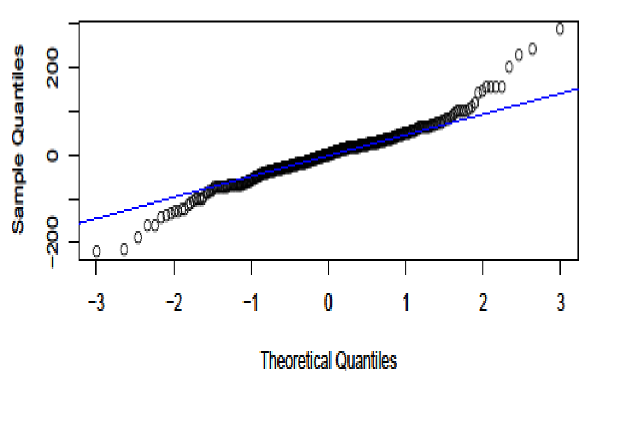


Figure 10: A Q-Q plot of the residuals of the ARFIMA (3, 0.4999, 3) model

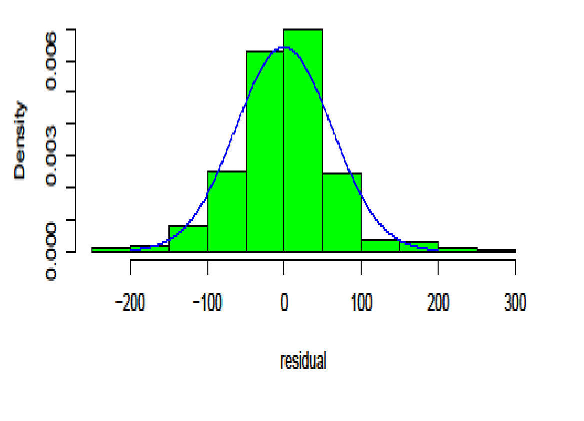


Figure 11: Histogram of the residuals of the ARFIMA (3, 0.4999, 3) model

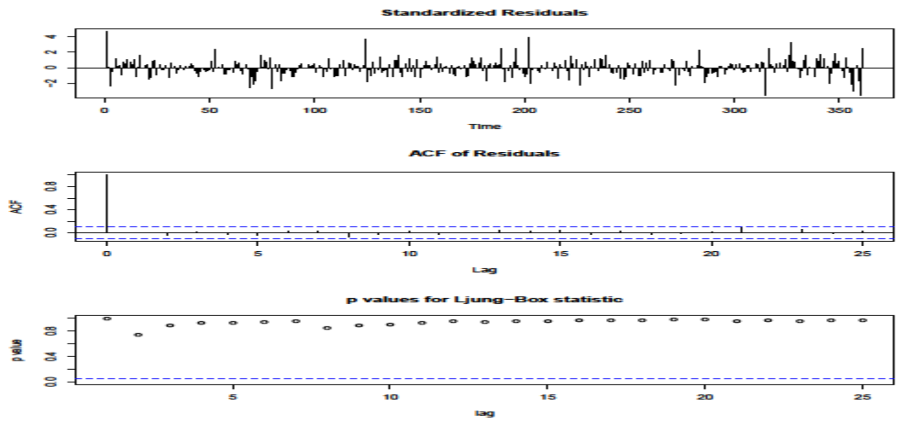


Figure 12: Graphical analysis of the behavior of the residuals of the fitting ARFIMA model

The dotted lines in the second plot correspond to the {$\pm 2/\sqrt{N}$} significance level for the autocorrelations while the dotted lines in the third plot denote the probability (p) of obtaining the indicated result by chance.

6. Forecasting Results

After obtaining appropriate (that is, fitting) models and determining their parameters, ability of the most fitting model (i.e., the ARFIMA (3, 0.4999, 3) model) to forecast the DMEL time series was tested. We made 20-steps ahead forecast with the ARFIMA (3, 0.499, 3) model. Figure 13 unfolds fluctuations in the future forecasts of the DMEL time series obtained from this model.

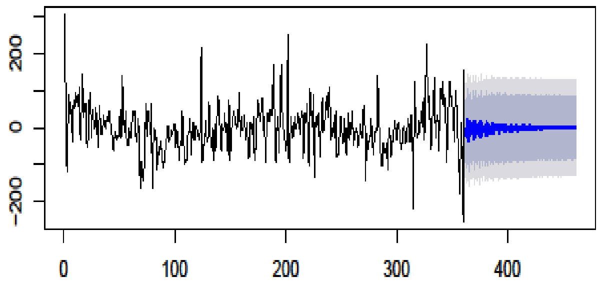


Figure 13: Point forecasts of the ARFIMA (3, 0.4999, 3) model and the associated 80% and 95% prediction intervals

The far right, blue-colored part of Figure 13 points out a series of future predictions represented by a straight line extending to the end. This line indicates that the DMEL values become constant and do not change afterwards.

Figure 14 is a plot of the measured DMEL values and this forecasted by the ARFIMA (3, 0.4999, 3) model. It can be seen that the most fitting predictions are those generated by the ARFIMA model. It is noticed in Figure 14 that the DMEL forecasts of the ARFIMA (3, 0.4999, 3) model are very close to the measured DMEL values (DMEL0M).

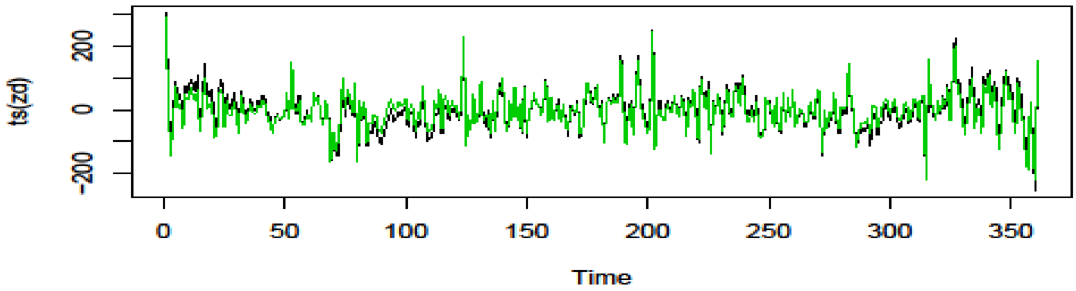


Figure 14: The measured DMEL values (DMEL0M) and the corresponding values forecasted by the ARFIMA (3, 0.4999, 3) model and the ARIMA (0, 1, 2) model

7. Conclusion

In this paper, we have shown how to fit ARFIMA models to real DMEL time series. Our experiment confirmed that the ARFIMA model is a good traffic model that is capable of capturing the properties of real traffic. This can be attributed to the fact that the ARFIMA processes are highly flexible and capable of simultaneously modeling both the long-range and the short-range dependent behavior of a time series.

The stochastic structure of the DMEL data with zero mean (DMEL0M) at West Tripoli Electricity Station in Libya has been analyzed by using a long memory ARFIMA model. The ARFIMA (3, 0.4999, 3) model was used to fit the same DMEL0M data. Even though this model fit the data well, we calculated the fractionally-differenced time series by using the value of 0.4999 for the fractional_differencing parameter, *d*, as suggested by Whittle (1953). In order to develop an ARFIMA model on underlying series and determine the model parameters, we followed the rule of Box and Jenkins and applied the ACF and PACF of differenced time series on the DMEL0M data. Then, we estimated the parameters of the AR and MR operators of this model. We then used these estimates to make 20-steps ahead forecasts with the ARFIMA (3, 0.4999, 3) model. The forecasting results highlighted that the DMEL values predicted by the ARFIMA (3, 0.4999, 3) model were very close to the measured DMEL values.

REFERENCES

1- Akaike, H., (1973) "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models". *Biometrika*, 60(2), 255-265.
2- Baillie, R.T., (1996) "Long memory processes and fractional integration in econometrics"- *J. Econometrics*, 73: 5-59.
3- Baillie, R., & Chung, S.-K. (2002). "Modeling and forecasting from trend-stationary long memory models with applications to climatology". *International Journal of Forecasting 18* , 215-226.
4- Brockwell, P.J. and Davis, R.A., (1991) "Time Series: Theory and Methods". *Springer*-Verlag: New York,.
5- Dickey, D. A. and Fuller, W. A. (1981). "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root". *Econometrica* **49**, 1057-1072.
6- Erfani,A.and Samimi, A.J. (2009).Long Memory Forecasting of Stock Price Index Using a Fractionally Differenced Arma Model, *Journal of Applied Sciences Research*,Vol.5,No.10, pp.1721-1731.
7- Granger, C.and Joyeux R, (1980) "An Introduction to Long-Memory Time Series Models and Fractional Differencing". *Journal of Time Series Analysis*, 1..15-30.
8- Hosking, J.R. (1981) "Fractional differencing". *Biometrika*, 68, 165–176.
9- Kai Liu, Chen,Y. and Zhang., Xi (2017) "An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average)"; *American Society of Mechanical Engineers*: New York, NY, USA,.
10- *Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root". J Econometrics. 54 (1–3): 159–178.*
11- *Phillips, P. C. B.; Perron, P. (1988). "Testing for a Unit Root in Time Series Regression". Biometrika. 75 (2): 335–346.*
12- Shimotsu, K., and Phillips, P.C.B. (2006), "Local whittle estimation of fractional integration and some of its variants," *J Econometrics*, 130, 2, 209-233.
13- Velasco, C., Robinson, P.M., (2000) "Whittle pseudo-maximum likelihood estimation for nonstationary time series", *Journal of the American Statistical Association* 95, 1229-1243.
14- Whittle., P (1953) "The Analysis of Multiple Stationary Time Series" . *Journal of the Royal Statistical Society. Series B (Methodological)* Vol. 15, No. 1, pp. 125-139 .