International Journal of Basic Science and Technology
August, Volume 11, Number 3, Pages 220 - 228 http://www.ijbst.fuotuoke.edu.ng/220
https://doi.org/10.5281/zenodo.16873546 **ISSN 2488-86**

Comparison of Models (SARIMA, ARFIMA, AND SARFIMA) in the Presence of Long Memory

[1*]Aruah K.E., Okereke, W.E., [1]Asogwa, O.C., [1]Okonkwo, C.I. and [1]Chimezie I. E.
[1]Department of Mathematics/Statistics Alex-Ekwueme Federal University Ndufu-Alike
[2]Michael Okpara University of Agriculture Umudike.

Article Information

Article # 100283
Received: 4th April. 2025
1st Revision 11th May. 2025
2nd Revision:26th May 2025
Acceptance:29th July 2025
Available online:
14th August 2025.

Keywords
SARFIMA model
SARIMA model
ARFIMA model
Long memory
.

Abstract

One of the greatest challenges faced in data analysis is fitting the most appropriate model to a dataset. In reality, misspecification of the model has resulted in several wrong decisions in data science. This work compares the efficiency of modelling a time series with seasonal long memory properties with SARIMA, ARFIMA and SARFIMA models. Monthly average global temperature data was used for the this illustration. The temperature data displayed signs of long memory as the ACF plot decayed slowly on further scrutiny, the Hurst exponent produced by R/S analysis we confirmed the presence of long memory.The ACF exhibited exponential decay and a sinusoidal pattern, suggesting non-stationarity and seasonality. Test for stationarity and seasonality were done to confirm these assertions from the plot. Finally, the AIC and BIC were used to evaluate the efficiency of all three models and the result shows that, in the presence of seasonality and long memory, the SARFIMA model was more efficient.

***Corresponding Author**: Kelechi, E.A., aruah.kelechi@funai.edu.ng

Introduction

Several modelling approaches have been used to forecast different datasets, among which are long memory models that allow a time series process to be fractionally integrated. When the time series is detected to have seasonality, removing the impact of any seasonality from a time series before analysis takes place leads to a loss of information. It is necessary, to investigate the efficiency of the model when the seasonal frequency is incorporated into the estimation procedure and the series is allowed to be integrated of a fractional order. This work compairs, the efficiency of a seasonal autoregressive integrated moving-average model, auto-regressive fractionally integrated moving-average model and a seasonal auto-regressive fractionally integrated moving-average model when there is evidence of seasonality and long memory in the series.

Long Memory Process: Long memory in time series can be defined as autocorrelation at long lags (Robinson, 1995). According to Jin and Frechette (2004), memory means that observations are not independent (each observation is affected by the events that preceded it). The autocorrelation function (acf) of a time series y_{Ti} is defined as:

$$\rho_k = cov\ (y_t, y_{t-1}) \ /\ var\ (y_t) \qquad (2.1)$$

where;

ρ_k is the autocorrelation function,
cov (y_t, y_{t-1}) is the covariance of a time series and the one succeeding it, and
var (y_t) is the variance of the time series
For integer lag k, a covariance stationary time-series process is expected to have autocorrelations such that $\lim_{k\to\infty} \rho_k = 0$. Most of the well-known classes of stationary and invertible time series have autocorrelations that decay at an exponential rate, so that $\rho k \approx |m|^k$, where $|m| < 1$, and this property is true, for example, for the well-known stationary and invertible ARMA(p,q) process. For long memory processes, the autocorrelations decay at a hyperbolic rate, which is consistent with $\rho_k \approx Ck^{2d-1}$, as k increases without limit, where C is a constant and d is the long memory parameter.

ARFIMA model:
Fractional integration is a generalisation of integer integration, under which time series are usually presumed to be integrated of order zero or one. For example, an autoregressive moving-average process integrated of order d[denoted by ARFIMA(p, d, q)] can be represented as:

$$- L)^d \phi(L) y_t = \theta(L) \mu_t \qquad (2.2)$$

where, u_t is an independently and identically distributed (i.i.d.) random variable with zero mean and constant variance, L denotes the lag operator; and $\phi(L)$

August, Volume 11, Number 3, Pages 220 - 228 http://www.ijbst.fuotuoke.edu.ng/
https://doi.org/10.5281/zenodo.16873546 ISSN 2488-86

and $\theta(L)$ denote finite polynomials in the lag operator with roots outside the unit circle. For $d = 0$, the process is stationary, and the effect of a shock $u_{(t)}$ on $y_{(t+j)}$ decays geometrically as j increases. For $d=1$, the process is said to have a unit root, and the effect of a shock $u_{(t)}$ on $y_{(t+j)}$ persists into the infinite future. In contrast, fractional integration defines the function $(1 - L)^d$ for non-integer values of the fractional differencing parameter d.

SARFIMA model:

Let $\{X_t\}$ be a zero-mean seasonal fractionally integrated process defined as

$$\left(1-B\right)^{d}\left(1-B^{s}\right)^{D} X_t = \varepsilon_t \qquad (2.3)$$

For $t=1,\ldots,n$, where $0<d, D<0.5$, $\varepsilon_t \in Z$ is a white-noise process with zero mean, variance, B is the usual backshift operator and s is the seasonal period. The process specified by (2.3) is the seasonal fractionally integrated processes denoted here by SARFIMA$(0,d,0)\times(0,D,0)_s$. A more general class of seasonal models can be generated by allowing $\{X_t\}$ to be a stationary and invertible seasonal ARMA process.

Materials and Methods

The data used for this study is secondary data, collected from the net; climate.gov. The data comprises the monthly average global temperature from 2001 through 2018 and is measured in degrees Fahrenheit

The time plot of the temperature series is shown in Figure 4.1 indicates that the series is seasonal.

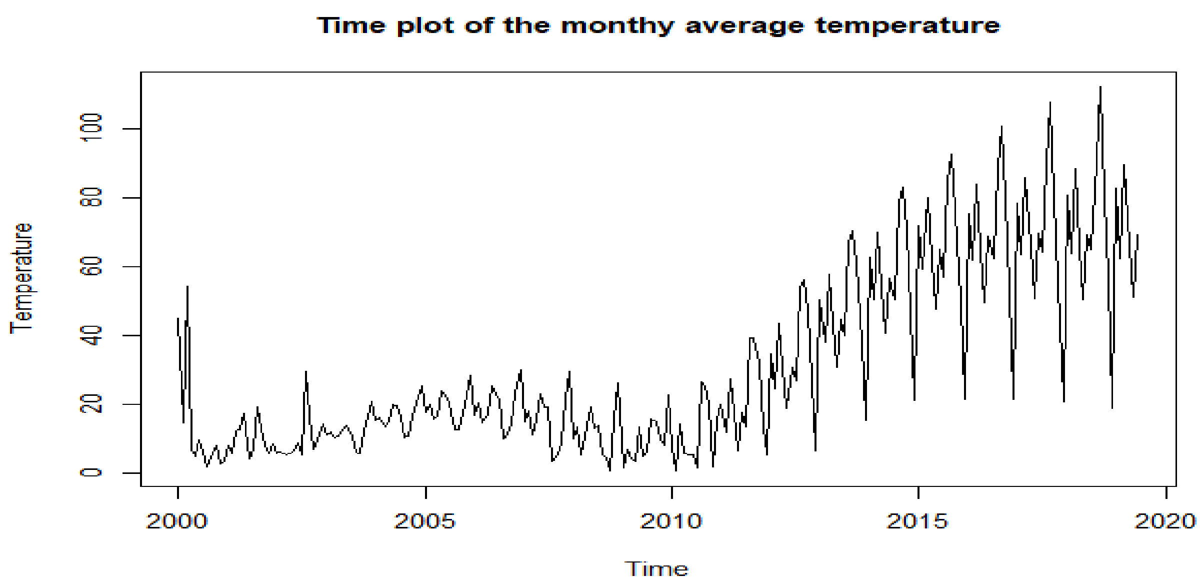


Figure 4.1: Time Plot of the Temperature Series of the Average Global temperature measured in degrees Fahrenheit

This however, does not guarantee that there is neither a secular unit root nor a seasonal unit root. As such, more investigations were done by exploring the ACF and PACF plots.

The ACF plot shows a repeated sine, downward movement and a slow exponential while the PACF plot in Figure 4.2, shows both upward and downward movement, suggesting that the series is not stationary. The slow exponential decay of the ACF indicates a long memory property of the series. These plots strengthen the suspicion that the series may have the presence of secular and seasonal variation. here was a significant upward trend from year 2010.

International Journal of Basic Science and Technology
August, Volume 11, Number 3, Pages 220 - 228 http://www.ijbst.fuotuoke.edu.ng/222
https://doi.org/10.5281/zenodo.16873546 ISSN 2488-86

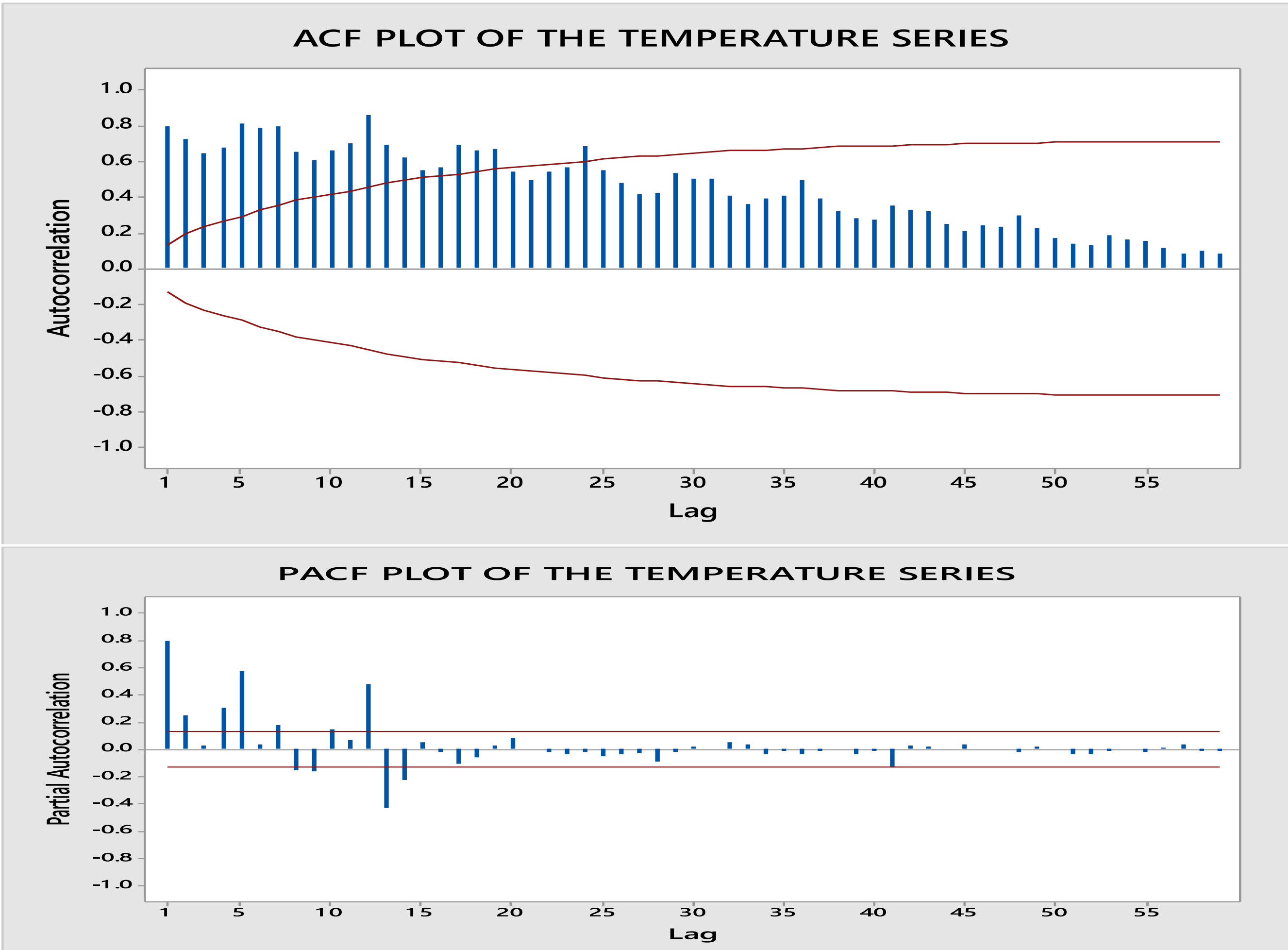


Figure 4.2: ACF and PACF Plot of the Temperature Series before Differencing

Testing for Long Memory

The presence of long memory was further tested using Hurst exponent (H) produced by the Rescaled range analysis. All steps of the calculation of H was done using a program in R. The value of H was obtained to be 0.71394, indicating that average temperature data has a long memory structure since $0.5 < H < 1$

Testing for Stationarity and Seasonality

For confirmation and verification of these assertions from the plot, if there is presence of unit root in the series, the Augmented Dickey Fuller test is mostly used as described in Fuller (1967). The Augmented Dickey-Fuller test for unit root and the Wo-test for seasonality were adopted to investigate the presence of trend and seasonality in the series.

The ADF test examines the null hypothesis that the time series Y_t is stationary against the alternative that it is non-stationary, likewise, the Wo test which examines the null hypothesis that there is a presence of seasonality in the series against the alternative that there is no presence of seasonality in the series.

The Augmented Dickey-Fuller test with test statistic -3.6504 and p-value 0.02923 rejects the alternative hypothesis of stationarity. This therefore shows strong evidence and indicates that the series is not stationary at 1% significance level. The Wo-test developed by Webel and Ollech (2019), for the original series with a test statistic value of 0.0000012 and a p-value 0.000 do not reject the alternative hypothesis of non-seasonality. This indicates a strong presence of a seasonal component in the series. Therefore, there is strong evidence to conclude that the series is not stationary and as such should be subjected to seasonal and secular differencing.

August, Volume 11, Number 3, Pages 220 - 228 http://www.ijbst.fuotuoke.edu.ng/223
https://doi.org/10.5281/zenodo.16873546 ISSN 2488-86

Table 4.1 Summary of the secular and seasonal variation test

Method	differencing order	lag order	test statistic	p-value
Augmented Dickey Fuller	0	6	-3.6504	0.02923
Augmented Dickey Fuller	1	6	-9.2027	0.01
Wo-test	0	1	1.2×10^{-6}	0.000
Wo-test	12	1	0.0018	0.0417
Wo-test	24	1	0.9652	0.9982

R version 3.5.3

After the first differencing, the augmented Dickey-Fuller test was rerun and was found to be stationary. We therefore conclude that the model, after first differencing, became stationary.
Similarly, the Wo-test was rerun for the series after seasonal differencing. The wo-test statistic gave a Re: result of 0.0018 and a p-value 0.0417.This result as shown in table 4.1, indicates that the series was still seasonal after first differencing at 1% significance level but was isolated at 5% significance level. Thus, we can conclude that the series is slightly seasonal after first differencing
.

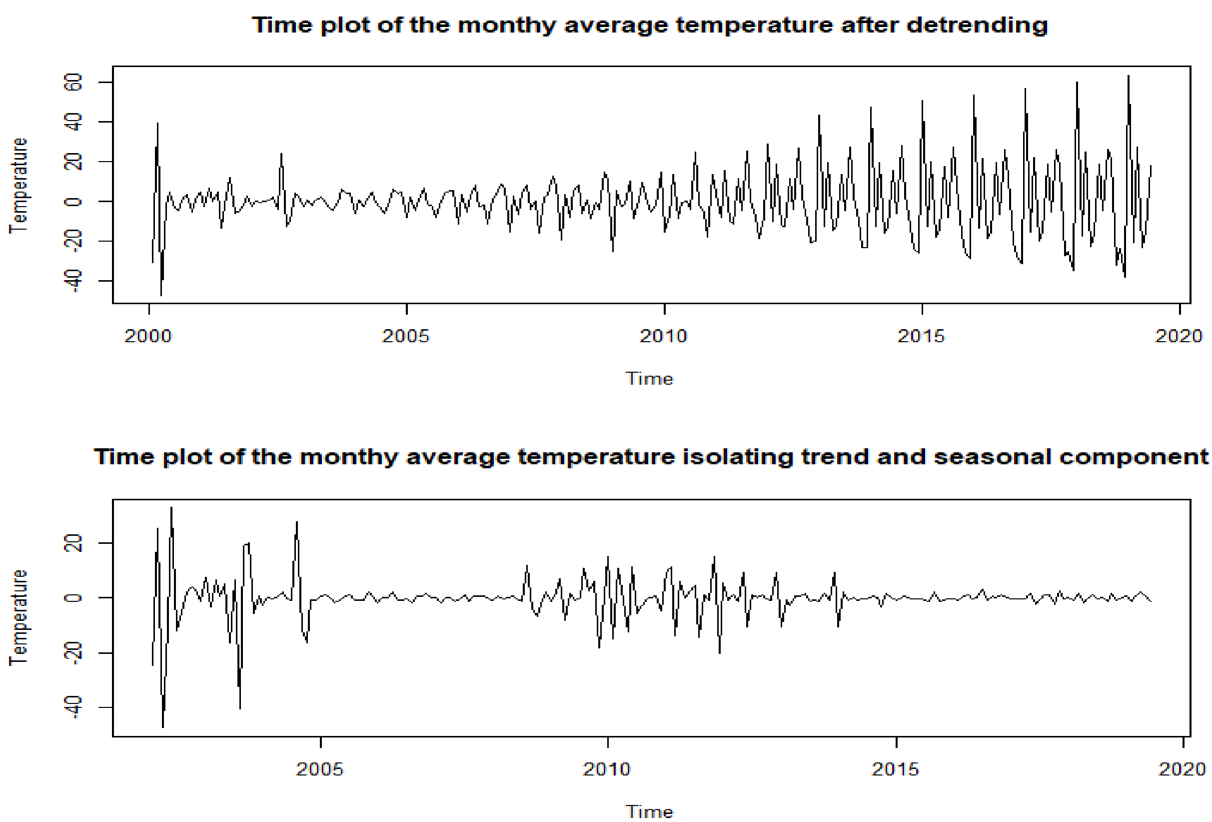


Figure 4.3 Time plot of the series after secular and seasonal decomposition

In an attempt to get the best model, we adopted the Box and Jenkins (1976) approach to fit a SARIMA model. This approach involves determining the order of the AR and MA processes from the PACF and ACF plots, respectively. As such, the ACF and PACF plots for the series differenced at different lags is presented in Figure 4.4 below

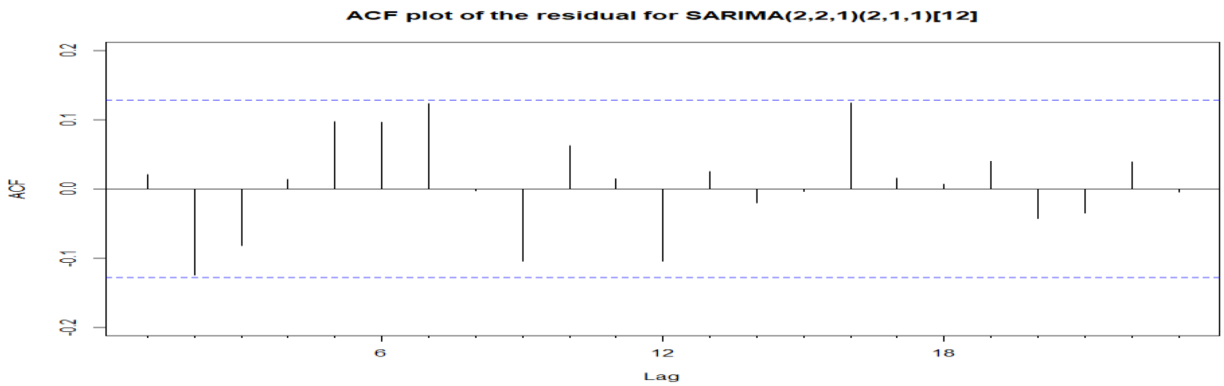
.

Figure 4.4 ACF and PACF plot after secular and seasonal decomposition

From the ACF and PACF plots, we see a significant cut-off in both plots. From the Box and Jenkins approach for identifying a model, we ascertained a possible SARIMA model of order $(2,2,1)(2,1,1)_{12}$. This may not be the most appropriate model that gives the best fit for a SARIMA model, because of the long property observed in the ACF plot, so we diagnose the SARIMA$(2,2,1)(2,1,1)_{12}$ to ascertain if the model is appropriate by fitting the ACF of the Residual.

```
Model Estimation

Call:
arima(x = tsdata1, order = c(2, 2, 1), seasonal = list(order = c(
2, 1, 1)))

Coefficients:
          ar1      ar2      ma1     sar1    sar2    sma1
      -0.5338  -0.2829  -0.9998  -0.0861  0.1187  0.5057
s.e.   0.0656   0.0695   0.0218   0.5293  0.2319  0.5164

sigma^2 estimated as 675.32:  log likelihood = -714.32,  aic = 14
42.63

Training set error measures:
                     ME      RMSE       MAE        MPE      MAPE
MASE       ACF1
Training set -0.9426463 5.843794 3.269185 -27.99146 44.97579 0.26
92348 0.02069044
```

International Journal of Basic Science and Technology
August, Volume 11, Number 3, Pages 220 - 228 http://www.ijbst.fuotuoke.edu.ng/226
https://doi.org/10.5281/zenodo.16873546 ISSN 2488-86



Though the SARIMA model is adequate, it may however not be the most efficient. Therefore, we adopt the auto fit in R package to get the model with the best fit for ARFIMA and SARFIMA model.

Model Fitting and Comparison

The auto fitting functions present in the R package, were used to obtain the optimal model based on Alkaike Information Criterion and Bayesian Information Criterion. The *auto.arima* function in the R package *forecast* was used to fit the SARIMA model while the arfima function was used for the ARFIMA and SARFIMA model based on the algorithm presented by Olatayo and Adedotun (2014). Now we shall discuss the result of these model estimations and identify the model that best fit (in terms of efficiency, based on the least standard error, AIC and BIC).
Table 4.2 shows the best models in terms of AIC model selection with estimators. NA indicates the corresponding parameter is not estimated and is set to be 0. The numbers in parentheses in the column of AIC (BIC) denote the ranking of models in terms of AIC (BIC)

Table 4.3 Summary of model estimation for the best models

SARIMA (2,2,1)(2,1,2)12										
	AR1	**AR2**	**AR3**	**AR4**	**MA1**	**SAR1**	**SAR2**	**SMA1**	**D**	**D**
Estimate	-0.5338	-0.2829	NA	NA	-0.9998	-0861	0.1187	05057	1	1
Se	00656	0.0695	NA	0.0033	0.0218	0.5293	0.2319	0.5164		
ARFIMA (2,-0.0468,1)										
	AR1	**AR2**	**AR3**	**AR4**	**MA1**	**MA2**	**MA3**	**MA4**	**d**	**D**
Estimate	1.2323	-0.235	NA	NA	0.8615	NA	NA	NA	-0.0468	NA
Se	0.0971	0.0964	NA	NA	0.0452	NA	NA	NA		
SARFIMA (2,0,2)(2,0.381,1)12										
	AR1	**AR2**	**AR3**	**AR4**	**MA1**	**MA2**	**MA3**	**MA4**	**d**	**D**
Estimate	0.8192	-0.922	NA	NA	0.5950	-0.9992	NA	NA	NA	0.381
Se	0.0423	0.034	NA	NA	0.0265	0.0000	NA	NA		

Table 4.3 presents the residual comparison of the three estimated model. These includes the standard error, AIC and BIC of the models. Based on the least standard error, Alkaike Information Criterion, Bayesian Information criterion, and absolute log likelihood criterion, the model comparison selects the SARFIMA(2,0,2)(2,0.381,1) as the model with the best fit.

Table 4.3 Model Summary

Model	**Standard Error**	**AIC**	**BIC**	**-loglikelihood**
SARIMA(2,2,1)(2,1,2)12	(3) 675.32	(3) 1442.63	(3) 1473.57	(3) -714.32

International Journal of Basic Science and Technology
August, Volume 11, Number 3, Pages 220 - 228 http://www.ijbst.fuotuoke.edu.ng/227
https://doi.org/10.5281/zenodo.16873546 ISSN 2488-86

ARFIMA(2,-0.0468,1)	(2) 198.40	(2) 1248.10	(2) 1268.83	(2)-618.048
SARFIMA(2,0,2)(2,0.381,1)$_{12}$	(1) 118.33	(1) 1218.4	(1) 1242.59	(1) -602.201

R version 3.5.3

The best model SARFIMA(2,0,2)(2,0.381,1)$_{12s}$ was finally used to make a forecast of the temperature from 2019 to 2023, and the forecast as shown in Figure 4.5 suggests that the temperature is expected to decline slowly in the next five years. The residual diagnostic was further conducted using the normal QQplot and the histogram of the residual. Both plots show that the residuals follow a normal distribution.

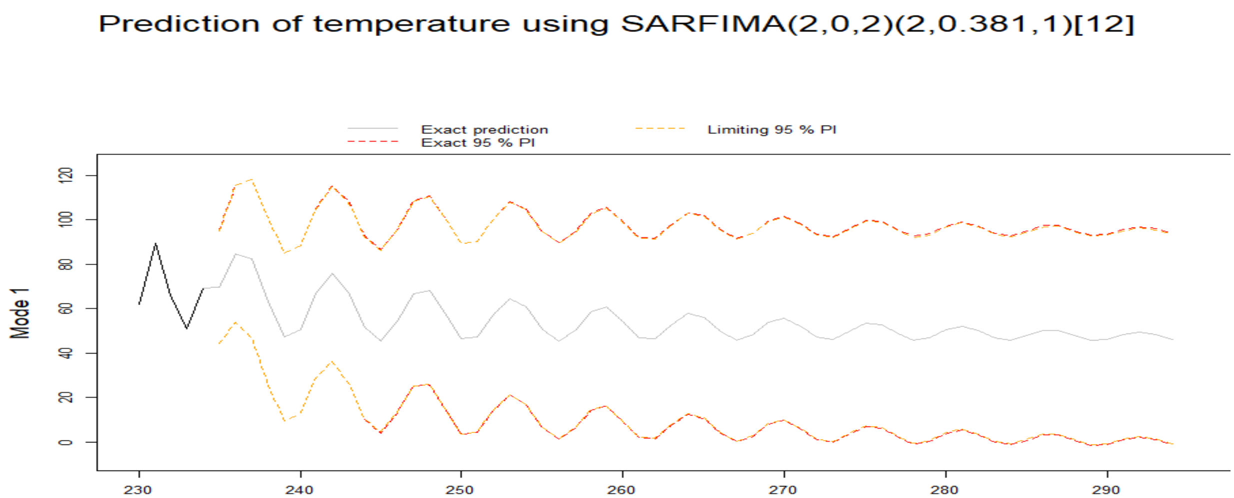
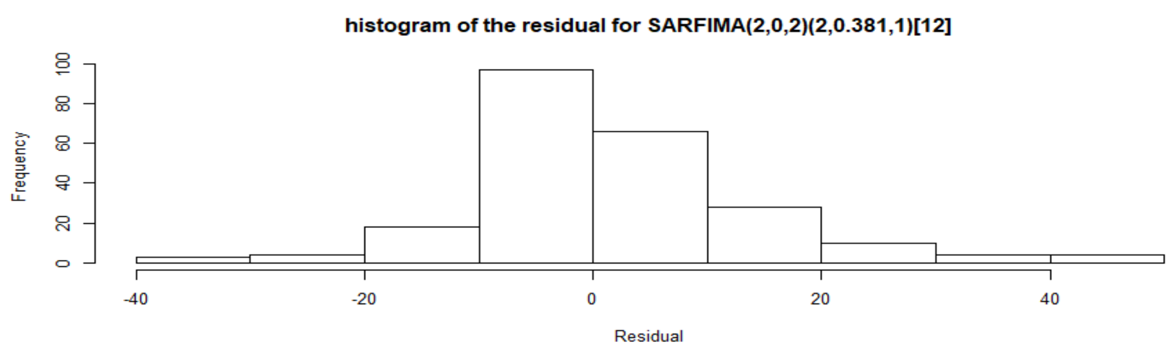
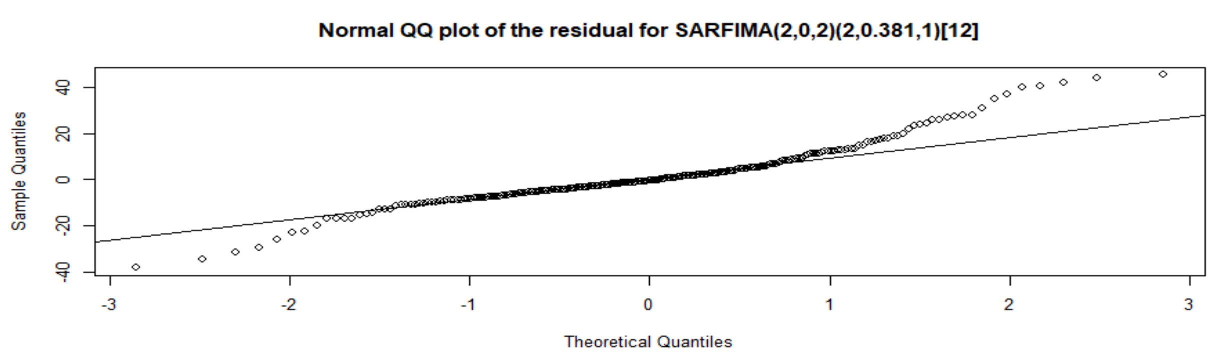


Figure 4.5 Prediction plot of the SARIMA(2,0,2)(2,0.381,1)$_{12}$





August, Volume 11, Number 3, Pages 220 - 228 http://www.ijbst.fuotuoke.edu.ng/228
https://doi.org/10.5281/zenodo.16873546 ISSN 2488-86

Figure 4.5 Residual Diagnostics plot for Normality

Conclusion

In this study, we considered a monthly average global temperature from the period of 2000 through 2018. The ACF plot showed a very slow decay, suggesting a long memory and the presence of trend and seasonal components in the time series.

Based on the properties of the series, we compared the estimates of three models: Sarima $(2,2,1)(2,1,2)_{12}$, ARFIMA(2,-0.0467,1) and SARFIMA$(2,0,2)(1,0.381,1)_{12}$. The results indicate that the SARFIMA model performed better and best fits the temperature series. The comparison based on the least standard error, AIC, BIC and negative log likelihood all chose SARFIMA $(2,0,2)(2,0.381,1)_{12}$ as a better model. Finally, the Normal QQ plot showed that the residual of the fit is normally distributed.

The best model SARFIMA $(2,0,2)(2,0.381,1)_{12}$ was finally used to make a forecast of the temperature from 2019 to 2023, and the forecast as shown in Figure 4.5 suggests that the global temperature is expected to decline slowly in the next five years.

References

Aidoo, E. (2010). Modeling and forecasting inflation rates in Ghana: Anapplication of SARIMA models. Högskolan Dalarna, School of Technology and Business Studies.

Aladag, C. H., Egrioglu, E., and Kadilar, C. (2012). Improvement in forecasting accuracy using the hybrid model of ARFIMA and feed forward neural network. *American Journal of Intelligent Systems*, *2*(2), 12-17.

Alaton, P., Djehiche, B., and Stillberger, D. (2002). On modelling and pricing weather derivatives. *Applied mathematical finance*, *9*(1), 1-20.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *74*(1), 3-30.

Campbell, S. D., and Diebold, F. X. (2005). Weather forecasting for weather derivatives. *Journal of the American Statistical Association*, *100*(469), 6-16.

Cao, M., and Wei, J. (2004). Weather derivatives valuation and market price of weather risk. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *24*(11), 1065-1089.

Cao, M., Li, A., and Wei, J. Z. (2003). Weather derivatives: A new class of financial instruments. *Available at SSRN 1016123*.

Couchman, J., Gounder, R., and Su, J. J. (2006). Long memory properties of real interest rates for 16 countries. *Applied Financial Economics Letters*, *2*(1), 25-30.

Doornik, J.A. and Ooms, M. (2004) Inference and Forecasting for ARFIMA Models, with an Application to US and UK Inflation. *Studies in Nonlinear Dynamics and Econometrics, 8, 1218.* *http://dx.doi.org/10.2202/1558-3708.1218*

Foster, K. (2003). The trouble with normalisation. *Energy and Power Risk Management*, *7*, 22-23.

Hurvich, C.M. (2002). Multi-step forecasting of long memory series using fractional exponent model. *International Journal of Forecasting* 18:167129, *464*.

National Academy of Sciences (2010). Advancing the Science of Climate Change.Accessed

Qi, C. (2023). SARFIMA model prediction for infectious diseases: application to hemorrhagic fever with renal syndrome. *Advanced Techniques in Biology & Medicine*

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, *23*(5), 1630-1661.

Roustant, O., Laurent, J. P., Bay, X., and Carraro, L. (2003). Model risk in the pricing of weather derivatives. *Banque and Marche*.

Webel,K. and Ollech,D. (2019). *An overall seasonality test. Deutsche Bundesbank's* Discussion Paper series.

Zeng, L. (2000). Weather derivatives and weather insurance: concept, application, and analysis. *Bulletin of the American Meteorological Society*, *81*(9), 2075-2082